

УДК 575.112

<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.

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>.

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.

✉ Часовских Наталья Юрьевна, nch03@mail.ru

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>.

ВВЕДЕНИЕ

Современные исследования в области геномики дают возможность оперировать большим количеством данных. При этом должна решаться задача поиска биологического смысла в полученной информации. Основной вопрос – как сопоставить данные о последовательностях с известными результатами функционального анализа генов и как получить новые знания о функциях генов? На помощь приходят инструменты онтологии – согласованного описания конкретной предметной области, в данном случае – геномики и молекулярной биологии [1].

Онтология с точки зрения информатики – это модель представления объектов, свойств объектов и отношений между ними [2]. Данная модель в рамках конкретной предметной области содержит набор понятий (терминов) с определениями и атрибутами, соответствующее множество аксиом и правил вывода [3]. При этом обычно базы знаний предметной области включают базы, содержащие общие знания о множестве классов понятий, свойствах и отношениях между ними, а также базы знаний об индивидуальных объектах, их свойствах и связях с другими объектами (т. е. с конкретными данными). В базе знаний обе компоненты взаимосвязаны [1].

Применительно к области геномики, для разработки онтологии функций генов также создавался словарь терминов с определениями, устанавливалась иерархия взаимосвязей между ними. Большим плюсом явилось то, что одна и та же онтология позволяла изучать функциональные аспекты геномов раз-

личных организмов, а также аннотировать функции новых последовательностей [4].

При разработке генной онтологии особое внимание уделялось вопросам, которые необходимо описывать в терминах: место экспрессии гена в организме и субклеточная локализация продукта гена; время экспрессии гена (в понятиях онтогенеза организма); функция генного продукта и место в иерархии процессов; какими генами регулируется активность продукта данного гена. Первоначально решение данных задач было реализовано применительно к базе данных геномов дрозофилы FlyBase [4]. При этом в качестве основных измерений классификации использовались понятия молекулярной функции и биологического процесса. Впервые созданная таким образом генная онтология была описана в статье М. Ashburner и соавт. [5].

На примере первого сравнения двух полных эукариотических геномов (дрожжи *Saccharomyces cerevisiae* и червь *Caenorhabditis elegans*) было показано, что значительная часть генов в этих организмах является ортологами. При этом 12% генов червей кодируют белки, биологическое значение которых может быть экстраполировано из их сходства с ортологами в дрожжах (27% генов последних). Оказалось, что такие белки вовлечены в общие для всех эукариот биологические процессы (например, репликация ДНК, транскрипция и метаболизм). Дальнейшее сравнение геномов дрожжей, червей и плодовых мушек также подтвердило наличие у них гомологов [6]. Показано также, что гены и белки, которые вовлечены в основные биологические процес-

сы, с большой долей вероятности являются ортологами, что уже подтверждено для млекопитающих и модельных организмов (например, дрожжей) [7–12].

Высокая степень подобия и функциональной консервативности для геномов разных видов, показанная в дальнейшем, открыла новые возможности для автоматизированного переноса биологических аннотаций из экспериментально исследованных модельных организмов в менее исследованные организмы, тем самым упрощая формирование генной онтологии.

КОНСОРЦИУМ ГЕННОЙ ОНТОЛОГИИ

Изначально генная онтология (консорциум генной онтологии, GO Consortium) являлась совместным проектом трех баз данных модельных организмов: FlyBase16, Mouse Genome Informatics17,18 (MGI) и Saccharomyces Genome Database19 (SGD). Позднее к ним присоединились и другие базы данных организмов [5].

Цель Консорциума – создание структурированного, точно определенного, общего, контролируемого словаря для описания ролей генов и генных продуктов в любом организме – остается неизменной, несмотря на прогресс в ее реализации и динамичном развитии системы терминологии. Каждый элемент в GO связан с другими видами информации, базами данных генов и белков, таких как SwissPROT [13], Gen-Bank [14], DDBJ [15], PIR [16], MIPS [17], YPD и WormPD [18], Pfam [19], SCOP [20] и ENZYME [21], что позволяет учитывать постоянное пополнение и изменение знаний об объектах [5].

Основная терминология GO описывает гены и их продукты в рамках трех категорий (аспектов). Биологический процесс включает одну или несколько упорядоченных совокупностей молекулярных функций, подразумевающую часто химическое или физическое преобразование. Примеры высокоуровневых, общих терминов биологических процессов (рисунок) – *GO:0065007 biological regulation* (биологическая регуляция) или *GO:0050896 response to stimulus* (ответ на стимул), низкоуровневых – *GO:0042770 signal transduction in response to DNA damage* (передача сигнала в ответ на повреждение ДНК) или *GO:0072331 signal transduction by p53 class mediator* (передача сигнала медиаторами класса p53). Молекулярная функция описывает биохимическую активность продукта гена. Примеры общих терминов «фермент», «транспортер» или «лиганд», а более узких – «аденилатциклаза» или «лиганд Toll-рецептора». Клеточный компонент – место активности продукта гена в клетке, примеры – «рибосома», «ядерная мембрана» или «аппарат Гольджи» [5].

По состоянию на 2025 г. в онтологии содержалось 26037 терминов в аспекте «биологический процесс», 10154 – «молекулярная функция», 4023 – «клеточный компонент» [22].

Структура GO является графом, в котором каждый термин – узел, а связи между узлами – ребрами, а также присутствует иерархия т. е. дочерний термин более детализирован, чем родительский (рисунок), узел может иметь более одного родительского термина. В GO также заложены определения и категории для отношений между терминами. Часто используемыми отношениями являются: *is a* (является подтипом); *part of* (часть); *has part* (имеет часть); *regulates* (регулирует), *negatively regulates* (отрицательно регулирует), *positively regulates* (положительно регулирует) [23].

Особенности отношений между терминами:

- *is a* формирует базовую структуру GO. Если *A is a B*, подразумевается, что узел *A* является подтипом узла *B*. Например, митотический клеточный цикл *is a* (является подтипом) клеточного цикла;

- *part of* используется для представления отношений «часть–целое», *part of* добавляется между *A* и *B* только в том случае, если *B* обязательно является частью *A*, и присутствие *B* подразумевает присутствие *A*;

- *has part* представляет собой отношение «часть–целое» с точки зрения родителя. Как и в случае с *part of*, данное отношение используется только в случаях, когда *A* всегда содержит *B* как часть, т. е. когда *A* обязательно имеет часть *B*. Если *A* существует, *B* всегда будет существовать;

- *regulates* используется для описания отношений, когда один процесс напрямую влияет на другой процесс, т. е. регулирует его. Применяется для обозначения обязательной регуляции: если присутствуют *A* и *B*, *B* всегда регулирует *A*, но *A* не всегда может регулироваться *B*) [23].

Все термины (за исключением корневых, представляющих каждый аспект) имеют отношение подкласса к другому термину. Так, *GO:0023052 signaling* (сигналинг) *is a* (является подтипом) *GO:0050789 regulation of biological process* (регуляция биологического процесса) (см. рисунок).

На сегодняшний день принят ряд способов (в соответствии с терминологией, основанной на графе) ссылаться на логические отношения и представлять их в GO. Подразумевается, что узел ссылается на термин GO, при этом родительский узел ссылается на узел, который ближе к корню графа, а дочерний узел – на узел, который ближе к конечным узлам. Для отношений *is a* и *part of* родительский узел будет более общим термином GO, а дочерний – более конкретным термином [23]. Стрелка указывает направление отношения (на рисунке черная стрелка *is a*

направлена от *GO:0023052 signaling* к *GO:0050789 regulation of biological process*, а пунктирные линии обозначают предполагаемое отношение, т.е. то, которое не было явно указано в онтологии.

Узлы (то есть термины) в графе GO могут иметь любое количество и тип отношений с другими узлами. Подобно иерархиям – генеалогическому древу или таксономии видов – узел может иметь связи с более чем одним дочерним (с более конкретизированным описанием) узлом, но также может иметь более одного родительского (с более общим описанием) узла и различные отношения с его различными родителями [5, 23, 24]. На рисунке термин *GO:0051716 cellular response to stimulus* (клеточный ответ на стимул) имеет два родительских узла: является подтипом *GO:0050896 response to stimulus* (ответ на стимул) и в то же время подтипом *GO:0009987 cellular process* (клеточный процесс).

Каждый термин GO включает следующие элементы:

Accession (идентификационный номер – уникальный семизначный идентификатор с префиксом GO:. Например, GO:0005739, GO:1904659 или GO:0016597).

Name (наименование термина, понятное человеку – например, митохондрия).

Ontology (обозначает, к какой из трех категорий принадлежит термин). Указывается как *molecular_function* (MF, молекулярная функция), *biological_process* (BP, биологический процесс) и *cellular_component* (CC, клеточный компонент).

Synonyms (синонимы).

Alternate Ids (альтернативный идентификатор).

Definition (определение – текстовое описание термина и ссылка на источник информации).

Comment (комментарии).

History (история).

Chem. react. (участие в химических реакциях).

Subset (подмножество).

В тех случаях, когда информации для какого-либо элемента нет, указывается ее отсутствие (None). Обязательно прописывается, как термин соотносится с другими терминами в онтологии [5].

Помимо терминов в GO также включен блок аннотаций – утверждений, связывающих продукт гена с конкретным термином онтологии. Таким образом, все связанные с определенным геном аннотации позволяют комплексно охарактеризовать его биологическую роль. На 2025 г. GO включает 8 683 287 аннотаций, из них для категории биологического процесса – 3 064 581, молекулярной функции – 2 791 054, клеточного компонента – 2 791 054 [22].

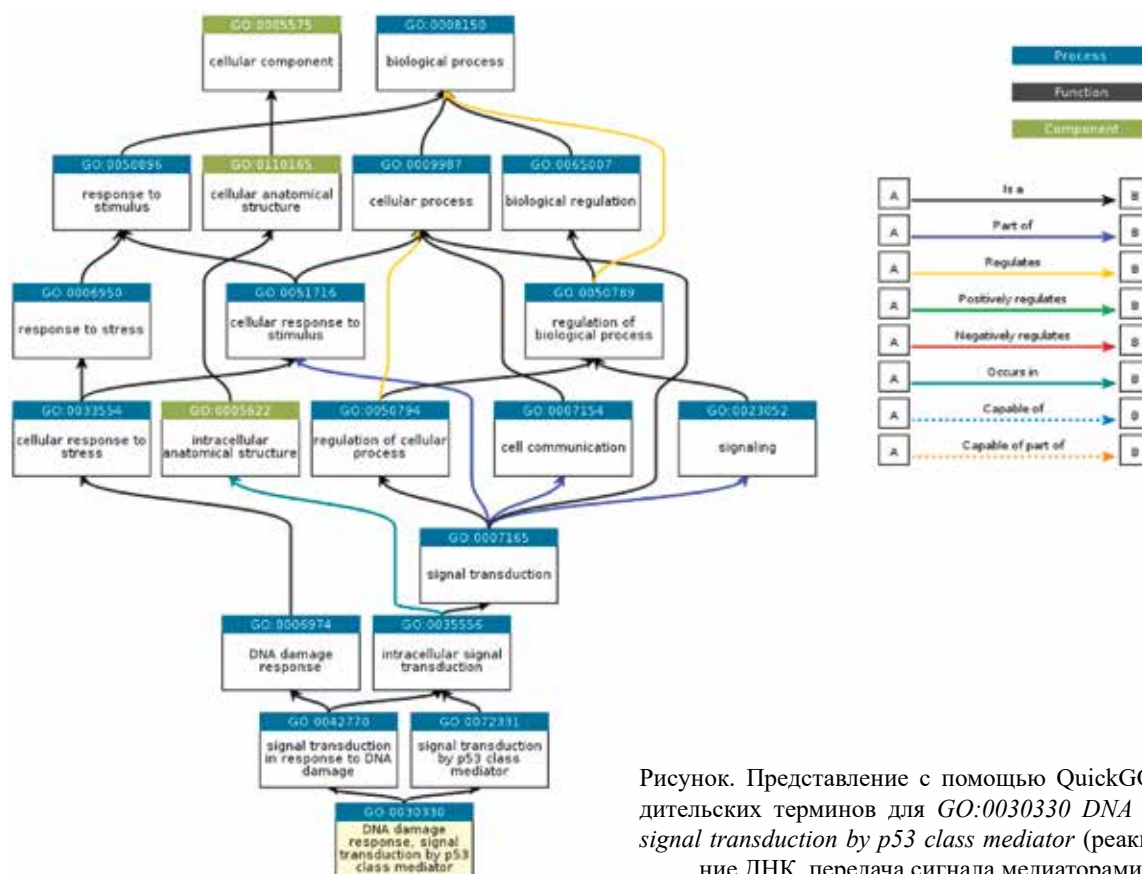


Рисунок. Представление с помощью QuickGO диаграммы родительских терминов для *GO:0030330 DNA damage response, signal transduction by p53 class mediator* (реакция на повреждение ДНК, передача сигнала медиаторами класса p53)

Онтология доказательств и заключений (The Evidence and Conclusion Ontology, ECO) служит для описания различных типов доказательств, возникающих в ходе научного исследования и используемых для поддержки утверждений. Это могут быть: доказательства, полученные в результате экспериментальных исследований; филогенетические доказательства, в основе которых анализ функций генов в определенных ветвях филогенетического дерева и вывод о взаимосвязи между генами; доказательства, полученные в результате вычислительных методов; заявления, сделанные авторами (имеют доказательства либо нет); выводы, сделанные по литературе [25]. Также указывается, вручную или автоматически была произведена аннотация. Пользователи могут выбирать данные, связанные с определенными типами доказательств. На 2025 г. в GO содержится 3 971 399 филогенетических доказательств, 1 049 834 доказательства, полученных из экспериментов, и 2 460 718 – автоматических [22].

Как показано выше, в GO процесс аннотации фиксирует активность и локализацию генного продукта с помощью терминов, предоставляя ссылку и вид доказательств (в том числе указывая код доказательства). Формат вывода информации об аннотации в GO стандартизован – это файл ассоциаций генов. Каждая строка в текстовом файле представляет ассоциацию между генным продуктом и термином GO, содержит коды доказательств, ссылки и другую информацию [23].

Ресурсы GO также включают программное обеспечение для функционирования базы знаний, веб-доступ к онтологии и аннотациям и аналитические инструменты [23]. Кроме того, GO, как динамическая онтология, поддерживает актуальное состояние информации, пересматривая ее, расширяясь и обновляясь по мере накопления биологических знаний.

СЕРВИСЫ

Помимо наполнения базы знаний и аннотации важное место в поддержании GO играет разработка и совершенствование инструментов онтологии. На сегодняшний день существуют различные способы доступа к данным GO – через веб-порталы, загружаемые файлы и API. Веб-интерфейсы позволяют с помощью обычных веб-браузеров получать доступ к данным GO. Ниже представлены наиболее популярные сервисы.

AmiGO (<http://amigo.geneontology.org>) [24] – официальный инструмент работы с онтологиями генов и аннотациями, собранными из баз данных модельных организмов, UniProtKB, InterPro и других источни-

ков, всего 37 в настоящий момент. Инструмент предоставляет возможность загрузки собственных данных для анализа и разнообразные режимы поиска [25]. AmiGO является продуктом Консорциума GO и официальным каналом распространения наборов данных GO.

QuickGO (<http://www.ebi.ac.uk/QuickGO>) – проект Gene Ontology Annotation (GOA) в Европейском институте биоинформатики [26], является продуктом Европейской лаборатории молекулярной биологии (EMBL-EBI), также предоставляет широкие возможности для поиска данных по запросам.

DAVID (<https://davidbioinformatics.nih.gov/>) – база данных для аннотации генов и ее визуализации, позволяет проводить функциональный анализ для интерпретации больших наборов данных по экспрессии генов в контексте терминов GO [27, 28].

Blast2GO (<https://www.blast2go.com/>) – осуществляет полный анализ генома, присваивает термины генной онтологии путем интеграции данных о сходстве последовательностей генов с функциональными аннотациями, прогнозирует молекулярные функции [29].

PANTHER (<https://www.pantherdb.org/>) – курируемая база данных семейств генов и белков, используемая для определения функций продуктов генов, является частью GO [30].

Одна из важнейших задач, решаемых с помощью GO, – это анализ обогащения наборов генов. Данный метод с использованием доступных аннотаций позволяет выяснить, какие термины GO представлены (или нет) для конкретной совокупности генов. Веб-сайт GO позволяет напрямую связать пользователей с инструментом анализа обогащения из системы классификации PANTHER [31].

Файлы данных GO содержат текущие и долгосрочные результаты онтологии и аннотации, причем в ряде случаев удобнее извлекать необработанные данные непосредственно из этих файлов с помощью различных инструментов. Наиболее часто используют файлы онтологии и файлы ассоциаций.

Существует три различные версии GO в порядке возрастания сложности – go-basic, go и go-plus [23]. Версия go-basic – базовая редакция GO, аннотации могут распространяться вверх по графу. Отношения между элементами описываются как *is_a*, *part_of*, *adjustable*, *negatively_regulates* и *positively_regulates* (расшифрованы в предыдущем разделе статьи). В связи с тем, что многие устаревшие инструменты продолжают пользоваться данной версией, она пока остается доступной.

Версия go – основная редакция GO, включает дополнительные типы отношений в иерархии, такие как

has_part (имеет часть) и comes_in (входит), соединяющие несвязанные в go-basic элементы. Версия go-plus – наиболее содержательная версия GO, включает больше отношений, чем предыдущие, и связей с внешними онтологиями, в том числе с ChEBI [32], Uberon anatomy [33] и Plant Ontology для растений [34].

В современных исследованиях часто приходится анализировать большие наборы генов, в том числе применять методы геной аннотации. Однако осмысление таких результатов ввиду их значительно объема и сложности взаимоотношений элементов является трудной задачей. Для ее упрощения применяют методы визуализации больших списков терминов GO.

В ряде случаев при исследованиях необходимо охарактеризовать десятки или сотни генов, экспрессия которых может меняться при различных состояниях организма. При этом с точки зрения раскрытия механизмов происходящих процессов важно понимать, участвуют ли эти гены в одних и тех же метаболических или сигнальных путях, биохимических процессах. Данные наборы генов подвергаются статистическим тестам на обогащение для выявления различных функциональных категорий [35].

Анализ обогащения терминами GO может приводить к формированию длинных и избыточных списков значимых терминов, что также затрудняет интерпретацию. Именно в этом случае методы визуализации помогают выявить биологический смысл, обобщая основные тенденции охарактеризованных данных. Как правило, для этого используют группировку сходных терминов GO [35].

Возможности визуализации предоставляются браузерами для интерактивного просмотра (AmiGO и QuickGO), инструментами построения сетей, которые не являются специфическими для GO, но могут отображать любой вид графика, включая GO или его часть (Cytoscape [36], Gephi [37] и Pajek [38]). Также для Cytoscape разработаны плагины, специализированные на обработке групп терминов GO, например, EnrichmentMap [39], BINGO [40], ClueGo [41]. Кроме того, для визуализации применяются методы, отображающие особенности группировки терминов GO в «облаке тегов» (с текстом разных цветов и размеров) или в древовидной карте иерархической организации (REVIGO [42], GOSummaries [43]).

Обработку терминов GO и аннотаций GO можно осуществлять с помощью Python, а также других языков – Java, R, Perl и Matlab. На веб-сайте консорциума Gene Ontology содержится список программных библиотек, доступных из ftp://ftp.geneontology.org/pub/go/www/GO.tools_by_type.software.shtml.

ПРИМЕНЕНИЕ

Генная онтология играет важную роль в развитии биомедицинских исследований и получении новых биологических знаний. Основные области применения GO связаны с анализом данных геномики и протеомики с помощью следующих подходов.

Функциональная аннотация. Позволяет классифицировать в трех категориях (биологический процесс, молекулярная функция и клеточный компонент) гены, что помогает оценить их биологическое назначение. Таким образом можно идентифицировать гены, участвующие в определенных метаболических путях, передаче сигнала и других процессах, лучше понять их роль, а также роль соответствующих белковых продуктов в реализации функций клеток и организма в целом. Аннотации GO активно применяются для идентификации генов, связанных с определенными заболеваниями и для прояснения механизмов этих заболеваний. Выделяя ключевые гены, вовлеченные в пути, связанные с развитием патологического процесса, исследователи могут определять потенциальные терапевтические мишени. Такие исследования проводятся при анализе различных молекулярных механизмов в организме человека [44–49], животных [50–52], растений [53–56]. Аннотация GO стала незаменимой для анализа больших объемов данных экспрессии генов. При этом кластеризация и классификация профилей экспрессии генов позволяют определять биологически значимые тенденции в зависимости от условий эксперимента [57].

Анализ обогащения путей, при котором термины GO можно использовать для сопоставления генов с определенными биологическими путями (клеточного цикла, апоптоза, иммунного ответа и другими) и выявлять гены, вовлеченные в реализацию разных этапов одного пути. Полученная таким образом информация позволяет идентифицировать потенциальные механизмы заболеваний на клеточном уровне [58–61]. Данный подход статистически оценивает представленность тех или иных биологических путей или процессов в определенных наборах генов. Это позволяет анализировать наборы данных об экспрессии генов в контексте функций. Используя термины GO, можно идентифицировать биологические процессы или молекулярные функции, которые значительно обогащены в различных экспериментальных условиях, что помогает интерпретировать биологическое значение изменений экспрессии генов [62–65].

Также аннотации GO помогают осуществлять сравнение функций генов у разных видов с помо-

пью согласованных терминов ГО. Такой подход позволяет оценить, как эволюционируют функции гомологичных генов, отследить эволюционные связи. При этом термины ГО позволяют предсказывать функции генов на основе гомологии последовательностей у видов плохо аннотированных геномов [66–69].

ЗАКЛЮЧЕНИЕ

Рассмотренная выше система генной онтологии и генной аннотации постоянно развивается, обновляя базу знаний и данных, откликаясь на появление новых экспериментальных результатов. ГО является открытым общественным проектом, концепция которого призвана сделать, по возможности, гибкой и динамичной аннотацию гомологичных последовательностей генов и белков во многих организмах с использованием общего словаря. При этом обращение к различным внешним базам знаний, курация аннотаций обеспечивают охват большинства известных на сегодняшний день данных в этой предметной области. Генная онтология и генная аннотация открывают новые возможности для исследовательского сообщества по изучению больших объемов данных геномики и протеомики, идентификации и осмыслению тенденций биологических процессов, возникающих при различных состояниях организма, в том числе по поиску терапевтических мишеней. В целом генная онтология является базисом для формирования системы знаний, осмысления и пересмотра данных биологии, генетики и протеомики в динамике, предоставляя новые возможности исследователям и в будущем.

СПИСОК ИСТОЧНИКОВ

1. Подколотный Н.Л., Подколотная О.А. Онтологии в биоинформатике и системной биологии. *Вавиловский журнал генетики и селекции*. 2015;19(6):652–660. 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 pro-

- teins 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 enrichment 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 classification 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. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res.* 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.* 2009;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. *Bioinformatics.* 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. GOSummary: 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-Lazetić 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., Qorbane 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: 10.1007/978-1-4939-9736-7_12.

Информация об авторе

Часовских Наталья Юрьевна – д-р мед. наук, доцент, зав. кафедрой медицинской и биологической кибернетики, СибГМУ, г. Томск, nch03@mail.ru, <https://orcid.org/0000-0001-6077-0347>

✉ **Часовских Наталья Юрьевна**, nch03@mail.ru

Поступила в редакцию 22.05.2025;
одобрена после рецензирования 30.05.2025;
принята к публикации 04.09.2025