



UDC: 57.088.7 : 577.214 + 577.217

SYSTEMS BIOLOGY: ESSENTIAL PRINCIPLES AND OMICS APPROACHES

R. Stoika

*Institute of Cell Biology, NAS of Ukraine, 14–16, Drahomanov St., Lviv 79005, Ukraine
Ivan Franko National University of Lviv, Biology Faculty, 4, Hrushevskyi St., Lviv 79005, Ukraine
Danylo Halytsky Lviv National Medical University, 69, Pekarska St., Lviv 79010, Ukraine
e-mail: stoika@cellbiol.lviv.ua*

*Stoika R. Systems biology: essential principles and OMICS approaches. **Studia Biologica**, 2020: 14(1); 73–88 • DOI: <https://doi.org/10.30970/sbi.1401.609>*

The systems biology is a new branch of modern biology aimed at studying various systems existing in nature on the level of their biological components such as biomolecules (ex. DNA, RNA, proteins, carbohydrates, lipids, etc.), cells and organelles, tissues and organs, organisms, or species. It differs from the synthetic biology aimed at the creation and study of the biological systems that do not exist in nature for: a) development and combining different functional modules in one system, b) better understanding of processes taking place in the living organisms, c) development of new methods for influencing functions of the living organisms. These two branches of modern biology have similar methodological approaches based on the achievements of molecular biology and bioinformatics. The behavior of different systems in the living organisms is very complex, however, it might be defined using specific approaches of the systems biology based on studying qualitative and quantitative characteristics of all components of a specific biological system through the OMICS approaches. In this review, specific OMICS for various biological systems are briefly described. The application of the systems biology and the OMICS approaches demonstrated their great potentials for medicine, pharmaceuticals, biotechnology, namely for the development of “smart” biomaterials.

Keywords: systems biology, synthetic biology, OMICS approaches

Advantages and drawbacks of molecular biology methods. Applying the computational and mathematical modeling of complex biological systems resulted in development of novel biological discipline, the systems biology. It is aimed at studying various

systems existing in nature on the level of their biological components such as biomolecules (ex. DNA, RNA, proteins, carbohydrates, lipids, etc.), cells and organelles, tissues and organs, organisms, or species. The systems biology is the basis of modern integral medicine (normal and pathological). Besides, it led to a creation of novel materials and biotechnologies of the biomedical use.

While the approaches of classical genetics are based on a study of characteristics of the phenotype and using the results of such study for exploring the genotype peculiarities, the reverse genetics (in fact, the molecular biology) acts oppositely via studying the genotype features for defining the phenotype characteristics (Fig. 1).

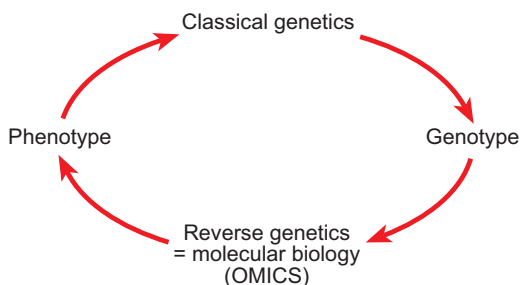


Fig. 1. Interrelations between the classical genetics *versus* reverse genetics and phenotype *versus* genotype

Рис. 1. Взаємозв'язок між класичною генетикою і зворотною генетикою, з одного боку, та фенотипом і генотипом, з іншого боку

The molecular biology methods of studying the DNA (basis of genotype) have several advantages over classical biochemical methods of studying proteins (basis of phenotype). These advantages are as following:

- 1) DNA is much more stable biomaterial (resists short-term boiling) comparing to most proteins;
- 2) DNA sequence is much easier to analyze comparing with protein sequence;
- 3) DNA investigation requires much less quantity of starting material comparing with the amount of protein needed for study (due to capabilities of the polymerase chain reaction).

However, there are also the drawbacks in the DNA-based molecular biology approaches, comparing to protein-based biochemical approaches, since the presence of the nucleic acid, either DNA or mRNA, in cell does not obligatory mean its expression and appearance of functional protein. The knowledge about all genes present in human genome do not allow answering a question which of genes are expressed and how these genes collaborate in the living cells.

The principal approaches used to confirm a postulate that a specific gene detected in genome is realized (expressed) in a specific protein with a specific function are as following: 1) gene overexpression that is usually leading to an increase in specific protein and its function [19]; 2) gene "knock-out" (abbreviation – KO) that is a genetic technique leading to making the gene inoperative ("knocked out" of the genome) [9]. Researchers should compare the difference between the knockout organism and normal one. Knocking out two genes simultaneously is also possible (double knockout). The triple knockouts and quadruple knockouts are used to study the role of three or four inoperative genes in the organism. In all these cases, it is necessary to distinguish between heterozygous and homozygous KOs; 3) while the gene knockout is the total removal or

permanent inactivation of a gene through genetic engineering, another mechanism known as gene knock-down is responsible for silencing of gene expression through the RNA interference (RNAi) [18]. It allows a specific inhibition of functioning of any specific gene, thus, being a valuable instrument for identification of genes involved in disease processes and for treatment purposes through inactivation of such genes (functional genomics). The method is based on the introduction of the double stranded RNA into a mammalian cell that triggers a shutdown of gene transcription and protein translation. The designed small (21–23 nucleotides long) interfering RNAs (siRNAs) are effective and highly specific suppressors of gene expression in mammalian organisms.

Human Genome Program. The idea of sequencing the entire human genome was first discussed at scientific meetings organized by the US Department of Energy in between 1984–1986 [12]. That concept was accepted by the US National Research Council in 1988. A decision was taken to start a broad scientific programme that includes the creation of genetic, physical and sequence maps of the human genome, with parallel studies in various experimental model organisms such as bacteria, yeast, worms, flies, and mice. The development of the same biotechnologies in order to support the accepted objectives was recommended. Besides, at the very beginning, keeping of all investigators involved in the research to the ethical, legal, and social issues appearing at realization of human genome project had been agreed. The programme has started in the US as a joint initiative of the Department of Energy and the National Institutes of Health. Research institutions of other countries (UK, France, Germany, other countries of the European Community, Japan, China, others) joined the program at different stages. The Human Genome Organization (HUGO) was founded to support the international coordination in the genomic research.

By June 2000, the HUGO centres were producing raw DNA sequence related to human genome at a rate equivalent to 1,000 nucleotides per second, 24 hours per day, seven days per week. The 1st version of the draft sequence of human genome as the map and sequence data was available on October 7, 2000. The following peculiarities of human genome were revealed in those studies (reviewed in [12]):

1) GC-rich and GC-poor regions were experimentally detected in human genome via density gradient separation. Subsequently, it was found that these regions may have different biological properties, such as gene density, composition of repeat sequences, correspondence with cytogenetic bands, and recombination rate;

2) The molecular biology puzzle for found that genome size does not correlate precisely with organismal complexity. This is known as the C-value paradox. The largest genome (686,000 Mb) is found in amoeba, a one-cell organism, and it is 200 fold larger than the human genome, and 20,000 fold larger than the yeast genome. Most excess DNA is a repetitive DNA that apparently lacks a function (selfish DNA). Its role in genome evolution is unknown [13].

3) In human, the coding sequences comprise <5% of the genome, whereas repeat sequences account for >50% of the genome [13]. These repeats can be divided into five categories: (1) transposon-derived repeats; (2) inactive (partially) retroposed copies of cellular genes referred to as pseudogenes; (3) simple sequence repeats, consisting of direct repetitions of relatively short k -mers such as $(A)_n$, $(CA)_n$ or $(CGG)_n$; (4) segmental duplications, consisting of blocks of around 10–300 kb; and (5) blocks of tandemly repeated sequences, such as at centromeres, telomeres, the short arms of acrocentric chromosomes, and ribosomal gene clusters;

4) Initially, after excluding the above noted repeat sequences, the draft human genome was counted for 30,000–40,000 protein-coding genes, but then, for the finished genome, it was revised to be 20,000–25,000. More precise numbers of genes in human genome were also informed. It is considered that the total number of functional protein coding genes in human genome is slightly larger than 21,000 [20] that is only about twice as many as in worm or fly;

5) Human genes are more complex due to the presence of higher amount of the alternative splicing which generates a larger number of protein products. Hundreds of human genes could appear due to a horizontal transfer from bacteria at some point in the vertebrate lineage, and dozens of genes could have been derived from transposable elements. More than 1.4 million single nucleotide polymorphisms (SNPs) have been identified in human genome;

6) Human proteome contains 61 % of the apparent homologues with the fly proteome, 43 % – with the worm proteome, and 46 % – with the yeast proteome. If compared with the two invertebrates (worm and fly), humans have more proteins involved in cytoskeleton, defense and immunity, and regulation of transcription and translation that might be explained by the peculiarities of vertebrate physiology. Human contains greater numbers of genes, domain and protein families, paralogues, multidomain proteins with multiple functions, and domain architectures. Thus, higher complexity of human proteome is not simply a consequence of its larger size, but it is probably due to a large-scale protein innovation;

7) After publication of the results of Human Genome program, there were speculations regarding their future applications in biology and medicine. However, very soon, many direct applications of those results were proposed, for example, in a search for the disease genes that can be identified by *in silico* approaches using public databases of human genomic sequence. At least 30 disease genes had been cloned, and their role in disease pathogenesis had been defined. Progress in the pharmaceutical industry strongly depends upon a search for drug targets that are necessary to develop new therapies of the addressed action. Besides, there are many potential applications of the results of Human Genome program in the basic physiology and cell biology. As an example, could be mapping similar traits in both mice and human genomes that allowed identification of common molecular targets such as G-protein coupled receptors used in the action of more than 1/3 of all existing medicines.

8) Since the first stage of the human sequencing covered only about 96 % of the euchromatic regions of the genome, it is important to fill up the remaining gaps.

Summarizing the results of realization of The Human Genome program, one should state that the genome sequence served as a foundation for biomedical research in the years ahead. The remaining gaps are filled constantly with an ambitious wish to resolve them as quickly as possible using the increasingly automated protocols of DNA sequencing. These unclosed gaps have been also sized by FISH techniques for specific chromosomes and other methods. 99.99 % accuracy of sequencing has been achieved. Future human genome studies will involve a three-step program: 1) large-scale identification of regulatory regions; 2) sequencing of additional large genomes; 3) completing the catalogue of human variations. The final goal in this program is to define the ways from the genome sequence to functions. That will require improved techniques and databases for the global analysis of a) RNA and protein expression, b) protein localization, c) protein–protein interactions and d) chemical inhibition of these pathways. The

OMICS approaches based on new experimental and computational techniques are used to achieve success in such analyses. In more detail, the characteristics of these approaches are presented below.

Up to 2019, there were about 500,000 genomes publicly available, and the genetic information on the millions of other people has been collected for other studies [5]. In previous estimates, some scientists believed that there might be up to 100,000 genes in human genome, however, the accomplishment of the Human Genome Project showed that, in fact, humans have below 25,000 genes. This number was quite a surprise to many scientists, since many other organisms, such as rice and water fleas, actually have much more genes than human do. Thus, an important lesson for genetics states that the complexity of an organism is not necessarily correlated with how many genes it has. Another lesson concerns a problem that the human genome contains many highly repetitive sequences (e.g., AGAGAGA or TTTTTT). The 3rd lesson says that in about 3 billion bases that the human genome consists of, all humans share about 99.9 % of this genome, and only 0.1 % (3 million bases) of 3 billion bases is variable. A spot in the genome that can differ between people is called a single nucleotide polymorphism, or SNP. The version of a SNP in a person is called the genotype, and these small genetic differences are part of what makes people unique.

The links to the main DNA sequence database banks are listed:

GenBank, <https://www.ncbi.nlm.nih.gov/genbank/>

European Molecular Biology Lab, <https://www.embl.de/>

DNA Data Bank of Japan, <http://www.ddbj.nig.ac.jp>

China National GeneBank, <https://db.cngb.org/>

Brief history of development of molecular biology. Taking into account great importance of the molecular biology in explanation of the mechanisms of functioning of the organisms, below, a brief history of this branch of biology is described. That history was divided in five periods (in brackets, the names of scientists awarded with Nobel prize and the year of their award are indicated):

- 1) pre-historical period (the main genetic conceptions were proposed in 1865 by the Czech monk Gregor Mendel, however, these conceptions stayed unknown till the beginning of 20th century);
- 2) Origination of the molecular biology foundations (50–60th years in 20th century):
 - Structure of DNA double helix (J. Watson and F. Crick, 1962);
 - Genetic code (M.W. Nirenberg, J.H. Matthaei, S. Ochoa, H.G. Khorana, R. Holley);
 - Mechanisms of protein biosynthesis (P.C. Zamecnik, J. Watson, F. Crick, others);
 - Genetic operon (F. Jacob, A.M. Lwoff, J. Monod, 1965);
 - 3-dimensional structure of proteins (J. Kendrew, M. Perutz, 1962).
- 3) Development of principle molecular biology methods (70–80th years of 20th century):
 - Chemical synthesis of alanine transport RNA (R.W. Holley, H.G. Khorana, M.W. Nirenberg, 1968);
 - Isolation and application of restriction endonucleases (H. Smith, D. Nathans, W. Arber, 1978);
 - Molecular cloning of genes and polymerase chain reaction (K.B. Mullis, M. Smith, 1993);
 - Sequencing of DNA (P. Berg, W. Gilbert, F. Sanger, 1980);
 - Monoclonal antibodies (N.K. Jerne, G.J.F. Köhler, C. Milstein, 1984).

- 4) Establishment of molecular biology as an important branch of modern biology (90th – 2000 years):
 - Development of methods of gene engineering;
 - Sequencing of genomes of several primitive organisms (nematode worm, *Drosophila* insect, yeast, *Arabidopsis thaliana* plant, others);
- 5) Modern period of development of molecular biology (started in 2001, [5, 12]):
 - Accomplishment of “Human Genome” program with sequencing >96 % of human genome (2001);
 - Development and biomedical application of OMICS approaches;
 - Stem cell technologies and regenerative medicine;
 - Genetic cloning of complex organisms (plants and animals);
 - CRISPR/Cas9 technology for gene editing.

OMICS: Essential principles and classificataion. Molecular biology is operating with following principal definitions: 1) genotype (genome) is a complete set of DNA, including all of its genes in a specific organism: 2) phenotype is an expression of features and other characteristics of specific organism that appears as a result of interaction of genetic structure of the organism with the surrounding environment. OMICS is the most advanced approach for genome characterization and its phenotypic realization in different biological systems [6, 10, 17, 21]. The OMICS hierarchy is shown in Fig. 2. As noted above, the 1st version of human genome was presented in 2001. The genome investigation is carried out via using genomics methods, such as DNA arrays and RNA-Seq. Together with the transcriptome and the methods its investigation, such as transcriptomics, the genome and transcriptome define the potentials (genes and messenger RNA) of the genetic expression. While the real functions in the organisms are determined by the products of gene expression, namely, proteins who functions are additionally modulated by the carbohydrates (glycome and glycomics), lipids (lipidome and lipidomics), posttranslational modifications, for example, phosphorylation (phosphorome and phosphoromics), as well as interaction with other biomolecules (interactome and interactomics for proteins, nucleic acids, and others). All above noted OMICS determine cell behavior, such as cell cycling, motility, contractility, adhesion, secretion, sensing, transport, cell signaling, and metabolic activity.

Below, some of the OMICS approaches shown in Fig. 2, are characterized in more detail. It is reasonable to start from the **genome and the genomics** as a method for genome study focused on the genomes' structure, function, evolution, mapping, and editing [12].

Transcriptomics is the study of the transcriptome that is a complete set of RNA transcripts produced by the genome, under specific circumstances (activation, differentiation, malignant transformation, other disease), or in a specific tissue cells [14]. The principal techniques used in genomics are DNA microarrays and RNA-Seq. The microarrays measure the abundances of a defined set of the transcripts via their hybridisation to an array of complementary probes. This technique allows a simultaneous assay of thousands of transcripts at a greatly reduced cost comparing to their individual identification. That also provides a significant saving of laboratory labour. Microarrays consist of short nucleotide oligomers (“probes”) that are arrayed in a grid on a glass slide. Transcript abundance is determined by hybridisation of fluorescently labelled transcripts to these probes, and the fluorescence intensity at the location of each probe on the array indicates the transcript abundance for that probe of the DNA sequence. Usually, the comple-

mentary DNA zonds are labeled with red or green fluorescent dyes for comparing patterns of nucleic acids from different sources (ex. different tissues, or normal and pathological tissue, etc.) While these colors characterize tissue-specific gene expression, the orange color indicates the overlapping of gene expression. Serial and cap analysis of gene expression (SAGE/CAGE) is another approach used in transcriptomics studies.

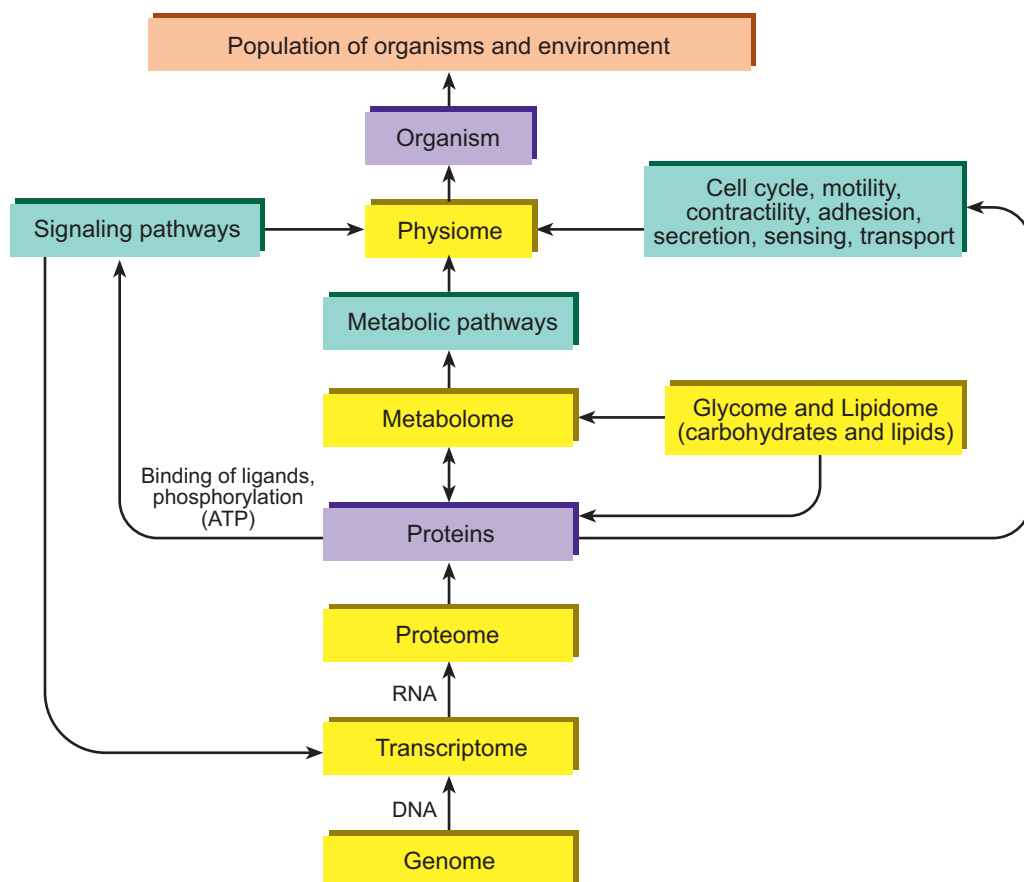


Fig. 2. Hierarchy of OMICS and some of their elements

Рис. 2. Ієрархія OMICS і деякі їїні елементи

The DNA arrays can be applied for identification of different number of transcripts (messenger RNA) – from 24 or 100 to 1,000 or even 15,000 [14]. It can be designed for a specific set of transcripts, for example, for measuring level of expression of mRNA coding for the cytokines, protein products involved in signal transduction, and others. We have measured mRNA coding for the cytokines and cell signalling proteins in murine macrophages J774.2 at their direct contact with different target cells, namely murine transformed fibroblasts of L929 line, or murine normal fibroblasts of NIH-3T3 line, or both these types of cells [11].

In parallel, the cytotoxic action of the macrophages towards studied cells of L929 and NIH-3T3 lines was investigated using cytomorphological (fluorescent microscopy)

and biochemical (Western blot analysis of the proapoptotic cellular proteins and components of the transcription factor NF- κ B) methods. The bystander effect of the normal fibroblasts on the macrophages that contact with the transformed fibroblasts was revealed. It was mediated by the transcription factor NF- κ B and a secreted mediators (IL-1b) whose expression was changed at such interaction (Table 1 and Fig. 3, with kind permission of Dr. Nataliya Kashchak [11]).

Table 1. List of genes whose protein products are associated with signal transduction pathways in the macrophages interacting with normal and transformed fibroblasts (the DNA Array GEMatrix™ KIT from SuperArray Inc., Bethesda was used)

Таблиця 1. Список генів, чий білкові продукти задіяні у шляхах сигнальної трансдукції у макрофагах під час їхньої взаємодії з нормальними і трансформованими фібробластами (використано the DNA Array GEMatrix™ KIT from SuperArray Inc. in Bethesda)

Genes	Localization
Activating transcription factor 2 (CRE-BP1 transcription factor)	(1,A) (1,B)
Bcl2-associated X protein	(1,C) (1,D)
Mouse lymphocyte differentiation antigen Ly-1 (CD5 antigen)	(1,E) (1,F)
Mouse <i>c-fos</i> oncogene (FBJ osteosarcoma oncogene)	(2,A) (2,B)
<i>Jun</i> oncogene	(2,C) (2,D)
Mouse normal <i>c-myc</i> gene and translocated homologue from J558 plasmocytoma cells	(2,E) (2,F)
<i>Mus musculus</i> mRNA for aromatase P450 (Cytochrome P450, 19, aromatase)	(3,A) (3,B)
Mouse Egr-1 (Early growth response 1)	(3,C) (3,D)
<i>Mus musculus</i> p53 responsive (EI24) mRNA	(3,E) (3,F)
Fas ligand	(4,A) (4,B)
DNA-damage inducible transcript	(4,C) (4,D)
<i>Mus musculus</i> mRNA for heat shock transcription factor 1	(4,E) (4,F)
Mouse heat shock protein 86	(5,A) (5,B)
<i>Mus musculus</i> I-kappa B alpha chain	(5,C) (5,D)
Mouse interleukin-2	(5,E) (5,F)
Inducible nitric oxide synthase	(6,A) (6,B)
Nuclear factor of kappa light chain gene enhancer in B-cells 1, p105	(6,C) (6,D)
Mouse CDK4 and CDK6 inhibitor p16ink4a	(6,E) (6,F)
Cyclin-dependent kinase inhibitor p19	(7,A) (7,B)
Cyclin-dependent kinase inhibitor p21Waf1	(7,C) (7,D)
Mouse cyclin-dependent kinase inhibitor p27Kip1	(7,E) (7,F)
Transformation related protein 53 Tumor antigene	(8,A) (8,B)
Cyclin-dependent kinase inhibitor p57Kip2	(8,C) (8,D)
Cytoplasmic beta-actin	(3,G) (4,G)
Glyceraldehyde-3-phosphate dehydrogenase	(5,G) (6,G) (7,G)
Bacterial plasmid	(8,E) (8,F) (8,G)
	(1,G) (2,G)

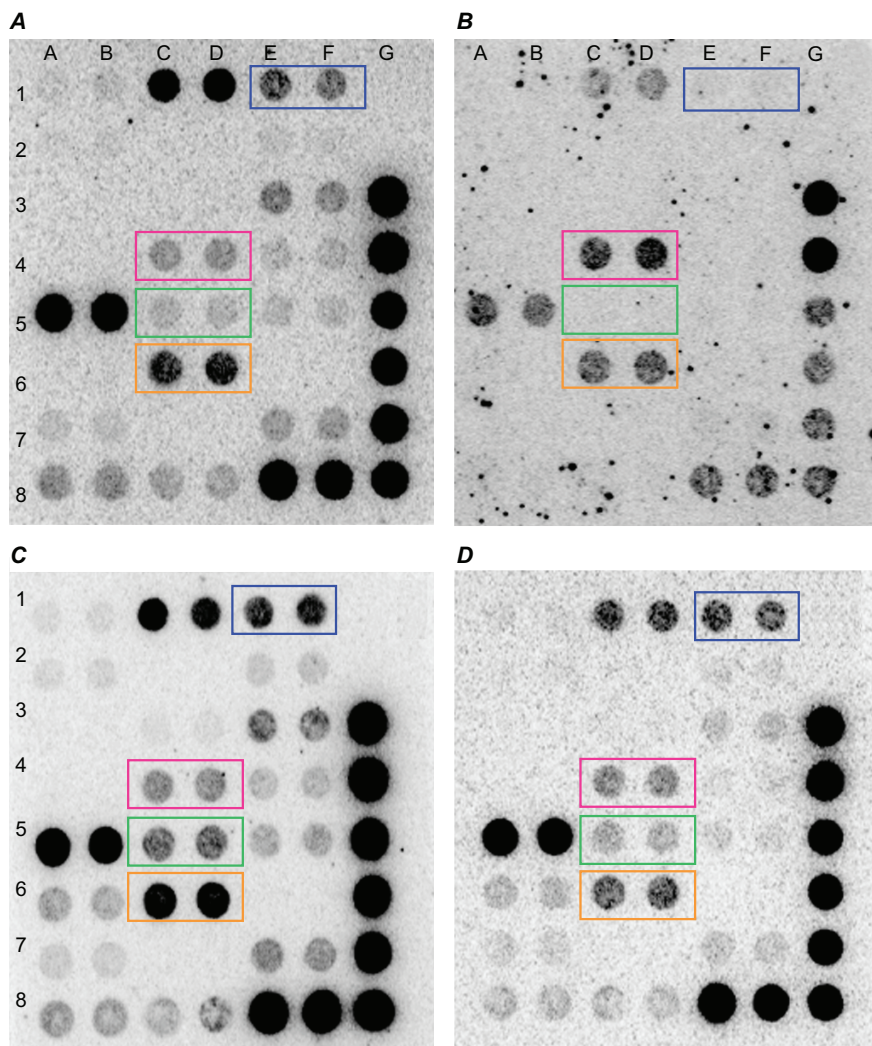


Fig. 3. Expression profile of genes associated with signal transduction pathways: **A** – murine monocyte/ macrophage J774.2 cells cultured alone; **B** – J774.2 cells isolated from the co-culture with transformed murine fibroblast L929 cells; **C** – J774.2 cells isolated from the co-culture with pseudonormal murine fibroblast NIH 3T3 cells; **D** – J774.2 cells isolated from the co-culture with NIH 3T3 and L929 cells

■ Mouse lymphocyte differentiation antigen Ly-1 (CD5 antigen) (>2); ■ DNA-damage inducible transcript (>2); ■ *Mus musculus* I-κB α-chain (>2); ■ Nuclear factor of κ-light-chain gene enhancer in B-cells 1, p105 (>2). In brackets, the indicator of change (>2) in gene expression is noted

Рис. 3. Профіль експресії генів, чиї продукти асоційовані зі шляхами трансдукції регуляторних сигналів: **A** – Культивовані мишачі моноцити/макрофаги лінії J774.2; **B** – мишачі моноцити/макрофаги лінії J774.2, отримані зі спільної культури з трансформованими мишачими фібробластами лінії L929; **C** – мишачі моноцити/макрофаги лінії J774.2, отримані зі спільної культури із псевдонормальними мишачими фібробластами лінії NIH 3T3; **D** – мишачі моноцити/макрофаги лінії J774.2, отримані зі спільної культури з клітинами ліній L929 і NIH 3T3

■ Антиген CD5, антиген Ly-1 диференціювання лімфоцитів миші (>2); ■ Індукційний транскрипт пошкодження ДНК (>2); ■ α-Ланцюг I-κB *Mus musculus* (>2); ■ p105, ядерний фактор легкого ланцюга κ-генного інхансера В-клітин (>2). У дужках вказано зміни (>2) рівня експресії генів

The **proteome** (1st note appeared in 1994) is the entire set of proteins expressed in cell or its organelle, tissue, or organism under defined conditions at a certain time, and **proteomics** (1st note appeared in 1997) studies the proteome. Proteome and proteomics are most important for the molecular diagnostics in medicine, as well as for development of “smart” drugs of the addressed action [2, 3, 15]. The goal of proteomics is a quantitative description of proteins’ expression and its changes under various perturbations such as a disease or drug action. In all proteomics studies, there should be a pair of the biological objects for comparison, for example, normal *versus* malignant cells, healthy *versus* disease cells, tissues, or organs, inactive *versus* activated cells, non-differentiated *versus* differentiated cells. Principle methods in proteomics are: 1) 2-dimentional electrophoresis of proteins; 2) mass-spectrometry of proteins. The auxiliary methods in proteomics are: 1) Western-blot analysis of proteins; 2) immuno-enzymatic analysis (ELISA); 3) determination of specific biological activity of proteins (ex. enzymatic).

It is only a very small amount of the intravenously injected medicine that reaches its molecular or cellular target in the organism, while the most of it is a potential source for negative side effects [22]. Thus, the addressed delivery of the medicine to its molecular target (mostly proteins) in the organism can eliminate side effects. The addressed delivery of the medicine can also be a basis for the personalized treatment of patients, while targeting a specific protein involved in the impaired function.

There are several complications in the proteomics, namely, post-translational modifications of proteins, such as phosphorylation, ubiquitination, methylation, acetylation, glycosylation, oxidation, nitrosylation, and other modifications. These complications have to be taken into account when the identification of a specific protein is not certain.

Below, an example of the proteomics based search for proteins involved in PTTG-dependent regulation of T-lymphocytes activation is shown (Fig. 4 and Table 2, with permission of Dr. Yevhen Filyak [7]).

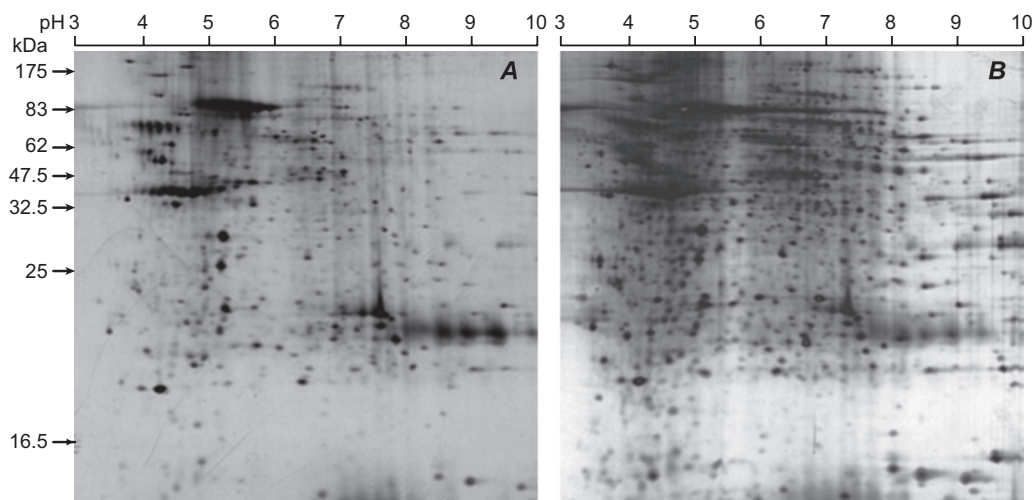


Fig. 4. Two-dimentional (electrophoresis and isoelectrofocusing) electrophoregram of proteins of quiescent (A) and anti-CD3-activated (B) murine perypheral blood lymphocytes

Рис. 4. Двовимірна (електрофорез та ізоелектрофокусування) електрофореграма білків неактивованих (A) і активованих за допомогою анти-CD3-антитіл (B) лімфоцитів периферичної крові миші

Table 2. The list of proteins whose expression is changed in murine lymphocytes under anti-CD3/TCR activation. Protein identification was conducted using proteomics based on the results of two-dimensional electrophoresis combined with the MALDI-TOF mass-spectrometry

Таблиця 2. Перелік білків, рівень експресії яких змінюється у мишачих лімфоцитах, активованих анти-CD3/TCR. Ідентифікацію білків здійснювали за допомогою протеоміки, що базувалася на результатах двовимірного електрофору в поєднанні з мас-спектрометрією MALDI-TOF

Protein	Indicator	Verification (%)
Myelin proteolipid protein	AAA41897.1	98
Leukocyte-associated immunoglobulin-like receptor-1d	AAR32127.1	100
Complement receptor CR2	AAA37451.1	100
Unnamed protein product	BAB29181.1	97%
Complement receptor CR2	AAA37451.1	100%
Phosphoprotein phosphatase (EC 3.1.3.16) 2A alpha regulatory chain - mouse (fragment)	A47187	98%
Serum amyloid A - mouse (fragment)	I71951	100%
Phosphoprotein phosphatase (EC 3.1.3.16) 2A alpha regulatory chain - mouse (fragment)	A47187	99%
Similar to DYRK2 protein	XP_354565.1	100%
Immunoglobulin heavy chain variable region	AAO23318.1	100%
Hepcidin antimicrobial peptide 2	AAO73588.1	100%
Serum amyloid A - mouse (fragment)	AAO73588.1	100%
Poly(A)-binding protein cytoplasmic 5	CAC42820.1	99%
Chic 1	CAD33951.1	100%
Abca4 protein	AAH43937.1	100%
Similar to dJ132F21.1 (A pupative novel protein)	XP_357231.1	100%
Retinoic acid, EGF, and NGF upregulated; Retinoic acid, EGF, and NGF upregulated gene cDNA sequence AF465352	NP_694783.1	100%
Lrpap1 protein	AAH59887.1	100%
Unnamed protein products	BAC38872.1	100%
Similar to dJ132F21.1 (A pupative novel protein)	XP_357231.1	98%
Aminoadipate-semialdehyde dehydrogenase – phosphopantetheinyl transferase	AAH49851.1	100%
Clecsf 6 protein	AAH06623.1	98%
9130022A01Rik protein	AAH61228.1	100%
Zinc finger protein mkr5	AAA37120.1	93%
Keratin complex 2, basic, gene 8; cytokeratin8 [Mus Musculus]	NP_112447.1	99%
Casein kinase 1, alpha 1	AAH48081.1	86%

End the Table 2

Unnamed protein product Dimethylaniline monooxygenase [N-oxide forming] 3 (Hepatic flavin-containing monooxygenase 3) (FMO3) (Dimethylaniline oxidase 3)	BAC34077.1	100%
Similar to Heterogeneous nuclear ribonucleoprotein A1 (Helix-destabilizing protein) (Single strand binding protein) (hnRNP core protein A1) (HDP)	XP_356697.1	100%
Cationic amino acid transporter 2B	CAC82478.1	100%
RIKEN cDNA 1110065P20	XP_149496.3	100%
Chain A, 1.8a Crystal Structure Of The Monomeric Gtpase Rab5c (Mouse)	1HUQ	98%
Transcription factor Mlr 1 protein (Mblk1-related protein-1)	AAH66151.1	100%
Lectin, galactose binding, soluble 1	AAH02063.1	100%
Biliary glycoprotein	CAA47695.1	100%
Unnamed protein product	BAB26620.1	99%
Ccni protein	AAH07177.1	100%
Pyrophosphatase	AAH10468.1	99%
Similar to cytoplasmic beta-actin put. beta-actin (aa 27–375)	XP_358871.1 CAA27396.1	100%
Capping protein alpha subunit Capping protein (actin filament) muscle Z-line, alpha-1	AAC00566.1 AAH16232.1	100%
Unnamed protein product	BAB28436.1	72%
Pitpnb protein phosphatidylinositol transfer protein beta isoform	AAH34676.1 AAA87593.1	28%
Calpain 9	Q9D805	100%
Unnamed protein product	BAC37121.1	100%
Histone deacetylase11	AAH16208.1	99%
Similar to actin, gamma, cytoplasmic	XP_134663.2	100%

The performed proteomics analysis revealed 1,135 protein spots on the electrophoregram of quiescent murine lymphocytes and 1,433 protein spots – on the electrophoregram of anti-CD3-activated lymphocytes (Fig. 4). 185 of those proteins demonstrated changed (>2 times) amount on the electrophoregram, and they were selected for the MALDI-TOF mass-spectrometric identification. Computer search in the protein database showed that 3 proteins belong to transcription factors, 2 proteins are cytokines, 6 proteins are regulators of cell cycling, proliferation, and migration.

Glycome and Glycomics. The glycoelements present on cell surface play a significant role in many biological processes, such as [21]: a) cell-cell communication; b) cell immunity; c) cancer development; d) recognition of dying cells by the macrophages for further clearance of dying cells from the organism; e) cell interaction with the infection agents; f) control of cell movement; g) discrimination between different cell types; h) functioning of the receptors of biologically active substances (ex. cytokines);

j) as a glyco-component integrated in plasma membrane (ex. receptors) or soluble in the cytosol (ex. Interferon).

Glycosylation of the biomolecules occurs in Golgi apparatus that is connected with the endoplasmatic reticulum. Lectins are proteins that specifically recognize free sugar or glyco-component in the glyco-conjugate. Sugar conjugation is performed via hydroxyl group of serine/threonine/tyrosine (O-glycosylation) or via amino-group of the asparagine (N-glycosylation). Cell surface glyco-pattern is specific for most cells, and it is an object of glycomics study.

Metabolomics is a study of the set of metabolites within an organism, cell, or tissue [6, 8, 10, 16, 17, 21]. It is focused at measurement and biological interpretation of the low molecular weight (~50–1,500 Da) biochemicals or “metabolites”. A system of metabolomics helps solving many clinical and research problems. In the translational science, these are firstly biomarker identification, as well as study of disease pathogenesis and treatment mechanisms. Other bio-application of metabolomics are directed production of specific metabolites, tissue engineering (ex. bone and skin regeneration), pharmacology and toxicology, dietology and nutrition, metabolism in plants and micro-organisms.

Interactomics studies interactions between proteins and other molecules within a cell, as well as the consequences of those interactions [6, 17, 21]. Among the mostly studied interactions are: a) protein/DNA (ex. transcription factors); b) protein/protein (ex. action of protein kinases).

Phenomics deals with characteristics phenomes as the physical and biochemical traits of organisms, also investigating their change in response to genetic mutation and/or environmental influences [17, 21]. Phenomics is tightly related to functional genomics, pharmaceutical research, metabolic engineering, and phylogenetics. Its main methods are: a) study of consequences of gene over-expression; b) gene silencing through gene knock-out (deletion) or functional knock-down with the interfering RNA. At cellular level, typical phenomics characteristics are cell growth, survival, proliferation, differentiation, and functional activity.

Physiome is an integrated modeling category that used computer informative methods of monitoring biochemical, biophysical, and anatomical information about the cells, tissues and organs for the purpose of their testing [16, 21].

Of course, this is not the full list of all known OMICS, and current development of the OMICS approaches demonstrates that OMICS phenomenon is endless.

Gene regulatory network is a set of molecular regulators that interact with each other and with other substances in the cell to govern the expression levels of mRNA and proteins. Such networks play a central role in morphogenesis and the creation of body structures [5]. The molecular regulators can be DNA, RNA, proteins, and their complexes. The interaction can be direct or indirect, for example, through a transcribed RNA or translated proteins. These proteins can be structural, and accumulate in plasma membrane or particular structures within the cell, or be an enzyme catalyzing certain reactions, such as the breakdown of compounds present in food or toxin. The function of other proteins – the transcription factors – is to activate specific genes, and they play the main role in gene regulatory networks, also named cascades. The transcription factors bind to the promoter (regulatory) region of genes in order to turn them on and initiate the production of another protein. Some transcription factors possess an inhibitory activity and turn off the expression of the corresponding genes.

Virtual Physiological Rat (VPR) is an impressive product of application of the OMICS approaches. The VPR Project is conducted within the international collaboration program supported by the National Institute of General Medical Sciences as a National Center for Systems Biology [4]. The project was motivated by the need in deep study of cardiovascular physiology that includes both physiological and genomic data from animal models of disease, since it is necessary to understand how multiple genes and environmental factors interact to determine cardiovascular phenotype. Within the VPR Project, novel computational tools to define the functions of physiological systems and their pathophysiological perturbations caused by disease development. These tools are being developed and validated based on experimental characterization of physiological function across a number of organ systems in rat strains engineered to show relevant disease phenotypes. Computer simulation tools are used to integrate disparate data (genomic, anatomic, physiological, etc.) in order to predict function and translate the study of animal models to new information on specific human diseases, such as hypertension, kidney disease, heart failure, or metabolic syndrome. As an example of using VPR based approaches, could be an application of the artificial intelligence for basic management, data collection, automated monitoring and managing intracranial pressure, seizures, hemodynamics, and ventilation in the complicated patients with the neurologic injury [1].

1. Al-Mufti F., Kim M., Dodson V., Sursal T., Bowers C., Cole C., Scurlock C., Becker C., Gandhi C., Mayer S.A. Machine Learning and Artificial Intelligence in Neurocritical Care: a Specialty-Wide Disruptive Transformation or a Strategy for Success. **Curr Neurol Neurosci Rep**, 2019; 19(11): 89.
[DOI: <https://doi.org/10.1007/s11910-019-0998-8>; PMID: 31720867; Google Scholar]
2. Anderson N.G. Proteome and proteomics: new technologies, new concepts, and new words. **Electrophoresis**, 1998; 19(11): 1853–1861
[DOI: <https://doi.org/10.1002/elps.1150191103>; Google Scholar]
3. Aslam B., Basit M., Nisar M.A., Khurshid M., Rasool M.H. Proteomics: Technologies and Their Applications. **J. Chromatogr. Sci.**, 2017; 55(2): 182–196.
[DOI: <https://doi.org/10.1093/chromsci/bmw167>; Google Scholar]
4. Beard D.A., Neal M.L., Tabesh-Saleki N., Thompson C.T., Basingthwaight J.B., Shimoyama M., Carlson B.E. Multiscale modeling and data integration in the virtual physiological rat project. **Ann. Biomed. Eng.**, 2012; 40(11): 2365–78.
[DOI: <https://doi.org/10.1007/s10439-012-0611-7>; Google Scholar]
5. Broekema R.V., Bakker O.B., Jonkers I.H. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. **Open Biol.**, 2020; 10(1): 190221.
[DOI: <https://doi.org/10.1098/rsob.190221>; Google Scholar]
6. Costa A., Franco O.L. Impact and influence of “omics” technology on hypertension studies. **Int. J. Cardiol.**, 2017; 228: 1022–1034.
[DOI: <https://doi.org/10.1016/j.ijcard.2016.11.179>; Google Scholar]
7. Filyak Y.Z., Filyak O.S., Souchelnitskyi S.I., Stoika R.S. Proteomics of activation of murine T lymphocytes with pttg-knockout. **Reports of the National Academy of Sciences of Ukraine**, 2007, N 5, P. 172–179.
http://nbuv.gov.ua/UJRN/dnanu_2007_5_31
8. Haug K., Salek R.M., Steinbeck C. Global open data management in metabolomics. **Curr. Opin. Chem. Biol.**, 2017; 36: 58–63.
[DOI: <https://doi.org/10.1016/j.cbpa.2016.12.024>; Google Scholar]
9. Houdebine L.-M. Design of Vectors for optimizing transgene expression. In Chapter 17. **Transgenic Animal Technology** (Third Edition), 2014: 489–511.
[DOI: <https://doi.org/10.1016/B978-0-12-410490-7.00017-7>; Google Scholar]

10. Joyce A.R. Palsson B.O. The model organism as a system: integrating “omics” data sets. **Nature Reviews in Molecular Cell Biology**, 2006; 7(3): 198–210.
[DOI: <https://doi.org/10.1038/nrm1857>; Google Scholar]
11. Kashchak N., Tsaryk R., Stoika R. Bystander effect of normal fibroblasts for macrophages co-cultured with susceptible transformed target cells. **Cell Biol Int**, 2005; 29(1): 41–50.
[DOI: <https://doi.org/10.1016/j.cellbi.2004.11.009>; Google Scholar]
12. Lander E.S., Linton L.M., Birren B. et al. Initial sequencing and analysis of the human genome. **Nature**, 2001; 409(6822): 860–921.
[DOI: <https://doi.org/10.1038/35057062>; Google Scholar]
13. Latorre Amparo and Silva Francisco J. The size of the genome and the complexity of living beings. *Metode* 32, 2013.
14. Mammalian Gene Collection (MGC) Program Team. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. **Proc. Nat. Acad. Sci. USA**, 2002; 99(26): 16899–16903.
[DOI: <https://doi.org/10.1073/pnas.242603899>]
15. Martens L., Vizcaino J.A. A golden age for working with public proteomics data. **Trends Biochem. Sci.**, 2017; 42(5): 333–341.
[DOI: <https://doi.org/10.1016/j.tibs.2017.01.001>; Google Scholar]
16. Menni C, Zierer J, Valdes AM, Spector TD. Mixing omics: combining genetics and metabolomics to study rheumatic diseases. **Nat. Rev. Rheumatol.**, 2017; 13(3): 174–181.
[DOI: <https://doi.org/10.1038/nrrheum.2017.5>; Google Scholar]
17. Millius A., Ueda H.R. Systems biology-derived discoveries of intrinsic clocks. **Front. Neurol.**, 2017; 8(25): 1–19.
[DOI: <https://doi.org/10.3389/fneur.2017.00025>; Google Scholar]
18. Mocellin S. Provenzano M. RNA interference: learning gene knock-down from cell physiology. **J. Transl. Med.**, 2004; 2: 39.
[DOI: <https://doi.org/10.1186/1479-5876-2-39>; Google Scholar]
19. Prelich G. Gene Overexpression: Uses, Mechanisms, and Interpretation. **Genetics**, 2012; 190(3): 841–854.
[DOI: <https://doi.org/10.1534/genetics.111.136911>; Google Scholar]
20. Saccone C., Pesole G. **Handbook of Comparative Genomics: Principles and Methodology** / John Wiley and Sons. 2003. Chapter 2. Eukaryotes. P. 36–84.
21. Schneider M.V., Orchard S. Omics technologies, data and bioinformatics principles. Chapter 1 In: Mayer B. (Ed.) **Bioinformatics for OMICS data: Methods and protocols, Methods in molecular biology**. Springer Science+Business Media, LLC, 2011; 719: 3–30.
[DOI: https://doi.org/10.1007/978-1-61779-027-0_1; Google Scholar]
22. Silverman R.B., Holladay M.W. **The organic chemistry of drug design and drug action**. 3rd Edition. Elsevier – Academic Press, Amsterdam, Boston, Heidelberg et al. 2014: 536 p. In Chapter 3: p.125. **eBook ISBN: 9780123820310**

СИСТЕМНА БІОЛОГІЯ: ОСНОВНІ ПРИНЦИПИ ТА ПІДХОДИ ОМІКИ

Р. Стойка

Інститут біології клітини НАН України, вул. Драгоманова, 14/16, Львів 79005, Україна

Львівський національний університет імені Івана Франка

вул. Грушевського, 4, Львів 79005, Україна

Львівський національний медичний університет імені Данила Галицького

вул. Пекарська, 69, Львів 79010, Україна

e-mail: stoika@cellbiol.lviv.ua

Системна біологія – це нова галузь сучасної біології, метою якої є дослідження різних існуючих у природі систем на рівні їхніх біологічних компонентів, таких як

біо-молекули (наприклад, ДНК, РНК, протеїни, вуглеводи, ліпіди тощо), клітини й органели, тканини й органи, організми або види. Вона відрізняється від синтетичної біології, метою якої є створення і вивчення біологічних систем, яких немає у природі, для: а) розробки і поєднання різних функціональних модулів у одній системі; б) кращого розуміння процесів, які відбуваються в живих організмах; в) розроблення нових методів впливу на функції живих організмів. Разом із тим, ці дві галузі біології мають подібні методологічні підходи, здебільшого засновані на досягненнях молекулярної біології та біоінформатики. Поведінка різних систем у живих організмах є дуже складною, проте вона може бути визначена за допомогою спеціальних підходів системної біології, скерованої на вивчення якісних і кількісних характеристик усіх компонентів конкретної біологічної системи, для чого використовують підходи ОМІКи. У цьому огляді коротко описано специфічні ОМІКи для вивчення окремих компонентів різних біологічних систем. Застосування системної біології та підходів ОМІКи продемонструвало їхній великий потенціал для медицини, фармацевтики, біотехнології, зокрема, для розвитку “розумних” біоматеріалів.

Ключові слова: системна біологія, синтетична біологія, підходи ОМІКи