

Review

Sequencing of Nucleic Acids: from the First Human Genome to Next Generation Sequencing in COVID-19 Pandemic

Borut Furlani,^{1,§} Katarina Kouter,^{2,§} Damjana Rozman^{3,*}
and Alja Videtič Paska^{2,*}

¹ University of Ljubljana, Faculty of Medicine, Vrazov trg 2, 1000 Ljubljana, Slovenia

² University of Ljubljana, Faculty of Medicine, Institute of Biochemistry and Molecular Genetics, Medical Centre for Molecular Biology, Vrazov trg 2, 1000 Ljubljana, Slovenia

³ University of Ljubljana, Faculty of Medicine, Institute of Biochemistry and Molecular Genetics, Centre for Functional Genomics and Bio-Chips, Zaloška 4, 1000 Ljubljana, Slovenia

§These two author equally contributed to this work.

** Corresponding author: E-mail: alja.videtic@mf.uni-lj.si; Tel. + 386 1 543 76 61; Fax. + 386 1 543 76 41, damjana.rozman@mf.uni-lj.si; Tel. + 386 1 543 75 91; Fax. + 386 1 543 76 41*

Received: 01-26-2021

Abstract

Despite being around for more than 40 years, DNA sequencing is regarded as young technology in clinical medicine. As sequencing is becoming cheaper, faster and more accurate, it is rapidly being incorporated into clinical laboratories. In 2003, the completion of the first human genome opened the door to personalized medicine. Ever since it has been expected for genomics to widely impact clinical care and public health. However, many years can pass for genomic discoveries to reflect back and benefit the patients. DNA sequencing represents a less biased approach to diagnostics. It is not only a diagnostic tool, but can also influence clinical management and therapy. As new technologies rapidly emerge it is important for researchers and health professionals to have basic knowledge about the capabilities and drawbacks of the existing sequencing methods, and their use in clinical setting and research. This review provides an overview of nucleic acid sequencing technologies from historical perspective and later focuses on clinical utilization of sequencing. Some of the most promising areas are presented with selected examples from Slovenian researchers.

Keywords: Clinical sequencing; DNA; nucleic acid; genomics; next generation sequencing; precision medicine

1. Introduction

More than forty years ago, two papers described the first method for determining the sequence of nucleotide bases in DNA^{1,2} and ever since, the development of new sequencing methods has been exponential. In 2003, the Human Genome Project (HGP) was finalized, presenting the complete version of the Human Genome.³ Once the reference human genome was established, scientists tried to explain disease mechanisms or susceptibility for certain diseases through population resequencing and determination of disease-causing genomic variants. Recently, DNA sequencing has been moving into the clinical setting to be

implemented in diagnostics and clinical management. Due to the rapid development of novel technologies, sequencing of nucleic acids will contribute to the discovery of the genomic, transcriptomic and epigenomic basis of unsolved diseases, improved diagnostics, and personalized therapies.

Advances in sequencing technologies have contributed to a significant reduction of the sequencing costs in the last 15 years (Figure 1). In 2004, the National Human Genome Research Institute started an initiative to reduce the whole genome sequence cost to US\$1000⁴ accelerating the development of cheaper and faster sequencing methods. A deviation of sequencing cost from 'Moore's law' oc-

curred around year 2008 which coincides with the transition from Sanger's sequencing of nucleic acids to the next generation sequencing technologies (NGS) resulting in a rapid fall of sequencing cost. Human genome sequencing for \$1000 was achieved a few years ago⁵ and with novel technologies we are quickly approaching a \$100 human genome. With today's enormous sequencing outputs and lower cost per base the biggest challenge remains meaningful interpretation and informative reporting of sequencing results.

This review provides an overview of nucleic acid sequencing technologies and examples of clinical utilization of sequencing, also from Slovenian researchers.

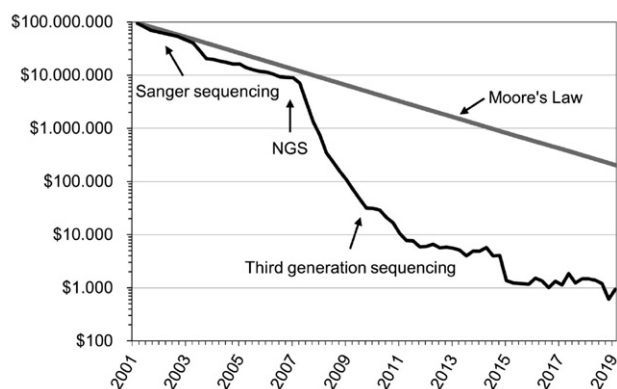


Figure 1: Sequencing cost for the human genome. The departure of sequencing cost curve from Moore's law coincides with the emergence of next generation sequencing (NGS). Moore's law originates in the computer hardware industry that involves doubling of 'computing power' every two years. It is considered that technologies that follow the law are regarded as successful.⁶ It thus represents a useful relationship to compare technology advances. Data shown in the figure was obtained from the National Human Genome Research Institute.⁶

2. Short Review of the Sequencing Technologies

In 1977, researchers developed two methods which enabled sequencing of nucleic acids of several hundred base pairs (bp), the Sanger's "dideoxy method" and the Maxam-Gilbert's method^{1,2}, causing revolution in biology.

2. 1. First-Generation DNA Sequencing

The *Sanger sequencing* is known today as the first-generation DNA sequencing and was based on the use of polymerase chain reaction. DNA polymerase is an enzyme that can elongate an existing DNA molecule by adding deoxynucleotides (dNTPs; a nucleobase, deoxyribose, and phosphate groups) to the DNA 3'-end *via* a phosphodiester bond between the 3' carbon atom hydroxy group of the DNA incorporated deoxynucleotide and the

5' carbon atom phosphate group of a joining dNTP. What made the sequencing possible is the addition of labelled dideoxynucleotides (ddNTP); dNTPs lacking the deoxyribose 3' carbon atom hydroxy group needed to form the phosphodiester bond. After the incorporation of a ddNTP, DNA polymerase could no longer add new nucleotides as the phosphodiester bonds can no longer be formed. This results in the production of DNA fragments that vary in length; however, all fragments end with a labelled ddNTP. Initially, sequencing was performed using radioactively labelled ddNTPs. About a decade later Sanger sequencing was automated and commercialized using fluorescent-labelled ddNTPs⁷ and capillary electrophoresis, providing a single base pair resolution.⁸ Using automated Sanger sequencing, researchers were able to read up to 75,000 bp per day.⁹ This method presented the foundation for the HGP, the biggest collaborative biological project that was officially completed in 2003 (it took 13 years and cost almost US\$3 billion to obtain the complete version of the human genome).^{3,6,10} The next step was to identify genomic differences/variants among people to explain disease mechanisms and/or disease susceptibility. Such projects required genomes of many individuals to be sequenced, however, Sanger sequencing was far too time-consuming and expensive.

2. 2. Next (or Second) Generation Sequencing (NGS)

The "hunt" began for alternative DNA sequencing methods, ultimately resulting in the emergence of *NGS technologies*. NGS is based on massive parallel sequencing, meaning that billions of short DNA fragments are sequenced simultaneously producing short sequence "reads".¹¹ Reads are computationally aligned to the reference sequence to assemble the consensus DNA sequence (Figure 2). NGS technologies significantly increased sequencing throughput, decreased labour, and sequencing cost.¹² In 2004, the company 454 Life Sciences released the first commercially available NGS platform¹³, which is no longer used today. 454 Life Sciences technology was based on pyrosequencing; using detection of light to determine the DNA sequence. The basic principle of pyrosequencing consists of using enzymes to build the complementary DNA strand and detect the base order of DNA strands mobilized to beads located on a titer plate. Each addition of dNTPs to the growing DNA strand, catalysed by DNA polymerase, results in the release of a pyrophosphate. Pyrophosphate is then catalysed to adenosine triphosphate by sulfate adenyltransferase. Enzyme luciferase later utilizes the adenosine triphosphate to generate light by converting luciferin to oxyluciferin. Only one of the four dNTPs is added at a time, and unused dNTPs are degraded by the enzyme apyrase before the addition of new dNTP. Intensity of the light is detected by sensors, indicating if and how many dNTP were added to the complementary DNA strand.¹⁴

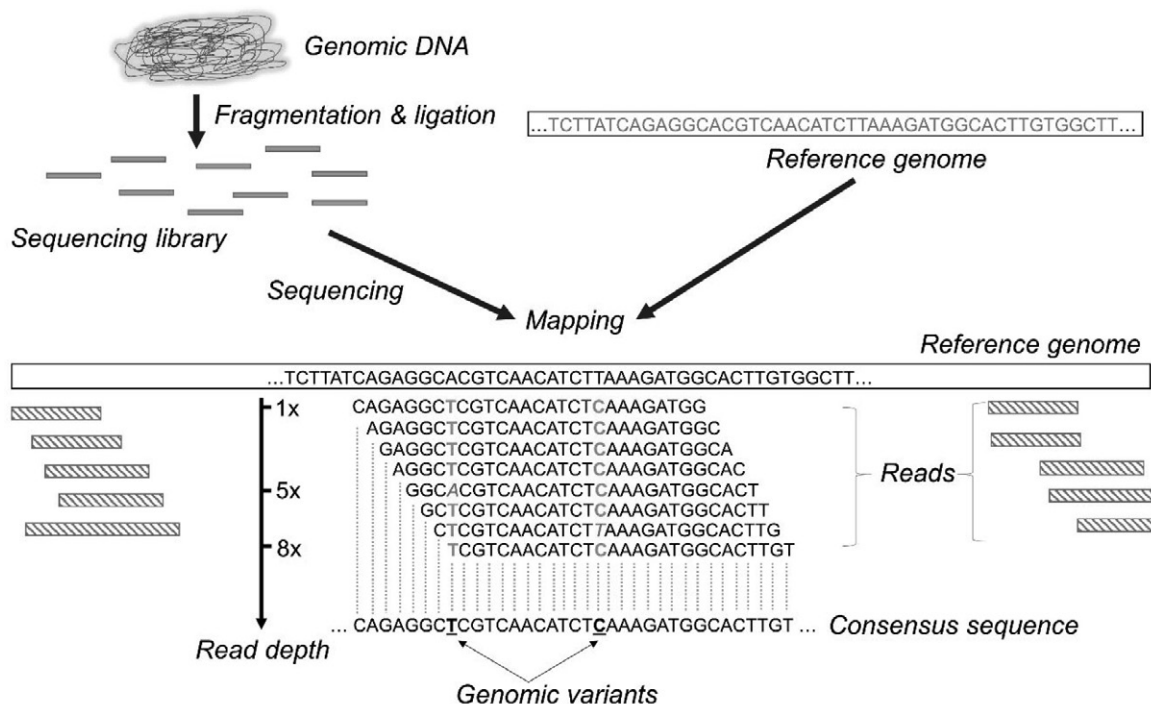


Figure 2: Schematic representation of NGS workflow. Steps common to different NGS platforms are DNA fragmentation (that is required or not depending on the library type), library preparation, massive parallel sequencing, bioinformatics analysis, and variant annotation and interpretation.²⁷ Sequencing generates billions of “reads” that are computationally aligned to the *reference genome* to assemble linear consensus sequence. The number of reads in which a particular base/variant appears is known as a read depth and determines the confidence with which a particular base/variant is called.¹² A particular single nucleotide variant should appear in more than 10 reads (meaning read depth to be at least 10 x) to be regarded as a genuine genomic variant.¹²

Illumina, Ion Torrent, and Beijing Genomics Institute (BGI) are currently the main NGS companies on the market.¹⁵ Today, over 90% of all sequencing in the laboratories around the world is performed on Illumina’s platforms.¹⁶ Illumina devices work on the principle of “sequencing by synthesis”. Sequencing takes place on the surface of the glass slide flow cell. The single-stranded DNA sequences of the DNA library to be sequenced bind by hybridization to oligonucleotides located at the surface of the flow cell. DNA polymerase completes the missing strand, resulting in a double-stranded DNA sequence, and then the source strand is removed. Only the newly formed DNA strand remains bound to the flow cell. Next, DNA strands are subjected to clonal amplification, denaturation and clustering. Antisense DNA sequences are removed from the surface of the flow cell. The next step is sequencing, which takes place simultaneously on all bound sequences. With each sequencing cycle, a single complementary dNTP is added to the bound sequence. A blocker is located at the 3’ of deoxyribose, thus allowing addition of only one dNTP per cycle. Various modes of use are available, with four channel chemistry, two channel chemistry and less often used but still available, one channel chemistry. For four channel chemistry based devices, each of the dNTPs in the mixture is labelled with a specific fluorescent dye (dATP with red, dGTP with blue, dTTP with green, and dCTP with yellow). After each multiplica-

tion cycle, the device determines the inserted dNTP using lasers and four filters to be able to distinguish all four possible bases. The number of multiplication cycles determines the length of the reading. Within an individual cluster, all sequences are identical; thus, in each multiplication cycle the whole cluster glows with the same colour. As for two channel chemistry, only two different fluorescent dyes are needed to label dNTPs; after laser excitation, the device detects dTTP as a green signal, dCTP as a red signal, dATP as a combination of the signal of both dyes, and dGTP is not marked and the device detects them as the absence of a signal. One channel chemistry, as the name suggests, uses a single dye to detect all four bases. Compared to glass slide flow cells, one channel chemistry uses CMOS (complementary metal-oxide semiconductor) chips. After each sequencing cycle, newly incorporated bases are detected using two chemistry steps and combination of two images. In the first chemistry steps, dATPs and dTTPs are labelled with the dye and the first image is taken. Next, during the second chemistry steps, added reagent removes the label from dATP (dATPs have a cleavable linker allowing the removal of the dye) and adds the dye to dCTPs (dCTPs have a linker group that allows dye binding). The second image is taken and the combination of both signals then determines, which of the four bases was incorporated at each sequencing cycle.¹⁷ Additionally, Illumina offers also patterned flow cells. Their

surface area is organized in billions of evenly spaced nanowells with fixed locations, and in each nanowell a distinct cluster is generated. Therefore it offers more efficient use of the flow cell surface area, increased data output, reduced costs, and faster run times. Due to the structured organization patterned flowcells provide significant advantages over non-patterned cluster generation. It is more tolerant to a broader range of library densities, due to nanowell positioning there is no need to map cluster sites, thus saving sequencing running time. With higher cluster density also data per flow cell are better usable, affecting the reduction of the cost per gigabase (Gb).¹⁸

Illumina offers numerous sequencing platforms for different applications such as MiSeq FGx system that is the first validated benchtop sequencer designed specifically for forensic science or MiSeqDx and NextSeq550Dx which are both Food and Drug Administration-approved platform for *in vitro* diagnostic testing.¹⁹ NGS has substantially increased sequencing output; the production-scale sequencing platform from Illumina NovaSeq 6000 System can read up to 6 tera-bases in 2 days and is ideal for ultra-deep sequencing of the entire genome.¹⁵

Compared to previously mentioned technologies, Ion Torrent detection of sequencing by synthesis is not based on optics and uses unlabelled dNTPs. As each new dNTP is being incorporated into a growing DNA strand, a pyrophosphate and a hydrogen ion are being released. Detection is therefore based on the change in pH by an ion-sensitive field-effect transistor sensor inside a CMOS layer.²⁰

BGI sequencing platforms enable nanoball sequencing, a mechanism that bypasses the requirement for PCR amplification during the library preparation. DNA to be sequenced is first fragmented until a desired length is achieved. Next, fragments are end repaired and specific adaptor sequences and split oligo sequences are added. This enables the single stranded DNA fragments to then be circularized—forming a single stranded circular DNA shape—and replicated many times using a modified rolling circle amplification using Phi 29 polymerase, until a long single-stranded DNA is formed. DNA then forms into a nanoball of a few hundred nanometres in diameter. DNA nanoballs are adhered onto the patterned array flow cell. Similar to Illumina sequencing, one dNTP is incorporated per cycle and the sequence order is determined by laser excitation. BGI released the DNBSEQ-T7 machine that can produce 1–6 tera-bases of high-quality data per day for a wide range of applications. Interestingly, BGI also offers whole genome sequencing for only \$600, including sample processing, sequencing, and data analysis.²¹ The high sequencing output and the rapid decrease of the sequencing cost made it possible for such platforms to be adopted by many clinical laboratories.^{22,23} Exponential genomic data generation has accelerated translational research and development of new genomic tests.²⁴ However, NGS has also a few pitfalls. The read length is important for the accuracy of the generated sequence; technologies

that utilize longer reads generally produce longer and high-quality assemblies.⁹ However, most of the NGS utilizes short reads (35–600 bp) due to the nature of sequencing chemistry.^{9,12} Many genomic regions contain repetitive sequences⁹ much longer than sequencing reads which may lead to misassemblies and sequencing gaps.^{16,25} Moreover, short-read technologies less accurately detect larger structural variations that are frequently clinically relevant.²⁶ Another disadvantage is that most of the NGS platforms use PCR in the amplification step (to increase signal strength), which tends to be less accurate in genomic areas that are high in guanine-cytosine content, which can result in errors during DNA “photocopying”.⁹

2. 3. Third Generation Sequencing

The main characteristic of third generation sequencing is utilization of long (10,000–2 million bp) sequencing reads.¹⁵ Long read lengths provide better resolution of repetitive genomic regions and structural variants²⁸ and allow the assembly of complex genomes.¹⁵ Further, third generation technologies do not require library amplification to increase signal strength and enable real-time sequencing.^{9,29} In 2011, Pacific Biosciences released technology named *Single-Molecule Real-Time (SMRT) sequencing*³⁰ and in 2014, *Nanopore sequencing* (Oxford Nanopore Technologies) was introduced.³¹ Compared to Illumina, SMRT technology uses a differently labelled dNTPs; as there is no blocker bound to the deoxyribose and the label is located on the phosphate group, dNTPs can be added sequentially without the additional step of terminating blocker removal, hence measuring it in real-time. Nanopore sequencing is based on the tiny changes in current. During sequencing, DNA strands pass through protein nanopores of about 1.8 nanometres in diameter, which are embedded in a polymer membrane. As DNA strands enter the pore, current changes according to the DNA bases that are located inside the pore (about 6 DNA bases at a time). Nanopore sequencing platforms are distinguished by great portability (MinION is in the size of a USB key³²), ultralong reads, and simple library preparation. Such devices can be used in virtually any environment; for example, to identify infectious disease outbreaks, as already demonstrated in several studies.^{33,34} Most recently, Nanopore sequencing has been used for the accurate and comprehensive detection of SARS-Cov-2 during COVID-19 pandemic.^{35–37} Third generation sequencing allows direct determination of epigenetic modifications³⁸ and base modifications in RNA sequencing.³⁹ The use of methylation profiling by Nanopore sequencing has already been reported by Euskirchen et al.⁴⁰ where the power of this technology for rapid tumour classification has been illustrated. However, one drawback of Nanopore sequencing is the higher error rate compared to short-read technologies.⁴¹ Despite many intriguing possibilities of long-read technologies, lower output and accuracy limit their entry into the clinical environment for the time being.¹⁵ Some shortcom-

ings of both second and third generation technologies can be compensated for by using them in combination – known as hybrid sequencing.^{42,43} Recent Nanopore sequencing platforms already offer an improved sequencing throughput; PromethION 48 can run up to 48 flow cells at once, producing up to 7,6 tera-bp output yield.⁴⁴ Third generation technologies represent a new revolution in genomics as they enable identification of yet undetermined or poorly determined genomic regions.^{45,46} For example, ultra-long-read nanopore sequencing enabled the complete resolution of human X-chromosome.⁴⁷ and SMRT enabled to resolve 2.25-kb-long stretches of short tandem repeats, implicated in Fragile X syndrome.⁴⁸ As the output and accuracy of third generation technologies further increase, they will likely play an important role in the clinical setting for the identification of important structural variants that are poorly determined by short-read technologies.

3. Genomic Sequencing in the Clinical Setting

The *precision medicine* is based on treating the patient as an individual, evolving current clinical practice to more individualized health care.⁴⁹ It seeks application of genomics as a major strategy in tailoring care to maximise health and minimize harm to patients. As the genome is the best source of information about what makes an individual unique (e.g. more/less susceptible to neurodegenerative diseases) the inclusion of genomic sequencing in the clinical practice accelerates the advancement of precision medicine.

3. 1. Sequencing Approaches

Sequencing presents a less biased approach to diagnostics and has a great potential for ending the diagnostics odyssey for patients with rare diseases.⁵⁰ Clinicians choose between whole-genome sequencing (WGS), whole-exome sequencing (WES), transcriptome sequencing, and target-panel sequencing. The most comprehensive, but also the most expensive, is WGS as it interrogates the entire genome. On average, it detects 3 million genomic variants^{11,51} most of which belong to non-coding regions of the genome making interpretation difficult.⁵⁰ An extensive bioinformatics analysis is thus required to narrow all genomic variants to only a few that might be related to the patient's phenotype. WGS poses problems of data interpretation and storing due to a large amounts of information.⁵² Its use has been studied in neonatal intensive care units in critically ill new-borns^{53,54} and oncology patients,⁵⁵ but is still mainly used in the scientific research. WES focuses on more manageable portion of the genome, known as exome that codes for proteins (only 1–2% of the genome). Variants from the exome are thus easier to interpret. It is also cheaper, faster and in most cases enough informative for

the clinical practice. One disadvantage, however, is lower accuracy in certain areas of genes that are relevant for some medical conditions; which can ultimately result in false-negative results.^{15,54,56} Another disadvantage of the WES is also the limited detection of clinically relevant copy number variations (CNVs) and structural variations. Currently there are no accepted standard protocols or quality control measures for CNVs identification in NGS data, and in many cases microarrays are used over WES.⁵⁷ WES achieves the diagnostic rate in the range of 25–35%, compared to WGS, which is in the range of 40–60%.^{12,58} Due to its better diagnostic yield, WGS is expected to become clinically more important as genomic variant interpretation improves and sequencing cost further decreases.⁵⁸ Today, most of the clinical sequencing utilizes target gene panels, which interrogate selected regions of the genome associated with a disease. Targeted sequencing is highly reliable in the identification of variants for disease-related genes, is cost effective, and sequencing results are easier to interpret compared to WES and WGS.¹⁵ There are numerous predesigned targeted gene sequencing panels available from Illumina, for example, AmpliSeq or TrueSeq.¹⁹ Based on numerous studies, small NGS panels focusing on a limited number of actionable genes are expected to become a standard diagnostic tool in oncology.^{59–61} A comparison of different sequencing methods can be found in Table 1.

An illustrative example of how the genomic sequencing contributed to the development of minimally-invasive diagnostics (that stems from oncology) is use of *liquid biopsy*, in which circulating-free DNA (cfDNA) from non-solid biological tissues, primarily blood, is analysed.⁶² Circulating tumor DNA (ctDNA) represents only a small fraction (<0.5%) of cfDNA⁶³ and is to some extent representative of the primary tumor DNA. Liquid biopsy is particularly suitable for tumors that are anatomically inaccessible to perform biopsy, in cases of metastatic and advanced stage cancers, and in minimize the number of recurrent biopsies (as a part of patient's follow up after diagnosis had already been made).⁵ CfDNA sequencing was shown to be useful also to monitor response to targeted therapy and to detect new resistance mutations, for example in epidermal growth factor receptor (EGFR) gene.⁶⁴ Due to its short half life ctDNA is appropriate for the assessment of tumor dynamics especially in more advanced disease stages.⁶² Liquid biopsy has also promising use in the population screening and early cancer diagnostics; high sensitivity and specificity were reported in early lung cancer diagnostics⁶⁵ and some mutations were detected 2 years before disease onset.⁶⁶ Besides, it also allows studying epigenetic modifications^{67,68} and methylation pattern (methylome) which can aid to disease classification. Mutations combined with epigenomic, proteomic, and even demographic data, would present unique tumor molecular profile from individual patient and present an important step forward in personalized medicine.

Table 1: Comparison of next and third generation sequencing platforms.

Sequencing technology	Platform	Data type	Runtime	Read length	Accuracy	Output range	Applications
Illumina	NextSeq 550	Paired end	12–30h	2 × 150bp	>99.9%	120 Gb	Targeted sequencing (amplicon-based, gene panel). Transcriptome sequencing (total RNA-Seq, mRNA-Seq, gene expression profiling). Arrays.
Oxford Nanopore Technologies	NovaSeq 6000	Paired end	13–44h	2 × 250bp (maximum)	>99.9%	80–6000 Gb	Whole-genome sequencing, read length) exome sequencing, whole transcriptome sequencing, methylation sequencing.
	MinION	Long reads	72h	Longest > 4 Mbp	87–98%	Up to 50 Gb per Flow cell	Whole-genome sequencing, targeted sequencing, RNA sequencing, epigenetics.
PacBio	Sequel II	High fidelity reads	30h	10–25kb	>99%	15–30 Gb per Flow cell	Whole genome sequencing, RNA sequencing, targeted sequencing, population seq, epigenetics.
BGI	MGISEQ-T7	Paired end	24h	150bp	Q30 >80%	1.5–6Tb	Whole-genome sequencing, deep exome sequencing, transcriptome sequencing, targeted panel sequencing.

Q30 designates probability of an incorrect base call 1 in 1000 times.

Nevertheless, the cfDNA sequencing is becoming an important tool also as non-invasive prenatal diagnostics. It has been proven as effective and safe screening method for trisomies 21, 18, and 13, and sex chromosome aneuploidies compared with traditional prenatal screening. The tests also proved to be fast and financially sustainable.⁶⁹

3. 2. Integration of Genomic Sequencing in the Clinical Practice

For a successful implementation in the clinical practice, sequencing must provide reliable results to the clinician. NGS requires end-to-end validation from DNA extraction to bioinformatic analysis to minimize the occurrence of false-positive and false-negative results.⁷⁰ Complete implementation also requires integration of genomic data into the *electronic health records* (EHRs), which can be quite challenging for smaller healthcare institutions.⁷¹ Successful implementation of personalized medicine will increasingly rely on EHRs to store vast amounts of genomic data and to appropriately integrate relevant genomic information into clinical care.⁷¹ One of the barriers for integrating sequencing results in EHRs is that they are frequently entered as a summary rather than raw data which limits data accession and reanalysis.⁷² Although genomic data is static, its interpretation is not. As new knowledge arises, interpretation can yield additional diagnoses, and data reanalysis was found to be a cost-effective approach.⁷³ Routine reanalyses were shown to improve diagnostic rates^{74–76} due to establishment of new

disease-gene associations, improved bioinformatics tools, and data sharing.⁷⁷ The integration of artificial intelligence (AI) is also promising, however, translating technical success in AI-driven analytics into meaningful clinical impact remains a challenge.⁷⁸ The proliferation of genomic sequencing also requires medical professionals who will be adept at understanding and returning genomic results to the patient. Thus, education for the next generation of health care providers is of great importance.⁷⁹ Another factor regarding implementation is the coverage of sequencing costs by health insurance companies, which unfortunately lags behind the advances in the sequencing technology.⁸⁰ The clinician must sometimes provide notes on how the genomic testing will affect the course of the disease or its management.⁸¹ Increasing cost coverage by the health insurance companies will be catalysed by a further drop in sequencing cost and accumulating evidence in studies of clinical usefulness.⁸¹ Several studies^{82–84} demonstrated high diagnostic yield and cost-effectiveness of genomic sequencing which was maximized by its early application in the diagnostic pathway.^{81,82}

3. 3. National Genomes Sequencing

To elucidate genetic background of a certain population, more and more countries are opting for studies of *national genomes*. The genomic data can be of great support to a healthcare system. If we are acquainted with common alleles in the healthy population, it is easier to identify disease related variants. National genome projects are

one of the fundamental elements in establishing effective identification and rapid diagnosis of rare diseases as they allow distinguishing between potentially causative genetic variants and rare, benign genetic variants that are unique for the original population. These rare genetic variants are the main source of false-positive results of genetic testing that can directly affect the clinical diagnosis and the course of treatment. Some projects are focused on rare diseases or cancer, whilst others have pursued population-based projects. In the UK, for example, the goal of 100,000 genomes was reached and served to establish the infrastructure needed for the integration of sequencing in the clinical setting.⁸⁵ The Chinese Academy of Sciences (CAS) launched the country's Precision Medicine Initiative with the goal of sequencing 1000 million human genomes by 2030.

In Europe, the 1+ Million European Genomes Project began in 2018, in which Slovenia also participates.⁸⁶ Slovenian Genome Project began at the end of 2019. This presents a pilot project which will outline the key directions of the future genomic projects, develop bioinformatic tools for data exchange, prepare the legal and ethical base, educate healthcare professionals and general public, and carry out a pilot genome sequencing project of Slovenian patients with rare diseases and healthy Slovenian population. In Slovenia, NGS methodology is already well established as its advances have been used in fields such as clinical neurology (dementia,⁸⁷ multiple sclerosis⁸⁸) paediatrics (metabolism of new-borns,⁸⁹ hearing loss⁹⁰), clinical oncology,⁹¹ clinical microbiology (microbiota associated with preterm birth,⁹² and age and gender⁹³), pharmacogenomics⁹⁴ and forensic medicine.⁹⁵

The use of genomic screening in preventive health is interesting, although it brings some concerns about its widespread implementation in routine clinical practice.^{96,97} Current estimates predict that 3–5% of people present a medically actionable variant.⁹⁸ This assumption is based on the DiscovEHR project, which involved over 50,000 adult participants aimed at connecting high-performance sequencing to an integrated health system. Among ~ 4.2 million rare single nucleotide variants, including insertions and deletions, adverse variants in 76 clinically relevant genes were found in about 3.5% of individuals. This study set a basis of individual tailored medicine that is based on therapeutic discoveries guided by genomics.⁹⁸ In the cases of “medically actionable genes”, genomic screening of asymptomatic population could have a significant public health impact.⁹⁹ Finding new population specific actionable genes that may save lives and/or majorly impact the quality of lives remains among the goals of the national genome sequencing projects.

3. 4. Sequencing During the COVID-19 Pandemics

Sequencing technologies can be used to detect and identify pathogens, determine their resistance to antibiot-

ics, construct phylogenetic trees, or epidemiologically track disease outbreaks.¹⁰⁰ Genome sequencing can typify microbial strains with greater accuracy compared to classical microbiological methods.¹⁰¹ As of April 2021, COVID-19 pandemic affected over 136 million people worldwide and resulted in death of almost 3 million people.¹⁰² On January 5, 2020, next-generation meta-transcriptomic sequencing allowed researchers to obtain the first and complete viral genome of SARS-CoV-2 from a patient in Wuhan, China.¹⁰³ Soon, several hundred genomes became publicly available (<https://www.gisaid.org/>), allowing rapid development of diagnostic tests¹⁰⁴ vaccines and antivirals¹⁰⁵, and disease tracking.¹⁰⁶ Few concepts of sequencing have been used for SARS-CoV-2.¹⁰⁷ Most studies have used the Illumina platform, however the Oxford Nanopore Technologies has been utilized for aforementioned shotgun metatranscriptomics¹⁰⁸ which enables de novo genome assembly without prior knowledge of the sequence.¹⁰⁹ Another method is amplicon-based sequencing, limited to specific parts of the viral genome. Such libraries can be sequenced on benchtop platforms with a mid-throughput (Illumina NexSeq, MiSeq, Ion torrent).¹⁰⁷ RNA sequencing using Oxford Nanopore Technologies and DNA nanoball sequencing was applied in a recent study³⁶ to reveal finished representation of SARS-CoV-2 transcriptome and epitranscriptome.

Comparative genomics revealed, that SARS-CoV-2 was a member of *Betacoronavirus* and fell into a subgenus *Sarbecovirus* that also includes Sars-Cov¹¹⁰ and allowed the “hunt” for its zoonotic origins.¹⁰³ Sequencing is vital in aspects of finding novel viral hosts to block interspecies transmission. Bat coronavirus, RaTG13, sampled in Yunnan province, is at the nucleotide level approximately 96% similar to SARS-CoV-2; however, there were major differences in key genomic features important for infectivity of the virus.^{103,111} Moreover, due to the ecological separation of humans and bats, it is probable that some other species acted as an intermediate host.¹⁰³ Recent research reports that viruses in Malayan pangolins are closely related to SARS-CoV-2¹¹² and these animals are of great interest because of involvement in illegal animal trafficking.¹⁰³ Pangolin-CoV was shown to be 91.02% identical to SARS-Cov-2 at the which makes him the second closest relative behind RaTG13.¹¹³ It seems that betacoronaviruses exist in the number of mammalian species and it is thus imperative to perform a wider sampling of animals from wet markets and those who live close to human populations to block potential interspecies transmission.¹⁰³ Furthermore, NGS allows tracking of the viral strains. RNA viruses continuously accumulate mutations.¹¹⁴ New mutations in SARS-CoV-2 genome will continuously arise over time and space and result in branching of the original “reference genome”. NGS allows us to observe evolutionary pattern of SARS-CoV-2, which is crucial for efficient disease prevention and control, for example, to reveal new routes of infection. Tracking mutations and variable regions of the viral is thus

imperative for development of effective therapy in the eyes of viral diversity and consequent drug resistance.¹¹⁵ As of April 2021, more than 1 million coronavirus-related genome sequences have been uploaded to EpiFluTM (GISAID) world-wide and the number is rapidly increasing. NGS is used to trace interpersonal transmission of the virus.¹¹⁶ High resolution genomic epidemiology is thereby becoming an effective tool for public health surveillance and disease control.¹¹⁷ Knowing SARS-COV2 genome also helps to achieve a more effective disease strategy, to investigate cases with unclear sources of infection within a short turnaround time.^{110, 118}

Furthermore, NGS technology was applied to investigate mechanisms of SARS-CoV-2 infection.¹¹⁶ For example, RNA sequencing has been used to determine susceptible organs with higher expression of angiotensin-converting enzyme 2 (ACE2) receptor, which serves as a receptor for SARS-CoV-2.¹¹⁹ NGS and single-cell RNA sequencing were used to determine expression of ACE2 receptor in numerous organs and cells after infection with SARS-CoV-2¹¹⁶, which will most definitely benefit diagnostic and therapeutic target identification.

4. Future Outlook

Researches tend to sequence as many human genomes as possible, known as population-scale resequencing, capturing not just genomic but also epigenomic data. Genomics and epigenomics are frequently studied separately but to fully reach the potential of precision medicine it will be necessary to study them together. Only integrated data from various big “omics” will enable us to fully understand disease mechanisms and substantially increase the sensitivity of genomic sequencing. One of the obstacles, however, is the storage and processing of large amounts of data, which is why the parallel development of bioinformatic analysis is required. The question which obtained information is of importance requires the existence of large genomic databases. Integrated analysis of medical “big-data” will benefit from artificial intelligence technologies, such as machine learning and its subset, deep learning.¹²⁰ Novel long-read technologies will enable routine resequencing, allowing better determination of repetitive regions and structural variants. Epidemiologists will follow the outbreak of an infectious disease through microbial sequencing from various samples, such as wastewaters. Therefore, it will be possible to detect a disease outbreak at an early stage and possibly even prevent it. Small, portable devices will be useful in such situations. Oxford Nanopore has recently started developing device called SmidgION, which is even smaller than MinION.⁴⁴ It will be used with smartphones and other portable devices. The portability, high output, and simplicity of such machines have infinite on-site applications: in ecology, forensics, population screening, epidemiology, to name a few.

However, although the technology shows immense breakthrough, education of staff has to become the central concern of all countries employing next generation technologies. Well trained medical geneticists and consultants with narrow specializations on next generation sequencing will be the ones enabling meaningful interpretation of the results, thus providing use of the result in clinical setting.

Acknowledgments

The study was partially financed by Slovenian Research Agency programme grant No. P1-0390.

5. References

1. A. M. Maxam, W. Gilbert, *PNAS* **1977**, *74*, 560–564. DOI:10.1073/pnas.74.2.560
2. F. Sanger, S. Nicklen, A. R. Coulson, *PNAS* **1977**, *74*, 5463–5467. DOI:10.1073/pnas.74.12.5463
3. J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, et al., *Science* **2001**, *291*, 1304–1351.
4. J. A. Schloss, *Nat. Biotechnol.* **2008**, *26*, 1113–1115. DOI:10.1038/nbt1008-1113
5. E. L. Van Dijk, H. Auger, Y. Jaszczyszyn, C. Thermes, *Trends Genet.* **2014**, *30*, 418–426. DOI:10.1016/j.tig.2014.07.001
6. National Human Genome Research Institute, <https://www.genome.gov/>, (accessed: April 5, 2020)
7. J. M. Prober, G. L. Trainor, R. J. Dam, F. W. Hobbs, C. W. Robertson, R. J. Zagursky, et al., *Science* **1987**, *238*, 336–341. DOI:10.1126/science.2443975
8. H. Swerdlow, R. Gesteland, *Nucleic Acids Res.* **1990**, *18*, 1415–1419. DOI:10.1093/nar/18.6.1415
9. E. L. van Dijk, Y. Jaszczyszyn, D. Naquin, C. Thermes, *Trends in Genet.* **2018**, *34*, 666–681. DOI:10.1016/j.tig.2018.05.008
10. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, et al., *Nature* **2001**, *409*, 860–921. DOI:10.1038/35057062
11. R. Drmanac, A. B. Sparks, M. J. Callow, A. L. Halpern, N. L. Burns, B. G. Kermani, et al., *Science* **2010**, *327*, 78–81.
12. K. R. Kumar, M. J. Cowley, R. L. Davis, *Semin. Thromb. Hemostasis* **2019**, *45*, 661–673. DOI:10.1055/s-0039-1688446
13. M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, et al., *Nature* **2006**, *441*, 120–120. DOI:10.1038/nature04774
14. C. T. Harrington, E. I. Lin, M. T. Olson, J. R. Eshleman, *Arch. Pathol. Lab. Med.* **2013**, *137*, 1296–1303. DOI:10.5858/arpa.2012-0463-RA
15. G. Gao, D. I. Smith, *Clinical Chemistry* **2020**, *66*, 77–88. DOI:10.1373/clinchem.2019.303305
16. S. Goodwin, J. D. McPherson, W. R. McCombie, *Nat. Rev. Genet.* **2016**, *17*, 333. DOI:10.1038/nrg.2016.49
17. Illumina, CMOS Chip and One-Channel SBS Chemistry, <https://www.illumina.com/science/technology/next-genera->

- tion-sequencing/sequencing-technology/semiconductor-sequencing-cmos.html, (accessed: December 2, 2020)
18. Illumina, Patterned flow cells, <https://emea.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/patterned-flow-cells.html>, (accessed: April 14, 2021)
 19. Illumina, <https://www.illumina.com/>, (accessed: March 20, 2020)
 20. J. M. Rothberg, W. Hinze, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, et al., *Nature* **2011**, *475*, 348–352. DOI:10.1038/nature10242
 21. BGI Genomics, <https://www.bgi.com/global/>, (accessed: March 22, 2020)
 22. G. O. S. Coyne, N. Takebe, A. P. Chen, *Curr. Probl. Cancer* **2017**, *41*, 182–193. DOI:10.1016/j.currprobcancer.2017.02.001
 23. Y. O. Alekseyev, R. Fazeli, S. Yang, R. Basran, T. Maher, N. S. Miller, et al., *Academic Pathology* **2018**, *5*, 2374289518766521. DOI:10.1177/2374289518766521
 24. R. Luthra, K. P. Patel, M. J. Routbort, R. R. Broaddus, J. Yau, C. Simien, et al., *J. Mol. Diagn.* **2017**, *19*, 255–264. DOI:10.1016/j.jmoldx.2016.09.011
 25. S. L. Salzberg, J. A. Yorke, *Bioinformatics* **2005**, *21*, 4320–4321. DOI:10.1093/bioinformatics/bti769
 26. J. Weischenfeldt, O. Symmons, F. Spitz, J. O. Korbel, *Nat. Rev. Genet.* **2013**, *14*, 125–138. DOI:10.1038/nrg3373
 27. D. Qin, *Cancer Biol. Med.* **2019**, *16*, 4–10.
 28. A. Rhoads, K. F. Au, *Genomics Proteomics Bioinf.* **2015**, *13*, 278–289. DOI:10.1016/j.gpb.2015.08.002
 29. E. E. Schadt, S. Turner, A. Kasarskis, *Hum. Mol. Genet.* **2010**, *19*, 227–240. DOI:10.1093/hmg/ddq416
 30. M. A. Quail, M. Smith, P. Coupland, T. D. Otto, S. R. Harris, T. R. Connor, et al., *BMC Genomics* **2012**, *13*, 341. DOI:10.1186/1471-2164-13-341
 31. L. Steinbock, A. Radenovic, *Nanotechnology* **2015**, *26*, 074003. DOI:10.1088/0957-4484/26/7/074003
 32. T. Laver, J. Harrison, P. O'Neill, K. Moore, A. Farbos, K. Paszkiewicz, et al., *Biomol. Detect. Quantif.* **2015**, *3*, 1–8. DOI:10.1016/j.bdq.2015.02.001
 33. J. Quick, N. J. Loman, S. Duraffour, J. T. Simpson, E. Severi, L. Cowley, et al., *Nature* **2016**, *530*, 228–232.
 34. N. R. Faria, E. C. Sabino, M. R. Nunes, L. C. J. Alcantara, N. J. Loman, O. G. Pybus, *Genome Med.* **2016**, *8*, 97. DOI:10.1186/s13073-016-0356-2
 35. M. Wang, A. Fu, B. Hu, Y. Tong, R. Liu, Z. Liu, et al., *Small* **2020**, *16*, 2002169. DOI:10.1002/smll.202002169
 36. D. Kim, J.-Y. Lee, J.-S. Yang, J. W. Kim, V. N. Kim, H. Chang, *Cell* **2020**, *181*, 914–921.e10. DOI:10.1016/j.cell.2020.04.011
 37. S. Marquez, B. Prado-Vivar, J. J. Guadalupe, B. G. Granja, M. Jibaja, M. Tobar, et al., *medRxiv* **2020**.
 38. B. A. Flusberg, D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, et al., *Nat. Methods* **2010**, *7*, 461–465. DOI:10.1038/nmeth.1459
 39. D. R. Garalde, E. A. Snell, D. Jachimowicz, B. Sipos, J. H. Lloyd, M. Bruce, et al., *Nat. Methods* **2018**, *15*, 201–206. DOI:10.1038/nmeth.4577
 40. P. Euskirchen, F. Bielle, K. Labreche, W. P. Kloosterman, S. Rosenberg, M. Daniau, et al., *Acta Neuropathol.* **2017**, *134*, 691–703. DOI:10.1007/s00401-017-1743-5
 41. M. Jain, J. R. Tyson, M. Loose, C. L. Ip, D. A. Eccles, J. O'Grady, et al., *F1000Research* **2017**, *6*, 760. DOI:10.12688/f1000research.11354.1
 42. J. R. Miller, P. Zhou, J. Mudge, J. Gurtowski, H. Lee, T. Ramaraj, et al., *BMC Genomics* **2017**, *18*, 541. DOI:10.1186/s12864-017-3927-8
 43. R. Minei, R. Hoshina, A. Ogura, *BMC Genomics* **2018**, *19*, 700. DOI:10.1186/s12864-018-5067-1
 44. Oxford Nanopore Technologies, <https://nanoporetech.com/>, (accessed: April 5, 2020)
 45. M. J. Chaisson, J. Huddleston, M. Y. Dennis, P. H. Sudmant, M. Malig, F. Hormozdiari, et al., *Nature* **2015**, *517*, 608–611. DOI:10.1038/nature13907
 46. J. S. Seo, A. Rhie, J. Kim, S. Lee, M.-H. Sohn, C.-U. Kim, et al., *Nature* **2016**, *538*, 243–247. DOI:10.1038/nature20098
 47. K. H. Miga, S. Koren, A. Rhie, M. R. Vollger, A. Gershman, A. Bzikadze, et al., *Nature* **2020**, *585*, 79–84.
 48. M. H. Schmidt, C. E. Pearson, *DNA repair* **2016**, *38*, 117–126. DOI:10.1016/j.dnarep.2015.11.008
 49. F. S. Collins, H. Varmus, *N. Engl. J. Med.* **2015**, *372*, 793–795. DOI:10.1056/NEJMp1500523
 50. M. B. Neu, K. M. Bowling, G. M. Cooper, *Curr. Opin. Pediatr.* **2019**, *31*, 732–738. DOI:10.1097/MOP.0000000000000815
 51. E. A. Vidal, T. C. Moyano, B. I. Bustos, E. Pérez-Palma, C. Moraga, E. Riveras, et al., *Sci. Rep.* **2019**, *9*, 1–11. DOI:10.1038/s41598-019-39391-z
 52. A. Nekrutenko, J. Taylor, *Nat. Rev. Genet.* **2012**, *13*, 667–672. DOI:10.1038/nrg3305
 53. C. J. Saunders, N. A. Miller, S. E. Soden, D. L. Dinwiddie, A. Noll, N. A. Alnadi, et al., *Sci. Transl. Med.* **2012**, *4*, 135–154. DOI:10.1126/scitranslmed.3004041
 54. M. M. Clark, A. Hildreth, S. Batalov, Y. Ding, S. Chowdhury, K. Watkins, et al., *Sci. Transl. Med.* **2019**, *11*, eaat6177.
 55. H. Nakagawa, M. Fujita, *Cancer Sci.* **2018**, *109*, 513–522. DOI:10.1111/cas.13505
 56. S. W. Kong, I.-H. Lee, X. Liu, J. N. Hirschhorn, K. D. Mandl, *Genet. Med.* **2018**, *20*, 1617–1626. DOI:10.1038/gim.2018.51
 57. J. Y. Hehir-Kwa, R. Pfundt, J. A. Veltman, *Expert Rev. Mol. Diagn.* **2015**, *15*, 1023–32. DOI:10.1586/14737159.2015.1053467
 58. J. S. Mattick, M. Dinger, N. Schonrock, M. Cowley, *Med. J. Aust.* **2018**, *209*, 197–199. DOI:10.5694/mja17.01176
 59. J. F. Laes, P. Aftimos, P. Barthelemy, J. Bellmunt, G. Berchem, C. Camps, et al., *Oncotarget* **2018**, *9*, 20282–20293. DOI:10.18632/oncotarget.24757
 60. T. Giardina, C. Robinson, F. Grieco-Iacopetta, M. Millward, B. Iacopetta, D. Spagnolo, et al., *Pathology* **2018**, *50*, 389–401. DOI:10.1016/j.pathol.2018.01.005
 61. K. Sunami, H. Ichikawa, T. Kubo, M. Kato, Y. Fujiwara, A. Shimomura, et al., *Cancer Sci.* **2019**, *110*, 1480–1490. DOI:10.1111/cas.13969

62. E. Crowley, F. Di Nicolantonio, F. Loupakis, A. Bardelli, *Nat. Rev. Clin. Oncol.* **2013**, *10*, 472. DOI:10.1038/nrclinonc.2013.110
63. R. Mead, M. Duku, P. Bhandari, I. Cree, *BJC.* **2011**, *105*, 239–245. DOI:10.1038/bjc.2011.230
64. T. K. Sundaresan, L. V. Sequist, J. V. Heymach, G. J. Riely, P. A. Jänne, W. H. Koch, et al., *Clin. Cancer Res.* **2016**, *22*, 1103–1110. DOI:10.1158/1078-0432.CCR-15-1031
65. J. Phallen, M. Sausen, V. Adleff, A. Leal, C. Hruban, J. White, et al., *Sci. Transl. Med.* **2017**, *9*. DOI:10.1126/scitranslmed.aan2415
66. E. Gormally, P. Vineis, G. Matullo, F. Veglia, E. Caboux, E. Le Roux, et al., *Cancer Res.* **2006**, *66*, 6871–6876. DOI:10.1158/0008-5472.CAN-05-4556
67. W. Liang, Y. Zhao, W. Huang, Y. Gao, W. Xu, J. Tao, et al., *Theranostics* **2019**, *9*, 2056. DOI:10.7150/thno.28119
68. L. Giannopoulou, M. Zavridou, S. Kasimir-Bauer, E. S. Lianidou, *Transl. Res.* **2019**, *205*, 77–91. DOI:10.1016/j.trsl.2018.10.003
69. Ontario Health Technology, *Ontario health technology assessment series* **2019**, *19*, 1–166.
70. S. Yohe, B. Thyagarajan, *Arch. Pathol. Lab. Med.* **2017**, *141*, 1544–1557. DOI:10.5858/arpa.2016-0501-RA
71. N. S. Abul-Husn, E. E. Kenny, *Cell* **2019**, *177*, 58–69. DOI:10.1016/j.cell.2019.02.039
72. B. H. Shirts, J. S. Salama, S. J. Aronson, W. K. Chung, S. W. Gray, L. A. Hindorff, et al., *J. Am. Med. Assoc.* **2015**, *22*, 1231–1242. DOI:10.1093/jamia/ocv065
73. L. J. Ewans, D. Schofield, R. Shrestha, Y. Zhu, V. Gayevskiy, K. Ying, et al., *Genet. Med.* **2018**, *20*, 1564–1574. DOI:10.1038/gim.2018.39
74. S. M. Hiatt, M. D. Amaral, K. M. Bowling, C. R. Finnilla, M. L. Thompson, D. E. Gray, et al. *Clin. Genet.* **2018**, *94*, 174–178. DOI:10.1111/cge.13259
75. A. M. Wenger, H. Guturu, J. A. Bernstein, G. Bejerano, *Genet. Med.* **2017**, *19*, 209–214. DOI:10.1038/gim.2016.88
76. C. F. Wright, J. F. McRae, S. Clayton, G. Gallone, S. Aitken, T. W. FitzGerald, et al., *Genet. Med.* **2018**, *20*, 1216–1223. DOI:10.1038/gim.2017.246
77. A.-L. Bruel, A. Vitobello, F. T. Mau-Them, S. Nambot, Y. Duffourd, V. Quéré, et al., *Genet. Med.* **2019**, *21*, 1657–1661. DOI:10.1038/s41436-018-0383-z
78. Editorial, *Lancet* **2018**, *390*, 2739. DOI:10.1016/S0140-6736(17)31540-4
79. M. Champion, C. Goldgar, R. J. Hopkin, C. A. Prows, S. Dasgupta, *Genet. Med.* **2019**, *21*, 2422–2430. DOI:10.1038/s41436-019-0548-4
80. P. A. Deverka, D. Kaufman, A. L. McGuire, *JAMA* **2014**, *312*, 1857–1858. DOI:10.1001/jama.2014.14915
81. D. R. Adams, C. M. Eng, *N. Engl. J. Med.* **2018**, *379*, 1353–1362. DOI:10.1056/NEJMra1711801
82. T. Y. Tan, O. J. Dillon, Z. Stark, D. Schofield, K. Alam, R. Shrestha, et al., *JAMA Pediatrics* **2017**, *171*, 855–862. DOI:10.1001/jamapediatrics.2017.1755
83. L. Meng, M. Pammi, A. Saronwala, P. Magoulas, A. R. Ghazi, F. Vetrini, et al., *JAMA Pediatrics* **2017**, *171*, 173438–173438.
84. L. Farnaes, A. Hildreth, N. M. Sweeney, M. M. Clark, S. Chowdhury, S. Nahas, et al., *NPJ Genomic Med.* **2018**, *3*, 10. DOI:10.1038/s41525-018-0049-4
85. The 100,000 Genomes Project, <https://www.genomicsengland.co.uk/about-genomics-england/the-100000-genomes-project/>, (accessed: July 22, 2020)
86. European Commission, <https://ec.europa.eu/digital-single-market/en/european-1-million-genomes-initiative/>, (accessed: July 22, 2020)
87. B. Zalar, A. Maver, A. Kovanda, A. Peterlin, B. Peterlin, *Psychiatr. Danubina* **2018**, *30*, 216–219. DOI:10.24869/spsih.2018.216
88. L. Vidmar, A. Maver, J. Drušević, J. Sepčič, I. Novaković, S. Ristič, et al., *Sci. Rep.* **2019**, *9*, 1–10. DOI:10.1038/s41598-019-45598-x
89. A. Smon, B. R. Lampret, U. Groselj, M. Z. Tansek, J. Kovac, D. Perko, et al., *Clin. Biochem.* **2018**, *52*, 48–55. DOI:10.1016/j.clinbiochem.2017.10.016
90. J. Kovač, G. Klančar, K. Trebušak Podkrajšek, S. Battelino, *Front. Genet.* **2017**, *8*, 95. DOI:10.3389/fgene.2017.00095
91. D. Šekoranja, E. Boštjančič, V. Salapura, B. Mavčič, J. Pižem, *Cancer Genet.* **2018**, *228*, 12–16. DOI:10.1016/j.cancergen.2018.07.001
92. K. Hočevar, A. M. Maver, M. Vidmar Šimic, A. Hodžič, A. Haslberger, T. Premru-Sršen, et al., *Front. Med.* **2019**, *6*, 201. DOI:10.3389/fmed.2019.00201
93. A. Mahnic, M. Rupnik, *PLoS One* **2018**, *13*, e0209209. DOI:10.1371/journal.pone.0209209
94. K. Hočevar, A. Maver, B. Peterlin, *Front. Pharmacol.* **2019**, *10*, 240–240. DOI:10.3389/fphar.2019.00240
95. K. Kouter, T. Zupanc, A. Videtic Paska, *J. Affect. Disord.* **2019**, *253*, 419–425. DOI:10.1016/j.jad.2019.04.077
96. J. P. Evans, B. C. Powell, J. S. Berg, *JAMA* **2017**, *317*, 1904–1905. DOI:10.1001/jama.2017.0432
97. M. F. Murray, *Ann. Intern. Med.* **2018**, *169*, 407–408. DOI:10.7326/M18-1722
98. F. E. Dewey, M. F. Murray, J. D. Overton, L. Habegger, J. B. Leader, S. N. Fetterolf, et al., *Science* **2016**, *354*, aaf6814.
99. J. S. Berg, M. J. Khoury, J. P. Evans, *Genet. Med.* **2011**, *13*, 499–504. DOI:10.1097/GIM.0b013e318220aaba
100. A. Kozińska, P. Seweryn, I. Sitkiewicz, *J. Appl. Genet.* **2019**, *60*, 103–111. DOI:10.1007/s13353-019-00482-2
101. A. Schürch, S. Arredondo-Alonso, R. Willems, R. V. Goering, *Clin. Microbiol. Infect.* **2018**, *24*, 350–354. DOI:10.1016/j.cmi.2017.12.016
102. Coronavirus Update, <https://www.worldometers.info/coronavirus/>, (accessed: April 11, 2021)
103. Y.-Z. Zhang, E. C. Holmes, *Cell* **2020**, *181*, 223–227. DOI:10.1016/j.cell.2020.03.035
104. V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. Chu, et al., *Eurosurveillance* **2020**, *25*, 2000045. DOI:10.2807/1560-7917.ES.2020.25.3.2000045
105. Y. Wang, F. Zhou, D. Zhang, J. Zhao, R. Du, Y. Hu, et al., *Trials* **2020**, *21*, 422.
106. N. D. Grubaugh, J. T. Ladner, P. Lemey, O. G. Pybus, A. Rambaut, E. C. Holmes, et al., *Nat. Microbiol.* **2019**, *4*,

- 10–19. DOI:10.1038/s41564-018-0296-2
107. M. Chiara, A. M. D'Erchia, C. Gissi, C. Manzari, A. Parisi, N. Resta, et al., *Brief. Bioinformatics* **2020**, *22*, 616–630. DOI:10.1093/bib/bbaa297
108. J. F.-W. Chan, S. Yuan, K.-H. Kok, K. K.-W. To, H. Chu, J. Yang, et al., *The Lancet* **2020**, *395*, 514–523. DOI:10.1016/S0140-6736(20)30154-9
109. J. Handelsman, *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 669–85. DOI:10.1128/MMBR.68.4.669-685.2004
110. R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, et al., *The Lancet* **2020**, *395*, 565–574. DOI:10.1016/S0140-6736(20)30251-8
111. B. Coutard, C. Valle, X. de Lamballerie, B. Canard, N. Seidah, E. Decroly, *Antiviral Res.* **2020**, *176*, 104742. DOI:10.1016/j.antiviral.2020.104742
112. T. T.-Y. Lam, N. Jia, Y.-W. Zhang, M. H.-H. Shum, J.-F. Jiang, H.-C. Zhu, et al., *Nature* **2020**, *583*, 282–285. DOI:10.1038/s41586-020-2169-0
113. T. Zhang, Q. Wu, Z. Zhang, *Curr. Biol.* **2020**, *30*, 1346–1351.e2.114. R. Sanjuan, M. R. Nebot, N. Chirico, L. M. Mansky, R. Belshaw, *J. Virol.* **2010**, *84*, 9733.
115. P. Ellis, F. Somogyvari, D. P. Virok, M. Nosedá, G. R. McLean, *Curr. Genet. Med. Rep.* **2021**, 1–12.
116. X. Chen, Y. Kang, J. Luo, K. Pang, X. Xu, J. Wu, et al., *Front. Cell. Infect. Microbiol.* **2021**, *11*. DOI:10.3389/fcimb.2021.632490
117. J. L. Gardy, N. J. Loman, *Nat. Rev. Gen.* **2018**, *19*, 9–20. DOI:10.1038/nrg.2017.88
118. X. Deng, W. Gu, S. Federman, L. du Plessis, O. G. Pybus, N. R. Faria, et al., *Science* **2020**, *369*, 582. DOI:10.1126/science.abb9263
119. M. M. Medina-Enriquez, S. Lopez-Leon, J. A. Carlos-Escalante, Z. Aponte-Torres, A. Cuapio, T. Wegman-Ostrosky, *Cell Biosci.* **2020**, *10*, 148. DOI:10.1186/s13578-020-00519-8
120. R. Hamamoto, M. Komatsu, K. Takasawa, K. Asada, S. Kaneko, *Biomolecules* **2020**, *10*, 62. DOI:10.3390/biom10010062

Povzetek

Določanje nukleotidnega zaporedja – sekvenciranje je že več kot 40 let stara tehnologija, a kot orodje v klinični medicini velja za relativno mlado. Z napredkom in razvojem tehnologije sekvenciranje postaja občutno cenejše, hitrejše in natančnejše; prav zato ga je mogoče vključevati v klinične laboratorije za rutinsko uporabo. Leta 2003 je bilo prvič določeno zaporedje človeškega genoma, s tem pa postavljeni temelji za uporabo genomike v klinični praksi ter razvoj personalizirane medicine. Poznavanje nukleotidnega zaporedja predstavlja nepristranski pristop k diagnostiki ter ima lahko pomembno vlogo tudi pri spremljanju bolezni in zdravljenju, a prehod iz raziskovalne uporabe v klinično prakso lahko traja tudi veliko let. V zadnjem desetletju je bil porast novih tehnologij sekvenciranja skokovit, kar pa pomeni, da se mora strokovni kader relativno hitro izobraževati na področju osnovnega znanja o zmožnostih in pomanjkljivostih obstoječih metod ter njihovi uporabi v kliničnih okoljih in raziskavah. V preglednem članku predstavljamo pregled tehnologij sekvenciranja z zgodovinskega vidika in mu dodamo klinično uporabo sekvenciranja. Nekatera najbolj obetavna področja so predstavljena z izbranimi primeri slovenskih raziskovalcev.



Except when otherwise noted, articles in this journal are published under the terms and conditions of the Creative Commons Attribution 4.0 International License