

Personalized medicine: the role of sequencing technologies in diagnostics, prediction and selection of treatment of monogenous and multifactorial diseases

**Oleg Glotov^{1,2}, Alexandr Chernov^{3,4}, Michael Fedyakov⁵,
Valentina Larionova^{3,6}, Andrey Zaretsky⁷, Maksim Donnikov⁸, and
Andrey Glotov^{2,9}**

¹Department of Experimental Medical Virology, Molecular Genetics and Biobanking, Pediatric Research and Clinical Center for Infectious Diseases, ul. Professora Popova, 9, Saint Petersburg, 197022, Russian Federation

²Department of Genomic Medicine, D. O. Ott Research Institute of Obstetrics, Gynecology and Reproductology, Mendeleevskaya liniya, 3, Saint Petersburg, 199034, Russian Federation

³Institute of Experimental Medicine, ul. Akademika Pavlova, 12, Saint Petersburg, 197376, Russian Federation

⁴Bioenergetics Department of Life Sciences, Ben-Gurion University, Beer Sheva, 84105, Israel

⁵Genetics Laboratory, City Hospital No. 40, ul. Borisova, 9, Saint Petersburg, 197706, Russian Federation

⁶North Western State Medical University named after I. I. Mechnikov, ul. Kirochnaya, 41, Saint Petersburg, 191015, Russian Federation

⁷Department of Molecular Technologies, Research Institute of Translational Medicine, Pirogov Russian National Research Medical University, ul. Ostrovityanova, 1, Moscow, 117997, Russian Federation

⁸Medical Institute, Surgut State University, ul. Energetikov, 22, Surgut, 628412, Russian Federation

⁹Laboratory of Biobanking and Genomic Medicine, Institute of Translation Biomedicine, Saint Petersburg State University, Universitetskaya nab., 7–9, Saint Petersburg, 199034, Russian Federation

Address correspondence and requests for materials to Oleg Glotov, olgotov@mail.ru

Citation: Glotov, O., Chernov, A., Fedyakov, M., Larionova, V., Zaretsky, A., Donnikov, M., and Glotov, A. 2022. Personalized medicine: the role of sequencing technologies in diagnostics, prediction and selection of treatment of monogenous and multifactorial diseases. *Bio. Comm.* 67(4): 266–285. <https://doi.org/10.21638/spbu03.2022.403>

Authors' information: Oleg Glotov, PhD, Head of Department, orcid.org/0000-0002-0091-2224; Alexandr Chernov, PhD, Researcher, orcid.org/0000-0003-2464-7370; Michael Fedyakov, Specialist, orcid.org/0000-0002-3291-3811; Valentina Larionova, PhD, Professor, orcid.org/0000-0002-3128-8102; Andrey Zaretsky, Researcher, orcid.org/0000-0002-7778-6617; Maxim Donnikov, PhD, Leading Researcher, orcid.org/0000-0003-0120-4163; Andrey Glotov, Dr. of Sci. in Biology, Head of Department, orcid.org/0000-0002-7465-4504

Manuscript Editor: Anna Malashicheva, Laboratory of Regenerative Biomedicine, Institute of Cytology, Russian Academy of Sciences, Saint Petersburg, Russia; Laboratory of Molecular Cardiology, Almazov National Medical Research Centre, Saint Petersburg, Russia

Received: August 8, 2022;

Revised: September 22, 2022;

Accepted: September 25, 2022.

Copyright: © 2022 Glotov et al. This is an open-access article distributed under the terms of the License Agreement with Saint Petersburg State University, which permits to the authors unrestricted distribution, and self-archiving free of charge.

Funding: Supported by the Ministry of Science and Higher Education of Russian Federation (project "Multicenter research bioresource collection "Human Reproductive Health" contract No. 075-15-2021-1058 from September 28, 2021). The reported study was funded by Foundation for Scientific and Technological Development of Yugra according to the research project No. 2022-05-04.

Ethics statement: This paper does not contain any studies involving human participants or animals performed by any of the authors.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The review highlights various methods for deciphering the nucleotide sequence (sequencing) of nucleic acids and their importance for the implementation of the three main principles of personalized medicine: prevention, predictability and personalization. The review, along with its own practical examples, considers three generations of sequencing technologies: 1) sequencing of cloned or amplified DNA fragments according to Sanger and its analogues; 2) massive parallel sequencing of DNA libraries with short reads (NGS); and 3) sequencing of single molecules of DNA and RNA with long reads. The methods of whole genome, whole exome, targeted, RNA sequencing and sequencing based on chromatin immunoprecipitation are also discussed. The advantages and limitations of the above methods for diagnosing monogenic and oncological diseases, as well as for identifying risk factors and predicting the course of socially significant multifactorial diseases are discussed. Using examples from clinical practice, algorithms for the application and selection of sequencing technologies are demonstrated. As a result of the use of sequencing technologies, it has now become possible to determine the molecular mechanism of the development of monogenic, orphan and multifactorial diseases, the knowledge of which is necessary for personalized patient therapy. In science, these technologies paved the way for international genome projects — the Human Genome Project, the HapMap, 1000 Genomes Project, the Personalized Genome Project, etc.

Keywords: sequencing technologies, whole genome sequencing, whole exome sequencing, advantages, limitations, personalized medicine, human genomic projects, diagnostics, disease prediction, personalized therapy

Introduction

Advances in biological and medical science and technology in the late 20th and early 21st century ushered the introduction of innovative technologies for early diagnostics and detection of numerous target protein and gene-based markers, accelerating the implementation of advanced screening strategies and target therapy protocols. These developments improved our understanding of disease mechanisms behind rare monogenic disorders and allowed to enhance prevention and efficiency of treatment of socially significant multifactorial diseases, eventually resulting in improved health-related quality of life and longer life expectancy in economically developed countries. These successes helped transform the entire healthcare paradigm from an impersonal approach towards predictive and preventive personalized medicine (PM) and treatment, relying on individual gene expression profiles and disease-specific molecular markers. This allowed for risk assessment, disease prediction, forecasting of individual health outcomes and treatment responses, based on a patient's age, gender, disease type and stage. Rapid development of personalized medicine is in line with the forecasted growth of its market, reaching 87.7 bln. USD by 2023 (Newswire, 2016). This process was driven by the advances in genetics and IT which have brought up new disciplines, like genomics (proteomics, metabolomics, transcriptomics, pharmacogenomics) and bioinformatics with its innovative technologies, including new generation sequencing (NGS) first and foremost.

The review, along with its own practical examples, considers various sequencing technologies, their advantages and disadvantages for the detection of diagnostic molecular genetic markers of monogenic and socially significant multifactorial human diseases. The methods of sequencing are consecrated from the point of view of their application in order to prevent the development of diseases and their complications. Particular emphasis is placed on the role of sequencing methods for the selection of effective personalized (targeted) therapy, taking into account the molecular mechanism of the development of the disease.

DNA sequencing technologies and their principles

First generation methods: Sanger sequencing, Maxam — Gilbert sequencing, pyrosequencing

The first method for direct enzymatic DNA sequencing was the method proposed by F. Sanger and D. Coulson in 1975 (Sanger and Coulson, 1975). A single-stranded DNA fragment was used as a template in the polymerase copying reaction, and the Klenow fragment of DNA

polymerase I (PolI) from *E. coli* was used as an enzyme, (Klenow fragment is a fragment of DNA polymerase I and catalyzes DNA synthesis on a DNA template in 5' → 3' direction). Synthetic oligonucleotides or natural sub-fragments obtained by hydrolysis with restriction endonucleases were used as primers. This method was called the “plus-minus” method, and phage φX174 short DNA sequencing, consisting of 5386 nucleotide pairs, was carried out by this method.

In 1977, Frederick Sanger proposed another method of enzymatic sequencing, which was called the method of terminating triphosphate analogues (the “chain termination” method) (Sanger and Coulson, 1975; Sanger, Nicklen and Coulson, 1977).

The manual method was based on enzymatic copying using the Klenow fragment of DNA polymerase I from *E. coli*. Synthetic oligonucleotides were used as primers. Specific termination of the synthesis was ensured by adding to the reaction mixture, in addition to the four types of dNTPs (one of which was radioactively labeled at the alpha position of the phosphate), also one of the 2',3'-dideoxynucleoside triphosphates (ddATP, ddTTP, ddCTP, or ddGTP), which can be included in the growing DNA chain, but is not able to provide further copying due to the lack of a 3'-OH group. Thus, to determine the primary structure of the studied DNA fragment, it was necessary to carry out four copying reactions: one type of terminator in each of the reactions. After that, the separation of the sets of DNA fragments which form the “sequencing ladder” was carried out by electrophoresis in polyacrylamide gel under denaturing conditions. Upon completion of electrophoresis, the gel was exposed to X-ray film, and after some time (usually from one to two days) it was possible to “read” the nucleotide sequence of the sequenced DNA region from the developed film, starting from the bottom of the gel and sequentially rising up along these four tracks, corresponding to one piece of DNA.

The method proposed in 1977 by Frederick Sanger has been automated. The most successful sequencer of the early 90s was ABI 373 (Applied Biosystems). Automation of the Sanger method and its combination with the capillary electrophoresis format made it possible to implement the international project “Human Genome” in 1990–2003 and to determine the nucleotide sequence of human DNA, that is, to sequence the genome. High precision and optimal length of reads made these methods a gold standard for sequencing (Rabbani et al., 2012).

For many years, the Sanger method has served as a method for sequencing individual sections of the genome in order to analyze mutations and polymorphisms in various genes. The use of this method made it possible to study the spectrum of mutational damage in the studied genes and to establish the cause of many hereditary diseases. Determining the molecular nature

of a particular disease and the functional class of the mutation(s) at the present stage allows to predict the course of the disease and personalize therapy. Identification of the carriage of mutations in various genes that are the cause of the majority of hereditary diseases with autosomal recessive inheritance in future parents makes it possible to prevent the birth of a sick child through the use of reproductive technologies. Thus, the Sanger method contributes to the implementation of the three main principles of personalized medicine: predictability, prevention, personalization.

American geneticists Craig Venter and Francis Collins were the first to implement human genome shotgun sequencing. The method involves DNA amplification, breaking its copies into overlapping fragments using non-specific endonucleases, then ligating genomic DNA fragments into a vector (e.g. bacterial chromosomes) to construct a genome library, followed by Sanger sequencing, obtaining nucleotide reads using software tools and reconstruction of the original DNA sequence (Table 1) (Staden, 1979; Weber and Myers, 1997). The original DNA reassembly is performed in steps beginning from the assembly of the sequence reads into contigs, i.e. contiguous fragments of the original DNA sequence assembled into larger scaffolds and separated by gaps. Pairwise end sequencing, known as double-barrel shotgun sequencing, is modified shotgun sequencing using double-stranded template DNA. If in paired-end sequencing the size of the gap ranges between 5–20,000 b.p., the insert DNA is amplified using PCR and then sequenced. If the gap between contigs is over 20,000 b.p., the DNA fragment is cloned in bacterial artificial chromosome, followed by genomic DNA vector sequencing (Osoegawa et al., 2001). This method allows to sequence longer DNA fragments. Subsequently, this methodology became the basis for working with the results of second-generation sequencing (Table 1) (Gonzalez-Garay, 2014).

Parallel to Sanger sequencing (developed in 1977), Walter Gilbert and Allan Maxam published a paper describing a different method for sequencing DNA using chemical degradation followed by base-specific restriction (Maxam and Gilbert, 1977). This technique is known as Maxam — Gilbert sequencing, or chemical cleavage method, and requires radioactive labeling at one 5'-end of the double-stranded DNA fragments, DNA purification with subsequent chemical treatment to break nucleotide bases and generate a set of radiolabeled DNA fragments. The fragments are then electrophoresed in denaturing acrylamide gels for size separation and visualized using autoradiography. Due to its technical complexity and the radioactive labelling requirement, Maxam — Gilbert technique has fallen out of favor.

Pyrosequencing technique was developed by Pål Nyrén and Mostafa Ronaghi from the Royal Institute

of Technology in Stockholm in 1996 and later modified by Jonathan Mark Rothberg (Connecticut, USA). The method employs a CCD camera for luminometric quantification of the pyrophosphate that is released as a result of the conversion of luciferin to its oxo-derivative upon polymerase-catalyzed DNA chain elongation. The ensuing light output is proportional to the number of nucleotide bases incorporated into the sequence. The pyrogram, reflecting the nucleotide sequence of a DNA fragment, is then interpreted using dedicated software (Ronaghi, Uhlen and Nyren, 1998; Margulies et al., 2005). A single pyrosequencing cycle is capable of generating hundreds of thousands DNA sequence reads at a cost cheaper than Sanger sequencing. The main problems with pyrosequencing include inability to sequence long nucleotide sequences and low accuracy in detecting variants within long homopolymer stretches. In addition, this technology does not allow unmistakable sequencing of long sequences consisting of the same nucleotide (the “homopolymer problem”).

Second generation methods: high throughput sequencing with short reads

Massively parallel signature sequencing (MPSS) was introduced in 1992 by Sydney Brenner and Sam Eletr at Lynx Therapeutics, a biotech company where Sam Eletr was CEO. MPSS is used to identify mRNA transcripts and quantify gene expression in the individual cell by capturing transcripts on individual microbeads through a complementary DNA signature sequence; bases of the mRNA are then read by hybridization to a fluorescently labeled coder and then removed. The result is an array of sequences, ranging from 17 to 20 bp. The count of mRNA transcripts, indicating the expression level, is determined by the number of transcripts present per million molecules. MPSS does not require that genes are identified before beginning the analysis. The sensitivity of MPSS is a few molecules of mRNA per cell. This method, however, is very complicated and was performed exclusively in Lynx Therapeutics (Brenner et al., 2000).

The pioneering work by George Church and colleagues (Massachusetts, USA) led to the development of a multiplex sequencing technique (polony sequencing), which combines emulsion PCR, enzymatic ligation and four-color imaging (Shendure et al., 2005). The platform is known as the Polonator sequencer (The Polonator G 007 (Open Source Next Generation Sequencing Technology)). This device made it possible to carry out a fairly accurate and high-performance reading of short DNA fragments — about 30 nucleotides. The manufacturing technology for this instrument and reagents, originally conceived in a fully open format (DIY — Do It Yourself), was subsequently licensed to Danaher Motion. The device was not a commercial success — less

than 10 copies of the sequencer were sold. The portfolio of patents related to the Polonator technology was purchased by Illumina, which decided not to develop this direction in sequencing.

A new version of ligation-based sequencing was proposed in 2009 by Life Technologies (USA) and was called SOLiD (sequencing by oligonucleotide ligation and detection) sequencing technology based on solid-phase ligation, and built the concept instrument called SOLiD™ Instruments (Table 1) (Valouev et al., 2008). The SOLiD technology is based on polony sequencing, accompanied by the library preparation for single-stranded DNA fragments, ligated to A1 and A2 adapters, followed by emulsion PCR; as a result, single microbeads, immobilized by single-stranded primers (either P1 or P2), are encapsulated in individual droplets and hybridize to the adapter sequence on the beads, with subsequent PCR amplification of the DNA template and a set of 5'-end fluorescently labelled probes composed by eight bases, competing for ligation to the sequencing primers. The sequencing procedure involved five rounds of primer reset, each round including 6 to 7 cycles. Only the first two nucleotides at the 3'-end of the probe match the two nucleotide bases of the target DNA; once bases 6 to 8 are separated, the phosphorothiolate linkage is broken, thus regenerating the 5' phosphate group for subsequent probe ligation. After each round the emitted fluorescence of the ligated probe is evaluated. Unlike other sequencing platforms, SOLiD System utilizes two-base-encoding, which interrogates each nucleotide base twice during sequencing, allowing for unmatched accuracy. Capable of sequencing hundreds of millions and billions of short-length sequences (of approximately 25 bp), SOLiD system reduces the cost of the whole genome sequencing to 100,000 USD (Gonzalez-Garay, 2014). As with the Polonator, this technology proved to be overly complicated for routine use and was halted by the manufacturer after several years of active but largely unsuccessful publicity.

The third variant of high-throughput ligation sequencing was proposed almost simultaneously with the SOLiD method. The pioneer of microchip technology Radoje Drmanac, who founded Complete Genomics Inc. (CGI; California, USA), became the author of the new solution. Combinatorial Probe-Ancor Synthesis (cPAS) developed by Complete Genomics company, is a combination of sequencing by hybridization and sequencing by ligation methods. A DNA fragment of 500 bp length is ligated to adapters complementary to oligonucleotide anchor primers to form a single-stranded DNA circle. Oligonucleotide primer extension is used for DNA rolling circle amplification that generates approximately 300–500 copies of the original DNA. The long strand of ssDNA folds upon itself to produce a DNA nanoball structure (DNB) of 220 nm in diameter

which is attached to a flow cell with 1 to 4 electrostatically charged lanes. Then the DNB adapter sequence is ligated to the primer (anchor) and 4 fluorescently labelled sample types with a specific 5'-end nucleotide. After hybridization and DNA ligation to the fluorescent nucleotide, a laser beam is used to excite the fluorophore, allowing for dye detection by a camera. The DNA sequencing involves 50 to 150 repeated hybridization-ligation cycles. Rolling circle amplification, where every subsequent cycle uses the original DNA circle template, leads to low error accumulation during PCR amplification of DNA, higher accuracy of reads up to 99.9%, reduced costs per test, and faster sequencing procedure (Huang et al., 2017). The throughput of Complete Genomics Inc. (CGI; California, USA) devices was claimed to be from 3.2 to 4.5 mln sequence variants per genome. The platform achieves astonishing sequence accuracy (1 false variant per 100 kilobases), low reagent consumption, affordable cost (4400 USD for whole genome sequencing), and efficient imaging, offering a powerful tool to detect rare genetic variants (Drmanac et al., 2010). Complete Genomics Inc has never sold its devices, and has only commercialized high-precision genome reading services (primarily for medical purposes). The economic efficiency of this company turned out to be rather limited, as a result of which the company closed, and the technology was bought by the Beijing Genomic Institute (BGI). Subsequently, the Chinese colleagues proposed MGI technology, a turnkey solution for high-throughput sequencing, which, according to them, incorporated all the best of Complete Genomics, but, according to some experts, was more reminiscent of the Solexa / Illumina method (see below).

In the early 2000s, Jonathan Mark Rothberg (Connecticut, USA) managed to create a high-performance version of pyrosequencing and started its commercialization on the basis of 454, later acquired by the Roche corporation (Margulies et al., 2005). It was the first example of a commercially successful second generation sequencer. In one cycle of the updated pyrosequencing, hundreds of thousands of reads can be obtained. Fourth generation pyrosequencing-based GS FLX Titanium technology was capable of generating reads of 400 bp and longer (Siqueira, Fouad, and Rôças, 2012). Using the 454 technology, the genomes of many living organisms, from humans to bacterial communities, have been successfully sequenced (Gonzalez-Garay, 2014). Since 2016, the 454 technology has not been supported by the manufacturer (it is impossible not only to upgrade or repair the device, but also to purchase consumables and reagents for it).

In 2010, the same author — J.M. Rothberg introduced the technology of ion semiconductor sequencing technology and founded Ion Torrent Systems Inc. (Gilford, USA) company. The technology utilized an

innovative semiconductor-based imaging system for the detection of hydrogen ions, which are released during the DNA microwell polymerization. Homopolymer nucleotide repeats in the template sequence cause the release of respective amount of hydrogen ion, which triggers a proportionate of the electron signal indicated by the ion sensor (Rothberg et al., 2011). In principle, the technology had much in common with pyrosequencing, but allowed for greater scaling. Ion Torrent was acquired by Life Technologies Corporation (now Thermo Fisher Scientific), and instruments based on this technology remain commercially available and in limited demand today.

A fundamentally different high-throughput sequencing technology was developed in the laboratory of Shankar Balasubramanian at the Faculty of Chemistry of Cambridge University (United Kingdom). The technology, based on parallel sequencing of short reads (75 bp), uses solid phase bridge sequencing by reversible terminators. A digital camera registers every incorporated nucleotide. In 2006 pioneer sequencer Genome Analyzer utilizing Solexa/Illumina technology was commercialized (Gonzalez-Garay, 2014). The Solexa next generation sequencing technology was soon acquired by the biotech giant Illumina. It is with Solexa/Illumina technology and the developments of Illumina that the methodological revolution is associated, which made high-throughput sequencing affordable and allowed it to be used to solve almost any molecular biological and medical genetic problems. Solexa/Illumina technology is represented by several variations (Hodkinson and Grice, 2015):

- 1) single-read sequencing, allowing to sequence 8 samples per single cell with read length 75 bp and sequencing depth of 100 mln reads/cell;
- 2) paired-end sequencing allows to sequence forward and reverse DNA strands from both respective ends of a long-length (200 to 500 bp) DNA fragment, ligated to two types of adapters at both ends. After bridge DNA amplification, adapter sequences interact consecutively with forward and reverse primers to synthesize complementary DNA strands. The sequencing depth is 200 mln reads, assuming the read length is 75 bp;
- 3) multiplex sequencing enables parallel sequencing of 96 indexed and adapter-ligated DNA samples per flow cell, pooling up to 12 samples per lane;
- 4) mate pair sequencing allows to sequence two distanced (5000 bp) DNA fragments as a pair of short reads. This is a powerful technology to identify mutations or perform *de novo* sequencing;
- 5) RNA sequencing.

According to Illumina forecasts, the cost of human genome sequencing may drop down to 100 USD in the

next 3 to 10 years (Herper, 2017), with the whole process completed within an hour's time (Fikes, 2017; Lightbody et al., 2019).

Third Generation Methods: Real-time Sequencing of Single Molecules with Long Reads

Helicos™ single molecule sequencing (SMS) was developed by Helicos Biosciences company. The method utilizes DNA fragments tailed with poly(A) adapters and hybridized to a flow cell surface. In each cycle one type of fluorescently labelled nucleotides is supplied to the flow cell and added to the DNA strand for the Heliscope sequencer to measure the fluorescence. Helicos SMS can generate short-length reads (35 bp) (Thompson and Steinmann, 2010). Sample preparation does not require PCR amplification, thus avoiding underlying biases in sequencing (Heather and Chain, 2016).

Pacific Biosciences® (PacBio Sequel systems; California, USA) company developed single-molecule real-time (SMRT) sequencing, based on real-time DNA synthesis and fluorescently labelled nucleotide detection, which allows PacBio RSII sequencers to generate long read lengths (5 kilobases) (Eid et al., 2009; PACBIO). The method can also detect nucleotide modifications (methylation) and is therefore utilized for gap-filling of sequenced DNA samples (Table 1) (English et al., 2012). PacBio sequencers are used for genome *de novo* assembly or to capture isoforms (Goodwin, McPherson and McCombie, 2016).

Nanopore sequencing was developed by Oxford Nanopore technologies (GridION™ System) and works by monitoring changes to an electrical current as nucleic acids are passed through an 8 nm electrically charged nanopore, embedded in an electro-resistant membrane. The magnitude of the electric current density and time differences depend on the composition of the nucleotide passing through a nanopore. The resulting parameters of ionic currents are decoded to provide data on the DNA length and composition. The technique does not require nucleotide modification and is performed in real time (Table 1) (Branton, Deamer and Marziali, 2008). Alongside with ion channels, other two nanopore sequencing technologies have been explored: solid-state nanopore sequencing and protein-based nanopore sequencing. Solid-state nanopore sequencing utilizes solid-state nanopores manufactured on silicon nitride or aluminum oxide that show high thermal stability and excellent mechanical properties (Goto et al., 2020). Protein-based nanopore sequencing utilizes membrane proteins, such as Mycobacterium porin, CsgG, or protein α -hemolysin, capable to detect individual nucleotides (Liu et al., 2012; Di Muccio et al., 2019).

Table 1. Overview of sequencing technologies and their parameters (Liu et al., 2012; Quail et al., 2012; Morganti et al., 2020)

Technology	Working principle	Platform	Read length, bp	Reads per slide/ panel	Run time	Cost per 1 billion bases (in USD)	Strength	Limitation
Sanger sequencing	Chain termination	3130xL-3730xL	400-900	96	20 min — 3 h	2,400,000	Suitable for numerous applications	Expensive; requires PCR or cloned plasmids
454 Life Sciences	Pyrosequencing	FLX System	700-1000	1 mln	18-23 h	10,000	Long read length; short sequencing run time	High cost; homopolymer length errors
Illumina- SOLEXA	Synthesis-based sequencing, using fluorescently-labelled nucleotides and reversible terminators.	HiSeq2000/ miSeq	MiniSeq, NextSeq 75-300 MiSeq 50- 600 HiSeq2500 50- 500; HiSeq 3/4000 50-300 HiSeq X 300 NovaSeq6000 150	MiniSeq / MiSeq: 1-25 mln; NextSeq: 130-00 mln; HiSeq 2500: 300 mln — 2 bln; HiSeq 3/4000 2.5 bln; HiSeq X: 5 bln; NovaSeq6000 20 bln	4-5 days 12-30 h 4-24 h 7 h — 6 days 1-3 h 36-44 h	5-150	High throughput, depending on the sequencer model	High cost of equipment; DNA concentration requirements; long sequencing run time
SOLID sequencing	Ligation of fluorescently- labelled oligonucleotide probes	5500xL SOLiD System	50 + 35 or 50 + 50	1.2-1.4 bln	10 days	60-130	Low cost	long sequencing run time; unable to read palindromic sequences; technical complexity of analysis
IonTorrent	Semiconductor-based detection of hydrogen ion release during incorporation of nucleotides into the growing DNA	PGM sequencer	up to 600	80 mln	2-4 h	66.8-950.0	Low cost; short sequencing run time	Homopolymer length errors

End of the Table 1

Technology	Working principle	Platform	Read length, bp	Reads per slide/ panel	Run time	Cost per 1 billion bases (in USD)	Strength	Limitation
Helicos	Fluorescently-labelled nucleotides with reversible terminators	Heliscope	25–30	35,000–75,000	1 h	2,000	Length of usable reads; short sequencing run time	Low throughput at the cost of reduced error rate; high cost
cPAS-BGI	Combinatorial probe-anchor synthesis	MGI	BGISEQ50 35–50 MGISEQ200 50–200; BGISEQ500 MGISEQ2000 50–300	BGISEQ-50: 160M; MGISEQ 200: 300M; BGISEQ-500: 1300M; MGISEQ2000: 375M, 1500M	1–9 days	5–120		
Pacific Biosciences (SMRT)	Single molecule real-time sequencing	PacBio RS	20,000–30,000. (N50); 100,000	4,000,000 per SMRT Sequel 2 cell, 100–200 Gbp	30 min — 4 h	7.2–43.3	Short sequencing run time; detects 4 μ CK, 5 μ C, 6 mA.	Moderate throughput; high cost of equipment.
Oxford Nanopore technologies	Changes to an electrical current as nucleic acids are passed through a nanopore	GridION™	Depends on the library preparation, (up to 900,000)	Depends on the user-controlled length of reads (up to 20 Gbp)	Real-time data streaming (1 min to 48 h)	7–100	Portable pocket-size device; long read length; low cost; no amplification or complicated chemistry required.	Low throughput

Types of Targeted NGS

Depending on its application, high-throughput sequencing can be split into whole genome sequencing (WGS), whole exome sequencing (WES), targeted sequencing (TS), RNA sequencing (RNA-seq), or ChIP-sequencing (ChIP-seq) (Table 2).

WGS allows to analyze the entire human genome, including both structural and regulatory genes, as well as roughly 3 bln bp of haploid cells.

The latest release GRCh38 human genome assembly, however, shows that structural protein-coding genes (the exome) encompass only approximately 3.09% (90 mln bp) of human genome, but harbor approximately 85% of all described disease-causing sequence variants (Majewski et al., 2011; Guo et al., 2017). In contrast to WGS, WES is best suited for identification of variants in protein-coding regions, since it provides enhanced coverage depth (22,000 genes) at a reduced cost, which makes WES a reliable tool to detect single nucleotide polymorphism (SNP), insertion and deletion sites, potentially causative of a disease (Table 2) (Gorski et al., 2016; Suwinski et al., 2019).

WES has proved to be a valuable tool to identify pathologic variants in disease genes, e.g. those reported as the cause of Miller syndrome (Chong et al., 2015). For its diagnostic value, since 2011 WES is often performed in US clinical genetics labs (Pierson et al., 2011).

NGS platforms capable of WES are based on a broad range of methods, e.g. sequencing by synthesis (Illumina/SOLEXA), sequencing by ligation (SOLiD), pyrosequencing (454/Roche), or ion semiconductor sequencing (Ion Torrent) (Buermans and den Dunnen, 2014; Kchouk, Gibrat and Elloumi, 2017). When performing WES, a key consideration factor is the selection of primers for the targeted hybridization of gene-coding proteins (i.e. the exome capture kit), not as much as the choice of a platform. Various commercial kits are available, such as Agilent SureSelect XT, Agilent SureSelect QXT, NimbleGen SeqCap EZ and Illumina Nextera Rapid Capture Exome. They use biotinylated DNA baits, which are hybridized to genomic fragment libraries. Yet they differ in the genomic fragmentation method, bait length, bait density, and target region selection (Suwinski et al., 2019). In addition, WES generates ~5–6 GB of data, which is substantially lesser than ~90 GB for WGS for the same sample (AllSeq, WGS vs. WES).

At the same time, considering the significant size of generated WGS or WES data, data processing and analysis becomes a bottleneck, making it challenging to differentiate small mutations from random errors generated during sequencing (Hofmann et al., 2017). In addition, a major limitation of WES is the uneven coverage of sequence reads over the exome targets, contributing to many low coverage regions, which affect the down-

stream analysis and hinder accurate variant annotation, causing missed variant calls (Wang et al., 2017).

WES data can include inconsistencies, such as anomalies and outliers, or inconsistent speed at which data is loaded into the repository, alongside with inherent limitations, such as the GC bias, difficulties in discriminating paralogous sequences or in phasing alleles, or linking sequence variants with biological data and phenotype. Translation of sequencing findings into easily understood medical standards, similarly to clinical diagnostic scoring, may present another potential limitation (Suwinski et al., 2019).

Targeted sequencing (TS) allows to focus on specific regions or genes of interest within the genome and better understand preliminary evidence of a pathogenic process (Lionel et al., 2018). For example, short read massive parallel sequencing has emerged as a standard tool used in the US, European and Australian clinical genetics labs (Ardui et al., 2018). Whole genome bisulphite sequencing (WGBS) data allowed to reveal the role of methylation at the interferon induced transmembrane protein 3 (*IFITM3*) gene in the pathogenesis of kidney diseases (Rackham et al., 2017). Compared to WGS and WES, TS is unarguably more cost-efficient and reduces sequencing and bioinformatic data processing time by 50-fold and over (Gonzalez-Garay, 2014).

RNA-seq is increasingly frequently used for quantitative measurement of all expressed genes (transcriptome), which includes identification of new SNV transcripts, deletions, insertions, merging genes, small non-coding RNAs (snRNAs) — transfer RNA (tRNA), ribosomal RNA (rRNA), microRNA (miRNA), small interfering RNA (siRNA), and Piwi-RNA (piRNA) in normal and pathological tissues (Table 2) (O'Brien, Hayde, Zayed, and Peng, 2018). In addition, RNA-seq does not require transcript-specific probes (Ronaghi, Uhlen and Nyren, 1998). Due to its vital role in gene regulation, microRNA has been proved to be associated with the development of many diseases and morbidities (O'Brien, Hayde, Zayed, and Peng, 2018). For example, microRNAs are implicated in aging and longevity (Kinser and Pincus, 2020).

Chromatin immunoprecipitation sequencing (ChIP-seq) is used to identify histone modifications and transcription factor binding sites which help regulate gene expression (Table 2) (Rabbani et al., 2016; Lightbody et al., 2019). The technique involves sequencing of the genomic DNA fragments that co-precipitate with a DNA-binding protein and are analyzed by single-end or paired-end sequencing to generate short-length overlapping DNA fragments (150 to 500 bp) (Nakato and Sakata, 2021). Numerous methods for ChIP-seq analysis are available, tailored to specific study aims, such as high resolution X-ChIP-Seq, ChIP-on-chip, DNA adenine methyltransferase identification (DamID), proximity ligation (ChIA-PET), proximity ligation-assisted

Table 2. Targeted NGS technologies by study aims

NGS technologies	Study aims	Working principle	Data size	Reference
WGS	Identification of genetic mutations and SNPs in coding and non-coding genome regions	DNA extraction, fragmentation, sequencing, data analysis, identification of relevant variants	90 GB	Corbett et al., 2020; Rabbani et al., 2016; Suwinski et al., 2019
WES	Identification of variants in protein-coding loci and genes (exome)	DNA extraction, fragmentation, target gene capture, sequencing, data analysis, variant annotation	5–6 GB	Rabbani et al., 2016; Schwarze, Buchanan, Taylor, and Wordsworth, 2018; Suwinski et al., 2019
TS	Screening for DNA variants affecting numerous genes	DNA extraction, fragmentation, target gene sequence capture, sequencing, data analysis, variant annotation	1–3 GB	Burgess, 2021; Rabbani et al., 2016; Suwinski et al., 2019
RNA-seq	Identification of gene expression profiles, new protein isoforms, detection of merging genes, SNPs, insertions, deletions, small non-coding RNAs (transcriptomes)	RNA and cDNA extraction, fragmentation, sequencing, data analysis, variant annotation	3–4 GB	Marco-Puche et al., 2019; Rabbani et al., 2016; Suwinski et al., 2019; Wang, Gerstein, and Snyder, 2009
ChIP-seq	Studying DNA-protein interactions, identification of histones and transcription factor binding sites	DNA fragmentation, binding beaded antibodies to target proteins, DNA purification, sequencing, identification of gene variants	1–2 GB	Nakato and Sakata, 2021; Rabbani et al., 2016; Suwinski et al., 2019
Bisulfite-seq	Identification of DNA methylation sites	DNA treatment with sodium bisulfite to determine methylation status at CpG dinucleotides.	1–2 GB	Olova et al., 2018; Suwinski et al., 2019

ChIP-seq (PLAC-seq), nano-ChIP-seq, ChIP-exo and ChIP-nexus, competition-ChIP, DNA/RNA immunoprecipitation sequencing (DRIP-seq), DRIVE-seq, high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (CLIP-seq or HITS-CLIP), covalent attachment of tags to capture histones (CETCh-seq), cleavage under targets and release using nuclease (CUT&RUN), cleavage under targets and tagmentation (CUT&Tag) (Orian et al., 2009; Adli and Bernstein, 2011; Lickwar, Mueller and Lieb, 2013; Hartonen et al., 2016; Juric et al., 2019; Sanz and Chédin, 2019; Capurso, Tang, and Ruan, 2020; Nakato and Sakata, 2021). Today ChIP-seq as such and ChIP-on-chip are the two most widely used ChIP-seq approaches. The latter involves chromatin immunoprecipitation and DNA hybridization with microarray chips, using DNA-protein complexes shearing by adding formaldehyde, DNA extraction, protein precipitation by antibodies, DNA-protein cross-link removal, and finally purification of protein-enriched DNA (Kaboord and Perr, 2008).

Bisulfite sequencing is used to detect DNA methylation sites and is based on sodium bisulfite reaction with cytosine in single-stranded DNA, whereby unmethylated cytosines are converted into uracil (Hayatsu, Wataya, Kai, and Iida, 1970). In contrast, methylated DNA cytosine within CpG dinucleotides remains unaffected after treatment of DNA with bisulfite.

The use of targeted NGS technologies (WGS, WES, RNA-seq, ChIP-seq) in clinical practice is hampered by high costs which include the cost of genomic data storage, transfer, processing, and bioanalytics, as well as the

cost of reagents, sequencing machines and instruments (Schmidt and Hildebrandt, 2017). For instance, depending on the number of target genes, the cost of panel sequencing in France ranges between 376 to 968 EUR (Marino et al., 2018). By analyzing the cost of ten sequencing studies in the USA, Tan and colleagues calculated the average panel sequencing cost of 1,609 USD per sample (488 to 3,443 USD) (Tan, Shrestha, Cunich, and Schofield, 2018). Clinical interpretation of sequencing results is another limitation. Sequencing data fail to comprehensively explain manifestations of multifactorial diseases, their underlying etiology and pathogenesis, which are largely affected by environmental factors and lifestyle, as well as genes (Said, Verweij, and Van Der Harst, 2018; Suwinski et al., 2019). Fig. 1 outlines the general practice in NGS data for annotating genetic variants associated with human diseases.

Applications of NGS Technologies

Research applications: human genome projects

Introduction of sequencing technologies paved the way towards human genome decoding. To achieve this goal, in 1990 the US National Institutes of Health launched the Human Genome Project (HGP), that brought together scientists from across the UK, Japan, France, Germany, Spain, China, as well as the USA. The project was finalized in 2003, when the US National Center for Biotechnology Information (NCBI) released the Build 35 (hg17) of the finished human genome assembly (International

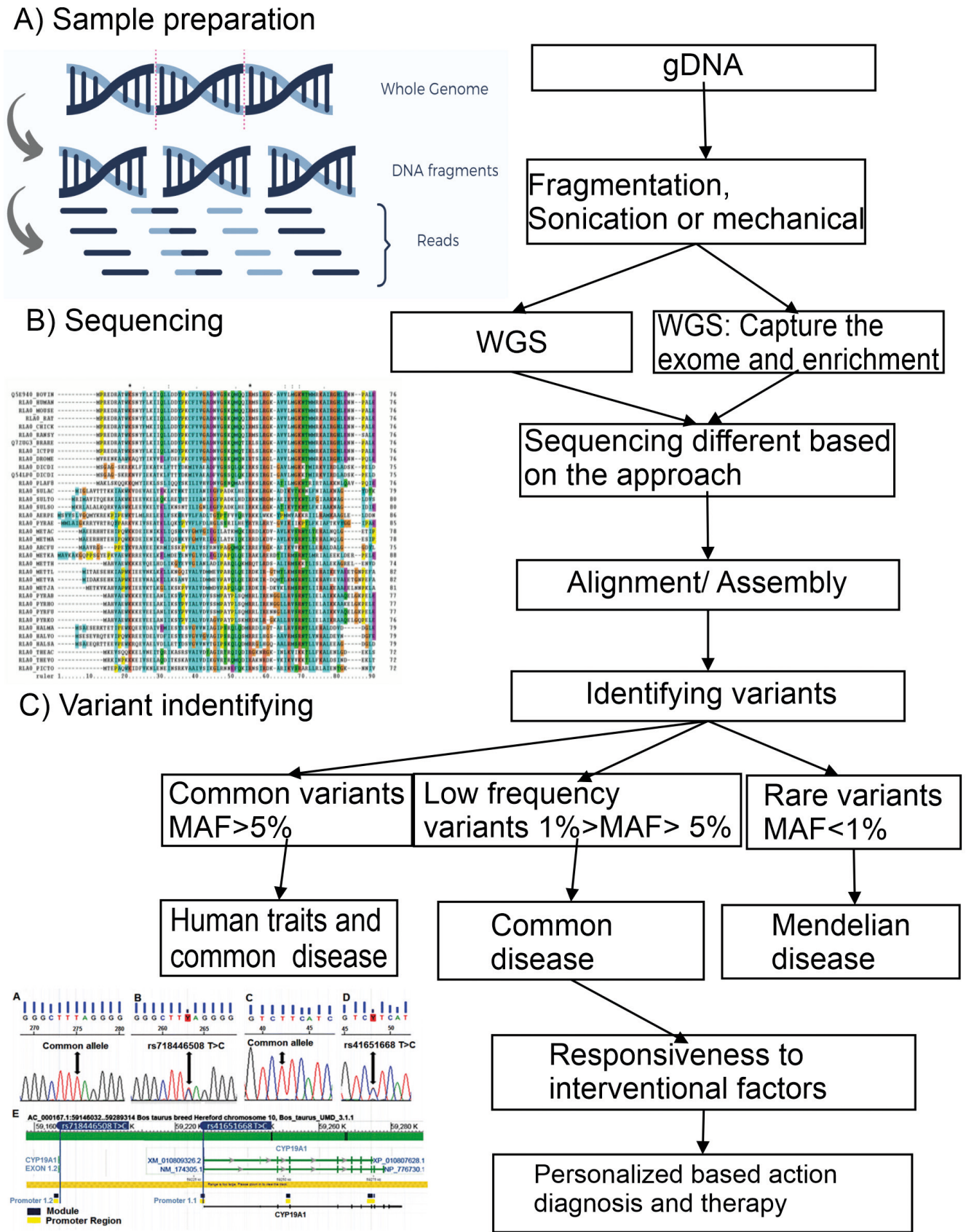


Fig. 1. NGS workflow steps to identify disease-causing genetic variants. WGS and WES techniques involve three main stages: A) sample preparation, B) sequencing, C) data analysis and variant identification (Rabbani et al., 2016).

Table 3. Human genome projects (Gonzalez-Garay, 2014)

Project name	License	Description	Reference
HapMap	Free access	HapMap project focuses on SNPs with a minor allele frequency of $\geq 5\%$	The International Hapmap Consortium, 2007
1000 Genomes	Free access	1000 Genome project captured up to 98% of the SNPs with a minor allele frequency of $\geq 1\%$ in 1092 individuals from 14 populations	Stankov, Benc, Draskovic, 2013.
The NHLBI (MD, USA) Exome Sequencing	Free access	The project is directed to discover protein-coding genes responsible for heart, lung and blood disorders. It analyses the allele frequency of each SNP.	NHLBI. Exome Sequencing Project...
The Personal Genome	Free access	The Personal Genome Project has the genomes of 174 individuals and the exomes of over 400 volunteers.	Personal Genome Project
NextCode Health	Commercial	The project has 40 million validated variants collected from the genotype of 140,000 volunteers from Iceland.	NextCode Health
CHARGE consortia	Free access	1000 whole exome data sets of well-phenotyped individuals from the CHARGE consortium	DNAexus. CHARGE project

Human Genome..., 2004). This release, however, contained numerous genomic gaps that were addressed and replaced later. On May 26, 2021 the Genome Reference Consortium (GRC) released the latest build GRCh38.p13 (Genome Reference Consortium).

Following the release of the human build, the next step was to develop a map of human genetic variation or a haplotype map (Haplotype Map or 'HapMap' project) by genotyping 270 samples from four populations with diverse geographic ancestry, see Table 3. In 2007 the Phase II HapMap was released, which characterizes 3.1 million human SNPs (International HapMap Consortium, 2007). A follow-up of the HapMap project was the project to study sequencing data from 1000 human genomes (1000 Genomes Project), see Table 3 (1000 Genomes Project Consortium, 2012). This project enabled the identification of 38 million SNV/SNPs, 1.4 million biallelic insertions or deletions (indels), and 14,000 large genomic deletions in 1092 individuals from 14 ethnicities. SNV detection provided a stepping stone for pathogenic variant annotation and a study of SNV-disease associations. HapMap and 1000 Genomes data was utilized to develop the research technique of genome-wide association studies (GWAS). GWAS allowed to study population frequencies for numerous SNVs/SNPs, associated with longevity (Deelen et al., 2019) or multifactorial diseases, such as type 1 diabetes mellitus (T1DM) (Stankov, Benc and Draskovic, 2013), type 2 diabetes mellitus (T2DM) (Sladek et al., 2007) and breast cancer (Fanale et al., 2012). The understanding of disease-specific inheritance patterns, implicated pathogenic mechanisms and SNV/SNP population frequencies enables pathogenic variant annotation (Rabbani et al., 2012). Moreover, HapMap and 1000 Genomes projects accelerated the detection of rare highly penetrant SNPs, which are assumed to cause rare monogenic human diseases (Freund et al., 2018).

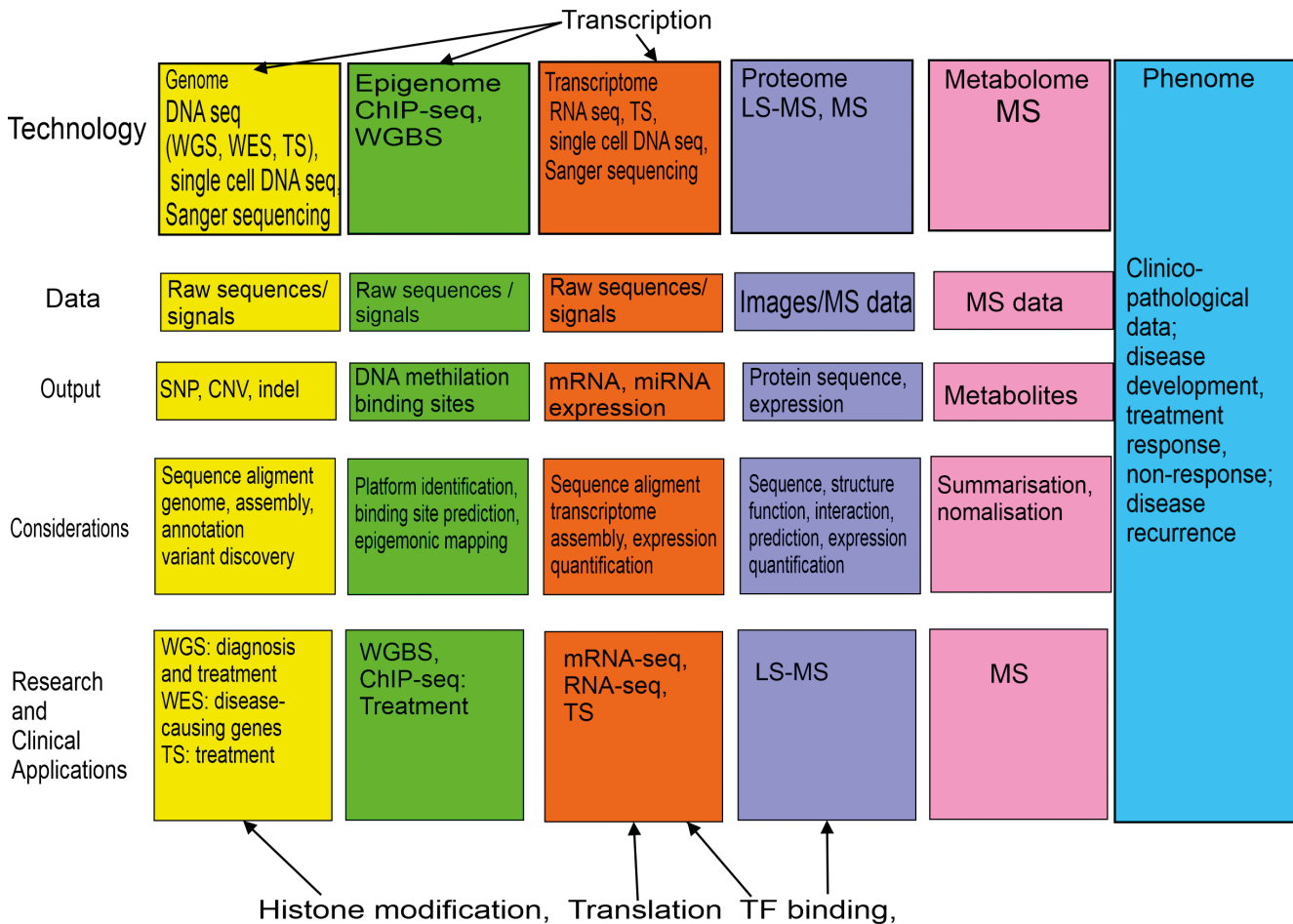
In addition to the above-mentioned projects, annotating causative variants for multifactorial and rare monogenic diseases is the main focus of projects on exome sequencing (The NHLBI (MD, USA) Exome Sequencing, CHARGE consortia), personal genome sequencing (Personal Genome Project), or genotyping of variants from Icelandic volunteers (NextCode Health), see Table 3.

Clinical applications: biomarkers in diagnostics, prevention and treatment

BIOMARKERS

The identified SNVs, along with disease-associated genes and coding proteins can be utilized as biomarkers for capturing pathogenic conditions. The National Institutes of Health USA in 1998 proposed the following definition of a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (Atkinson, Colburn, and Degruotola, 2001). Advances in omics technologies enables the use of biomarkers in combination with HTS approaches to detect all health-related variations at different levels of genetic data transfer — genome, epigenome, transcriptome, proteome and metabolome, see Fig. 2 (Zierer et al., 2015).

As already mentioned, SNPs, insertions, deletions, duplications, copy number variations (CNV) and fusion genes are the valuable biomarkers at the level of genome (Haraksingh and Snyder, 2013). Anywhere from 6 to 19% of each chromosome exhibit CNVs defined as a different number of DNA fragments copies (one kilobase or larger), compared to the reference genome (Feuk, Carson, and Scherer, 2006; Zarrei, MacDonald, Merico, and Scherer, 2015). By 2015 a total of 353,126 to 3,024,212 CNVs were validated (The Database of Ge-



MEDICAL GENETICS

Fig. 2. Characterization of biomarkers with respect to HTS technologies and their applications at different levels of genetic data transfer (Lightbody et al., 2019).

omic Variants). CNVs are implicated in the development of human diseases, their morbidity or drug treatment efficacy (Rabbani et al., 2016).

Among the promising complex biomarkers which are actively used primarily in oncology, we should mention the mutational burden (tumor mutational burden, TMB), as well as the genetic signatures of specific mutational processes, in particular, microsatellite instability (MSI) (Thibodeau, Bren, and Schaid, 1993; Morganti et al., 2020) — hypermutability of short tandem DNA repeats due to biallelic inactivation of one of the DNA mismatch repair genes (MMR — genes *MLH1*, *MSH2*, *MSH6* and *PMS2*), and deficiency of homologous recombination (HRD) (Moynahan, Chiu, Koller, and Jasin, 1999) — accumulation of specific long insertions and deletions due to biallelic inactivation of one of the genes for specific repair of DNA double-strand breaks (HR — genes *BRCA1* and *BRIP1*, *BRCA2* and *PALB2*, *RAD51C* and *RAD51D*, as well as genes of the FANC family). For example, SNPs in the vascular adhesion molecule (*VCAM1*) gene, ADP-ribosylation factor gene and guanine nucleotide exchange factor 2 (*ARFGEF2*), or a deletion in the cystic fibrosis (*CTFR*) gene, or CNV

in the human leukocyte antigen (*HLA*) locus are associated with sickle cell anemia (Dworkis et al., 2011), cystic fibrosis (White et al., 1990) and rheumatoid arthritis (Wellcome Trust Case Control Consortium, 2010).

Epigenetic biomarkers include DNA methylation status, histone modification, changes in DNA-protein interactions (Margulies et al., 2005; Lightbody et al., 2019). Such modifications alter (i.e. promote or inhibit) gene expression and are associated with predisposition or resistance to diseases and treatment (Bohacek and Mansuy, 2013).

Transcriptomic biomarkers include different RNA types — mRNA, tRNA, microRNA, piwi-RNA, siRNA, rRNA; proteomic biomarkers are referred to disease-associated proteins and peptides; metabolomic biomarkers imply amino-, keto, fatty and gallic acids, as well as amines, lipids, vitamins, anions, cations, microelements and blood soluble gases (carbon dioxide, nitrogen oxide), macroergic compounds (ATP, creatine phosphate), organic alcohols, etc. For example, long non-coding RNAs can be referred to aging-associated markers (Jin et al., 2019). Other potential biomarkers, speculated to present an aging-like phenotype, include nicotinamide adenine dinucleotide (NAD), α -ketoglutarate,

β -hydroxybutyrate, reduced nicotinamide adenine dinucleotide phosphate (NADP⁺), as well as mammalian target of rapamycin (mTOR) and AMP-activated protein kinase (AMPK) signaling pathways (Sharma and Ramanathan, 2020). Gene panels to predict preeclampsia in the third trimester pregnancy include fatty acid esters of hydroxy fatty acids (C18:0), lysophosphatidylethanolamine (C20:0), phosphatidylcholine (C19:0), sphingomyelins (SM C28:1, SM C30: 1) (Lee et al., 2020).

In addition, depending on their clinical application, all biomarkers break down into the following groups: diagnostic markers utilized to detect diseases; prognostic markers, providing data on disease outcomes, morbidity, treatment response or allowing to evaluate screening test data (Ronaghi, Uhlen and Nyren, 1998; Rabbani, Tekin, and Mahdieh, 2014).

DISEASE DIAGNOSTICS

Among other tasks, genetics aims to detect human genes and gene variants, associated with the disease or physiological conditions of interest (i.e. pregnancy or aging). As a powerful tool to detect genes and SNVs at early disease stages and thus contribute to healthcare costs reduction, WES and WGS guarantee early diagnostics, prevention, identification of mutation carrier state (genetic predisposition), associated with cardio-vascular, neurologic, metabolic (diabetes mellitus), or monogenic disorders, caused by low-frequency (1–5%) and rare (<1%) variants, alongside with cancer and physiological conditions (i.e. growth, aging) and parameters (i.e. BP, BMI) (Gonzalez-Garay, 2014; Rabbani et al., 2016). In addition, identification of disease-associated (condition-associated) genes and genetic variants, implicates a disease-specific pathogenic mechanism. In 2009 the team led by Jay Shendure produced the pioneer evidence of NGS capability to detect genetic aberrations. The researchers utilized WES approach to discover Miller Fischer syndrome, which is a rare recessive disorder within the spectrum of acute (immune-mediated) demyelinating polyneuropathies (AIDP) (Ng et al., 2009; Ng et al., 2010). Reproductive health represents another area for WGS and WES diagnostic application, offering preimplantation testing, prenatal testing for Down, Edwards, Patau, Turner and other aneuploidy syndromes, presymptomatic testing, fetal gender detection and Rhesus genotyping (Guy et al., 2019).

SEQUENCING-BASED DIAGNOSTIC AND SCREENING TESTS

The combination of NGS technologies and cell-free fetal DNA detection in maternal blood (Lo et al., 1997) enabled newly designed non-invasive prenatal testing (NIPT) for trisomy and other chromosomal abnormali-

ties in fetal tissue. For example, Tatyana Ivashchenko and colleagues (2019) from Ott Research Institute of Obstetrics, Gynecology and Reproductology (St. Petersburg, Russia) utilized mass parallel sequencing (MPS) and cell-free fetal DNA detection to detect aneuploidy in 21 fetal samples per total 149 screened pregnancies with 100% sensitivity and 99.9% specificity. The study identified Down syndrome in 12 samples, Edwards syndrome in 5 samples and Patau syndrome in two samples; in addition, a combination of trisomies of chromosomes 13 and 21 and trisomy X were detected in one sample each (Ivashchenko et al., 2019).

In 2016 the FDA published the guidelines on ‘Use of standards in FDA regulatory oversight of Next Generation Sequencing (NGS) used for diagnosing germline diseases’ (US Food and Drug Administration).

Diagnostic NGS-based testing is also used for other multifactorial diseases. Oleg Glotov and colleagues from Ott Research Institute of Obstetrics, Gynecology and Reproductology screened genetic variants using WES in a panel of 35 genes causative of maturity onset diabetes of the young (MODY) and permanent neonatal diabetes. A total of 38 genetic variants were identified in 33 out of 60 unrelated Russian children who developed diabetes before the age of 18 years. Of 33 patients, 81.8% had variants in MODY-related genes: glucokinase (*GCK*), transcription factor *HNF1A*, paired box protein *PAX4*, ATP binding cassette subfamily C member 8 (*ABCC8*), potassium channel subfamily J member 11 (*KCNJ11*), *GCK + HNF1A*, *GCK + proto-oncogene BLK* and *GCK + BLK + WFS1*. In other patients genetic variants causative of non-MODY monogenic diabetes were found. Those included GATA-binding protein 6 (*GATA6*), wolframin transmembrane glycoprotein of the endoplasmic reticulum (*WFS1*), eukaryotic translation initiation factor 2 alpha kinase 3 (*EIF2AK3*) and solute carrier family 19 member 2 protein (*SLC19A2*). Overall, the researchers detected 15 novel genetic variants in *GCK*, *HNF1A*, *BLK*, *WFS1*, *EIF2AK3* and *SLC19A2* to demonstrate a wide spectrum of genetic variants, causative of the non-T1DM in the studied patients (Glotov et al., 2019). Based on the NGS detection of CCTG and TG repeats in the cellular nucleic acid-binding protein (*CNBP*) gene, Variantyx Inc. (Massachusetts, USA) company developed a genetic test, performed using Illumina TruSeq Genomic Unity™ platform to identify diabetes-associated variants (Variantyx Inc. Genomic Unity Genetic Test). The Genetic Testing Registry (GTR®) web-site shows an NGS panel of 56 gene variants of diabetes and obesity to foster more powerful diagnostic modalities (The Genetic Testing Registry).

As vivid examples of the use of several methods to confirm the diagnosis, we present the following cases from our clinical practice.

To confirm the clinical diagnosis of anauketic dysplasia, an extremely rare form of autosomal recessive

skeletal chondrodysplasia, in a 6-year-old girl, we performed molecular genetic testing using NGS. A 6-year-old girl with severe short stature was admitted to the department of the Turner Institute for examination and treatment. During the dynamic examination (radiography of the hip joint at 6 months, 2 years, 4 years, 4.5 years, 6 years), a progressive deformity of the femur was revealed, which consists in a formed bend of the proximal part with a secondary inclination in the region of the main trochanter over the dysplastic epiphysis of the femur (MRI data). Data on biochemical blood parameters: serum and urinary oligosaccharides, mucopolysaccharides, serum lactate, pyruvate, creatine phosphokinase, alkaline phosphatase, calcium, phosphorus and vitamin D metabolism are normal. Karyotype 46, XX. The level of ACTH and STH is normal. Based on the data obtained, a diagnosis of anauketic dysplasia was suggested. To confirm the clinical diagnosis, as well as the prognosis of health and offspring, molecular genetic testing was carried out using high-throughput sequencing. During the analysis of sequencing data, we identified a variant of the nucleotide sequence (chr9:35657924–35657925delCTinsGC; rs387906533) in the first exon of the endoribonuclease mitochondrial RNase (RMRP) gene in the heterozygous state, leading to the replacement of two nucleotides n.91_92delinsGC. Due to the absence of the second mutation according to the new generation sequencing data, it was decided to perform direct automatic sequencing of the RMRP gene in the proband. The study confirmed the presence of the n.91_92delinsGC mutation in the heterozygous state. In addition, the n.-6_-5insTCTCAGCTTCAC (chr9:g.35658020–35658021insTCTCAGCTTCAC) substitution, previously not described in the literature, was identified in the promoter region of the gene. This substitution is an insertion of 12 nucleotides into the region between the TATA box and the start of transcription. The analysis of the RMRP gene in the parents of the proband was carried out by direct automatic sequencing. It was found that the n.-6_-5insTCTCAGCTTCAC mutation is of paternal origin, while the n.91_92delinsGC mutation is of maternal origin. As a result of the study, a new feature of the pathogenesis and course of the disease was revealed. It is important to note that MCMC-AD spectrum patients in the Russian population may differ from those abroad, which necessitates further studies of this pathology, including the use of the entire arsenal of molecular genetic methods, including NGS.

Another example demonstrates direct clinical utility of NGS-based approach in timely diagnostics of glutaric aciduria, type I (GA1, OMIM # 231670). The case involved a 1-year old child with retardation of motor development exacerbated by mental slowdown. At the first geneticist appointment both tandem mass spectrometry (TMS) and NGS examination were prescribed.

TMS had fast turnaround times, but its result was only suspecting GA1 with lowered free carnitine and increased glutarylcarnitine. Recommended neuroradiologic methods (CT, NMR) had blurred findings. And only after performing NGS the diagnosis became clear. The child was found having two rare pathogenic variants in gene *GCDH* (chr19: g.13002734AC>A (c.219del; p.Tyr74fs); rs1057516521) and chr19: g.13006883G>A (c.583G>A; p.Ala195Thr). The first variant (c.219del) encodes truncated version of the protein glutaryl-CoA dehydrogenase and was previously described in a patient with glutaric aciduria, control set frequency is quite low (0.0003978%), no homozygous cases detected. The second revealed variant (c.583G>A) in compound heterozygous state was previously found and described elsewhere as a pathogenic one in patients with glutaric aciduria and was absent in control sets of 1000Genomes, ESP6500, ExAC, gnomAD. Alanine substitution with threonine occurred in conservative position thus predicting to impair *GCDH* protein normal function. Both parents were carriers of each variant. The final result of this conundrum case was instant initiation of a specialized nutrition therapy with “Nutrigen 40 –trp, -lys” (Infaprim Ltd, Russia) which led to gradual improvement in both physical and mental development of the child observed by a geneticist and a pediatrician in charge.

To summarize, effective detection of causative gene variants is bound to promote novel treatment strategies for these pathogenic conditions.

DISEASE PREDICTION

Identification of genetic variants allows to predict the risk of developing a disease in asymptomatic individuals and ensure early treatment initiation in order to postpone the disease progression (Rabbani, Tekin, and Mahdiah, 2014). A most common example refers to NGS-based tests, recognized in the neonatal period in the USA and used for genetic diagnostics and prediction of Cornelia de Lange syndrome, Rubinstein-Taybi syndrome, CHARGE, Holt-Oram syndrome, Kabuki syndrome, Stickler syndrome, Zellweger syndrome, Alagille syndrome, Noonan syndrome (RASopathies), tuberous sclerosis, osteogenesis imperfecta, congenital adrenal hyperplasia due to 21-hydroxylase deficiency, phenylketonuria, galactosemia, cystic fibrosis, PURA-related disorder, neonatal diabetes mellitus, familial hyperinsulinism, epileptic encephalopathy, congenital central hypoventilation syndrome (Lalani, 2017). Recently, an international team used a genetic artificial intelligence model based on the evaluation of 5050 microscopic images of blastocysts on the 5th day after in vitro fertilization to predict the ploidy status (euploidy/aneuploidy) of human embryos. The endpoint was ploidy status (euploid or aneuploid) based on the results of preimplanta-

tion genetic testing for aneuploidy. Predictive accuracy was determined by calculating sensitivity (correct euploid prediction), specificity (correct aneuploid prediction), and overall accuracy. The sensitivity of euploidy prediction was 74.6%. The researchers observed a positive correlation between artificial intelligence scoring and the percentage of euploid embryos, with high scoring embryos (9.0–10.0) being 2 times more likely to be euploid than low scoring embryos (0.0–2.4). When using the genetic artificial intelligence model to distribute embryos in a group, the probability that the highest rated embryo will be euploid or one of the two best rated embryos will be euploid was 82.4 and 97.0%, respectively. The demographic data of patients, images of embryos on the 6th day of cultivation also fit well into the artificial intelligence model (Diakiw et al., 2022).

PERSONALIZED TREATMENT

Genetic alterations are among the most critical determinants of individual drug responses. Sequencing technologies provide a powerful tool to identify novel candidate genes and relevant SNPs affecting drug metabolism (pharmacokinetics) and mechanism of action (pharmacodynamics), which allows to evaluate drug doses, drug resistance, treatment responses and toxic effects (Rabbani, Tekin, and Mahdieh, 2014). Conversely, drugs also affect numerous aspects of enzymatic activities (proteome), metabolic pathways (metabolome) and gene expression (transcriptome), leading to multiple protein isoforms transcription (ion channels, enzymes, receptors, cytokines, growth factors, hormones), each of them demonstrating isoform-specific cellular activity (Bick and Dimmock, 2011). Therefore, gene expression profiling should be included in drug design to evaluate the combined effect from multiple genes and their variants (pharmacogenomics), specific for every single drug, disease and individual, representing a particular population or ethnicity. This potentiality is allowed by disease-specific personalized treatment protocols aiming to identify multi-gene biomarker and SNP panels by using WGS, WES, or RNA-seq techniques in whole-genome (exome, transcriptome, proteome) studies (Rabbani et al., 2016).

For example, the review by Mannino and colleagues (2019) evaluates the currently available pharmacogenetic evidences to identify 64 genes and 200 genetic variants, associated with response to the most common anti-diabetic drugs in T2DM patients: metformin, dipeptidyl peptidase 4 (DPP-4) inhibitor, glucagon like peptide 1 receptor (GLP-1R) agonists, thiazolidinediones and sulfonylureas/meglitinides. Metformin response is associated with the members of the organic cation transporter family *ATM* and *SLC2A2* loci; sulfonylurea response is determined by *CYP2C9*, *TCF7L2*, *ABCC8*, *KCNJ11* and *IRS1* genes; thiazolidinediones interact with *PPARG* lo-

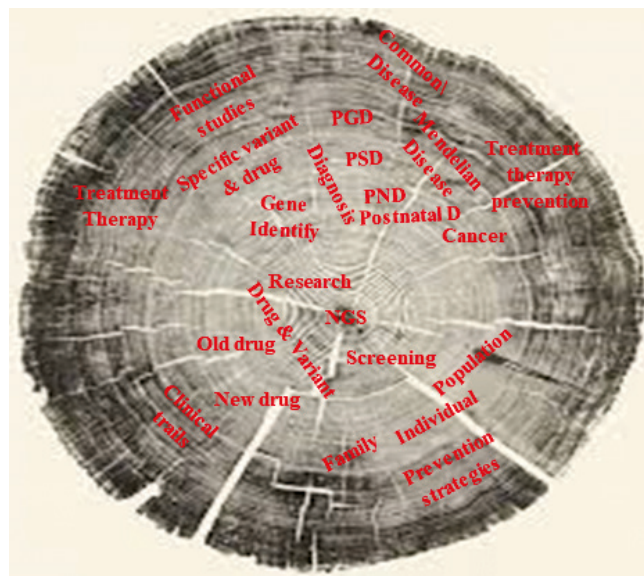


Fig. 3. The scheme shows how NGS leads to new advances in personalized medicine. Half apple symbolizes health management, including prevention and prediction. Totally, application of NGS approaches aims at better therapy, which is the goal of personalized medicine (consider the circles turning around) (Rabbani et al., 2016). Postnatal D — Postnatal Diagnosis, PND — Prenatal Diagnosis, PGD — Preimplantation Genetic Diagnosis.

cus; while DPP-4 inhibitors / GLP-1R agonists response affects *GLP1R* gene (Mannino, Andreozzi and Sesti, 2019).

Drug therapy selection based on patient-specific genetic variations becomes a great opportunity for physicians to multiply drug efficacy and safety (minimizing adverse drug reactions and toxic effects) in particular individuals (Malsagova et al., 2020). Fig. 3 shows how NGS technologies lead to new advances in personalized medicine.

Conclusion

Advances in sequencing technologies are driven by rapid technological progress, associated with accelerated throughput, whole-genome (exome) coverage of genes and gene variations, increased data size, improved accessibility and decreased real cost of sequencing. All of these were enabled by second, third and fourth generation sequencing technologies. These developments accelerated the discovery of new genes and gene variations, as well as the arrival of GWAS studies, large-scale global genetic projects, DNA banking and databases, leading to the identification of pathogenic and disease-associated gene variants, as well as new insights into the causes and mechanisms implicated in numerous congenital disorders, including monogenic and rare diseases. Sequencing technologies facilitated gene panel and screening test design, providing a powerful tool for early diagnostics, disease prevention, prediction and personalized treat-

ment. Such a huge leap forward translated into a complete rethinking of healthcare towards personalized medicine. These transformations are to improve health outcomes, sustain better health and increase life expectancy in a population through early prevention of severe socially significant diseases, early treatment initiation and the development of more effective target drugs.

References

- 1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., Kang, H. M., Marth, G. T., and McVean, G. A. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56–65. <https://doi.org/10.1038/nature11632>
- Adli, M. and Bernstein, B. E. 2011. Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nature Protocols* 6(10):1656–1668. <https://doi.org/10.1038/nprot.2011.402>
- AllSeq. WGS vs. WES. <http://allseq.com/kb/wgsvswes>
- Ardui, S., Ameer, A., Vermeesch, J. R., and Hestand, M. S. 2018. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Research* 46(5):2159–2168. <https://doi.org/10.1093/nar/gky066>
- Atkinson, A. J., Colburn, W. A., and Degruittola, V. G. 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics* 69(3):89–95. <https://doi.org/10.1067/mcp.2001.113989>
- Bick, D. and Dimmock, D. 2011. Whole exome and whole genome sequencing. *Current Opinion in Pediatrics* 23(6):594–600. <https://doi.org/10.1097/MOP.0b013e32834b20ec>
- Bohacek, J. and Mansuy, I. M. 2013. Epigenetic inheritance of disease and disease risk. *Neuropsychopharmacology* 38(1):220–236. <https://doi.org/10.1038/npp.2012.110>
- Branton, D., Deamer, D., and Marziali, A. 2008. The potential and challenges of nanopore sequencing. *Nature Biotechnology* 26(10):1146–1153. <https://doi.org/10.1038/nbt.1495>
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S. R., Moon, K., Burcham, T., Pallas, M., DuBridge, R. B., Kirchner, J., Fearon, K., Mao, J., and Corcoran, K. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* 18(6):630–634. <https://doi.org/10.1038/76469>
- Buermans, H. P. and den Dunnen, J. T. 2014. Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA) — Molecular Basis of Disease* 1842(10):1932–1941. <https://doi.org/10.1016/j.bbdis.2014.06.015>
- Burgess, D. J. 2021. Complex targeted sequencing in real time. *Nature Reviews Genetics* 22(2):67. <https://doi.org/10.1038/s41576-020-00324-6>
- Capurso, D., Tang, Z., and Ruan, Y. 2020. Methods for comparative ChIA-PET and Hi-C data analysis. *Methods* 170:69–74. <https://doi.org/10.1016/j.ymeth.2019.09.019>
- Chong, J. X., Buckingham, K. J., Jhangiani, S. N., Boehm, C., Sobreira, N., Smith, J. D., Harrell, T. M., McMillin, M. J., Wiszniewski, W., Gambin, T., Coban Akdemir, Z. H., Doheny, K., Scott, A. F., Avramopoulos, D., Chakravarti, A., Hoover-Fong, J., Mathews, D., Witmer, P. D., Ling, H., Hetrick, K., Watkins, L., Patterson, K. E., Reinier, F., Blue, E., Muzny, D., Kircher, M., Bilguvar, K., López-Giráldez, F., Sutton, V. R., Tabor, H. K., Leal, S. M., Gunel, M., Mane, S., Gibbs, R. A., Boerwinkle, E., Hamosh, A., Shendure, J., Lupski, J. R., Lifton, R. P., Valle, D., Nickerson, D. A., and Bamshad, M. J. 2015. The genetic basis of mendelian phenotypes: discoveries, challenges, and opportunities. *The American Journal of Human Genetics* 97(2):199–215. <https://doi.org/10.1016/j.ajhg.2015.06.009>
- Corbett, R. D., Eveleigh, R., Whitney, J., Barai, N., Bourgey, M., Chuah, E. J., Johnson, J., Moore, R. A., Moradin, N., Mungall, K. L., Pereira, S., Reuter, M. S., Thiruvahindrapuram, B., Wintle, R. F., Ragoussis, J., Strug, L. J., Herbrick, J. A., Aziz, N., Jones, S. J. M., Lathrop, M., Scherer, S. W., Staffa, A., and Mungall, A. J. 2020. A distributed whole genome sequencing benchmark study. *Frontiers in Genetics* 11:612515. <https://doi.org/10.3389/fgene.2020.612515>
- Deelen, J., Evans, D. S., Arking, D. E., Tesi, N., Nygaard, M., Liu, X., Wojczynski, M. K., Biggs, M. L., van der Spek, A., Atzmon, G., Ware, E. B., Sarnowski, C., Smith, A. V., Sepälä, I., Cordell, H. J., Dose, J., Amin, N., Arnold, A. M., Ayers, K. L., Barzilai, N., Becker, E. J., Beekman, M., Blanché, H., Christensen, K., Christiansen, L., Collerton, J. C., Cubaynes, S., Cummings, S. R., Davies, K., Debrabant, B., Deleuze, J. F., Duncan, R., Faul, J. D., Franceschi, C., Galan, P., Gudnason, V., Harris, T. B., Huisman, M., Hurme, M. A., Jagger, C., Jansen, I., Jylhä, M., Kähönen, M., Karasik, D., Kardia, S. L. R., Kingston, A., Kirkwood, T. B. L., Launer, L. J., Lehtimäki, T., Lieb, W., Lyttikäinen, L. P., Martin-Ruiz, C., Min, J., Nebel, A., Newman, A. B., Nie, C., Nohr, E. A., Orwoll, E. S., Perls, T. T., Province, M. A., Psaty, B. M., Raitakari, O. T., Reinders, M. J. T., Robine, J. M., Rotter, J. I., Sebastiani, P., Smith, J., Sørensen, T. I. A., Taylor, K. D., Uitterlinden, A. G., van der Flier, W., van der Lee, S. J., van Duijn, C. M., van Heemst, D., Vaupel, J. W., Weir, D., Ye, K., Zeng, Y., Zheng, W., Holstege, H., Kiel, D. P., Lunetta, K. L., Slagboom, P. E., and Murabito, J. M. 2019. A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nature Communications* 10(1):3669. <https://doi.org/10.1038/s41467-019-11558-2>
- Di Muccio, G., Rossini, A. E., Di Marino, D., Zollo, G., and Chinnappi, M. 2019. Insights into protein sequencing with an α -Hemolysin nanopore by atomistic simulations. *Scientific Reports* 9(1):6440. <https://doi.org/10.1038/s41598-019-42867-7>
- Diakiw, S. M., Hall, J. M. M., VerMilyea, M. D., Amin, J., Aizpurua, J., Giardini, L., Briones, Y. G., Lim, A. Y. X., Dacka, M. A., Nguyen, T. V., Perugini, D., and Perugini, M. 2022. Development of an artificial intelligence model for predicting the likelihood of human embryo euploidy based on blastocyst images from multiple imaging systems during IVF. *Human Reproduction* 37(8):1746–1759. <https://doi.org/10.1093/humrep/deac131>
- DNAexus. CHARGE project use case. <https://dnanexus.com/usecases-charge>
- Drmanac, R., Sparks, A. B., Callow, M. J., Halpern, A. L., Burns, N. L., Kermani, B. G., Carnevali, P., Nazarenko, I., Nilsen, G. B., Yeung, G., Dahl, F., Fernandez, A., Staker, B., Pant, K. P., Baccash, J., Borcharding, A. P., Brownley, A., Cedeno, R., Chen, L., Chernikoff, D., Cheung, A., Chirita, R., Curson, B., Ebert, J. C., Hacker, C. R., Hartlage, R., Hauser, B., Huang, S., Jiang, Y., Karpinchyk, V., Koenig, M., Kong, C., Landers, T., Le, C., Liu, J., McBride, C. E., Morenzoni, M., Morey, R. E., Mutch, K., Perazich, H., Perry, K., Peters, B. A., Peterson, J., Pethiyagoda, C. L., Pothuraju, K., Richter, C., Rosenbaum, A. M., Roy, S., Shafto, J.,

- Sharanovich, U., Shannon, K. W., Sheppy, C. G., Sun, M., Thakuria, J. V., Tran, A., Vu, D., Zaranek, A. W., Wu, X., Drmanac, S., Oliphant, A. R., Banyai, W. C., Martin, B., Ballinger, D. G., Church, G. M., and Reid, C. A. 2010. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 327(5961):78–81. <https://doi.org/10.1126/science.1181498>
- Dworkis, D. A., Klings, E. S., Solovieff, N., Li, G., Milton, J. N., Hartley, S. W., Melista, E., Parente, J., Sebastiani, P., Steinberg, M. H., and Baldwin, C. T. 2011. Severe sickle cell anemia is associated with increased plasma levels of TNF-R1 and VCAM-1. *American Journal of Hematology* 86(2):220–223. <https://doi.org/10.1002/ajh.21928>
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korfach, J., and Turner, S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323(5910):133–138. <https://doi.org/10.1126/science.1162986>
- English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D. M., Reid, J. G., Worley, K. C., and Gibbs, R. A. 2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* 7(11):e47768. <https://doi.org/10.1371/journal.pone.0047768>
- Fanale, D., Amodeo, V., Corsini, L. R., Rizzo, S., Bazan, V., and Russo, A. 2012. Breast cancer genome-wide association studies: there is strength in numbers. *Oncogene* 31(17):2121–2128. <https://doi.org/10.1038/onc.2011.408>
- Feuk, L., Carson, A. R., and Scherer, S. W. 2006. Structural variation in the human genome. *Nature Reviews Genetics* 7(2):85–97. <https://doi.org/10.1038/nrg1767>
- Fikes, B. 2017. New machines can sequence human genome in one hour, Illumina announces. *The San Diego Union-Tribune*. <https://www.sandiegouniontribune.com/business/biotech/sd-me-illumina-novaseq-20170109-story.html>
- Freund, M. K., Burch, K. S., Shi, H., Mancuso, N., Kichaev, G., Garske, K. M., Pan, D. Z., Miao, Z., Mohlke, K. L., Laakso, M., Pajukanta, P., Pasaniuc, B., and Arboleda, V. A. 2018. Phenotype-specific enrichment of mendelian disorder genes near GWAS regions across 62 complex traits. *The American Journal of Human Genetics* 103(4):535–552. <https://doi.org/10.1016/j.ajhg.2018.08.017>
- Genome Reference Consortium. Human Genome Overview. <https://www.ncbi.nlm.nih.gov/grc/human>
- Glotov, O. S., Serebryakova, E. A., Turkunova, M. E., Efimova, O. A., Glotov, A. S., Barbitoff, Y. A., Nasykhova, Y. A., Predeus, A. V., Polev, D. E., Fedyakov, M. A., Polyakova, I. V., Ivashchenko, T. E., Shved, N. Y., Shabanova, E. S., Tiselko, A. V., Romanova, O. V., Sarana, A. M., Pendiina, A. A., Scherbak, S. G., Musina, E. V., Petrovskaja-Kaminskaia, A. V., Lonishin, L. R., Ditkovskaya, L. V., Zhelenina, L. A., Tyrtova, L. V., Berseneva O. S., Skitchenko, R. K., Suspitsin, E. N., Bashnina, E. B., and Baranov, V. S. 2019. Whole-exome sequencing in Russian children with non-type 1 diabetes mellitus reveals a wide spectrum of genetic variants in MODY-related and unrelated genes. *Molecular Medicine Reports* 20(6):4905–4914. <https://doi.org/10.3892/mmr.2019.10751>
- Gonzalez-Garay, M. L. 2014. The road from next-generation sequencing to personalized medicine. *Personalized Medicine* 11(5):523–544. <https://doi.org/10.2217/pme.14.34>
- Goodwin, S., McPherson, J. D., and McCombie, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* 17(6):333–351. <https://doi.org/10.1038/nrg.2016.49>
- Gorski, M. M., Blighe, K., Lotta, L. A., Pappalardo, E., Gargiola, I., Mancini, I., Mancuso, M. E., Fasulo, M. R., Santagostino, E., and Peyvandi, F. 2016. Whole-exome sequencing to identify genetic risk variants underlying inhibitor development in severe hemophilia A patients. *Blood* 127(23):2924–2933. <https://doi.org/10.1182/blood-2015-12-685735>
- Goto, Y., Akahori, R., Yanagi, I., and Takeda, K. I. 2020. Solid-state nanopores towards single-molecule DNA sequencing. *Journal of Human Genetics* 65(1):69–77. <https://doi.org/10.1038/s10038-019-0655-8>
- Guo, Y., Dai, Y., Yu, H., Zhao, S., Samuels, D. C., and Shyr, Y. 2017. Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics* 109(2):83–90. <https://doi.org/10.1016/j.ygeno.2017.01.005>
- Guy, C., Haji-Sheikhi, F., Rowland, C. M., Anderson, B., Owen, R., Lacbawan, F. L., and Alagia, D. P. 2019. Prenatal cell-free DNA screening for fetal aneuploidy in pregnant women at average or high risk: Results from a large US clinical laboratory. *Molecular Genetics & Genomic Medicine* 7(3):e545. <https://doi.org/10.1002/mgg3.545>
- Haraksingh, R. R. and Snyder, M. P. 2013. Impacts of variation in the human genome on gene regulation. *Journal of Molecular Biology* 425(21):3970–3977. <https://doi.org/10.1016/j.jmb.2013.07.015>
- Hartonen, T., Sahu, B., Dave, K., Kivioja, T., and Taipale, J. 2016. PeakXus: comprehensive transcription factor binding site discovery from ChIP-Nexus and ChIP-Exo experiments. *Bioinformatics* 32(17):i629–i638. <https://doi.org/10.1093/bioinformatics/btw448>
- Hayatsu, H., Wataya, Y., Kai, K., and Iida, S. 1970. Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry* 9(14):2858–2865. <https://doi.org/10.1021/bi00816a016>
- Heather, J. M. and Chain, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics* 107(1):1–8. <https://doi.org/10.1016/j.ygeno.2015.11.003>
- Herper, M. 2017. Illumina promises to sequence human genome for \$100 — but not quite yet. *Forbes*. <https://www.forbes.com/sites/matthewherper/2017/01/09/illumina-promises-to-sequence-human-genome-for-100-but-not-quite-yet/#58262050386d>
- Hodkinson, B. P. and Grice, E. A. 2015. Next-Generation Sequencing: A review of technologies and tools for wound microbiome research. *Advances in Wound Care* 4(1):50–58. <https://doi.org/10.1089/wound.2014.0542>
- Hofmann, A. L., Behr, J., Singer, J., Kuipers, J., Beisel, C., Schraml, P., Moch, H., and Beerewinkel, N. 2017. Detailed simulation of cancer exome sequencing data reveals differences and common limitations of variant callers. *BMC Bioinformatics* 18(1):8. <https://doi.org/10.1186/s12859-016-1417-7>
- Huang, J., Liang, X., Xuan, Y., Geng, C., Li, Y., Lu, H., Qu, S., Mei, X., Chen, H., Yu, T., Sun, N., Rao, J., Wang, J., Zhang, W., Chen, Y., Liao, S., Jiang, H., Liu, X., Yang, Z., Mu, F., and Gao, S. 2017. A reference human genome dataset of the BGISEQ-500 sequencer. *Gigascience* 6(5):1–9. <https://doi.org/10.1093/gigascience/gix024>
- International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the hu-

- man genome. *Nature* 431(7011):931–945. <https://doi.org/10.1038/nature03001>
- International HapMap Consortium, Frazer, K. A., Ballinger D. G., Cox D. R. et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 449(7164):851–861. <https://doi.org/10.1038/nature06258>
- Ivashchenko, T. E., Vashukova, E. S., Kozyulina, P. Y., et al. 2019. The first experience of using NGS for NIPT. *Genetics* 55(10):1151–1157. (In Russian)
- Jin, L., Song, Q., Zhang, W., Geng, B., and Cai, J. 2019. Roles of long noncoding RNAs in aging and aging complications. *Biochimica et Biophysica Acta (BBA) — Molecular Basis of Disease* 1865(7):1763–1771. <https://doi.org/10.1016/j.bbadis.2018.09.021>
- Juric, I., Yu, M., Abnoui, A., Raviram, R., Fang, R., Zhao, Y., Zhang, Y., Qiu, Y., Yang, Y., Li, Y., Ren, B., and Hu, M. 2019. MAPS: Model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments. *PLOS Computational Biology* 15(4):e1006982. <https://doi.org/10.1371/journal.pcbi.1006982>
- Kaboord, B. and Perr, M. 2008. Isolation of proteins and protein complexes by immunoprecipitation. *2D PAGE: Sample Preparation and Fractionation. Methods in Molecular Biology* 424:349–364. https://doi.org/10.1007/978-1-60327-064-9_27
- Kchouk, M., Gibrat, J. F., and Elloumi, M. 2017. Generations of sequencing technologies: from first to next generation. *Biology and Medicine* 9:395.
- Kinsler, H. E. and Pincus, Z. 2020. MicroRNAs as modulators of longevity and the aging process. *Human Genetics* 139(3):291–308. <https://doi.org/10.1007/s00439-019-02046-0>
- Lalani, S. R. 2017. Current genetic testing tools in neonatal medicine. *Pediatrics and Neonatology* 58(2):111–121. <https://doi.org/10.1016/j.pedneo.2016.07.002>
- Lee, S. M., Kang, Y., Lee, E. M., Jung, Y. M., Hong, S., Park, S. J., Norwitz, E. R., Lee, D. Y., and Park, J. S. 2020. Metabolomic biomarkers in midtrimester maternal plasma can accurately predict the development of preeclampsia. *Scientific Reports* 10(1):16142. <https://doi.org/10.1038/s41598-020-72852-4>
- Lickwar, C. R., Mueller, F., and Lieb, J. D. 2013. Genome-wide measurement of protein-DNA binding dynamics using competition ChIP. *Nature Protocols* 8(7):1337–1353. <https://doi.org/10.1038/nprot.2013.077>
- Lightbody, G., Haberland, V., Browne, F., Taggart, L., Zheng, H., Parkes, E., and Blayney, J. K. 2019. Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. *Briefings in Bioinformatics* 20(5):1795–1811. <https://doi.org/10.1093/bib/bby051>
- Lionel, A. C., Costain, G., Monfared, N., Walker, S., Reuter, M. S., Hosseini, S. M., Thiruvahindrapuram, B., Merico, D., Jobling, R., Nalpathamkalam, T., Pellecchia, G., Sung, W. W. L., Wang, Z., Bikangaga, P., Boelman, C., Carter, M. T., Cordeiro, D., Cytrynbaum, C., Dell, S. D., Dhir, P., Dowling, J. J., Heon, E., Hewson, S., Hiraki, L., Inbar-Feigenberg, M., Klatt, R., Kronick, J., Laxer, R. M., Licht, C., MacDonald, H., Mercimek-Andrews, S., Mendoza-Londono, R., Piscione, T., Schneider, R., Schulze, A., Silverman, E., Siriwardena, K., Carter Snead, O., Sondheimer, N., Sutherland, J., Vincent, A., Wasserman, J. D., Weksberg, R., Shuman, C., Carew, C., Szego, M. J., Hayeems, R. Z., Basran, R., Stavropoulos, D. J., Ray, P. N., Bowdin, S., Meyn, M. S., Cohn, R. D., Scherer, S. W., and Marshall, C. R. 2018. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genetics in Medicine* 20(4):435–443. <https://doi.org/10.1038/gim.2017.119>
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. 2012. Comparison of Next-Generation Sequencing systems. *BioMed Research International* 2012:251364. <https://doi.org/10.1155/2012/251364>
- Lo, Y. M., Corbetta, N., Chamberlain, P. F., Rai, V., Sargent, I. L., Redman, C. W., and Wainscoat, J. S. 1997. Presence of fetal DNA in maternal plasma and serum. *Lancet* 350(9076):485–487. [https://doi.org/10.1016/S0140-6736\(97\)02174-0](https://doi.org/10.1016/S0140-6736(97)02174-0)
- Majewski, J., Schwartzenruber, J., Lalonde, E., Montpetit, A., and Jabado, N. 2011. What can exome sequencing do for you? *Journal of Medical Genetics* 48(9):580–589. <https://doi.org/10.1136/jmedgenet-2011-100223>
- Malsagova, K. A., Butkova, T. V., Kopylov, A. T., Izotov, A. A., Potoldykova, N. V., Enikeev, D. V., Grigoryan, V., Tarasov, A., Stepanov, A. A., and Kaysheva, A. L. 2020. Pharmacogenetic testing: A tool for personalized drug therapy optimization. *Pharmaceutics* 12(12):1240. <https://doi.org/10.3390/pharmaceutics12121240>
- Mannino, G. C., Andreozzi, F., and Sesti, G. 2019. Pharmacogenetics of type 2 diabetes mellitus, the route toward tailored medicine. *Diabetes/Metabolism Research and Reviews* 35(3):e3109. <https://doi.org/10.1002/dmrr.3109>
- Marco-Puche, G., Lois, S., Benítez, J., and Trivino, J. C. 2019. RNA-Seq perspectives to improve clinical diagnosis. *Frontiers in Genetics* 10:1152. <https://doi.org/10.3389/fgene.2019.01152>
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J.-B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., and Rothberg, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376–380. <https://doi.org/10.1038/nature03959>
- Marino, P., Touzani, R., Perrier, L., Rouleau, E., Kossi, D. S., Zhamin, Z., Charrier, N., Goardon, N., Preudhomme, C., Durand-Zaleski, I., Borget, I., and Baffert, S. 2018. Cost of cancer diagnosis using next-generation sequencing targeted gene panels in routine practice: a nationwide French study. *European Journal of Human Genetics* 26(3):314–323. <https://doi.org/10.1038/s41431-017-0081-3>
- Maxam, A. M. and Gilbert, W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences USA* 74(2):560–564. <https://doi.org/10.1073/pnas.74.2.560>
- Morganti, S., Tarantino, P., Ferraro, E., D'Amico, P., Viale, G., Trapani, D., Duso, B. A., and Curigliano, G. 2020. Role of Next-Generation Sequencing technologies in personalized medicine, NGS implementation in clinical practice. *P5 eHealth: An Agenda for the Health Technologies of the Future. Challenges and Limitations* 125–154. https://doi.org/10.1007/978-3-030-27994-3_8
- Moynahan, M. E., Chiu, J. W., Koller, B. H., and Jasin, M. 1999. Brca1 controls homology-directed DNA repair. *Molecular Cell* 4(4):511–518. [https://doi.org/10.1016/s1097-2765\(00\)80202-6](https://doi.org/10.1016/s1097-2765(00)80202-6)

- Nakato, R. and Sakata, T. 2021. Methods for ChIP-seq analysis: A practical workflow and advanced applications. *Methods* 187:44–53. <https://doi.org/10.1016/j.ymeth.2020.03.005>
- Newswire, P. 2016. Precision medicine market size to exceed \$87 billion by 2023: Global market insights Inc. <https://www.prnewswire.com/news-releases/precision-medicine-market-size-to-exceed-87-billionby-2023-global-market-insights-inc-599454691.html>.
- NextCode Health. <https://www.nextcode.com>
- Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A., Shendure, J., and Bamshad, M. J. 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics* 42(1):30–35. <https://doi.org/10.1038/ng.499>
- Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., Bamshad, M., Nickerson, D. A., and Shendure, J. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461(7261):272–276. <https://doi.org/10.1038/nature08250>
- NHLBI. Exome Sequencing Project (ESP) Exome Variant Server. <https://evs.gs.washington.edu/EVS>
- O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. 2018. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Frontiers in Endocrinology* 9:402. <https://doi.org/10.3389/fendo.2018.00402>
- Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrens, R. V., Branco, M. R., and Reik, W. 2018. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biology* 19(1):33. <https://doi.org/10.1186/s13059-018-1408-2>
- Open Source Next Generation Sequencing Technology. The Polonator G007. <https://www.polonator.org>
- Orian, A., Abed, M., Kenyagin-Karsenti, D., and Boico, O. 2009. DamID: a methylation-based chromatin profiling approach. *Chromatin Immunoprecipitation Assays. Methods in Molecular Biology* 567:155–169. https://doi.org/10.1007/978-1-60327-414-2_11
- Osoegawa, K., Mammoser, A. G., Wu, C., Frengen, E., Zeng, C., Catanese, J. J., and de Jong, P. J. 2001. A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Research* 11(3):483–496. <https://doi.org/10.1101/gr.169601>
- PacBio Sequel systems. <https://www.pacb.com/products-and-services/sequel-system>
- Personal Genome Project. <https://www.personalgenomes.org>
- Pierson, T. M., Adams, D., Bonn, F., Paola Martinelli, P., Cherukuri, P. F., Teer, J. K., Hansen, N. F., Cruz, P., Mullikin, J. C., Blakesley, R. W., Golas, G., Kwan, J., Sandler, A., Fajardo, K. F., Markello, T., Tiffit, C., Blackstone, C., Rugarli, E. I., Langer, T., Gahl, W. A., and Toro, C. 2011. Whole-exome sequencing identifies homozygous AFG3L2 mutations in a spastic ataxia-neuropathy syndrome linked to mitochondrial m-AAA proteases. *PLoS Genetics* 7(10):e1002325. <https://doi.org/10.1371/journal.pgen.1002325>
- Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., and Gu, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:341. <https://doi.org/10.1186/1471-2164-13-341>
- Rabbani, B., Mahdieh, N., Hosomichi, K., Nakaoka, H., and Inoue, I. 2012. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *Journal of Human Genetics* 57(10):621–632. <https://doi.org/10.1038/jhg.2012.91>
- Rabbani, B., Nakaoka, H., Akhondzadeh, S., Tekin, M., and Mahdieh, N. 2016. Next generation sequencing: implications in personalized medicine and pharmacogenomics. *Molecular BioSystems* 12(6):1818–1830. <https://doi.org/10.1039/c6mb00115g>
- Rabbani, B., Tekin, M., and Mahdieh, N. 2014. The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics* 59(1):5–15. <https://doi.org/10.1038/jhg.2013.114>
- Rackham, O. J., Langley, S. R., Oates, T., Vradi, E., Harmston, N., Srivastava, P. K., Behmoaras, J., Dellaportas, P., Botto, L., and Petretto, E. 2017. A Bayesian approach for analysis of whole-genome bisulfite sequencing data identifies disease-associated changes in DNA methylation. *Genetics* 205(4):1443–1458. <https://doi.org/10.1534/genetics.116.195008>
- Ronaghi, M., Uhlen, M., and Nyren, P. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281(5375):363–365. <https://doi.org/10.1126/science.281.5375.363>
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A., Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J. A., Namsaraev, E., McKernan, K. J., Williams, A., Roth, G. T., and Bustillo, J. 2011. An integrated semiconductor device enabling nonoptical genome sequencing. *Nature* 475(7356):348–352. <https://doi.org/10.1038/nature10242>
- Said, A. M., Verweij, N., and Van Der Harst, P. 2018. Associations of combined genetic and lifestyle risks with incident cardiovascular disease and diabetes in the UK biobank study. *JAMA Cardiology* 3(8):693–702. <https://doi.org/10.1001/jamacardio.2018.1717>
- Sanger, F. and Coulson, A. R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* 94(3):441–448. [https://doi.org/10.1016/0022-2836\(75\)90213-2](https://doi.org/10.1016/0022-2836(75)90213-2)
- Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA* 74(12):5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Sanz, L. A. and Chédin, F. 2019. High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. *Nature Protocols* 14(6):1734–1755. <https://doi.org/10.1038/s41596-019-0159-1>
- Schmidt, B. and Hildebrandt, A. 2017. Next-generation sequencing: big data meets high performance computing. *Drug Discovery Today* 22(4):712–717. <https://doi.org/10.1016/j.drudis.2017.01.014>
- Schwarze, K., Buchanan, J., Taylor, J. C., and Wordsworth, S. 2018. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genetics in Medicine* 20(10):1122–1130. <https://doi.org/10.1038/gim.2017.247>
- Sharma, R. and Ramanathan, A. 2020. The aging metabolome — biomarkers to hub metabolites. *Proteomics* 20(5–6):e1800407. <https://doi.org/10.1002/pmic.201800407>
- Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., Wang, M. D., Zhang, K., Mitra, R. D., and Church, G. M. 2005. Accurate multiplex

- polony sequencing of an evolved bacterial genome. *Science* 309(5741):1728–1732. <https://doi.org/10.1126/science.1117389>
- Siqueira, J. F. Jr, Fouad, A. F., and Rôças, I. N. 2012. Pyrosequencing as a tool for better understanding of human microbiomes. *Journal of Oral Microbiology* 4:1. <https://doi.org/10.3402/jom.v4i0.10743>
- Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S., Balkau, B., Heude, B., Charpentier, G., Hudson, T. J., Montpetit, A., Pshzhetsky, A. V., Prentki, M., Posner, B. I., Balding, D. J., Meyre, D., Polychronakos, C., and Froguel, P. 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445(7130):881–885. <https://doi.org/10.1038/nature05616>
- Staden, R. 1979. A strategy of DNA sequencing employing computer program. *Nucleic Acids Research* 6(7):2601–2610. <https://doi.org/10.1093/nar/6.7.2601>
- Stankov, K., Benc, D., and Draskovic, D. 2013. Genetic and epigenetic factors in etiology of diabetes mellitus type 1. *Pediatrics* 132(6):1112–1122. <https://doi.org/10.1542/peds.2013-1652>
- Suwinski, P., Ong, C. K., Ling, M. H. T., Poh, Y. M., Khan, A. M., and Ong, H. S. 2019. Advancing personalized medicine through the application of whole exome sequencing and big data analytics. *Frontiers in Genetics* 10:49. <https://doi.org/10.3389/fgene.2019.00049>
- Tan, O., Shrestha, R., Cunich, M., and Schofield, D. J. 2018. Application of next-generation sequencing to improve cancer management: A review of the clinical effectiveness and cost effectiveness. *Clinical Genetics* 93(3):533–544. <https://doi.org/10.1111/cge.13199>
- The Database of Genomic Variants. <http://dgv.tcag.ca/dgv/app>
- The Genetic Testing Registry (GTR). <https://www.ncbi.nlm.nih.gov/gtr/tests/509392/indication>
- The International Hapmap Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449(7164):851–861. <https://doi.org/10.1038/nature06258>
- Thibodeau, S. N., Bren, G., and Schaid, D. 1993. Microsatellite instability in cancer of the proximal colon. *Science* 260(5109):816–819. <https://doi.org/10.1126/science.8484122>
- Thompson, J. F. and Steinmann, K. E. 2010. Single molecule sequencing with a HeliScope genetic analysis system. *Current Protocols in Molecular Biology* 92:7.10.1–7.10.14. <https://doi.org/10.1002/0471142727.mb0710s92>
- US Food and Drug Administration. Use of standards in FDA regulatory oversight of Next Generation Sequencing (NGS) — based In Vitro Diagnostics (IVDs) used for diagnosing germline diseases. US Food and Drug Administration, 2016. <https://www.fda.gov/media/99208/download>
- Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek J. A., Costa, G., McKernan, K., Sidow, A., Fire, A., and Johnson, S. M. 2008. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Research* 18(7):1051–1063. <https://doi.org/10.1101/gr.076463.108>
- Varietyx Inc. Genomic Unity Genetic Test. <https://www.varietyx.com/wp-content/uploads/2019/05/Genomic-Unity-Sample-Report-STR.pdf>
- Wang, Q., Shashikant, C. S., Jensen, M., Altman, N. S., and Girirajan, S. 2017. Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. *Scientific Reports* 7(1):885. <https://doi.org/10.1038/s41598-017-01005-x>
- Wang, Z., Gerstein, M., and Snyder, M. 2009. RNA-seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10(1):57–63. <https://doi.org/10.1038/nrg2484>
- Weber, J. L. and Myers, E. W. 1997. Human whole-genome shotgun sequencing. *Genome Research* 7(5):401–409. <https://doi.org/10.1101/gr.7.5.401>
- Welcome Trust Case Control Consortium. 2010. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464(7289):713–720. <https://doi.org/10.1038/nature08979>
- White, M. B., Amos, J., Hsu, J. M., Gerrard, B., Finn, P., and Dean, M. 1990. A frame-shift mutation in the cystic fibrosis gene. *Nature* 344(6267):665–667. <https://doi.org/10.1038/344665a0>
- Zarrei, M., MacDonald, J. R., Merico, D., and Scherer, S. W. 2015. A copy number variation map of the human genome. *Nature Reviews Genetics* 16(3):172–183. <https://doi.org/10.1038/nrg3871>
- Zierer, J., Menni, C., Kastenmüller, G., and Spector, T. D. 2015. Integration of 'omics' data in aging research: from biomarkers to systems biology. *Aging Cell* 14(6):933–944. <https://doi.org/10.1111/acel.12386>