

## Postgenomic technologies for genomic and proteomic analysis in biological and medical research

S.A. Solodskikh<sup>1</sup>, M.V. Gryaznova<sup>1</sup>, Y.D. Dvoretzskay<sup>1</sup>, A. P. Gureev<sup>1</sup>, A.V. Panevina<sup>1</sup>,  
A.Y. Maslov<sup>1,2</sup>, O.V. Serzhantova<sup>1,3</sup>, A.A. Mikhailov<sup>1,3</sup>, C. Chinopoulos<sup>4</sup>, V.N. Popov<sup>1,5</sup>,

<sup>1</sup> Voronezh State University, Voronezh, Russian Federation

<sup>2</sup> Albert Einstein College of Medicine, Bronx, USA

<sup>3</sup> Voronezh Regional Clinical Oncological Dispensary, Voronezh, Russian Federation

<sup>4</sup> Semmelweis University, Budapest, Hungary

<sup>5</sup> Voronezh State University of Engineering Technologies, Voronezh, Russian Federation

**Received: 13.11.2019. Accepted 26.12.2019**

Over the 15 years since the decoding of the human genome a large number of individual genomes have been sequenced. Targeted sequencing – sequencing of select genome regions – has been widely used both in research and in medical practice. The use of various types of genetic analysis is starting to be used in daily clinical routine. At the same time, the price of sequencing decreases and as a result, the amount of genetic information available to researchers and physicians increases. These processes together determine the need for creation of databases for the centralized storage of genetic information which is crucial for synchronization and validation of the work of various institutions. One of the first such databases was the NCBI database created and supervised by the US National Center for Biotechnological Information (NCBI) in collaboration with the National Institute for Human Genome Research (NHGRI).

At the same time, the available methods for studying associations between DNA polymorphisms and various phenotypic manifestations do not cover the most important layer of regulation of biological processes – the proteome. The methods of high-throughput proteomic analysis that are to be developed will allow identifying driver mutations that make the greatest contribution to the phenotype of the studied object.

The application of an integrated analysis of the genome and proteome for the diagnosis and treatment of cancer pathologies is one of the most important research goals now. This approach will allow to identify new genetic biomarkers that could be used for reliable prediction of the treatment response, risks of the most important diseases, and the development of novel medications. This review shows recent advances in proteomic and genomic approaches to the development of more sensitive diagnostic and prognostic biomarkers that can be translated into improved clinical care and treatment of the disease.

**Keywords:** sequencing; genomics; transcriptomics; proteomics; socially significant diseases

### Introduction

Over the 15 years since the decoding of the human genome, a large number of individual genomes have been sequenced. Targeted sequencing – sequencing of select genome regions – has been widely used both in research and in medical practice. The use of various types of genetic analysis is starting to be used in daily clinical routine. At the same time, the price of sequencing decreases and as a result, the amount of genetic information available to researchers and physicians increases. These processes together determine the need for creation of databases for the centralized storage of genetic information which is crucial for synchronization and validation of the work of various institutions. One of the first such databases was the dbSNP database created and supervised by the US National Center for Biotechnological Information (NCBI) in collaboration with the National Institute for Human Genome Research (NHGRI). Currently, the database contains more than 23.7 million genetic polymorphisms, of which 14.5 million are validated.

For most diseases associated with changes in DNA, there is no currently clear understanding of the specific genetic mechanisms of its development. In other words, it can be reliably established that the disease is due to hereditary factors, but the specific mechanism of inheritance is still unexplored. The vague formulations found in official medical guidelines for diagnosing such diseases are based on this fact. For example, recommendations for the early diagnosis of type 2 diabetes (T2DM) from the International Diabetes Association suggest that T2DM is determined by hereditary factors. The authors recommend that patients to undergo genetic testing without specifying the nature, methodology, references and expected test results. Oncological diseases stand aside due to their specificity – there is exact knowledge of several mechanisms of cancer and key

genes whose dysfunction can cause cancer, which makes it much easier to establish the cause of the hereditary form of cancer and even develop a genetic testing method.

Even for those diseases for which the genetic mechanisms of predisposition and inheritance are somewhat reliably established, the clear practical guidelines are missing. Surprisingly, the most deadly class of the diseases - cardiovascular - despite studied fairly good (currently more than a hundred genetic polymorphisms that determine the inheritance of CVD are identified), are not covered by clinical guidelines. In developed countries, the mandatory procedures for assessing the risks of cardiovascular diseases, based on biochemical and questionnaire parameters, began to appear only during past two decades.

At the same time, clinical and scientific research are conducted in these cases. For a number of diseases, genetic predictor polymorphisms were discovered during the clinical trials and are stored in the ClinVar database with the "risk factor" flag. These risk factors fall into two classes - "pathogenic" and "likely pathogenic". There are also results of the most advanced and promising studies that complement and extend the clinical trial data.

Advances in genomic technologies have greatly facilitated the understanding of the genetic mechanisms of a number of diseases, and have also contributed to the discovery of new biomarkers. The combination of proteomic and genomic technologies is essential for the detection of biomarkers for the early diagnosis of diseases associated with DNA damage, as well as for general biological research, developments in the field of biotechnology and the food industry. Recent advances in the detection of socially significant diseases based on the human genome using advanced genomic technologies such as PCR and next generation sequencing (NGS) have shown promising results. Similarly, proteomics can lead to a revolution in the diagnosis and screening of socially significant diseases based on new proteomic databases, which include somatic variants and post-translational modifications. Thus, the developed proteomic technologies can be used as an addition to classical research methods (Panis, Pizzatti, Souza, & Abdelhay, 2016). Moreover, the use of several proteomic and genomic biomarkers, rather than one gene or protein, can significantly improve diagnostic accuracy and increase predictive ability, which can provide adequate monitoring of the response to treatment and can be an important milestone on the way to personalized medicine (Jackson & Chester, 2015; Larijani, Perani, Alburai'si, & Parker, 2015).

At the same time, the available methods for studying associations between DNA polymorphisms and various phenotypic manifestations do not cover the most important layer of regulation of biological processes - the proteome. The methods of high-throughput proteomic analysis that are to be developed will allow identifying driver mutations that make the greatest contribution to the phenotype of the studied object.

The application of an integrated analysis of the genome and proteome for the diagnosis and treatment of cancer pathologies is one of the most important research goals now. This approach will allow to identify new genetic biomarkers that could be used for reliable prediction of the treatment response, risks of the most important diseases, and the development of novel medications. This review shows recent advances in proteomic and genomic approaches to the development of more sensitive diagnostic and prognostic biomarkers that can be translated into improved clinical care and treatment of the disease (Tanase, Albulescu, & Neagu, 2015).

## Methods for studying genome

The personalized medicine can be defined in the following manner: all diagnostic tools, types and combinations of therapy, treatment procedures including surgery, medical recommendations, and, in the future, the development of new types of drugs that are created and applied based on knowledge about the individual characteristics of this patient. Until the decoding of the human genome, the very existence of such a medicine was impossible due to a lack of knowledge about the features of the genome, transcriptome, proteome, and human immunity mechanisms (Day & Siu, 2016).

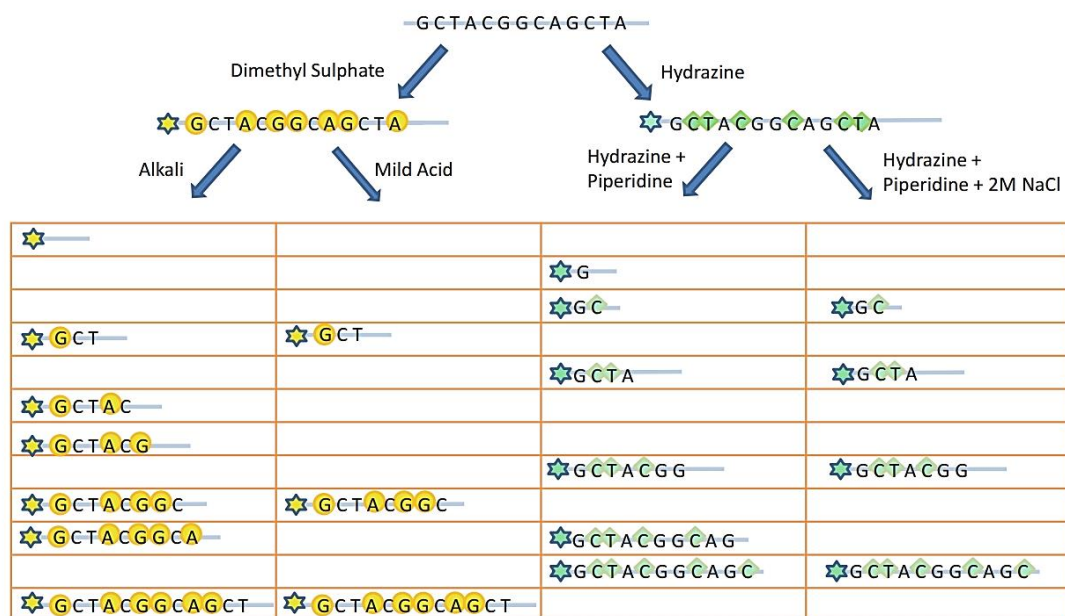
DNA sequencing has revolutionized molecular biology, medicine, genomics, and related fields. The first sequencing method proposed by Frederick Sanger in 1977 over the years has made possible the development of new and improved DNA sequencing platforms (Sanger, Nicklen, & Coulson, 1977). These technologies, along with a variety of computational tools for the analysis and interpretation of data helped researchers better understand the genomes of various organisms. They made sequencing a powerful yet feasible research tool that has evolved to the point where it can be easily used even in small laboratories with high efficiency, without the need for large sequencing centers.

### Classic sequencing methods

Even after the advent of next-generation sequencing methods, it is still considered that most of the DNA sequence data was obtained using first-generation technologies. Although these technologies are slower and more expensive (even after the automation), they are still used in studies where increased accuracy is required. The initial first-generation sequencing technologies were sequencing methods established by Maxam-Gilbert (Maxam & Gilbert, 1977) and the sequencing-by-synthesis method developed by Sanger (Sanger et al., 1977).

### Maxam-Gilbert sequencing method

This method first appeared in 1977 and is also known as the "chemical degradation" method. Chemical reagents act on the specific bases of existing DNA molecules, which lead to the subsequent cleavage. In this method, DNA is labeled with radioactive phosphorus at the 5' or 3' end. The next step is to obtain single-stranded DNA. This can be done by restriction cleavage, which produces sticky ends in DNA, or by denaturation at 90 °C in the presence of DMSO, which leads to the formation of single-stranded DNA. The sample is divided into 4 aliquots, after which a partial hydrolysis reaction is carried out in each part, leading to the occurrence of gaps at the sites of incorporation of 4 different nucleotides or their combinations.



**Fig. 1** Maxam-Gilbert sequencing method (Verma et al., 2016)

Currently, Maxam-Gilbert sequencing is not used in practice due to the low speed of analysis and overall laborious procedure.

### Sanger sequencing technology

This method is also known as chain termination sequencing. Sanger sequencing has played a crucial role in understanding the genetic landscape of the human genome. It was developed by Frederick Sanger in 1975, and commercialized in 1977 (Sanger et al., 1977).

The technique is based on the use of dideoxynucleoside triphosphates (ddNTP), in which the 3' hydroxyl group is missing. The process uses seven different components for performing sequencing. They include a single-stranded DNA template for sequencing, primers, Taq polymerase for amplification of the template, reaction buffer, deoxynucleotides (dNTP), fluorescently labeled ddNTPs and DMSO (used to denature secondary structures in the DNA chain). Since 3'-OH group is absent in the incorporated ddNTP, the phosphodiester bond between C3'-OH of the last base and C5' of the next dNTP is not formed, which leads to the termination of the chain at this point (Fig. 2a).

Electrophoresis in polyacrylamide gel is used to separate the products of each reaction by length in four parallel lanes (Fig. 2b). In 1987, Applied Biosystems, Inc. (ABI) released the first automated DNA sequencing machine, Model 370 ABI, developed by Leroy Hood and Mike Hankapiller, which could generate read lengths of up to 350 nucleotide pairs. In 1995, ABI released ABI PRISM 310 genetic analyzer, which made it possible to simplify the inconvenient and laborious process of preparing gels, component installation, and sample loading. Swerdlow and Gesteland developed the machine known as capillary sequencer, which uses capillaries filled with polyacrylamide gel, rather than using the gel plates. A general view of the obtained electrophoregrams is shown in Figure 2c. Currently available on the market sequencers use 4, 16, 48, 96 or 384 capillaries simultaneously. As the number of capillaries increases, the read length and, sequencing speed also increase (Verma et al., 2016).

### Next Generation Sequencing

The emergence of next-generation sequencing methods (NGS) has made the difficult task of sequencing much easier and faster. Fast and economically viable NGS technologies have become much more popular than the slow and laborious analogues of the first generation. In combination with bioinformatics technologies, these methods have significantly increased the speed of data collection and its amount. These methods made it possible to simplify DNA sample preparation for sequencing because the transformation of *E. coli* is no longer required (Kamps et al., 2017).

### 454 (Roche) pyrosequencing

454 sequencing was the first of the NGS techniques to be introduced in 2005. This process is called pyrosequencing, which is based on the emission of light due to a cascade of reactions that occur after the release of pyrophosphate. First, the DNA duplex is cut into smaller fragments followed by ligation of the adapters, which are complementary to the primer sequences, to the both sides of the DNA fragment. These adapters act as a primer-binding site and initiate the sequencing process. Each DNA fragment is connected to the emulsion microsphere, so that the ratio of DNA to microspheres in the sequencing reaction volume is 1:1. This is followed by amplification of each fragment using emulsion PCR and after several cycles, many copies of these DNA molecules per microsphere are synthesized (Ronaghi, 1998).

Immobilized enzymes (DNA polymerase, ATP-sulfurylase, luciferase and apyrase) are added to the wells in a microplate each containing one microsphere covered with copies of DNA fragment. Then, each of the four dNTPs is applied in turn. Complementary dNTP is integrated into the growing chain using DNA polymerase. This process is accompanied by the release of pyrophosphate (PPi), which is turned into ATP by ATP-sulfurylase. In the presence of generated ATP, the luciferase enzyme converts luciferin to oxyluciferin, which is accompanied by emission of light signal, the intensity of which is proportional to the amount of ATP. The intensity of the light signal is recorded by a CCD-camera. As soon as the signal is received and processed, the apyrase enzyme cleaves the existing nucleotides and ATP and then the next nucleotide is added (Fig. 3). Recently, this

technique has been further improved by introduction of paired-end sequencing. The adapters are linked to both ends of the fragmented DNA, which allows reading the fragment from both ends. The main advantage of this method is long reading length. In contrast to other NGS technologies, 454 sequencing gives a reading length of up to 400 base pairs and can generate more than 1,000,000 reads per cycle. This method is useful for the *de novo* assembly of genomes and the study of metagenomes (Petrosino et al., 2009).

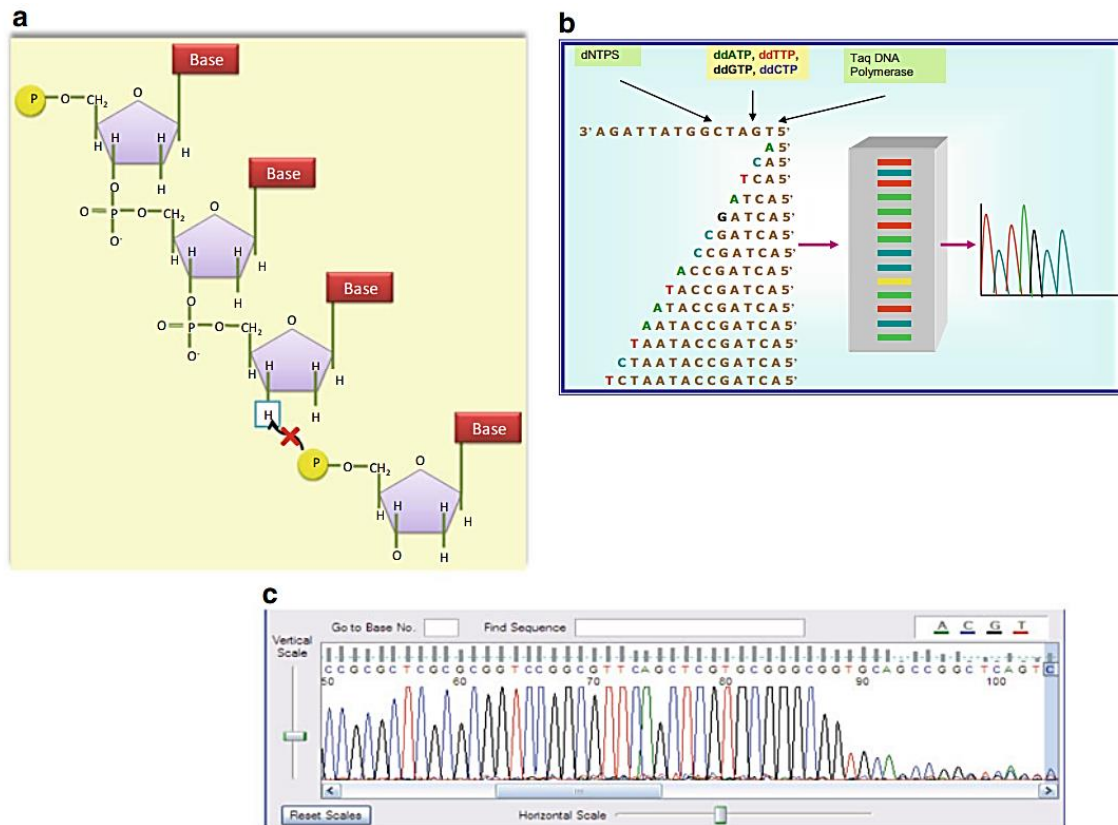


Fig. 2 Sanger sequencing technology (Verma et al., 2016)

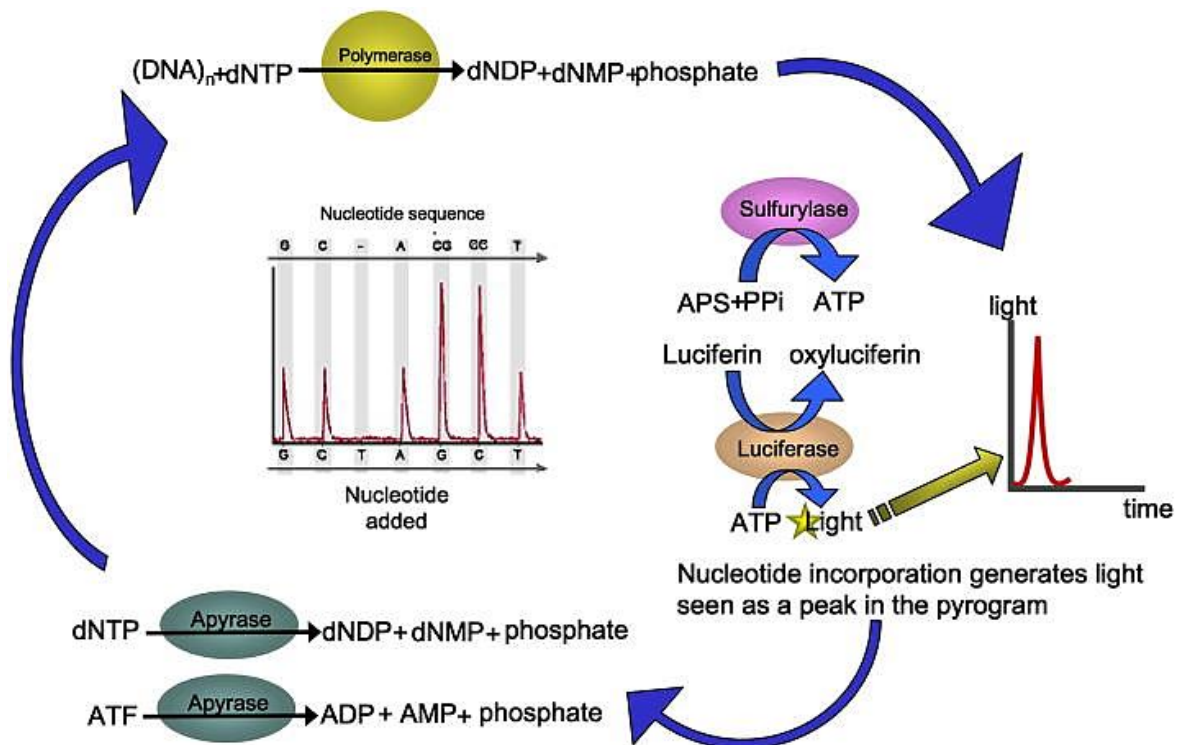


Fig. 3 Roche 454 pyrosequencing technology (Petrosino et al., 2009)



The main disadvantage of this sequencing technology is unreliable quality of reads containing homopolymers, because the amount of light produced by 8-10 repetitive nucleotides makes it impossible to accurately deduce the length of the homopolymer region.

### Illumina (Solexa)

The Solexa sequencing platform (later acquired by Illumina, Inc) was first developed by British chemists Shankar Balasubramanian and David Kleinerman and then was commercialized in 2006. It is based on the principle of sequencing by synthesis (Zhou et al., 2010). Genomic DNA is first fragmented and adapters are ligated to the both ends. Then, the barcoded and labeled DNA fragments are loaded onto a flow cell, where one end of the DNA fragment hybridizes to a complementary oligonucleotide that is covalently attached to the surface of the cell. The opposite end of each of the single-stranded DNAs hybridizes with the adjacent complementary oligonucleotide. After this, the fragment is amplified during the process called bridge PCR (Fig. 4).

For the next PCR cycle, the template chain and the newly synthesized complementary chain are denatured to start amplification again. After several amplification cycles, millions of dense clusters of duplex DNA are generated in each lane of the flow cell. After that, the cell is ready for sequencing (Fig. 5) (Lizard et al., 2017).

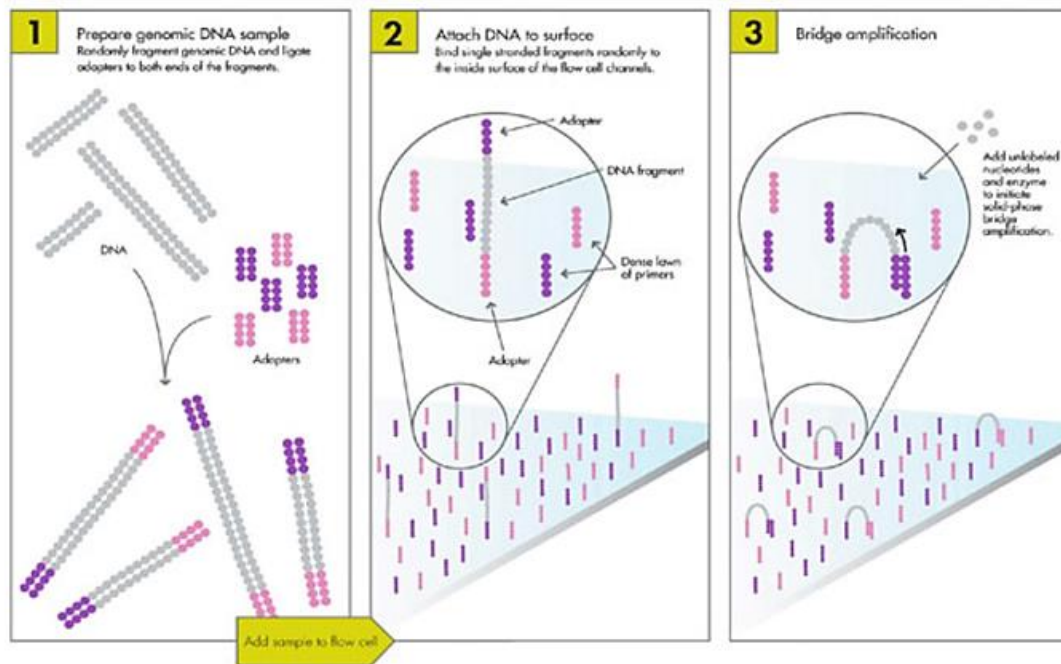


Fig. 4 Steps of library preparation for Illumina sequencing (Zhou et al., 2010)

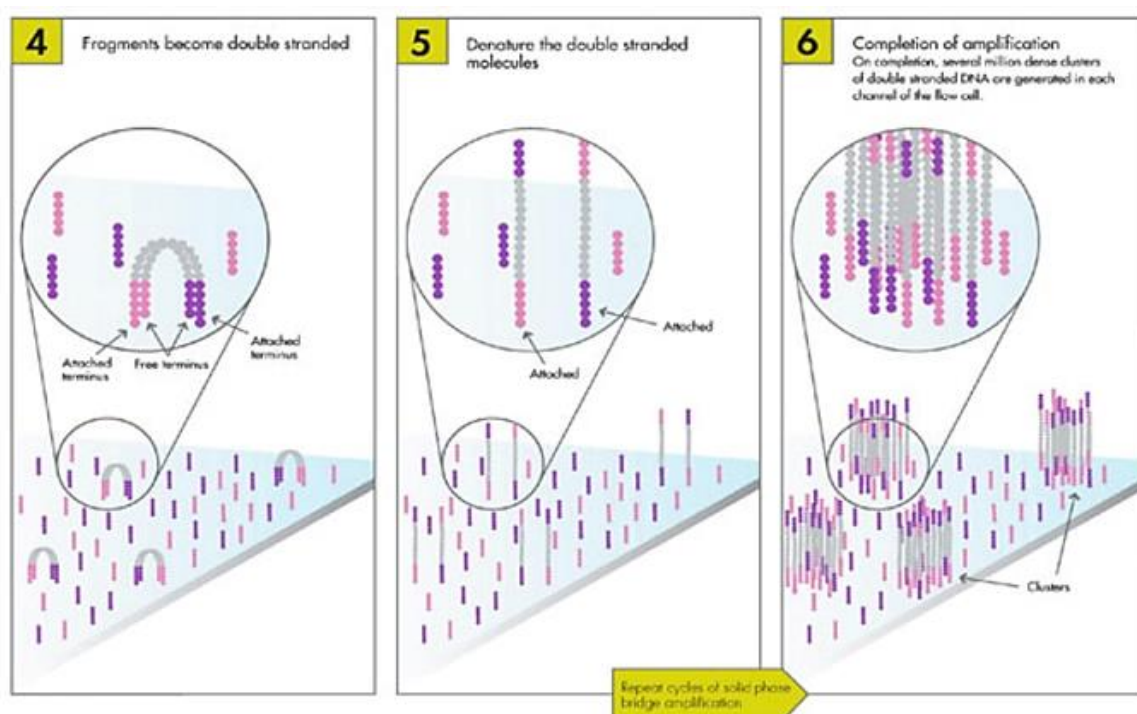
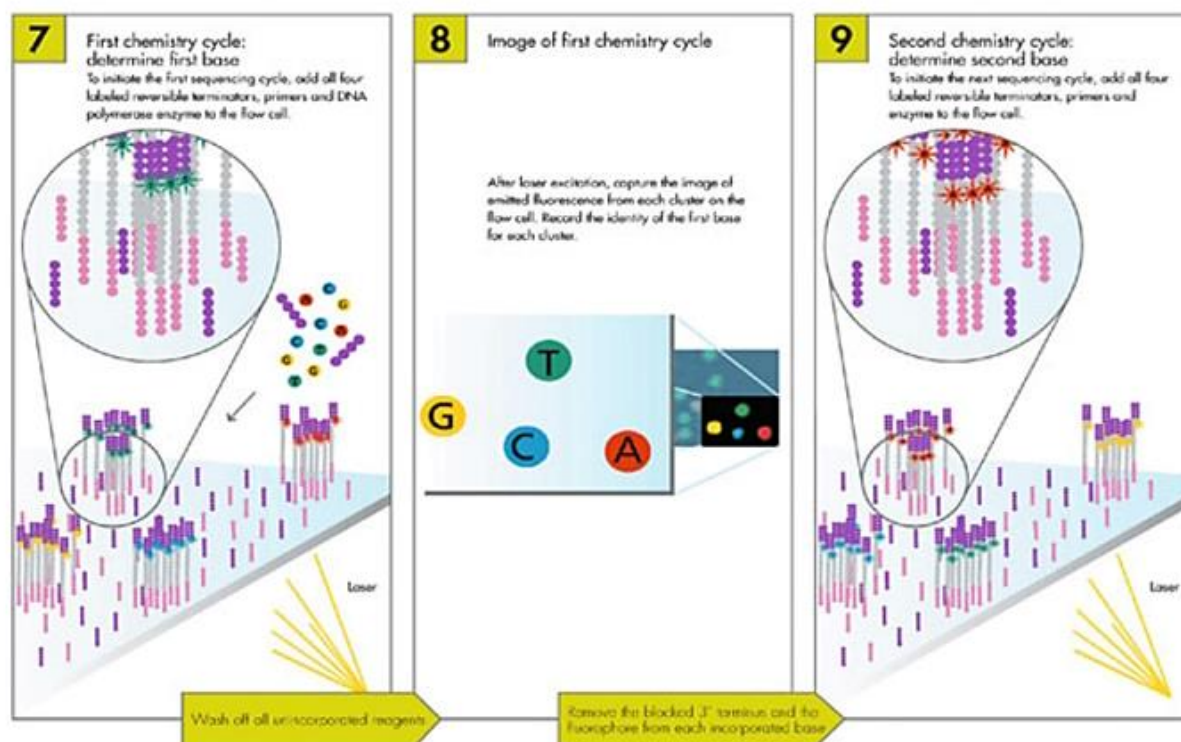


Fig. 5 Clonal amplification by bridge PCR (Zhou et al., 2010)

The amplified DNA is denatured, primers are attached and the second strand synthesis begins with the inclusion of complementary dNTPs. Each dNTP is labeled with different reversible fluorophores. Before the attachment of the next base, Tris-2-carboxyethyl-phosphine (TCEP) is added to remove the fluorophore fragment from the previous dNTP and to remove the block at the 3' end of the nucleotide (Fig. 6).



**Fig. 6** Signal detection during Illumina sequencing (Zhou et al., 2010)

Solexa launched the first commercial genome analyzer in 2006. The device processed up to 1 billion nucleotides in one run. Solexa was acquired by Illumina in 2007, which later developed new devices with increased throughput and ease of use. Currently, Illumina MiSeq sequencers, which are capable of producing up to 5 billion nucleotides per cell with v3 chemistry, are widely used alongside with the other sequencers (NextSeq and NovaSeq). The company later developed the in vitro diagnostics version of MiSeq – MiSeqDx, which was approved for clinical use by the FDA. Illumina sequencers are characterized by an extremely low error rate (less than 1%).

#### Ion Torrent

Ion Torrent sequencing is based on the process of the formation of covalent bonds in a growing DNA chain catalyzed by DNA polymerase. Incorporation of each nucleotide leads to the release of pyrophosphate and hydrogen ion  $H^+$ . This, in turn, decreases the pH of the medium, which is detected in order to determine the sequence of DNA (Malapelle et al., 2015). The sequencing semiconductor chip contains microcells, which are filled with microspheres carrying clonally amplified single-stranded template DNA molecules. Then the wells are sequentially filled with DNA polymerase and unmodified dNTP. If a complementary nucleotide is included in the growing chain, then a hydrogen ion will be released. An ion-sensitive field effect transistor (ISFET) is located below each microcell. These sensors determine the pH change by measuring the potential difference. Each change in pH is recorded. Before the start of the next cycle, unbound dNTP molecules are washed out. The following type of dNTP is added and the cycle repeats (Korlach et al., 2010).

In 2010, Ion Torrent Systems, Inc was founded. The company has developed a sequencer for small studies (Ion Torrent PGM), which is used for targeted sequencing of different sections of the genome, sequencing of small genomes (bacterial and viral), as well as sequencer for more extensive studies - Ion Proton. The advantage of the developed devices is a single platform for preparing libraries for sequencing, which is based on emulsion PCR. Sample preparation can also be automated by using the Ion Chef station, which is also developed by Ion Torrent.

The key advantage of the technology is the relatively low cost of launching the device and the great flexibility of the platform - a significant change in the sequencing protocols and the replacement of reagents is possible, which allows adapting the sequencing to the specific tasks of the researcher. The disadvantages of the technology include a relatively large rate of errors in the sequencing of homopolymer DNA regions. This is because the relation between the change in the pH of the medium and the length of the homopolymer site is nonlinear.

In 2012, Ion Torrent was acquired by Thermo Fisher Scientific, after which an improved version of Ion Proton (Ion GeneStudio S5) was developed. Currently, there are versions of both sequencers designed for clinical use.

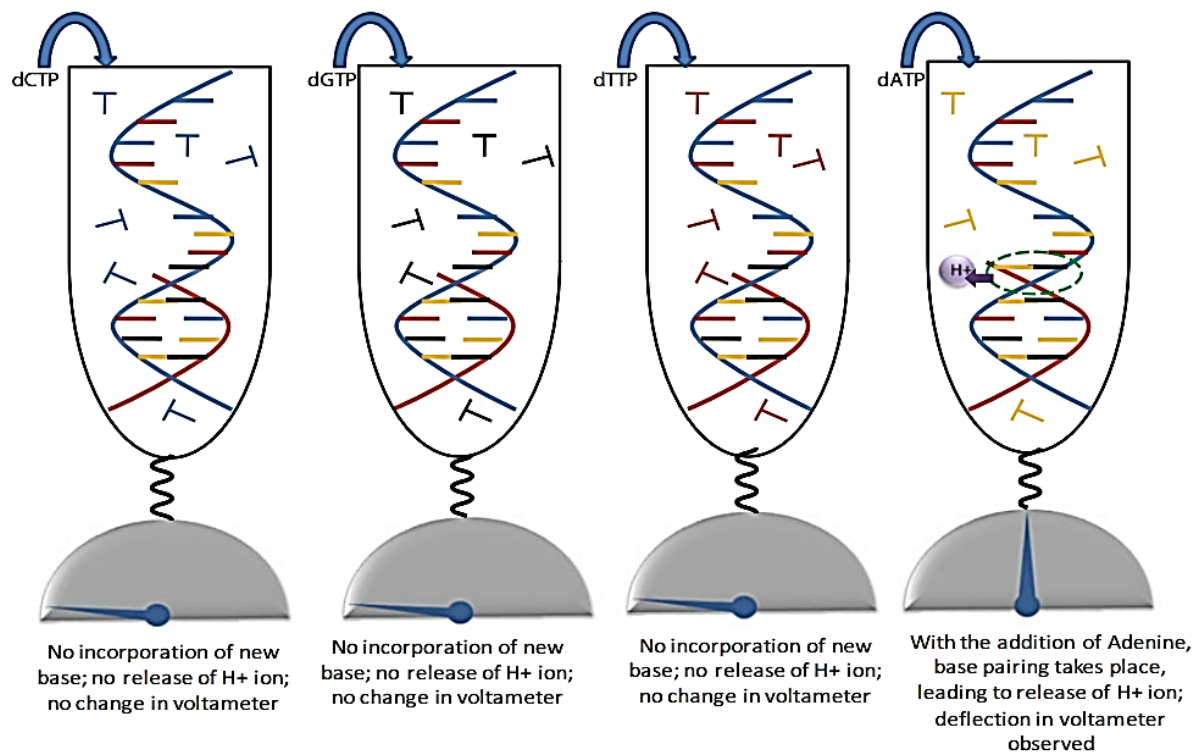


Fig. 7 Ion Torrent semiconductor sequencing principle (Verma et al., 2016)

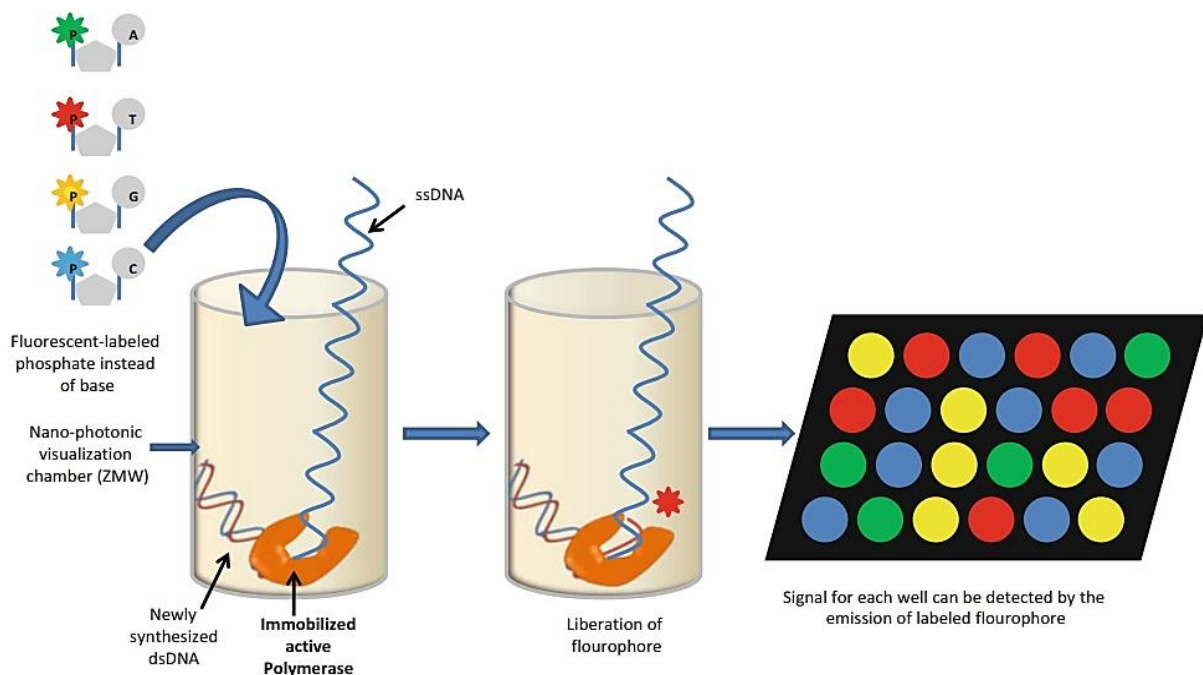


Fig. 8 Single DNA molecule sequencing principle used in Pacific Bioscience RS (Verma et al., 2016)

#### Pacific Bioscience RS (SMRT sequencing)

The SMRT (single molecule real-time) sequencing method was developed by Pacific Biosciences.

This method differs from other methods in two ways:

1. Instead of labeling the nucleic acid bases themselves, the phosphate end of the nucleotide is labeled.
2. The reaction takes place in a nanophotonic cell called ZMW (zero-mode waveguide). The sequencing reaction for the DNA fragment begins with DNA polymerase, which is located in the detection zone at the bottom of each ZMW. The DNA polymerase is fixed at the bottom of the cell and a single-stranded DNA chain is used as a template. When the nucleotide is integrated by DNA polymerase, the fluorescent label is cleaved. The device registers light that is emitted each time the nucleotide is integrated (Rhoads & Au, 2015).

Single molecule sequencing technologies are just beginning to enter the market. Pacific Biosciences is one of the suppliers of equipment and reagents for sequencing separate DNA molecules. The company was created in 2004 at the laboratories of



Cornell University in the United States and in 2011 released its first commercial sequencer PacBio RS. In the spring of 2013, the second version of the device - PacBio RSII - was released.

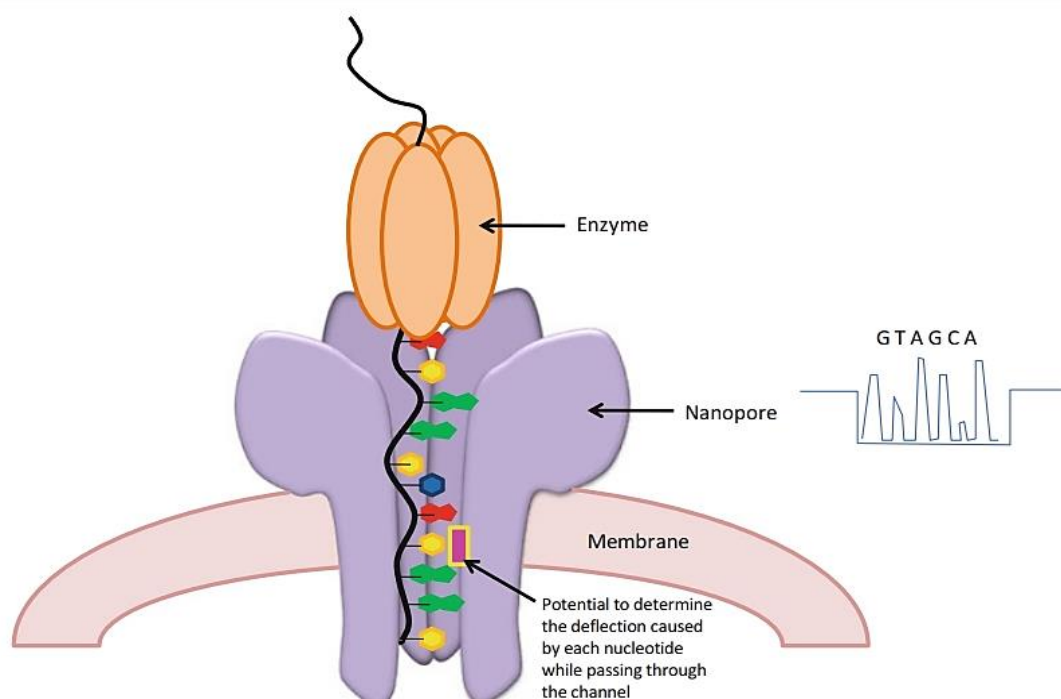
The main advantage of SMRT technology is the extremely long read length (from 8,000 to 30,000 nucleotides). This can significantly reduce the number of reading errors that are associated with PCR amplification of fragments as well as simplify the *de novo* assembly of genomes. The disadvantages of PacBio includes the weight and size of the device (it weighs more than a ton), as well as high cost of sequencing.

### Oxford Nanopore

The concept of nanopores and their use in sequencing was developed in the mid-1990s. After many years of research and development of the technology, Oxford Nanopore licensed it in 2008 (Zascavage et al., 2019). Nanopores are nanometer-wide channels that can be of three types:

- biological: pores that are formed by a pore forming protein in the membrane (e.g. alpha-hemolysin);
- solid: pores that are formed by synthetic material or obtained chemically (e.g. silicon and graphite);
- hybrid: pores that are formed by a biological agent such as a pore-forming protein encapsulated in a synthetic material.

In contrast to all the aforementioned sequencers, Oxford Nanopore does not require labeling or detection of nucleotides. This method is based on the principle of modulation of the ion current during the passage of a DNA molecule through nanopores. Since different nucleotides have different sizes, they block the ion current in different ways for a certain period. Having detected these changes, it is possible to determine the sequence of the necessary molecule (Fig. 9) (Haque et al., 2013).



**Fig. 9** Oxford Nanopore sequencing technology (Verma et al., 2016)

### Sequencing platforms comparison

Table 1 presents the comparative characteristics of different sequencing technologies. The most commonly used and representative sequencing instruments are analyzed.

**Table 1.** Comparative characteristic of different sequencing technologies

Sequencing technology	Device name	Cost of sequencing 1 kbp	Device price	Read length	Accuracy
Sanger	Applied Biosystems ABI PRISM 310	~ \$7000	\$77 000	700-800 bp	99,9999%
454 Roche	454 Genome Sequencer FLX	\$9.2	\$150 000	800 bp	99,9%
Illumina	HiSeq 2500	\$0.06	\$740 000	48-100 bp	99%
	MiSeq	\$0.07	\$125 000	48-300 bp	99%
Ion Torrent	Ion PGM	\$0.61	\$50 000	200-400 bp	99%
	Ion Proton	\$0.1	\$149 000	150-200 bp	99%
Pacific Biosciences	PacBio RS	\$46	\$350 000	Up to 30 000 bp	98%
Oxford Nanopore	Minlon	unknown	\$1 538	Up to 1 million bp	95%



### Methods for studying proteome

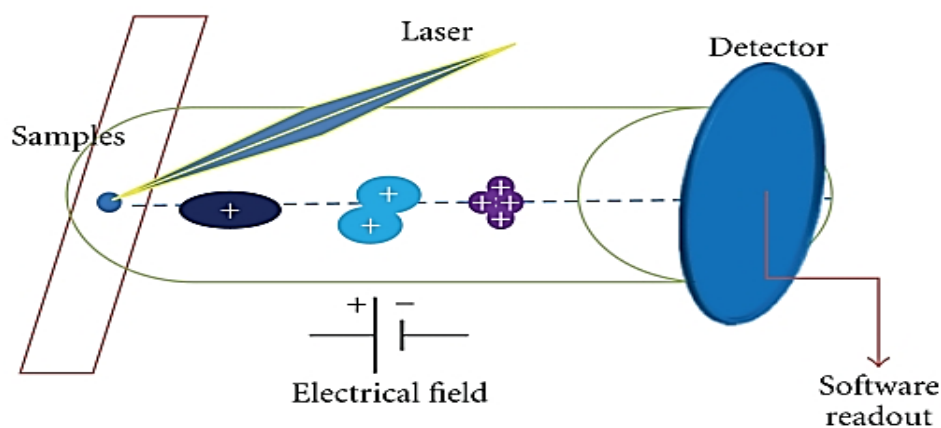
After successfully completing the Human Genome Project, the HUPO (Human Proteome Organization) officially launched the global Human Proteome Project (HPP), which aims to map the entire set of human proteins. The main efforts are aimed at the quantitative analysis of enzymes, their distribution, intracellular localization, as well as interaction with other biomolecules in different physiological conditions. As a general experimental strategy, the HPP research group has focused on three main methods (Omenn et al., 2017):

- mass spectrometric methods;
- methods based on the use of antibodies;
- analysis of proteomic databases.

In this review, the main proteomic platforms used to detect cancer biomarkers will be described.

### Proteomic assays based on mass spectrometry methods

Recently, the level of development of mass spectrometry technology allows it to be used to evaluate the whole human proteome. Mass-spectrometric methods played an important role in the discovery of protein biomarkers of cancer and other diseases (Cho, 2017). Initially, proteomic studies were based on two-dimensional gel electrophoresis with subsequent mass spectrometry. This sequential approach has greatly facilitated the identification of peptide sequences in proteins that were present in different amounts on gels. Subsequently, the analysis of proteomic samples was developed as one of the approaches to the discovery of biomarkers. Recognition of patterns of mass spectra allowed researchers and clinicians to use bioinformatics to diagnose cancer (Li et al., 2017). One of these methods is surface-enhanced laser desorption mass spectrometry (surface-enhanced laser spectrometry - SELDI-TOF-MS). This method can be used to analyze the mass of the protein directly without its enzymatic cleavage. Samples are dried, after which the laser ionizes the crystallized peptides. Then these ions are accelerated by an external electric field and sent to the tube. The detector measures the ions when they reach the end of the tube and the results are processed using specialized software (Fig. 10) (Liu, 2011).



**Fig. 10** SELDI-TOF-MS functional diagram (Liu, 2011)

Mass spectrometry can be combined with liquid chromatographic separation (liquid chromatography separation - LC); this combination is called LC/MS (liquid chromatography separation/mass spectrometry). In the LC/MS assay, whole proteins present in complex biological samples are broken down by enzymes into peptide fragments, and then the LC/MS method is used to identify thousands of proteins in biological samples, such as tissue, serum, plasma, or urine. LC/MS-based methods that can be used for a comprehensive analysis of cleaved peptides are called proteomic shotgun (Adaway et al., 2015; Lopes et al., 2017).

Mass spectrometry is a quantitative method. The advantage of the method is high accuracy and stability. The disadvantage is the relatively high cost of purchasing and maintaining equipment. The detection limit is from 1 pg of protein in 1 µl of liquid.

### Antibody-based proteomic assays

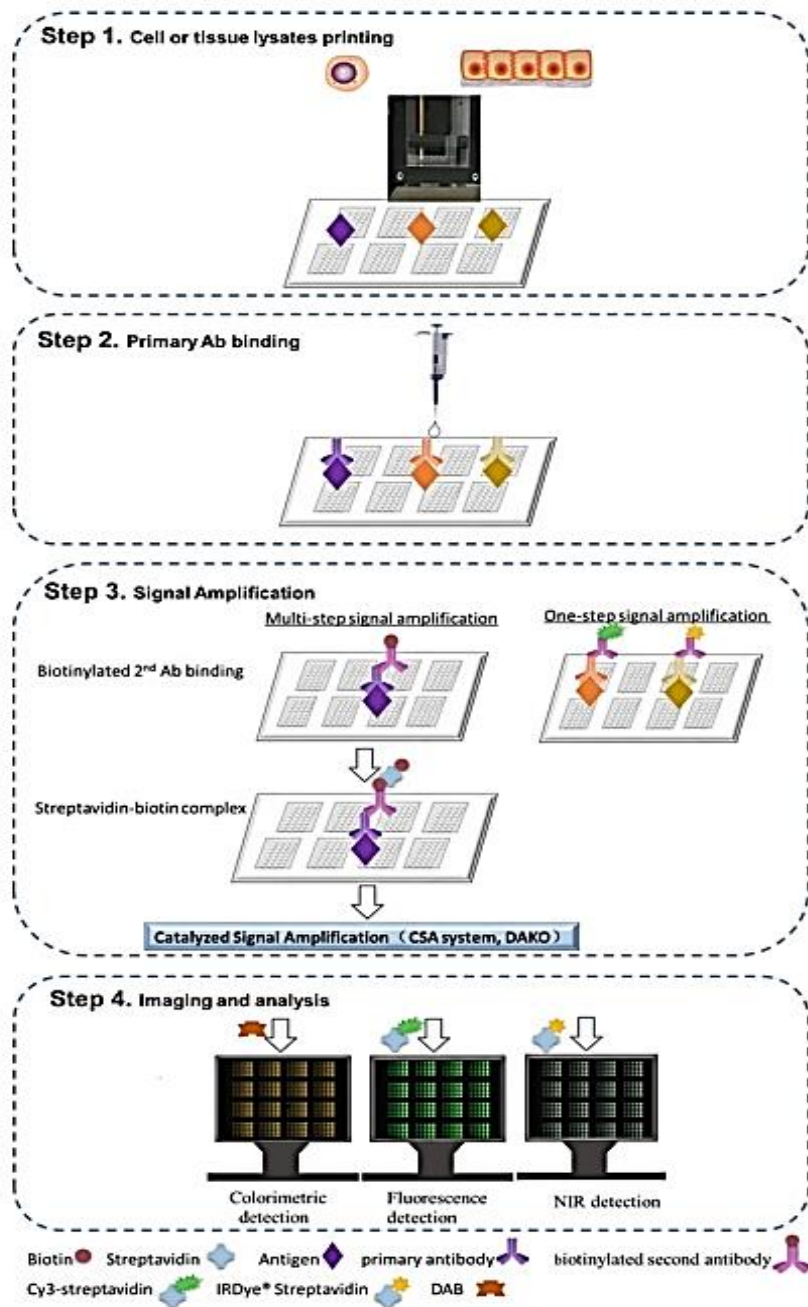
One of the objectives of proteomics is the creation of specific antibodies that can recognize each protein of a human proteome. Antibody proteome analysis plays a key role in the detection and confirmation of cancer biomarkers. In particular, this analysis contributes to the high throughput assessment of cancer biomarkers and provides a logical strategy for the systematic generation and use of specific antibodies for the study of the proteome. The Human Protein Atlas project was created to systematically generate specific antibodies on a global scale and use these antibodies to study the corresponding proteins and protein isoforms (Fagerberg et al., 2014). The use of antibodies for protein profiling on a global scale is an intuitive approach that should facilitate the systematic study of the cancer proteome. Approaches using antibodies can be used in combination with a wide range of high-throughput assays, such as immunohistochemistry (IHC), tissue microarrays (TMA) and protein microchips.

TMA is a method of assembling multiple tissue samples from a single paraffin block for the simultaneous evaluation of several biomarkers using IHC; TMA can potentially become a rapid molecular method of using a large-scale library of antibodies to study the relationship between molecular biomarkers and clinical results (Fagerberg et al., 2014).

Protein microarrays can be divided into two main classes: forward phase protein protein analysis (FPPA) and reverse phase protein protein analysis (RPPA). RPPA is a high-performance, antibody-based method for detecting protein expression in cell or tissue lysates. To some extent, this method is similar to Western blotting (Isik & Ercan, 2017). Western blot has historically been widely used to detect the expression of individual proteins; however, the need for a relatively large number of protein samples

per cycle makes this method unsuitable provided there are limited patient tissue samples for clinical studies. Therefore, there is an urgent need to improve the sensitivity of the detection strategy (Kim, 2017). In addition, to maximize the use of valuable clinical samples, it was necessary to develop high-throughput analysis. RPPA technology provides increased sensitivity, minimal sample requirements and multiplex analysis. RPPA is a promising method in the field of the hypersensitive detection of important proteins or markers in biological or clinical specimens. The advantages of RPPA are the possibility of personalized molecular profiling for patients with an automated high-performance system (Creighton & Huang, 2015).

This technology allows to detect protein samples taken from patients with a limited number of blood cells, or to carry out laser capture of biopsies, cell cultures, serum, urine, synovial fluid and vitreous humor. Depending on the type of microarray, from 20 pg to 1 ng of a protein sample can be used for analysis, and several thousand samples can be analyzed simultaneously on one slide (Fig. 11) (Creighton & Huang, 2015).



**Fig. 11** Reverse-phase protein assay functional scheme (Yuan et al., 2017)

There are various signal detection methods, and the most popular technologies include colorimetric methods, such as fluorescent catalytic signal amplification (CSA) and near infrared (NIR) methods, as shown in Figure 11. The obvious advantage of the colorimetric method is the simplicity of visualizing individual points on slide when a conventional flatbed scanner is sufficient. Fluorescence detection is beneficial in terms of the commercial availability of various fluorescent dyes, as well as high brightness and high sensitivity. NIR detection provides the largest dynamic range (up to 4 orders of magnitude) of the signal-to-noise ratio (Hendry et al., 2018).

Methods for proteomic analysis using antibodies are quantitative methods. The advantages of the method are high accuracy and stability, low cost. The disadvantage can be considered a longer analysis time. The detection limit is from 20 pg of protein in one µl of liquid.

### Proteome databases

Information about the molecular genetic mechanisms of cancer is accumulated on a large scale. The initial goals of large-scale research were aimed at sequencing the entire genome and mapping human transcriptome. Recently, information about human proteomes has attracted increasing attention. The molecular and functional complexity of human proteome creates problems for researchers, and this complexity requires bioinformatic resources specifically designed for collection and integration of currently available data. The Human Proteome Global Project provides a complete atlas of human proteins in their biological context. It generates publicly available data and information resources, which, in turn, further explore the human proteome. "Human Proteome" is built based on a knowledge-based database in order to integrate information obtained from the basic protein research methods described previously. With regard to knowledge-based proteomic approaches, the HPP working group decided to use the UniProtKB/Swiss-Prot, PRIDE, PeptideAtlas, GPMDB and Atlas Protein Atlas databases as main data sources (Thul & Lindskog, 2017).

## Conclusion

Currently, the key results of the Human Genome and Human Proteome projects, as well as a number of studies of genomic and proteomic disorders in various diseases carried out using high-throughput analysis platforms, allow us to obtain fundamentally new information about the manifestation of genomic disorders at higher levels - transcriptomic and proteomic.

Understanding the relationship between the genome and the proteome is necessary to develop new methods for diagnosing diseases associated with genetic disorders, research in the field of systems biology, applied research and development in the field of biotechnology and food products, as well as to create new methods of therapy and introduce new technologies into medical practice.

## Acknowledgments

This research and publication of its results were supported by Ministry of Science and Higher Education of the Russian Federation grant RFMEFI58618X0062 (V.N. Popov).

## References

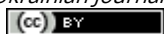
- Adaway, J. E., Keevil, B. G., & Owen, L. J. (2015). Liquid chromatography tandem mass spectrometry in the clinical laboratory. *Annals of Clinical Biochemistry*, 52(1), 18–38. <https://doi.org/10.1177/0004563214557678/>
- Cho, W. C. (2017). Mass spectrometry-based proteomics in cancer research. *Expert Review of Proteomics*, 14(9), 725–727. <https://doi.org/10.1080/14789450.2017.1365604/>
- Creighton, C. J., & Huang, S. (2015). Reverse phase protein arrays in signaling pathways: A data integration perspective. *Drug Design, Development and Therapy*, 9, 3519–3527. <https://doi.org/10.2147/DDDT.S38375/>
- Day, D., & Siu, L. L. (2016). Approaches to modernize the combination drug development paradigm. *Genome Medicine*, 8(1), 115. <https://doi.org/10.1186/s13073-016-0369-x/>
- Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., ... Uhlen, M. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular and Cellular Proteomics*, 13(2), 397–406. <https://doi.org/10.1074/mcp.M113.035600/>
- Haque, F., Li, J., Wu, H. C., Liang, X. J., & Guo, P. (2013). Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA. *Nano Today*, 8(1), 56–74. <https://doi.org/10.1016/j.nantod.2012.12.008/>
- Hendry, S., Byrne, D. J., Wright, G. M., Young, R. J., Sturrock, S., Cooper, W. A., & Fox, S. B. (2018). Comparison of Four PD-L1 Immunohistochemical Assays in Lung Cancer. *Journal of Thoracic Oncology*, 13(3), 367–376. <https://doi.org/10.1016/j.jtho.2017.11.112/>
- Isik, Z., & Ercan, M. E. (2017). Integration of RNA-Seq and RPPA data for survival time prediction in cancer patients. *Computers in Biology and Medicine*, 89, 397–404. <https://doi.org/10.1016/j.compbiomed.2017.08.028/>
- Jackson, S. E., & Chester, J. D. (2015). Personalised cancer medicine. *International Journal of Cancer*, 137(2), 262–266. <https://doi.org/10.1002/ijc.28940/>
- Kamps, R., Brandão, R. D., van den Bosch, B. J., Paulussen, A. D. C., Xanthoulea, S., Blok, M. J., & Romano, A. (2017). Next-generation sequencing in oncology: Genetic diagnosis, risk prediction and cancer classification. *International Journal of Molecular Sciences*, 18(2), 308. <https://doi.org/10.3390/ijms18020308/>
- Kim, B. (2017). Western Blot Techniques. *Methods in Molecular Biology*. Springer New York. [https://doi.org/10.1007/978-1-4939-6990-6\\_9/](https://doi.org/10.1007/978-1-4939-6990-6_9/)
- Korlach, J., Bjornson, K. P., Chaudhuri, B. P., Cicero, R. L., Flusberg, B. A., Gray, J. J., ... Turner, S. W. (2010). Real-time DNA sequencing from single polymerase molecules. *Methods in Enzymology*. Elsevier. <https://doi.org/10.1126/science.1162986/>
- Larijani, B., Perani, M., Alburai'si, K., & Parker, P. J. (2015). Functional proteomic biomarkers in cancer. *Annals of the New York Academy of Sciences*, 1346(1), 1–6. <https://doi.org/10.1111/nyas.12749/>

- Li, X., Wang, W., & Chen, J. (2017). Recent progress in mass spectrometry proteomics for biomedical research. *Science China Life Sciences*, 60(10), 1093–1113. <https://doi.org/10.1007/s11427-017-9175-2/>
- Liu, C. (2011). The Application of SELDI-TOF-MS in Clinical Diagnosis of Cancers. *Journal of Biomedicine and Biotechnology*, 2011, 1–6. <https://doi.org/10.1155/2011/245821/>
- Lizardi, P. M., Yan, Q., & Wajapeyee, N. (2017). Illumina Sequencing of Bisulfite-Converted DNA Libraries. *Cold Spring Harbor Protocols*, 2017(11), pdb.prot094870. <https://doi.org/10.1101/pdb.prot094870/>
- Lopes, A. S., Cruz, E. C. S., Sussulini, A., & Klassen, A. (2017). Metabolomic strategies involving mass spectrometry combined with liquid and gas chromatography. *Advances in Experimental Medicine and Biology*. Springer International Publishing. [https://doi.org/10.1007/978-3-319-47656-8\\_4/](https://doi.org/10.1007/978-3-319-47656-8_4/)
- Malapelle, U., Vigliar, E., Sgariglia, R., Bellevicine, C., Colarossi, L., Vitale, D., ... Troncone, G. (2015). Ion Torrent next-generation sequencing for routine identification of clinically relevant mutations in colorectal cancer patients. *Journal of Clinical Pathology*, 68(1), 64–68. <https://doi.org/10.1136/jclinpath-2014-202691/>
- Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(2), 560–564. <https://doi.org/10.1073/pnas.74.2.560/>
- Omenn, G. S., Lane, L., Lundberg, E. K., Overall, C. M., & Deutsch, E. W. (2017). Progress on the HUPO Draft Human Proteome: 2017 Metrics of the Human Proteome Project. *Journal of Proteome Research*, 16(12), 4281–4287. <https://doi.org/10.1021/acs.jproteome.7b00375/>
- Panis, C., Pizzatti, L., Souza, G. F., & Abdelhay, E. (2016). Clinical proteomics in cancer: Where we are. *Cancer Letters*, 382(2), 231–239. <https://doi.org/10.1016/j.canlet.2016.08.014/>
- Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A., & Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. *Clinical Chemistry*, 55(5), 856–866. <https://doi.org/10.1373/clinchem.2008.107565/>
- Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics*, 13(5), 278–289. <https://doi.org/10.1016/j.gpb.2015.08.002/>
- Ronaghi, M. (1998). A Sequencing Method Based on Real-Time Pyrophosphate. *Science*, 281(5375), 363–365. <https://doi.org/10.1126/science.281.5375.363/>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463/>
- Tanase, C., Albulescu, R., & Neagu, M. (2015). Proteomic Approaches for Biomarker Panels in Cancer. *Journal of Immunoassay and Immunochemistry*, 37(1), 1–15. <https://doi.org/10.1080/15321819.2015.1116009/>
- Thul, P. J., & Lindskog, C. (2017). The human protein atlas: A spatial map of the human proteome. *Protein Science*, 27(1), 233–244. <https://doi.org/10.1002/pro.3307/>
- Verma, M., Kulshrestha, S., & Puri, A. (2016). Genome Sequencing. *Methods in Molecular Biology*. Springer New York. [https://doi.org/10.1007/978-1-4939-6622-6\\_1/](https://doi.org/10.1007/978-1-4939-6622-6_1/)
- Yuan, Y., Hong, X., Lin, Z.-T., Wang, H., Heon, M., & Wu, T. (2017). Protein Arrays III: Reverse-Phase Protein Arrays. *Methods in Molecular Biology*. Springer New York. [https://doi.org/10.1007/978-1-4939-7231-9\\_21/](https://doi.org/10.1007/978-1-4939-7231-9_21/)
- Zascavage, R. R., Thorson, K., & Planz, J. V. (2019). Nanopore sequencing: An enrichment-free alternative to mitochondrial DNA sequencing. *Electrophoresis*, 40(2), 272–280. <https://doi.org/10.1002/elps.201800083/>
- Zhou, X., Ren, L., Meng, Q., Li, Y., Yu, Y., & Yu, J. (2010). The next-generation sequencing technology and application. *Protein and Cell*, 1(6), 520–536. <https://doi.org/10.1007/s13238-010-0065-3/>

---

**Citation:**

Solodskikh, S.A., Gryaznova, M.V., Dvoretzskay, Y.D., Gureev, A.P., Panevina, A.V., Maslov, A.Y., Serzhantova, O.V., Mikhailov, A.A., Chinopoulos, C., Popov, V.N. (2019). Postgenomic technologies for genomic and proteomic analysis in biological and medical research. *Ukrainian Journal of Ecology*, 9(4), 765–776.



This work is licensed under a Creative Commons Attribution 4.0. License