УДК 575.112
https://doi.org/10.20538/1682-0363-2026-1-176-184

Gene Ontology for Genomics and Biology

Chasovskikh N.Yu.

Siberian State Medical University
2 Moscovsky trakt, 634050 Tomsk, Russian Federation

ABSTRACT

The aim of the lecture was to consider the role of gene ontology (GO) and the GO Consortium in shaping the knowledge base for genomics, proteomics, and biology. GO organizes and continually updates data on the molecular functions and biological processes in which genes and their products are involved.

The structure of GO, the features of GO term hierarchy and the connections between them, as well as the elements of each term are considered. The features of services for working with basic knowledge and various ways to access civil defense data are given. In addition to term characteristics, GO pays great attention to annotations – statements that link a gene product to a certain ontology term. The annotation process captures the action and location of a gene product using terms, providing a reference and a type of evidence.

The areas of application of GO related to the analysis of genomics and proteomics data are considered. The main approaches used by researchers are functional annotation of genes and pathway enrichment analysis. Analysis of large volumes of data (for example, when assessing gene expression) allows to gain knowledge about the involvement of genes and their products in various processes, extract biological meaning, and evaluate the features of molecular mechanisms in various diseases. The increasing role of GO in the formation of new knowledge in the relevant field is shown.

Keywords: bioinformatics, gene ontology, functional annotation, biological process, molecular function, cellular component, GO annotation, gene function

Conflict of interest. The author declares the absence of obvious or potential conflict of interest related to the publication of this article.

Source of financing. The author states that no funding was received for the study.

For citation: Chasovskikh N.Yu. Gene Ontology for Genomics and Biology. *Bulletin of Siberian Medicine.* 2026;26(1):176–184. https://doi.org/10.20538/1682-0363-2026-1-176-184.

Генная онтология для геномики и биологии

Часовских Н.Ю.

Сибирский государственный медицинский университет (СибГМУ)
Россия, 634050, г. Томск, Московский тракт, 2

РЕЗЮМЕ

Цель исследования – рассмотреть роль генной онтологии (GO) и Консорциума GO в формировании базиса знаний для геномики, протеомики и биологии. Генная онтология позволяет систематизировать и постоянно обновляет данные о молекулярных функциях и биологических процессах, в которых участвуют гены и их продукты.

✉ *Chasovskikh Natalia Yu.,* nch03@mail.ru

Рассмотрена структура GO, особенности иерархии терминов GO и отношения между ними, элементы каждого из терминов. Приведены особенности сервисов, обеспечивающих возможности работы исследователей с базой знаний с помощью различных способов доступа к данным GO. Помимо характеристик терминов в GO большое внимание уделяется аннотациям – утверждениям, связывающим продукт гена с конкретным термином онтологии. Процесс аннотации фиксирует действие и локализацию генного продукта с помощью терминов, предоставляя ссылку и вид доказательств.

Рассмотрены направления применения генной онтологии, связанные с анализом данных геномики и протеомики. Основные подходы, используемые исследователями, – это функциональная аннотация генов, анализ обогащения путей. Анализ больших объемов данных (например, при оценке экспрессии генов) позволяет получить знания о вовлеченности тех или иных генов и их продуктов в различные процессы в организме, извлечь биологический смысл и оценить особенности молекулярных механизмов при различных заболеваниях. Показана возрастающая роль GO в формировании новых знаний в соответствующей области.

Ключевые слова: биоинформатика, генная онтология, функциональная аннотация, биологический процесс, молекулярная функция, клеточный компонент, GO аннотация, функция гена

Конфликт интересов. Автор декларирует отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Источник финансирования. Автор заявляет об отсутствии финансирования при проведении исследования.

Для цитирования: Часовских Н.Ю. Генная онтология для геномики и биологии. *Бюллетень сибирской медицины.* 2026;26(1):176–184. https://doi.org/10.20538/1682-0363-2026-1-176-184.

INTRODUCTION

Modern genomics research provides the opportunity to work with large volumes of data. However, it is essential to address the challenge of extracting a biological meaning from the obtained information. The primary question is how to map sequence data onto known results of gene functional analysis and how to derive new insights into gene functions. Ontology tools offer a solution – providing a consistent framework for describing a specific subject area, in this case, genomics and molecular biology [1].

From a computer science perspective, an ontology is a model used to represent objects, their properties, and the relationships between them [2]. Within a given subject area, this model comprises a set of concepts (terms) with definitions and attributes, along with a corresponding set of axioms and inference rules [3]. Typically, knowledge bases for a subject area include repositories containing general knowledge about classes of concepts, their properties, and relationships, as well as knowledge bases about individual objects, their properties, and links to other objects (e.g., specific data). Both components are interconnected within the knowledge base [1].

Applied to genomics, the development of a gene function ontology involved creating a controlled vocabulary of terms with clear definitions and establishing a hierarchical structure of relationships between these terms. A major advantage was that this ontology enabled the study of functional aspects of genomes from different organisms and facilitated the annotation of functions for novel sequences [4].

During the development of the gene ontology, special attention was given to questions that needed to be described in terms of: the location of gene expression within the organism and the subcellular localization of the gene product; the timing of gene expression (in the context of organismal ontogenesis); the function of the gene product and its position within the hierarchy of biological processes; genes regulating the activity of the given gene product. Initially, these challenges were addressed in the context of the Drosophila genome database, FlyBase [4]. The concepts of molecular function and biological process were employed as primary dimensions for the classification. The first gene ontology developed in this manner was described in the article by M.T. Ashburner et al. [5].

Using the first comparison of two complete eukaryotic genomes – the yeast *Saccharomyces cerevisiae* and the worm *Caenorhabditis elegans* – it was demonstrated that a significant proportion of genes in these organisms were orthologs. Furthermore, 12% of the worm genes encoded proteins whose biological significance can be inferred from their similarity to orthologs in yeast (which represents 27% of the yeast genes). It was found that such proteins participate

in biological processes common to all eukaryotes (for example, DNA replication, transcription, and metabolism). A subsequent comparison of the genomes of yeast, worms, and fruit flies also confirmed the presence of homologs among them [6]. It was also shown that genes and proteins involved in fundamental biological processes were highly likely to be orthologous, a fact already confirmed in mammals and model organisms (e.g., yeast) [7–12].

Since a high degree of similarity and functional conservation was subsequently demonstrated across the genomes of different species, this finding opened new opportunities for the automated transfer of biological annotations from experimentally studied model organisms to less characterized species, thereby facilitating the development of the gene ontology.

THE GENE ONTOLOGY CONSORTIUM

Initially, the Gene Ontology (the Gene Ontology Consortium, GO Consortium) was a collaborative project involving three model organism databases: FlyBase16, Mouse Genome Informatics17, 18 (MGI), and Saccharomyces Genome Database19 (SGD). Subsequently, other organism databases joined the consortium [5].

The Consortium goal to create a structured, precisely defined, common, and controlled vocabulary to describe the roles of genes and gene products in any organism remains unchanged, despite progress in implementation and the dynamic development of the terminology system. Each element in GO is linked to other types of information, including gene and protein databases, such as SwissPROT [13], Gen-Bank [14], DDBJ [15], PIR [16], MIPS [17], YPD and WormPD [18], Pfam [19], SCOP [20], and ENZYME [21], facilitating continuous updating and refinement of knowledge about these entities [5].

The core GO terminology describes genes and their products within three categories (aspects). A *Biological Process* encompasses one or more ordered collections of molecular functions, often involving a chemical or physical transformation. Examples of high-level, general biological process terms (Fig. 1) include *GO:0065007 biological regulation* or *GO:0050896 response to stimulus*; low-level terms, such as *GO:0042770 signal transduction in response to DNA damage* or *GO:0072331 signal transduction by p53 class mediator*. A *Molecular Function* defines the biochemical activity of a gene product.

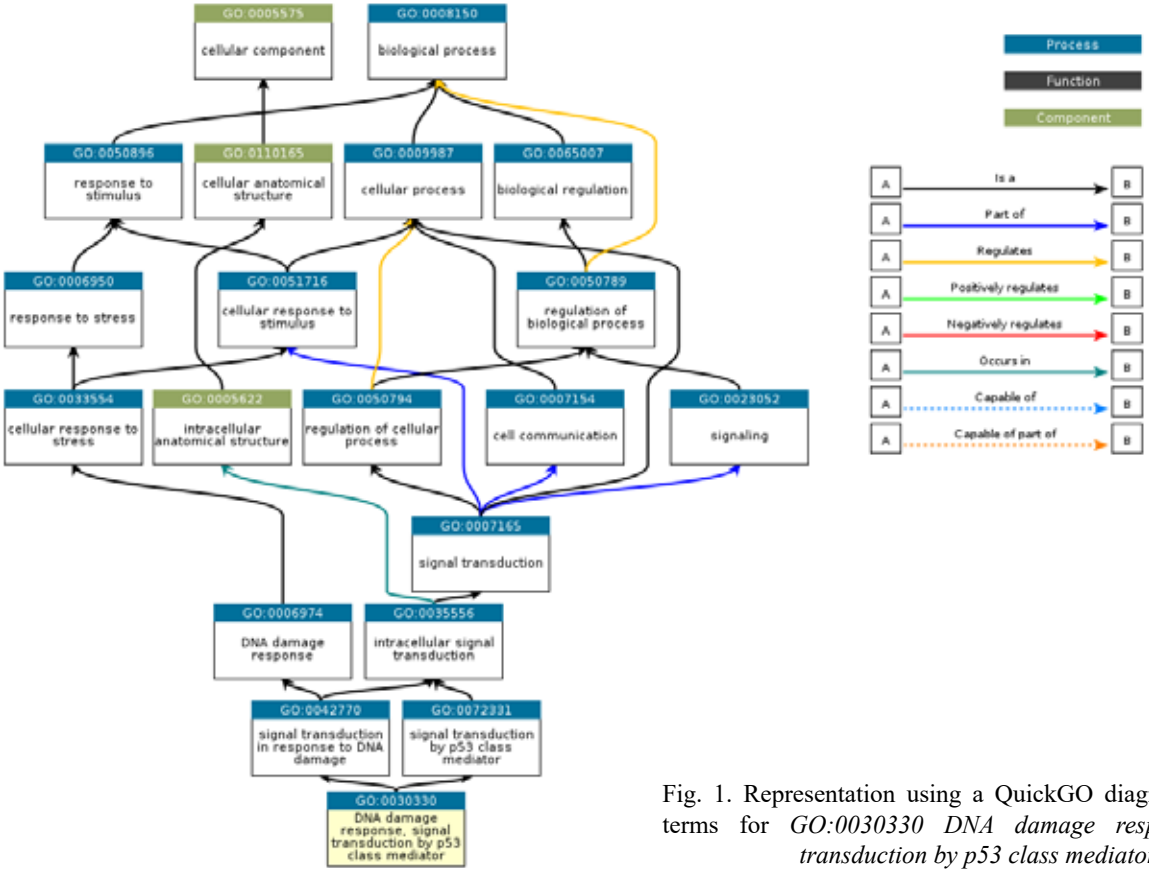

Fig. 1. Representation using a QuickGO diagram of parent terms for *GO:0030330 DNA damage response, signal transduction by p53 class mediator*

General examples include "*enzyme*", "*transporter*", or "*ligand*", while more specific terms include "*adenylate cyclase*" or "*Toll receptor ligand*". A *Cellular Component* indicates the location of gene product activity within a cell; examples are "*ribosome*", "*nuclear membrane*", or "*Golgi apparatus*" [5].

As of March 2025, the ontology contains 26,037 terms in the *Biological Process* aspect, 10,154 terms in the *Molecular Function*, and 4,023 items in the *Cellular Component* [22].

The GO structure is a graph in which each term represents a node, and the connections between nodes represent edges. It also possesses hierarchy, meaning that a child term is more specific than its parent (Fig. 1), and nodes may have multiple parent teams. GO also includes definitions and categories describing the relationships between terms. Commonly used relationships include: *is a*; *part of*; *has part*; *regulates*, *negatively regulates*, and *positively regulates* [23].

Features of the relationships between terms:

The term *"is a"* constitutes the fundamental structure of GO. If A *is a* B, it implies that node A refers to a subtype of node B. For example, mitotic cell cycle *is a* (subtype of) cell cycle.

The term *"part of"* represents part – whole relationships, used only if B is necessarily a part of A, and the presence of B implies the presence of A.

The term *"has part"* refers to the part – whole relationships from the parent's perspective. It is used only when A always contains B as a part; meaning A necessarily has B as a part. If A exists, B will always exist.

The term "*regulates*" describes relationships where one process directly influences another, i.e., regulates it. It denotes necessary regulation; if both A and B are present, B always regulates A, although A may not always be regulated by B [23].

All terms (except the root terms representing each aspect) have a subclass relationship to another term: for example, *GO:0023052 signaling is a* (refers to) *GO:0050789 regulation of biological process* (Fig. 1).

Currently, various methods consistent with graph-based terminology are employed for referencing and representing logical relationships in the GO. A node corresponds to a GO term; a parent node refers to a term closer to the root of the graph, and a child node – to a term closer to terminal nodes. For *is_a* and *part_of* relationships, the parent node refers to a more general GO term, and the child node – to a more specific term [23]. The direction of the relationship is indicated by an arrow (in Fig. 1, the black *is_a* arrow points from *GO:0023052 signaling* to *GO:0050789 regulation of biological process*), while dashed lines denote implied relationships, i.e., those not explicitly stated in the ontology.

Nodes (i.e., terms) in the GO graph can have any number and types of relationships with other nodes. Similar to hierarchies, such as a family tree or species taxonomy, a node can be connected to multiple more specific (child) nodes and may also have multiple parent (more general) nodes, maintaining various relationships with each parent [5, 23]. For example, in Fig. 1, the term *GO:0051716 cellular response to stimulus* has two parents: it is a subtype of *GO:0050896 response to stimulus* and simultaneously a subtype of *GO:0009987 cellular process*.

Each GO term consists of the following elements:

Accession: a unique seven-digit identifier prefixed by GO; e.g., GO:0005739, GO:1904659, or GO:0016597).

Name: the human-readable name of the term, e.g., mitochondrion.

Ontology: indicates the category to which the term belongs, specified as *molecular_function* (MF), *biological_process* (BP), and *cellular_component* (CC).

Synonyms.

Alternate IDs.

Definition: textual description of the term with reference to the source.

Comment.

History.

Chem. react.: participation in chemical reactions.

Subset.

If information for a particular element is unavailable, it is indicated as *None*. It is mandatory to specify how the term relates to other terms in the ontology [5].

In addition to terms, GO also includes annotations – statements linking a gene product to a specific ontology term. Collectively, annotations associated with a gene provide a comprehensive characterization of its biological role. As of 2025, GO contains 8,683,287 annotations: 3,064,581 for the Biological process, 2,791,054 for the Molecular function, and 2,791,054 for the Cellular component [22].

The Evidence and Conclusion Ontology (ECO) describes various evidence types arising during scientific research that support claims. These include: evidence from experimental studies; phylogenetic evidence, based on analysis of gene functions in phylogenetic branches and inference of gene relationships;

computational evidence; author statements (with or without evidence); and inferences from literature [25]. ECO also denotes whether annotation was performed manually or automatically. Users can select data associated with specific types of evidence. As of 2025, GO contains 3,971,399 phylogenetic evidence entries, 1,049,834 experimental evidence entries, and 2,460,718 automated entries [22].

As discussed, the GO annotation process captures gene product activity and localization using terms, providing references and evidence types (including evidence codes). The annotation format is standardized – a gene association file – where each line corresponds to an association between a gene product and a GO term, including evidence codes, references, and additional information [23].

GO resources also comprise software supporting the knowledge base functionality, web access to the ontology and annotations, and analytical tools [23]. Furthermore, GO functions as a dynamic ontology, continuously revising, expanding, and updating its content in accordance with accumulating biological knowledge.

SERVICES

In addition to maintaining the knowledge base and annotations, a significant role in supporting the GO is attributed to the development and enhancement of ontology tools. Currently, GO data can be assessed through various means, including web portals, downloadable files, and APIs.

Web interfaces allow users to access GO data via standard web browsers. The most popular services are described below.

AmiGO (http://amigo.geneontology.org) [24] is the official tool for working with gene ontologies and annotations complied from model organism databases, UniProtKB, InterPro, and other sources – a total of 37 to date. The tool provides functionalities, such as uploading user data for analysis and multiple search modes [25]. AmiGO is a product of the GO Consortium and serves as the official distribution channel for GO datasets.

QuickGO (http://www.ebi.ac.uk/QuickGO) is part of the Gene Ontology Annotation (GOA) project at the European Bioinformatics Institute (EBI) [26], developed by the European Molecular Biology Laboratory (EMBL-EBI), offering extensive capabilities for query-based data retrieval.

DAVID (https://davidbioinformatics.nih.gov/) is a database for gene annotation and visualization that facilitates functional analysis to interpret large gene expression datasets in the context of GO terms [27, 28].

Blast2GO (https://www.blast2go.com/) performs comprehensive genome analysis, assigning gene ontology terms by integrating gene sequence similarity data with functional annotations and predicting molecular functions [29].

PANTHER (https://www.pantherdb.org/) is a curated database of gene and protein families used to infer gene product functions; it forms an integral part of GO resources [30].

One of the most important applications of GO is gene set enrichment analysis. This approach, utilizing available annotations, enables identification of certain GO terms significantly represented (or underrepresented) within a specific gene set. The GO website provides direct linkage to the enrichment analysis tool from the PANTHER classification system [31].

GO data files contain up-to-date and historical ontology results and annotations. In some cases, it is more practical to extract raw data directly from these files using various tools. The most commonly used file types are ontology files and association files.

There are three different versions of GO, arranged in ascending order of complexity – go-basic, go, and go-plus [23]. Go-basic is the basic version of GO, where annotations can propagate upward through the graph. The relationships between elements are described as is_a, part_of, adjustable, negatively_ regulates, and positively_regulates (explained in the previous section of the article). Since many outdated tools still use this version, it remains available.

Go is the main version of GO, which includes additional relationship types within the hierarchy, such as has_part (has a part) and comes_in (is a part of), connecting elements that are not linked in go-basic. Go-plus is the most comprehensive version of GO, encompassing more relationships than the previous versions, as well as connections to external ontologies, including ChEBI [32], Uberon Anatomy [33], and Plant Ontology [34].

In modern research, it is often essential to analyze large gene sets, including by applying gene annotation methods. However, interpreting such results is a challenging task due to the large volume of data and the complexity of relationships between elements. To address this, visualization methods for large lists of GO terms are commonly employed.

In some cases, research involves characterizing dozens or hundreds of genes whose expression may

vary under different conditions within an organism. From the perspective of elucidating the mechanisms underlying these processes, it is important to determine whether these genes participate in the same metabolic or signaling pathways and biochemical processes. These gene sets are subject to statistical enrichment tests to identify relevant functional categories [35].

GO term enrichment analysis may produce long and redundant lists of significant terms, complicating interpretation. Visualization methods help to reveal the biological meaning by summarizing the main trends in the data. This typically involves grouping similar GO terms [35].

Visualization tools include browsers for interactive viewing (such as AmiGO and QuickGO), network construction software that – while not GO-specific – can display any type of graph, including GO or its subsets (Cytoscape [36], Gephi [37], and Pajek [38]). Additionally, Cytoscape plugins specialized in processing groups of GO terms have been developed, including EnrichmentMap [39], BINGO [40], and GlueGo [41]. Other visualization approaches represent GO term groupings as tag clouds (with colored and sized text) or hierarchical tree maps (REVIGO [42], GOSummaries [43]).

Processing of GO terms and GO annotations can be performed using Python as well as other programming languages – Java, R, Perl, and Matlab. The Gene Ontology Consortium website provides a list of available software libraries at:

ftp://ftp.geneontology.org/pub/go/www/GO.tools_by_type.software.shtml.

APPLICATION

GO plays a significant role in advancing biomedical research and acquiring new biological knowledge. The primary applications of GO relate to the analysis of genomics and proteomics data through the following approaches.

Functional annotation enables the classification of genes into three categories (biological process, molecular function, and cellular component), facilitating the assessment of their biological roles. This approach allows for the identification of genes involved in specific metabolic pathways, signal transduction, and other processes, thereby providing a deeper understanding of their functions and the functions of their corresponding protein products in cellular and organismal contexts. GO annotations are widely used to identify genes associated with particular diseases and to elucidate disease mechanisms. By highlighting key genes within pathways implicated in pathological processes, researchers can pinpoint potential therapeutic targets. Such analyses have been conducted across various molecular mechanisms in humans [44–49], animals [50–52], and plants [53–56]. GO annotation has become indispensable for analyzing large-scale gene expression data. Clustering and classification of gene expression profiles enable the identification of biologically meaningful trends dependent on experimental conditions [57].

Pathway enrichment analysis employs GO terms to map genes onto specific biological pathways (e.g., cell cycle, apoptosis, immune response, etc.) and to identify genes active at different stages within a pathway. This information aids in uncovering potential disease mechanisms at the cellular level [58–61]. This approach statistically assesses the representation of particular biological pathways or processes within defined gene sets, facilitating the functional interpretation of gene expression datasets. By using GO terms, one can detect biological processes or molecular functions that are significantly enriched under varying experimental conditions, thereby assisting in the elucidation of the biological significance of gene expression changes [62–65].

Comparative functional analysis using GO annotations enables the comparison of gene functions across different species via consistent GO terms. This approach supports the evaluation of homologous gene functions and tracking of evolutionary relationships. Furthermore, GO terms allow for the prediction of gene functions based on sequence homology in species with poorly annotated genomes [66–69].

CONCLUSION

The Gene Ontology and gene annotation system described above is continuously evolving, updating its knowledge base and adapting to new experimental findings. GO is an open, publicly accessible project, designed to make the annotation of homologous gene and protein sequences across multiple organisms as flexible and dynamic as possible by employing unified vocabulary. Concurrent referencing of various external knowledge bases and meticulous curation of annotations ensure comprehensive coverage of currently known data in this domain. Gene Ontology and gene annotation systems provide the research community with powerful tools to analyze large volumes of genomics and proteomics data and identify and interpret trends in biological processes under diverse organismal conditions, including the

identification of therapeutic targets. Overall, Gene Ontology serves as a foundational knowledge system that dynamically supports the interpretation and revision of biological, genetic, and proteomics data, thereby offering new possibilities for future research.

REFERENCES

1. Podkolodnyy N.L., Podkolodnaya O.A. Ontologies in Bioinformatics and Systems Biology. *The Vavilov Journal of Genetics and Breeding*. 2015;19(6):652–660. (In Russ.). DOI: 10.18699/VJ15.090.
2. Chandrasekaran B., Josephson J.R., Benjamins V.R. What are ontologies and why do we need them? *IEEE Intelligent Systems*. 1999;14(1):20–26. DOI: 10.1109/5254.747902.
3. Gruber T.R. Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Human-Computer Studies*.1995;43(5–6):907–928. DOI: 10.1006/ijhc.1995.1081.
4. Ashburner M. On the representation of "gene function" in databases [Internet]. EMBL – European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge; 1998 June 19. Availabel: 2025 May 15. DOI: 10.5281/zenodo.5504412.
5. Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet*. 2000;25(1):25–29. DOI: 10.1038/75556.
6. Rubin G.M., Yandell M.D., Wortman J.R., Gabor Miklos G.L., Nelson C.R., Hariharan I.K. et al. Comparative genomics of the eukaryotes. *Science.* 2000;287:2204–2215. DOI 10.1126/science.287.5461.2204.
7. Tang Z., Kuo T., Shen J., Lin R.J. Biochemical and genetic conservation of fission yeast Dsk1 and human SR protein-specific kinase 1. *Mol. Cell. Biol*. 2000;20:816–824. DOI: 10.1128/mcb.20.3.816-824.2000.
8. Vajo Z., King L.M., Jonassen T., Wilkin D.J., Ho N., Munnich A. et al. Conservation of the Caenorhabditis elegans timing gene clk-1 from yeast to human: a gene required for ubiquinone biosynthesis with potential implications for aging. *Mamm. Genome*. 1999;10:1000–1004. DOI: 10.1007/s003359901147.
9. Ohi R., Feoktistova A., McCann S., Valentine V., Look A.T., Lipsick J.S. et al. Myb-related Schizosaccharomyces pombe cdc5p is structurally and functionally conserved in eukaryotes. *Mol. Cell. Biol*. 1998;18:4097–4108. DOI: 10.1128/mcb.18.7.4097.
10. Bassett D.E. Jr., Boguski M.S., Spencer F., Reeves R., Kim S., Weaver T. et al. Genome cross-referencing and XREFdb: implications for the identification and analysis of genes mutated in human disease. *Nat. Genet.* 1997;15:339–344. DOI: 10.1038/ng0497-339.
11. Kataoka T., Powers S., Cameron S., Fasano O., Goldfarb M., Broach J. et al. Functional homology of mammalian and yeast RAS genes. *Cell.* 1985;40:19–26. DOI: 10.1016/0092-8674(85)90304-6.
12. Botstein D., Fink G.R. Yeast: an experimental organism for modern biology. *Science.* 1988;240:1439–1443. DOI: 10.1126/science.3287619.
13. Bairoch A., Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res*. 2000;28:45–48. DOI: 10.1093/nar/28.1.45.
14. Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Rapp B.A., Wheeler D.L. et al. *GenBank Nucleic Acids Res*. 2000;28:15–18. DOI: 10.1093/nar/28.1.15.
15. Tateno Y., Miyazaki S., Ota M., Sugawara H., Gojobori T. DNA data bank of Japan (DDBJ) in collaboration with mass sequencing teams. *Nucleic Acids Res*. 2000;28(1):24–26. DOI: 10.1093/nar/28.1.24.
16. Barker W.C., Garavelli J.S., Huang H., McGarvey P.B., Orcutt B.C., Srinivasarao G.Y. et al. The Protein information resource (PIR). *Nucleic Acids* Res. 2000;28:41–44. DOI: 10.1093/nar/28.1.41.
17. Mewes H.W., Frishman D., Gruber C., Geier B., Haase D., Kaps A. et al. MIPS: a database for genomes and protein sequences. *Nucleic Acids Res*. 2000;28:37–40. DOI: 10.1093/nar/28.1.37.
18. Costanzo M.C., Hogan J.D., Cusick M.E., Davis B.P., Fancher A.M., Hodges P.E. et al. The Yeast Proteome Database (YPD) and Caenorhabditis elegans Proteome Database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Res*. 2000;28:73–76. DOI: 10.1093/nar/28.1.73.
19. Bateman A., Birney E., Durbin R., Eddy S.R., Howe K.L., Sonnhammer E.L. The Pfam protein families database. *Nucleic Acids Res*. 2000;28:263–266. DOI: 10.1093/nar/28.1.263.
20. Lo Conte L., Ailey B., Hubbard T.J., Brenner S.E., Murzin A.G., Chothia C. SCOP: a structural classification of proteins database. *Nucleic Acids Res*. 2000;28:257–259. DOI: 10.1093/nar/28.1.257.
21. Bairoch A. The ENZYME database in 2000. *Nucleic Acids Res*. 2000;28:304–305. DOI: 10.1093/nar/28.1.304.
22. The Gene Ontology Resource. Release statistics. Consortium is funded by the National Human Genome Research Institute (US National Institutes of Health). Availabel: 2025 May 15. URL: https://geneontology.org/stats.html
23. Gaudet P., Škunca N., Hu J.C., Dessimoz C. Primer on the gene ontology. In: Dessimoz C, Škunca N (eds.). The gene ontology handbook. Methods in molecular biology, vol.1446. Humana Press, 2016: Chapter 3. DOI: 10.1007/978-1-4939-3743-1.
24. Carbon S., Ireland A., Mungall C.J., Shu S., Marshall B., Lewis S. et al. AmiGO: online access to ontology and annotation data. *Bioinformatics.* 2009;25(2):288–289. DOI: 10.1093/bioinformatics/btn615.
25. Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res*. 2015;43(Database issue):D1049–1056/ DOI: 10.1093/nar/gku1179.
26. Binns D., Dimmer E., Huntley R., Barrell D., O'Donovan C., Apweiler R. QuickGO: a web-based tool for Gen Ontology searching. *Bioinformatics.* 2009;25(22):3045–3046. DOI: 10.1093/bioinformatics/btp536.
27. Huang D.W., Sherman B.T., Lempicki R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 2018;4(1):44–57. DOI: 10.1038/nprot.2008.211.
28. Sherman B.T., Hao M., Qiu J., Jiao X., Baseler M.W., Lane H.C. et al. DAVID: a web server for functional enrich-

ment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res*. 2022;50(W1):W216–W221. DOI: 10.1093/nar/gkac194.
29. Conesa A., Götz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant. Genomics*. 2008;2008:619832. DOI: 10.1155/2008/619832.
30. Thomas P.D., Kejariwal A., Campbell M.J., Mi H., Diemer K., Guo N. et al. "PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification". *Nucleic Acids Res*. 2003;31(1):334–341. DOI: 10.1093/nar/gkg115.
31. Mi H., Muruganujan A., Casagrande J.T., Thomas P.D. Large-scale gene function analysis with the PANTHER classifi cation system. *Nat. Protoc*. 2013;8(8):1551–1566. DOI: 10.1038/nprot.2013.092.
32. Hastings J., de Matos P., Dekker A., Ennis M., Harsha B., Kale N. et al. (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013;41(Database issue):D456–463. DOI: 10.1093/nar/gks1146.
33. Mungall C., Torniai C., Gkoutos G., Lewis S., Haendel M. Uberon, an integrative multi species anatomy ontology. *Gen. Biol.* 2012;13(1):R5. DOI: 10.1186/gb-2012-13-1-r5.
34. Cooper L., Walls R.L., Elser J., Gandolfo M.A., Stevenson D.W., Smith B. et al. The Plant Ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol*. 2013;54(2):e1. DOI: 10.1093/pcp/pcs163.
35. Rivals I., Personnaz L., Taing L., Potier M.C. Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics.* 2007;23(4):401–407. DOI: 10.1093/bioinformatics/btl633.
36. Smoot M.E., Ono K., Ruscheinski J., Wang P.L., Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2011;27(3):431–432. DOI: 10.1093/bioinformatics/btq675.
37. Bastian M., Heymann S., Jacomy M. Gephi: an open source software for exploring and manipulating networks. Proceedings of the International AAAI Conference on Web and Social Media, 3(1):361–362. DOI:10.1609/icwsm.v3i1.13937.
38. Batagelj V. Exploratory social network analysis with Pajek (Structural analysis in the social sciences). Cambridge, Cambridge University Press, 2011:442. DOI: 10.1017/9781108565691.
39. Merico D., Isserlin R., Stueker O., Emili A., Bader G.D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*. 2010;5(11):e13984. DOI: 10.1371/journal.pone.0013984.
40. Maere S., Heymans K., Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*. 2005;21(16):3448–3449. DOI: 10.1093/bioinformatics/bti551.
41. Bindea G., Mlecnik B., Hackl H., Charoentong P., Tosolini M., Kirilovsky A. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. B*ioinformatics*. 2009;25(8):1091–1093. DOI: 10.1093/bioinformatics/btp101.
42. Supek F., Bošnjak M., Škunca N., Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One.* 2011;6(7):e21800. DOI: 10.1371/journal.pone.0021800.
43. Kolde R., Vilo J. GOsummaries: an R Package for Visual Functional Annotation of Experimental Data. *F1000Res*. 2015;4:574. DOI: 10.12688/f1000research.6925.1.
44. Bright L.A., Mujahid N., Nanduri B., McCarthy F.M., Costa L.R., Burgess S.C. et al. Functional modelling of an equine bronchoalveolar lavage fluid proteome provides experimental confirmation and functional annotation of equine genome sequences. *Anim. Genet*. 2011;42(4):395–405. DOI: 10.1111/j.1365-2052.2010.02158.x.
45. Chloe Li K.Y., Cook A.C., Lovering R.C. GOing forward with the cardiac conduction system using Gene Ontology. *Front Genet*. 2022;13:802393. DOI: 10.3389/fgene.2022.802393.
46. Vinterhalter G., Kovačević J.J., Uversky V.N., Pavlović-Lažetić G.M. Bioinformatics analysis of correlation between protein function and intrinsic disorder. *Int. J. Biol. Macromol*. 2021;167:446–456. DOI: 10.1016/j.ijbiomac.2020.11.211.
47. Luo Z., Li W., Li J., Zhang Y. A new Tec family-based clinical model predicts survival in differentiated thyroid cancer patients via machine learning. *Thyroid Res*. 2025;18(1):18. DOI: 10.1186/s13044-025-00234-x.
48. Wang W., Wang H.T., Liu X., Zhu L.M., Lin T.T. Proteomic analysis of meibomian gland carcinoma cells after overexpression of thrombospondin 1. *Zhonghua Yan Ke Za Zhi.* 2025;61(5):376–383. DOI: 10.3760/cma.j.cn112142-20240709-00294.
49. He S., Nie H., Yin X., Zhong Z. Identification of key extracellular proteins as the potential biomarkers in thyroid eye disease. *PLoS One*. 2025;20(4):e0322415. DOI: 10.1371/journal.pone.0322415.
50. Buza T.J., McCarthy F.M., Burgess S.C. Experimental-confirmation and functional-annotation of predicted proteins in the chicken genome. *BMC Genomics*. 2007;8:425. DOI: 10.1186/1471-2164-8-425.
51. Piovesan D., Profiti G., Martelli P.L., Fariselli P., Fontanesi L., Casadio R. SUS-BAR: a database of pig proteins with statistically validated structural and functional annotation. *Database (Oxford).* 2013;2013:bat065. DOI: 10.1093/database/bat065.
52. Zhu Z., McClintock T.S., Bieberich E. Transcriptomics analysis reveals potential regulatory role of nSMase2 (Smpd3) in nervous system development and function of middle-aged mouse brains. *Genes Brain Behav*. 2024;23(4):e12911. DOI: 10.1111/gbb.12911.
53. Lohse M., Nagel A., Herter T., May P., Schroda M., Zrenner R. et al. Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant Cell Environ*. 2014;37(5):1250–1258. DOI: 10.1111/pce.12231.
54. Wimalanathan K., Lawrence-Dill C.J. Gene Ontology Meta Annotator for Plants (GOMAP). *Plant Methods*. 2021;17(1):54. DOI: 10.1186/s13007-021-00754-1.
55. Foulger R.E., Denny P., Hardy J., Martin M.J., Sawford T., Lovering R.C. Using the Gene Ontology to Annotate Key Players in Parkinson's Disease. *Neuroinformatics*. 2016;14(3):297–304. DOI: 10.1007/s12021-015-9293-2.
56. Sessa E.B., Masalia R.R., Arrigo N., Barker M.S., Pelosi J.A. GOgetter: A pipeline for summarizing and visualizing GO slim annotations for plant genetic data. *Appl. Plant Sci*. 2023;11(4):e11536. DOI: 10.1002/aps3.11536.

57. Yue Q., Huang C., Song P., Wang S., Chen H., Wang D. et al. Transcriptomic analysis reveals the molecular mechanisms underlying osteoclast differentiation in the estrogen-deficient pullets. *Poult. Sci*. 2023;102(3):102453. DOI: 10.1016/j.psj.2022.102453.
58. Zhai J., Lyu T., Guo Y., An Y., Xiang Y., Xie L. et al. OTX2 expression contributes progression of gastric cancer in young adults. *Sci. Rep*. 2025;15(1):16146. DOI: 10.1038/s41598-025-99632-2.
59. Huang Z., Liu D., Zhang Y., Lu W., Hu L., Zhang J. et al. PITX1 as a grading, prognostic and tumor-infiltrating immune cells marker for chondrosarcoma: a public database-based immunoassay and tissue sample analysis. *Front. Oncol*. 2025;15:1477649. DOI: 10.3389/fonc.2025.1477649.
60. Sun X., Cheng Y.M., Sun M.W., Zhang X.D., Yu X.Y. et al. High expression of SOX10 is correlated with poor prognosis and immune infiltrates in skin cutaneous melanoma. *Front. Oncol*. 2025;15:1444670. DOI: 10.3389/fonc.2025.1444670.
61. Li B.Y., Li H.L., Zeng F.E., Luan X.Y., Liu B.Q, Wang Z.Z. et al. Identification of PD-L1-related biomarkers for selecting gastric adenocarcinoma patients for PD-1/PD-L1 inhibitor therapy. *Discov. Oncol*. 2025;16(1):689. DOI: 10.1007/s12672-025-02515-1.
62. Shakeri Abroudi A., Azizi H., Djamali M., Qorbanee A., Skutella T. Integration of Microarray and Single-Cell RNA-Seq Data and Machine Learning Allows the Identification of Key Histone Modification Gene Changes in spermatogonial stem cells. *Biology (Basel)*. 2025;14(4):387. DOI: 10.3390/biology14040387.
63. Zheng Y., Yu S.Y., Yan X., Li J.P., Zhang Q., Yuan X. [Gene expression profiling analysis of stress-sensitive genes and their potential functions in myoblasts]. *Shanghai Kou Qiang Yi Xue.* 2025;34(1):7–13. (In Chin.).
64. Wang Y., Li Q. Integrated multiomics analysis identifies potential biomarkers and therapeutic targets for autophagy associated AKI to CKD transition. *Sci. Rep.* 2025;15(1):13687. DOI: 10.1038/s41598-025-97269-9.
65. Qian X., Jia W., Li Y., Chen J., Zhang J., Sun Y. COL4A1 Promotes Gastric Cancer Progression by Regulating Tumor Invasion, Tumor Microenvironment and Drug Sensitivity. *Curr. Med. Chem.* 2025. DOI: 10.2174/0109298673351943250314074632.
66. Peng S., Zhang Q., Yang Y., Li Y., Feng W., Zhao D. et al. Genome-wide identification and expression profiling of MYB transcription factors in Artemisia argyi. *BMC Genomics.* 2025;26(1):384. DOI: 10.1186/s12864-025-11441-z.
67. Tang H., Finn R.D., Thomas P.D. TreeGrafter: phylogenetic tree-based annotation of proteins with Gene Ontology terms and other annotations. *Bioinformatics*. 2019;35(3):518–520. DOI: 10.1093/bioinformatics/bty625.
68. Gaudet P., Livstone M.S., Lewis S.E., Thomas P.D. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief Bioinform*. 2011;12(5):449–462. DOI: 10.1093/bib/bbr042.
69. Gaudet P., Dessimoz C. Gene ontology: pitfalls, biases, and remedies. *Methods Mol. Biol.* 2017;1446:189–205. DOI: 10.1007/978-1-4939-3743-1_14.

Author Information

Chasovskikh Natalia Yu. – Dr. Sci. (Med.), Associate Professor, Head of the Division of Medical and Biological Cybernetics, Siberian State Medical University, Tomsk, nch03@mail.ru, https://orcid.org/0000-0001-6077-0347

(✉) **Chasovskikh Natalia Yu.,** nch03@mail.ru

Received on May 22, 2025;
approved after peer review on May 30, 2025;
approved on September 04, 2025