The most common technologies and tools for functional genome analysis

Evelina Gasperskaja,

Vaidutis Kučinskas

Department of Human and Medical Genetics,
Faculty of Medicine, Vilnius University
Vilnius, Lithuania

Since the sequence of the human genome is complete, the main issue is how to understand the information written in the DNA sequence. Despite numerous genome-wide studies that have already been performed, the challenge to determine the function of genes, gene products, and also their interaction is still open. As changes in the human genome are highly likely to cause pathological conditions, functional analysis is vitally important for human health.

For many years there have been a variety of technologies and tools used in functional genome analysis. However, only in the past decade there has been rapid revolutionizing progress and improvement in high-throughput methods, which are ranging from traditional real-time polymerase chain reaction to more complex systems, such as next-generation sequencing or mass spectrometry. Furthermore, not only laboratory investigation, but also accurate bioinformatic analysis is required for reliable scientific results. These methods give an opportunity for accurate and comprehensive functional analysis that involves various fields of studies: genomics, epigenomics, proteomics, and interactomics. This is essential for filling the gaps in the knowledge about dynamic biological processes at both cellular and organismal level. However, each method has both advantages and limitations that should be taken into account before choosing the right method for particular research in order to ensure successful study. For this reason, the present review paper aims to describe the most frequent and widely-used methods for the comprehensive functional analysis.

Keywords: functional analysis, technologies, variants, genomics, transcriptomics, gene expression

Correspondence to: Evelina Gasperskaja, Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Santariškių 2, LT-08661 Vilnius, Lithuania. E-mail: evelina.gasperskaja@mf.vu.lt

INTRODUCTION

Although human genomes are about 99.9% identical, the remaining 0.1% is the reason of difference between people caused by different variants. Since 2003, the complete sequence of human genome, its annotation and increased advancement of sequencing technologies (i. e., Sanger and Next-generation-sequencing; NGS) have provided all the necessary conditions for the identification of all variants in human coding and non-coding sequence (1, 2). Although the technique for variant detection is now becoming a routine, the key question throughout many years concerns the function of detected variants. The resource of important information about functional genomics are several large-scale projects, for instance, the ENCODE project, the main goal of which was to identify all the functional elements, including regulatory elements in both coding and non-coding regions(3). According to another, the 1000 Genomes Project, there are about 20,000–23,000 variants in synonymous and nonsynonymous regions of the human genome. Even though not all of them are functionally meaningful, 530–610 of the variants have functional impact by causing inframe deletions and insertions, premature stop codons, frameshifts, or by disrupting splice sites (4). Despite numerous studies, scientists are still facing a huge challenge in unravelling what the sequence means and in deciding whether or not a found variant is pathogenic. A pathogenic variant can lead to disease or cause a number of disorders. However, understanding of pathogenic mechanisms creates an opportunity to prevent severe consequences by developing novel diagnostic tools and by designing highly effective treatments for the disease (5, 6). To achieve this aim it is necessary to perform large-scale functional genome analysis that involves different fields of study: genomics, epigenomics, transcriptomics, proteomics, and interactomics. In order to describe the functions of genes and proteins as well as to research the relationship between the genotype and the phenotype, a large number of various methods, including model systems (e. g., CRISPR-Cas9), can be used (6–8). However, every method has its advantages and disadvantages (they are summarized in the Table). For this reason, the present paper aims to give a brief overview of the most common technologies and tools that could be applied for functional genome analysis, mainly of transcriptome.

FROM VARIANT DETECTION TO FUNCTIONAL GENOME ANALYSIS

Variants in coding as well as non-coding genome sequence range from single nucleotide changes to large, microscopically visible, chromosomal aberration. These variants may have a huge impact on the function of gene. They can be either beneficial (e. g., single nucleotide polymorphism; SNP) with no negative effect on the phenotype, or pathogenic (e. g., nonsense variant) – resulting in a number of different disorders and diseases (9, 10). Depending on the variant type and locus, there are numerous different genetic methods and tools for the variant detection. For example, due to its simplicity the most frequent method for the analysis of a large (>5 Mb) chromosomal aberration is karyotype analysis by using the GTG banding technique (11). Other molecular genetic methods, such as microarray-based comparative genomic hybridization (aCGH) or fluorescent *in situ* hybridization (FISH), should be applied for a more accurate analysis. However, these methods have some significant limitations: the aCGH does not detect mosaicism, balanced translocations and inversions, while the FISH requires specific probes (12, 13). Moreover, for particular variant detection another molecular genetic methods might be applicable, which include restriction enzyme assay, Multiplex ligation-dependent probe amplification (MLPA), even though many of the tests are based on the Polymerase chain reaction (PCR) and its variants (e. g., multiplex PCR) (14).

Although researchers can easily plan their assay in the case of particular variants, they are facing some challenges in the study of unspecified variants (6). Sequencing is considered to be the "gold standard" method for the identification of known as well as unspecified variants in the genomic DNA. In accordance with the previous statement, the Sanger or Next-Generation Sequencing (NGS) techniques can be used (15, 16). The concept behind these two methods is similar. During the polymerase chain reaction, which consists of several cycles of sequential DNA replication, DNA polymerase catalyzes the complementary incorporation of fluorescently-labeled deoxyribonucleoside 5'-triphosphates (dNTPs) into the DNA template. For each cycle, a colour of the labeled DNA fragment is recorded by a detector, thus determining nucleotide in the sequence.

The main difference between the conventional (i. e., Sanger) technology and the NGS is that the latter is not limited to a single DNA fragment but analyzes millions of fragments in massively parallel sequencing technology (17, 18). These two sequencing methods are widely used all over the world. Even so, it is considered that in a small-scale project it is more eligible to use the Sanger sequencing system because of its accuracy. On the other hand, in large-scale projects this research method would be expensive and time-consuming, therefore the NGS needs to be applied (19, 20). General progress in technology achieved in some strategies of the next-generation DNA sequencing has a huge impact on genetic research (2). Recently, the most widely used platforms have been Roche/454 Life Science, Applied Biosystems SOLiD, and Illumina Genome Analyzer. Another DNA sequencing technology has been lately developed by Ion Torrent. Nevertheless, "sequencing-by-synthesis" used by Illumina currently is one of the most popular NGS platform. First of all, a randomly fragmented DNA is ligated with specific adaptors and amplified by the use of PCR. Secondly, a performed DNA library should be immobilized on the beads or arrays, thus generating clusters of identical DNA fragments. These clusters are then read by sequential cycles of nucleotide incorporation, washing, and detection, where the number of the cycles eventually determines the read length (7, 19–21). In order to understand the genome structure, function, or evolution, it is not enough to obtain the DNA sequencing data through the NGS: but there is also a need for deep and precise analysis using bioinformatics approaches. The key path to successful sequence analysis is to align the sequence of interest with another sequence whose function is known (usually termed as the reference genome). It might be very useful when the gene function is unknown but is evolutionary related to another gene whose function is defined. In such a case, it can be suspected that the unknown gene has the same or similar function. Furthermore, the sequences might be scanned in order to find the significant matches between the components of a sequence that have been previously described as having a huge impact on the genomics function (6, 22). In order to compare the data, it is necessary to search for information in different biomedical databases. One of the biggest sources of biomedical and genomic information is the NCBI (National Center for Biotechnology Infor-

mation), which provides access to other databases such as PubMed, Entrez Gene, OMIM, Variation Viewer, dbSNP, and others (23).

EPIGENOMICS

For functional analysis, it is important to take epigenetic modifications such as DNA methylation and histone modifications into account, because they affect gene expression without any changes in the underlying DNA sequence (24). DNA methylation, which usually occurs in the context of densely situated CpG dinucleotides (i. e., CpG islands), correlates with transcriptional suppression. In order to detect DNA methylation status, unmethylated cytosines are converted into uracil by using sodium bisulfite, because methylated cytosine is resistant to this impact. Additionally, methylation-dependent restriction enzymes (MDRE) are highly effective for DNA methylation analysis. These enzymes, e. g., *HpaII* and *MspI*, recognize and simply digest the methylated DNA. Usually, MDRE or even more frequently used bisulfite conversion is the first step for many subsequent methods such as methylation-specific PCR, sequencing, bead array, etc. (25–28).

Histone modifications – acetylation, phosphorylation, methylation, ubiquitination, and others – are another cause of epigenetically-regulated genes. Depending on the modification type and locus, gene expression can be either activated or repressed (29). The most common method for the investigation of histone modifications is chromatin immunoprecipitation (ChIP) based on the interaction between the antigen (of associated with DNA target protein) and the antibody (specific to target modified protein). After the precipitation, the genomic DNA is released for further research hinged on microarray analysis (ChIP-chip), sequencing (ChIP-seq), or quantitative PCR. Although these methods are high-throughput, the dependence on a specific antibody sometimes limits the use of the ChIP (21, 30).

TRANSCRIPTOMICS

When the human genome was fully sequenced, the focus of attention shifted towards identifying and annotating its functional DNA elements, including those that regulate gene expression.

Identification of such elements is a vitally important step towards elucidating pathogenic pathways that affect human health (3, 6).

All the RNA-level processes, including transcription activation or inhibition, mRNA processing, and its transport are regulated by different functional elements of the genomic DNA. Nevertheless, the highest regulation occurs at the transcriptional initiation level through several regulative elements, which are called the *cis*-acting regulatory sequence and *trans*-factors (6, 31). *Trans*-factors such as transcription factors (TF), activators, and repressors (including co-activators and co-repressors) interact with specific DNA regions, i. e., *cis*-acting regulatory sequence that includes core promoter (with a TATA box and other binding elements), proximal promoter, enhancer, silencer, insulator, and locus control region (LCR). Investigation of these regulatory elements may be a challenge for the scientists because of the difficulties in identifying the position of transcription start sites (TSSs) and transcription factors binding sites (TFBSs) in the core promoter. However, there are several experimental and bioinformatical approaches (31). First of all, a comparative bioinformatical approach is necessary for the study of the regulatory elements. This type of research is usually based on constructing alignments between orthologous sequences because sequence homology provides valuable evidences to gene function analysis (6, 32). Nevertheless, a deeper understanding of regulatory elements requires laboratory investigations. It is believed that every TFBS could be detected by the above-mentioned ChIP method. Theoretically, depending on immunoprecipitation of the target protein, the core promoters, enhancers, silencers, insulators and LCRs could be determined (31). Furthermore, epigenetic markers can be helpful in detecting TSSs in the core promoter and enhancer loci, because TSSs of actively transcribed genes are marked by H3K4me3 and H3K27ac, while enhancers by H3K4me1 and H3K27ac (33). Another very frequent functional assay of the regulatory element is based on the transgenesis of a specific reporter-gene (e. g., the gene of the green fluorescent protein – GFP or luciferase) into the target regulatory sequence. After the translation, activity of the reporter-gene is measured, e. g., by fluorescence of the GFP, with the purpose to determine if the examined region contains elements that alter reporter-gene expression (31, 32).

Substantial information about functional genomics can be obtained through the analysis of the messenger RNA (mRNR) or cDNA, which is copied from the mRNA by reverse transcription PCR. Therefore researchers often choose to test the mRNR or cDNA rather than DNA, because RNA analysis may be more eligible for a gene that has many small exons and it can also reveal abnormal splicing (6). For many years there have been some standard methods for measuring the mRNR expression: Northern blotting, serial analysis of gene expression (SAGE) as well as quantitative real-time PCR (qPCR) among them. The SAGE method is based on the conversion of an RNA molecule into a short unique tag, while Northern blotting – on hybridization with radioactive probe. This allows to perform quantitative analysis by counting the number of tags and measuring intensity of band, respectively. However, both these methods are characterized as low-throughput (34, 35). Nevertheless, for the mRNA quantitation and gene expression evaluation the "gold standard" is qPCR, which is fast, very sensitive, and highly reproducible. The principle of this method is that during the reverse transcriptional reaction, complementary single-stranded cDNA from the RNA template is synthesized. The cDNA is necessary for subsequent use in quantitative PCR (36). The aim of this reaction is to measure fluorescence intensity that is directly proportional to the amount of cDNA in the sample (37). There are two strategies for qPCR data analysis: absolute quantification (based on the calibration curve) and relative quantification (based on the comparison with reference sample) (38). For the relative gene expression level calculation, the most convenient way is comparative CT (or $2^{-\Delta\Delta CT}$) method. This method relies on comparing the CT values of the target and reference samples, using a reference (endogenous housekeeping) gene as the normalizer. Finally, the method results in the fold change of target gene expression relative to a reference sample, normalized to a housekeeping gene (39).

Acceleration of high-throughput technologies such as cDNA microarray and RNA sequencing (RNA-seq), which also provides the possibility of transcriptional characterization, very often replaces preceding methods (40, 41). Results obtained by

a cDNA microarray assay provide important genome-wide information about the changes of gene expression in various cell lines and in different stages of development. This method is based on hybridization of fluorescently labelled cDNA with the particular oligonucleotides (probe) on the specific microarray. The amount of hybridization recorded for a specific probe is proportional to the number of DNA fragments in the sample. In this way, the obtained absolute hybridization values give an opportunity to detect genetic variation in the human genome (41, 42). Despite the great advantages of cDNA microarray, high-throughput RNA sequencing based on different NGS systems is also increasingly used. The RNA sequencing results in a number of short reads. Aligned to a reference genome, they produce a specific transcription map that corresponds to the transcriptional structure and gene expression level (43). It means that this technique is appropriate for gene, transcripts (including alternative gene spliced transcripts), or allele-specific expression identification. Moreover, it is possible to accurately measure translation of transcripts. As each method has both advantages and disadvantages, the last one is not an exception. The problems in RNA-seq are often related with high sequence similarity between alternative spliced isoforms or difficulties in data analysis (44, 45).

PROTEOMICS AND INTERACTOMICS

From the functional point of view, analysis of proteomics and interactomics is as vitally important as previously described analysis of genomics, epigenomics, and transcriptomics, because some studies show that gene expression at DNA or mRNA levels is substantially unchanged, although it affects the protein function and vice versa (46, 47). Proteins perform a vast array of functions within organisms, though abnormal protein expression that occurs due to post-transcriptional modifications or protein interaction with another protein or nucleic acids disrupts cell function (48).

Depending on the intent of the experiment, there are two well-known strategies for protein quantification: immunoassays or antibody-free detection methods. Immunoassay, such as the enzyme-linked immunosorbent assay (ELISA), is a widely-used method due to its high sensitivity and strong specificity (49). However, sometimes researchers can face the problem when no antibody exists for the protein of interest. In such cases the solution is antibody-free methods. Firstly, compared to one-dimensional protein separation method, two-dimensional gel electrophoresis (2-DE), which separates protein by two properties in 2D gels is more effective (50, 51). However, the most common and comprehensive analytical tool for protein detection, identification, and quantification is mass spectrometry (MS) that measures mass-to-charge (m/z) ratio of ions. Advancement of MS gives an opportunity to achieve a greater throughput of samples with high accuracy and precision. Additionally, it is considered that MS methodology is rapid and reliable for large-scale studies (52–54). Furthermore, due to its advantages MS is very often combined with another technique. For instance, some studies consist of antibody-based purification and mass spectrometry analysis termed mass spectrometric immunoassay (MSIA) (55).

An important step towards characterization of the protein function is the identification of the protein interaction network consisting of different proteins. The most frequent system for detection of interacting proteins in living yeast cells is the two-hybrid system (Y2H). The aim of such investigation is to create genetically modified yeast strains on a selective medium. In such a system, two interacting proteins bound to specific domains switch on polymerase II, which subsequently activate the transcription of a reporter gene, whose transcription leads to a specific phenotype (e. g., changed colour) (56). Furthermore, proteins interact also with nucleic acids, DNA and RNA. In functional approach, the most important interactions are between DNA and transcription factors or regulatory elements. In the case of RNA, it is necessary to test interactions between this nucleic acid and ribosome, or other RNA binding proteins. The analysis of both DNA-protein and RNA-protein interactions is based on similar techniques (57, 58). Previously mentioned high-throughput immunoprecipitation of the nucleic acid and protein complex is increasingly becoming the method of choice for the detection of TFBSs and histone modification. Subsequent microarray, or NGS analysis, enables the identification of a particular locus, i. e., the region that is

specifically interacting with the protein of interest. However, the main limitation of the ChIP method is the dependence on antibody specificity (59).

FUNCTIONAL GENOMICS INTEGRATING MODEL SYSTEMS

Nowadays, by the use of high-throughput sequencing technologies it is possible to generate detailed catalogue of genetic variation. However, the main question concerns the relationship between the geno-type and phenotype. In order to answer this question, it is not possible to perform functional research directly on human beings due to some bioethical aspects. So there should be applied experimental studies of model systems such as *in vitro* cell culture or animal models for functional interpretation of genome sequence variants (60).

Animal models have long been applied in different studies for the investigation of biological and pathogenical mechanisms, as well as for the development of effective treatment. Depending on the purpose of study, different animal models can be used, although the mouse (*Mus musculus*), the fruit fly (*Drosophila melanogaster*), and the zebrafish (*Danio rerio*) are the most commonly used in functional genome studies. This study has numerous advantages. For example, mutation can be induced artificially and the mutant phenotype can be recognized easily, genes can be cloned using standard procedures, the animal produces a large number of offspring in a relatively short period of time (61). There are two main strategies of using animal model: "knock out" – suppression of the gene of interest, "knock in" – incorporation of the same mutation that is observed in a human. For instance, a number of studies have been conducted for creating animal models of human diseases by chemical mutagenesis (e. g., *N*-ethyl nitroso-urea; ENU) that causes random allelic point mutations in mice. However, the major limitation of animal models is the phenotype that often does not reflect human beings (62, 63).

At this very moment, a promising technology for obtaining more information about human diseases is the CRISPR-Cas9 (Clustered Regularly Interspaced Short Polindromic Repeats/CRISPR-associated) system, which was found to be a prokaryotic immune system against viruses. This system consists of a small cluster of *cas* genes (encoding CRISPR associated proteins) and specific DNA sequence, called CRISPR locus, which comprises short repeats that are separated by unique spacers (64). During a virus infection, its unique spacer integrates into the bacterial CRISPR locus. Subsequently, this locus is transcribed into precursor CRISPR RNA (pre-crRNA). After the processing, the mature crRNA can recognize and destroy target nucleic acid by interacting with Cas proteins (65). So CRISPR locus contains information about previous virus infections, thus giving an opportunity for bacteria to recognize and inactivate the virus in case of re-infection. Currently, some scientific studies show that it is possible to engineer the protein and RNA components of bacterial CRISPR system in order to recognize and cut DNA at the desired locus (66). Due to these properties, there is a possibility of applying this system *in vitro* in the human cell line, in order to study human diseases without any negative consequences (67).

CONCLUSIONS

It is believed that successful functional genome analysis discovers genetic basic for human health by filling the gaps in knowledge about pathogenic pathways between genes, proteins, and their interaction network. There are a lot of different methods and tools for accurate functional analysis. Despite huge analytical progress, these methods have certain limitations (see the Table). Thus, in order to extend the limits of current techniques, some high-throughput technologies such as quantitative real-time polymerase chain reaction, next-generation sequencing or mass spectrometry have been developed, which provide an opportunity to perform genome-wide functional analysis. Furthermore, model systems such as CRISPR-Cas9 or animal models are required for an extensive functional interpretation of genome sequence variants. However, in processing large amounts of data researchers are still facing the problem, that usually is very complicated and time consuming. For this reason, there is a need of continuous improvement in technology and development of more efficient analytical tools. It should be noted that for more comprehensive results it is essential to use complex methodologies that complement each other's shortcomings.

Table. Summary of the main advantages and limitations of the most common technologies used for functional genome analysis

Technique	Advantages	Disadvantages	References
Variants detection methods			
GTG band-ing	Effortless analysis of the chromosome number and structure, including balanced rearrangements	Low sensitivity and resolution (5–10 Mb)	(11–13)
FISH	Detection of minor structural cytogenetic abnormalities	Based on probes annealing to specific target	
aCGH	High sensitivity and specificity	Inappropriate for the detection of balanced chromosomal rearrangements	
Sanger	High throughput, quality and reproducibility	Time consuming for large-scale projects	(2, 19)
NGS	Does not require a priori knowledge about genomic features Requires low amount of DNA/RNA as input	Expensive equipment. Complicated data analysis in the case of unspecified variants	
Epigenomics			
Bisulfite conversion	Resolution at the DNA level. Effective method providing information about cytosine methylation	Impossible to distinguish methylated and hemimethylated cytosine	(26–28)
MDRE	Easy to use Availability and assortment of endonucleases	DNA methylation assay is circumscribed by the use of a particular enzyme	(28)
ChIP (including ChIP-chip, ChIP-seq)	Fast and well-studied. Compatible with array- or sequencing-based analysis, i. e., it is possible to perform genome-wide analysis	Relies on antibody specificity Microarray assay relies on particular probes	(21)
Transcriptomics			
Northern Blot	Quantitative and inexpensive method. No specialized equipment is needed There is a possibility of accurate display of the size and amounts of small RNA	Radioactive probes Lower sensitivity and lower throughput	(34)
SAGE	Direct and quantitative method. *A priori* knowledge about the gene sequences is not required. SAGE library requires a small amount of RNA as input. Simple data analysis	Low-throughput	(35)
qPCR	Fast, accurate, sensitive and highly reproducible method for mRNA quantification. Ability to detect the amount of mRNR in real time	The risk of bias	(36–38)
cDNA microarray	Well-studied, high-throughput and quantitative method Based on fluorescence (no need of radioactive probes)	Complicated data analysis	(39, 40)

Table. (Continued)

Technique	Advantages	Disadvantages	References
Transcriptomics			
RNA-Seq	Direct, quantitative and high through-put method. Does not require a priori knowledge about the genomic features. Appropriate for gene, transcripts (including alternative gene spliced transcripts) or allele-specific expression identification	High sequence similarity between alternative spliced isoforms	(21, 42)
Transgenesis of reporter gene	"Gold standard" and accurate method for functional analysis of regulatory elements. Gene expression is easily detectable by fluorescence	Regulatory elements are widely dispersed through the genome that may cause some difficulties in detection	(31, 32)
Proteomics and interactomics			
ELISA	High sensitivity and specificity	Relies on antibody specificity	(47)
2-DE	Efficiently separates proteins by two properties	Poor separation of highly hydrophobic proteins. Inability to analyze very large or very small proteins	(48, 49)
MS	High-throughput method that rightly identifies and quantifies proteins	The sample should be high-quality and homogenous. Sometimes the dissociation efficiency of complex protein is lower	(50, 52)
Y2H	The two-hybrid technique is relatively simple. Appropriate for the first step in identifying interacting protein partners	The rate of false positive results is relatively high. The need of confirmatory test. Impossible interaction between two proteins at a time	(7)
Model systems			
Chemical mutagenesis (in animal model)	Mutation can be induced artificially and mutant phenotype can be recognized easily. Genes can be cloned using standard procedures	Phenotype not always reflects human beings	(61–63)
CRISPR-Cas9	The possibility to engineer the protein and RNA components of bacterial CRISPR system in order to recognize and cut DNA at desired locus	Work requires highly sterile conditions	(66)

References

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004; 431: 931–45.

2. Kircher M, Kelso J. High-throughput DNA sequencing-concepts and limitations. Bioessays. 2010; 32(6): 524–36.

3. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489(7414): 57–74.

4. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature. 2010; 467 (7319): 1061–73.

5. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum Genet. 2013; 132: 1077–30.

6. Strachan T, Read AP. Human molecular genetics. 4th ed. New York: Garland Science/Taylor & Francis Group; 2011.

7. Bunnik EM, Le Roch KG. An Introduction to Functional Genomics and Systems Biology. Adv Wound Care. 2013; 2(9): 490–8.

8. Shen H, McHale CM, Smith MT, Zhang L. Functional genomic screening approaches in mechanistic toxicology and potential future applications of CRISPR-Cas9. Mutat Res Rev Mutat Res. 2015; 764: 31–42.

9. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006; 7(2): 85–97.

10. Conrad DF, Hurles ME. The population genetics of structural variation. Nat Genet. 2007; 39(7 Suppl): S30–6.

11. Barkholt L, Flory E, Jekerle V, Lucas-Samuel S, Ahnert P, Bisset L, et al. Risk of tumorigenicity in mesenchymal stromal cell-based therapies-bridging scientific observations and regulatory viewpoints. Cytotherapy. 2013; 15(7): 753–9.

12. Bishop R. Applications of fluorescence in situ hybridization (FISH) in detecting genetic aberrations of medical significance. Bioscience Horizons. 2010; 3(1): 85–95.

13. Bejjani BA, Shaffer LG. Application of Array-Based Comparative Genomic Hybridization to Clinical Diagnostics. Journal of Molecular Diagnostics. 2006; 8(5): 528–33.

14. Sudbery P. Human molecular genetics (Cell and Molecular Biology in Action Series). Essex: Addison Wesley Longman Limited. 1998.

15. Bakker E. Is the DNA sequence the gold standard in genetic testing? Quality of molecular genetic tests assessed. Clin Chem. 2006; 52(4): 557–8.

16. Schuster SC. Next-generation sequencing transforms today's biology. Nat Methods. 2008; 5(1): 16–8.

17. Ihle MA, Fassunke J, König K, Grünewald I, Schlaak M, Kreuzberg N, et al. Comparison of high resolution melting analysis, pyrosequencing, next generation sequencing and immunohistochemistry to conventional Sanger sequencing for the detection of p.V600E and non-p.V600E BRAFmutations. BMC Cancer. 2014; 14: 13.

18. Nakazato T, Ohta T, Bono H. Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive. PLoS One. 2013; 8(10): e77910.

19. Shendure J, Hanlee J. Next-generation DNA sequencing. Nat Biotechnol. 2008; 26(10): 1135–45.

20. Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. Genomics. 2008; 92(5): 255–64.

21. Shendurel J, Aiden EL. The expanding scope of DNA sequencing 2012. Nat Biotechnol. 2012; 30(11): 1084–94.

22. Brooker R. Genetics analysis and principles. 3rd ed. Minneapolis: University of Minnesota; 2009.

23. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2007; 35 (Database issue): D5–12.

24. Ko YA, Susztak K. Epigenomics: The science of no-longer-"junk" DNA. Why study it in chronic kidney disease? Semin Nephrol. 2013; 33(4): 1–15.

25. Zilberman D, Henikoff S. Genome-wide analysis of DNA methylation patterns. Development. 2007; 134(22): 3959–65.

26. Genereux DP, Johnson WC, Burden AF, Stöger R, Laird CD. Errors in the bisulfite conversion of DNA: modulating inappropriate- and failed-conversion frequencies. Nucleic Acids Res. 2008; 36(22): e150.

27. Kurdyukov S, Bullock M. DNA methylation analysis: choosing the right method. Biology. 2016; 5(1): 1–21.

28. Laird PW. Principles and challenges of genome-wide DNA methylation analysis. Nat Rev Genet. 2010; 11(3): 191–203.

29. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011; 21(3): 381–95.

30. Collas P. The current state of chromatin immunoprecipitation. Mol Biotechnol. 2010; 45(1): 87–100.

31. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet. 2006; 7: 29–59.

32. Loots GG. Genomic identification of regulatory elements by evolutionary sequence comparison and functional analysis. Adv Genet. 2008; 61: 269–93.

33. Kimura H. Histone modifications for human epigenome analysis. J Hum Genet. 2013; 58(7): 439–45.

34. Pall GS, Hamilton AJ. Improved northern blot method for enhanced detection of small RNA. Nat Protoc. 2008; 3(6): 1077–84.

35. Hu M, Polyak K. Serial analysis of gene expression. Nat Protoc. 2006; 1(4): 1743–60.

36. Freeman WM, Walker SJ, Vrana KE. Quantitative RT-PCR: pitfalls and potential. Biotechniques. 1999; 26(1): 112–22, 124–5.

37. Pabinger S, Rodiger S, Kriegner A, Vierlinger K, Weinhausel A. A survey of tools for the analysis of quantitative PCR (qPCR) data. Biomolecular Detection and Quantification. 2014; 1(1): 23–33.

38. Filion M. Quantitative Real-time PCR in Applied Microbiology. Norfolk: Caister Academic Press; 2012.

39. Rao X, Huang X, Zhou Z, Lin X. An improvement of the 2^(−delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. Biostat Bioinforma Biomath. 2013; 3(3): 71–85.

40. Smith CJ, Osborn AM. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. FEMS Microbiol Ecol. 2009; 67(1): 6–20.

41. Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. BMC biology. 2011; 9: 34.

42. Gresham D, Dunham MJ, Botstein D. Comparing whole genomes using DNA microarrays. Nat Rev Genet. 2008; 9(4): 291–302.

43. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008; 18(9): 1509–17.

44. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009; 10(1): 57–63.

45. Williams AG, Thomas S, Wyman SK, Holloway AK. RNA-seq data: challenges in and recommendations for experimental design and analysis. Curr Protoc Hum Genet. 2014; 83: 11.13.1–20.

46. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 1999; 19(3): 1720–30.

47. Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, Aebersold R. Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. Mol Cell Proteomics. 2002; 1(4): 323–33.

48. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. 2008; 582(14): 1977–86.

49. Prieto JM, Balseiro A, Casais R, Abendano N, Fitzgerald LE, Garrido JM, et al. Sensitive and specific enzyme-linked immunosorbent assay for detecting serum antibodies against Mycobacterium avium subsp. paratuberculosis in fallow deer. Clin Vaccine Immunol. 2014; 21(8): 1077–85.

50. Rabilloud T, Lelong C. Two-dimensional gel electrophoresis in proteomics: a tutorial. J Proteomics. 2011; 74(10): 1829–41.

51. Bunai K, Yamane K. Effectiveness and limitation of two-dimensional gel electrophoresis in bacterial membrane protein proteomics and perspectives. J Chromatogr B Analyt Technol Biomed Life Sci. 2005; 815(1–2): 227–36.

52. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng. 2009; 11: 49–79.

53. Stanczyk FZ, Clarke NJ. Advantages and challenges of mass spectrometry assays for steroid hormones. J Steroid Biochem Mol Biol. 2010; 121(3–5): 491–5.

54. van Duijn E. Current limitations in native mass spectrometry based structural biology. J Am Soc Mass Spectrom. 2010; 21(6): 971–8.

55. Nelson RW, Krone JR, Bieber AL, Williams P. Mass spectrometric immunoassay. Anal Chem. 1995; 67(7): 1153–8.

56. Bruckner A, Polge C, Lentze N, Auerbach D, Schlattner U. Yeast two-hybrid, a powerful tool for systems biology. Int J Mol Sci. 2009; 10(6): 2763–88.

57. Ascano M, Gerstberger S, Tuschl T. Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. Curr Opin Genet Dev. 2013; 23(1): 20–8.

58. Helwa R, Hoheisel JD. Analysis of DNA-protein interactions: from nitrocellulose filter binding assays to microarray studies. Anal Bioanal Chem. 2010; 398(6): 2551–61.

59. Hoffman BG, Jones SJ. Genome-wide identification of DNA-protein interactions using chromatin immunoprecipitation coupled with flow cell sequencing. J Endocrinol. 2009; 201(1): 1–13.

60. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014; 508(7497): 469–76.

61. Meneely P. Genetic Analysis: Genes, Genomes, and Networks in Eukaryotes. 1st ed. Oxford: Oxford University Press; 2014.

62. Claij N, Peters DJ. Teaching molecular genetics: chapter 2-Transgenesis and gene targeting: mouse models to study gene function and expression. Pediatr Nephrol. 2006; 21(3): 318–23.

63. Guénet JL. Chemical mutagenesis of the mouse genome: an overview. Genetica. 2004; 122(1): 9–24.

64. van der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Microbiol. 2014; 12(7): 479–92.

65. Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-Cas immune system: biology, mechanisms and applications. Biochimie. 2015; 117: 119–28.

66. Gasiunas G, Barrangou R, Horvdath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A. 2012; 109(39): E2579–86.

67. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013; 339(6121): 823–6.

Evelina Gasperskaja, Vaidutis Kučinskas

DAŽNIAUSIOS TECHNOLOGIJOS, TAIKOMOS FUNKCINĖJE GENOMO ANALIZĖJE

Santrauka

Nors žmogaus genomo seka yra visiškai nuskaityta, išlieka pagrindinis klausimas dėl realizuojamos genetinės informacijos supratimo. Nepaisant daugelio atliktų genetinių tyrimų, vis dar vienas pagrindinių mokslininkų tikslų – nustatyti genų ir genų produktų funkcijas bei jų tarpusavio sąveikas. Ši funkcinė analizė gyvybiškai svarbi žmogaus sveikatai, kadangi pokyčiai, įvykę genome, gali lemti įvairius patogeninius procesus.

Jau daugelį metų plėtojamos technologijos ir priemonės funkcinei genomo analizei atlikti. Pastaruosius dešimtmečius buvo labai sparčiai tobulinamos technologijos ir sukurti nauji didelio našumo metodai. Dažniausiai taikomos realaus laiko polimerazės grandininės reakcijos, naujos kartos sekoskaitos arba masių spektrometrijos technologijos sukėlė proveržį medicinos srityje. Pažymėtina, kad siekiant gauti patikimus rezultatus tiksli bioinformacinė analizė yra ne mažiau svarbi nei laboratoriniai tyrimai. Šie metodai suteikia galimybę tiksliai ir išsamiai atlikti funkcinę genomo analizę apimant įvairias mokslo sritis, įskaitant genomiką, epigenomiką, proteomiką ir interaktomiką. Tokia funkcinė analizė padeda suprasti biologinius mechanizmus tiek ląstelių, tiek organizmo lygmenyje. Tačiau kiekvienas metodas turi privalumų ir apribojimų, todėl siekiant užtikrinti tyrimo sėkmę svarbu išsamiai aptarti kiekvieno tyrimo metodo principą. Dėl šios priežasties straipsnyje siekiama nurodyti ir apibūdinti dažniausiai ir plačiausiai naudojamus funkcinės analizės metodus.

Raktažodžiai: funkcinė analizė, technologijos, variantai, genomika, transkriptomika, genų raiška