



МИРОВЫЕ РЕПОЗИТОРИИ, СОДЕРЖАЩИЕ СВЕДЕНИЯ О РЕЗУЛЬТАТАХ ПОПУЛЯЦИОННЫХ ИММУНОЛОГИЧЕСКИХ И ГЕНЕТИЧЕСКИХ ИССЛЕДОВАНИЙ

А.Г. Титова[✉], Г.А. Трусов, А.В. Баёв, Д.В. Сосин, Д.Н. Нечаев, А.Н. Ломов, В.В. Макаров, В.С. Юдин, С.М. Юдин

Центр стратегического планирования и управления медико-биологическими рисками здоровью Федерального медико-биологического агентства, Москва, Россия

Введение. Благодаря активному развитию мультиомиксных технологий исследователи получают все больше сведений о генетических и иммунологических данных человека. Для систематизации и хранения такого рода информации используются репозитории, что ускоряет поиск и использование сведений для научных исследований и решения прикладных задач в области медицины.

Цель. Провести анализ мирового опыта эксплуатации репозитория генетических и иммунологических данных человека для определения их функциональных особенностей и роли в развитии популяционной иммунологии и генетики.

Обсуждение. Проанализированы функциональные особенности репозитория генетических и иммунологических данных. Данные о репозиториях, включенных в исследование, были получены из открытых источников. Процесс отбора репозитория включал три этапа: подбор научных публикаций, дедупликация, фильтрация по критериям отбора. Основные критерии последующей оценки репозитория генетических и иммунологических данных человека: объем данных; доступность данных; форматы данных. Поиск сведений о репозиториях и биобанках на территории Российской Федерации проводился по поисковым запросам в сети Интернет. В исследовании было проанализировано 15 крупнейших репозитория генетических и иммунологических данных, из которых 37,5% аффилированы с Великобританией, 43,75% — с США. Для создания и поддержания крупных репозитория, как правило, формируются международные и межинституциональные консорциумы. Доступность репозитория генетической информации обеспечивается комбинацией технологических, организационных и правовых механизмов. Наиболее частыми источниками финансирования репозитория являются государственный бюджет, средства частных фондов и благотворительных организаций, инвестиции фармацевтических компаний. Основные риски, возникающие при эксплуатации репозитория, можно разделить на четыре группы: этические, правовые, биологические и технологические, связанные с обеспечением конфиденциальности данных. В Российской Федерации генетические исследования являются одним из наиболее активно развивающихся направлений науки. В этой связи становятся актуальными задачи безопасного хранения, этичного использования и правовой защиты получаемых данных. Рассмотрены возможные направления для дальнейшего развития национальных репозитория генетических и иммунологических данных, а также возможности дополнительного регулирования обращения с генетическими данными на законодательном уровне.

Выводы. На основании анализа данных определены возможные риски, связанные с функционированием репозитория, предложены различные подходы к их минимизации и оптимизации развития репозитория. В качестве одного из наиболее перспективных направлений рассматривается разработка интеграционных модулей на основе искусственного интеллекта для обработки и аннотирования данных, представленных в стандартизированных протоколах.

Ключевые слова: репозиторий; генетические данные; иммунологические данные; научные исследования; риски; биобанк

Для цитирования: Титова А.Г., Трусов Г.А., Баёв А.В., Сосин Д.В., Нечаев Д.Н., Ломов А.Н., Макаров В.В., Юдин В.С., Юдин С.М. Мировые репозитории, содержащие сведения о результатах популяционных иммунологических и генетических исследований. *Медицина экстремальных ситуаций*. 2025;27(3):328–340. <https://doi.org/10.47183/mes.2025-277>

Финансирование: исследование выполнено без спонсорской поддержки.

Потенциальный конфликт интересов: С.М. Юдин — член редакционной коллегии журнала «Медицина экстремальных ситуаций».

✉ Титова Анастасия Германовна ATitova@cspfmba.ru

Статья поступила: 21.02.2024 **После доработки:** 03.06.2025 **Принята к публикации:** 15.07.2025 **Online first:** 20.08.2025

INTERNATIONAL DATA REPOSITORIES OF POPULATION-BASED IMMUNOLOGICAL AND GENETIC RESEARCH

Anastasia G. Titova[✉], George A. Trusov, Arthur V. Bayov, Dmitry V. Sosin, Denis N. Nechaev, Alexey N. Lomov, Valentin V. Makarov, Vladimir S. Yudin, Sergey M. Yudin

Centre for Strategic Planning and Management of Biomedical Health Risks of the Federal Medical and Biological Agency, Moscow, Russia

Introduction. Due to the active development of multiomics technologies, more and more information about human genetic and immunological research is becoming available. Data repositories are used to systematize and store such information, which facilitates the search and use of information for carrying out scientific research and solving applied problems in the area of medicine.

Objective. To analyze the global experience of using repositories of human genetic and immunological data to define their functional features and role in the development of population immunology and genetics.

Discussion. Functional features of genetic and immunological data repositories were analyzed. The data on the repositories included in the study was obtained from open sources. The selection process for repositories included three stages: selection of scientific publications, de-duplication, and filtering based on selection criteria. The main criteria for the subsequent evaluation of human genetic and immunological data repositories were as follows: data volume, data accessibility, and data formats. The search for information about repositories and biobanks in the Russian Federation was conducted using online search queries on the Internet. The study analyzed 15 largest genetic and immunological data repositories, of which 37.5% are affiliated with the UK and 43.75% are affiliated with the USA. The task of creating and maintaining large

© А.Г. Титова, Г.А. Трусов, А.В. Баёв, Д.В. Сосин, Д.Н. Нечаев, А.Н. Ломов, В.В. Макаров, В.С. Юдин, С.М. Юдин, 2025

repositories is solved, as a rule, by forming international and inter-institutional consortia. The availability of genetic data repositories is ensured by a combination of technological, organizational, and legal mechanisms. The most common sources of repository funding are state budgets, funds from private foundations and charitable organizations, and investments from pharmaceutical companies. The main risks associated with the operation of a repository can be divided into four groups: ethical, legal, biological, and technological risks related to data privacy. In the Russian Federation, genetic research is one of the most rapidly developing scientific directions. As a result, the challenges of secure storage, ethical use, and legal protection of data are acquiring particular importance. The presented review discusses possible directions for further development of national genetic and immunological data repositories, as well as the possibilities of additional regulation of genetic data handling at the legislative level.

Conclusions. The conducted review has identified possible risks associated with repository functioning and proposed various approaches to minimizing these risks and optimizing the development of data repositories. One of the most promising areas is the development of AI-based integration modules for processing and annotating data presented in standardized protocols.

Keywords: repository; genetic data; immunological data; scientific research; risks; biobank

For citation: Titova A.G., Trusov G.A., Bayov A.V., Sosin D.V., Nechaev D.N., Lomov A.N., Makarov V.V., Yudin V.S., Yudin S.M. International data repositories of population-based immunological and genetic research. *Extreme Medicine*. 2025;27(3):328–340. <https://doi.org/10.47183/mes.2025-277>

Funding: the study was carried out without sponsorship.

Potential conflict of interest: Sergey M. Yudin is a member of the Editorial Board of *Extreme Medicine*. The other authors declare no conflict of interest.

✉ Anastasia G. Titova ATitova@cspfmmba.ru

Received: 21 Feb. 2024 **Revised:** 3 June 2025 **Accepted:** 15 July 2025 **Online first:** 20 Aug. 2025

ВВЕДЕНИЕ

В настоящее время сохраняется тенденция к интенсивному развитию и внедрению в рамках научно-исследовательских работ методических решений для проведения популяционных иммунологических и генетических научных исследований человека. Генетические и иммунологические популяционные исследования представляют собой междисциплинарную область, объединяющую генетику, иммунологию, биостатистику и биоинформатику. Иммунология является неотъемлемой частью таких популяционных исследований, поскольку генетические факторы играют ключевую роль в формировании иммунного ответа [1]. Кроме того, изучение генетических вариаций позволяет определять механизмы развития аутоиммунных и инфекционных заболеваний [2]. Актуальным направлением исследований является оценка связи генетической гетерогенности и здоровья человека, включая предрасположенность к развитию различных заболеваний [3].

Основными направлениями популяционных иммунологических и генетических исследований являются выявление генетических особенностей популяции человека (например, адаптивные), определение генетических и молекулярных детерминант заболеваний человека, изучение влияния генетических вариаций на восприимчивость к лекарственным препаратам (фармакогенетика), механизмов формирования иммунного ответа человека и оценка динамики изменений его показателей. В процессе исследований используются данные клинических анализов, методы молекулярной биологии, иммунологии и генетики. Результаты исследований являются значимыми прежде всего для развития такого направления прикладной науки, как персонализированная медицина, включающая разработку новых методов диагностики и оценки индивидуальных рисков, создание персонализированных лекарственных препаратов для терапии и формирование программ по профилактике заболеваний исходя из индивидуальных генетических и иммунологических особенностей человека. На текущий

момент ведущими методическими подходами в области популяционной иммунологии и генетики остаются секвенирование ДНК и РНК, в том числе на уровне отдельных клеток; генотипирование; иммунофенотипирование; биоинформатический анализ и т.д.

Популяционные исследования генетических и иммунологических маркеров являются приоритетным направлением развития медицины, особенно персонализированной. В настоящее время генетический анализ используется для прогнозирования рисков заболеваний, а мониторинг иммунного статуса — для оценки эффективности терапии, причем эти данные можно стандартизировать. Для комплексной оценки здоровья пациента необходимо не только использование генетических и иммунологических данных, но и возможность их интегрировать в единую систему анализа данных по переменным биохимическим и физиологическим показателям. Для интеграции различных количественных/качественных медико-биологических показателей в единые модели оценки здоровья необходима разработка новых методов и стандартов по переменным параметрам (даже от одного и того же пациента), обеспечивающих возможность интерпретации и сравнения сводных данных. Будущее развитие этой области во многом связано с технологическим прогрессом: современные методы полногеномного секвенирования (включая анализ длинных фрагментов и повторов) в сочетании с мощными вычислительными ресурсами позволяют детально исследовать вклад генетической изменчивости и ее взаимодействия с различными факторами в развитие заболеваний. Это не только улучшает наше понимание биологии человека, но и открывает возможности для расширения функционала биомедицинских репозиторий, включающих специализированные базы данных пациентов.

Актуальность темы обусловлена высоким объемом накопленных в результате исследований генетических и иммунологических данных человека, необходимостью их систематизации и обеспечения доступности для научного сообщества. Так, на территории Российской Федерации, по данным домена «Наука



Рисунок подготовлен авторами с помощью ресурса «Домен «Наука и инновации»» (<https://gisnauka.ru>)

Рис. Динамика научно-исследовательских работ в период 2010–2025 гг.

и инновации»¹, в период 2010–2025 гг. было зафиксировано около 15 000 научно-исследовательских работ, в которых упоминались генетические технологии, и около 3000 научно-исследовательских работ, упоминающих секвенирование; соответствующие данные представлены на рисунке.

В настоящее время по всему миру формируются репозитории для хранения генетических и иммунологических данных. Основная задача репозитория заключается в аккумулировании и систематизации данных, связанных с генетической изменчивостью, для последующего анализа и разработки новых методов диагностики и оценки индивидуальных рисков, создания персонализированных лекарственных препаратов для терапии и формирования программ по профилактике заболеваний исходя из индивидуальных генетических и иммунологических особенностей человека.

Создание крупных генетических репозиториях сопряжено с комплексом взаимосвязанных проблем. Во-первых, остро стоят этические вопросы: необходимо получить информированное согласие пациента, обеспечить строгую конфиденциальность его данных, а также информировать пациента о мерах предосторожности для защиты генетической информации на его личных устройствах. Во-вторых, существуют аналитические и технические задачи: постоянно растущие объемы данных требуют непрерывного совершенствования методов их интерпретации и аннотирования, а также разработки подходов для комплексного анализа в рамках различных научных концепций. Это, в свою очередь, диктует необходимость решения проблем управления данными, включая формирование четких правил доступа для квалифицированного персонала и обеспечение надежной безопасности хранения информации. Наконец, при интерпретации результатов важно учитывать биологическую сложность: реализация генетической информации в фенотипические признаки всегда модулируется множеством факторов окружающей среды.

Цель исследования — провести анализ мирового опыта эксплуатации репозиториях генетических и иммунологических данных человека для определения их функциональных особенностей и роли в развитии популяционной иммунологии и генетики.

Для достижения поставленной цели авторами сформулированы основные задачи: анализ мирового

опыта практического применения сведений, хранящихся в репозиториях генетических и иммунологических данных человека для определения функциональных особенностей таких репозиториях; определение возможных рисков, в том числе этических рисков, рисков нарушения конфиденциальности персональных данных и несанкционированного использования, правильности интерпретации и достоверности данных; определение возможных путей устранения рисков; формирование предложений для повышения уровня внедрения результатов генетических и иммунологических исследований в практическую медицину.

МАТЕРИАЛЫ И МЕТОДЫ

В настоящем исследовании представлен анализ деятельности существующих репозиториях генетических и иммунологических данных по сведениям, полученным из открытых источников. Согласно полученным из открытых источников данным было отобрано 15 репозиториях на основании наибольших показателей реферируемости в профессиональной среде, частоты упоминания в научных публикациях, данных консорциумов и в результате анализа официальных веб-ресурсов. Процесс отбора репозиториях включал трехэтапную стратегию: подбор научных публикаций, дедупликацию, фильтрацию по критериям отбора.

На основе данных PubMed и Google Scholar была сформирована подборка, включающая более 100 научных публикаций за период 2018–2024 гг. по ключевым запросам: «genetic data repository», «immunological data repository».

Исходный пул публикаций был проанализирован и отсортирован по частоте упоминаний, в результате чего был сформирован предварительный перечень репозиториях и удалены дублирующиеся репозиториях.

Далее были исключены репозиториях с недостаточным описанием или закрытым/платным доступом. На следующем этапе отбора учитывали следующие критерии: объем данных (более 500 значений), географическая репрезентативность (в исследование включены репозиториях из 4 стран: США, Великобритания, Испания, Нидерланды), степень доступности данных (открытый/закрытый/платный доступ или только для сотрудников организации-разработчика), применение международных стандартов для хранения

¹ Домен «Наука и инновации». <https://gisnauka.ru/> (дата обращения: 10.04.2025).

данных (репозитории, содержащие данные в наиболее распространенных форматах, например vcf).

Для дальнейшего анализа выделены основные критерии оценки репозитория генетических и иммунологических данных человека, а именно: объем данных (количество и разнообразие хранимых данных); доступность данных (открытость доступа, лицензирование); форматы данных (поддерживаемые форматы файлов).

Для анализа деятельности репозитория отобраны ресурсы, сформированные на территориях различных стран и действующие на момент проведения исследования.

Поиск сведений о репозиториях и биобанках на территории Российской Федерации проводился в сети Интернет по поисковым запросам: «репозиторий генетических данных», «репозиторий иммунологических данных».

Несмотря на наличие множества репозитория, включенные в исследование ресурсы представляли собой наиболее значимые и востребованные для научного сообщества системы. Ограничение анализируемой выборки связано с необходимостью включения в нее информационных ресурсов только репозитория с хорошо документированными данными. Итоговая выборка охватывает глобальные и нишевые платформы, релевантные для дальнейшего анализа генетической или иммунологической информации в рамках научных исследований.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В список отобранных репозитория было включено 15 информационных ресурсов: IGSR, AFND, GWAS, gnomAD, NCBI GEO, BBMRI-NL, FHLdb, AIRR Data Commons, EGA, IEDB, OMIM, NIDDK Central Repository, ArrayExpress, ENA, dbGaP [4–16]. Среди них 1 (6,25%) репозиторий был создан в Нидерландах, 2 (12,5%) — в Испании, 6 (37,5%) — в Великобритании и 7 (43,75%) — в США.

Для создания и поддержания крупных репозитория, как правило, формируются международные и межинституциональные консорциумы. Наиболее частыми источниками финансирования репозитория являются средства государственного бюджета (представляющие значительную часть), частных фондов и благотворительных организаций, инвестиции фармацевтических компаний. Так, в США финансирование таких исследований осуществляется Национальным институтом здравоохранения (NIH), свой вклад вносят частные фонды и благотворительные организации. Фармацевтические компании также выделяют финансирование, так как они заинтересованы в результатах генетических исследований, которые могут помочь в разработке новых лекарственных препаратов.

Генетические и иммунологические популяционные исследования являются фундаментальной основой для развития современной медицины: разработки новых методов диагностики, терапии и профилактики заболеваний.

Доступность данных в репозиториях — это одно из ключевых условий для проведения научных исследований, поскольку их использование позволяет обеспечивать: воспроизводимость исследований,

проведение сравнительного анализа, разработку новых инструментов и методов. Именно открытый доступ к данным стимулирует создание новых алгоритмов и программного обеспечения для их анализа. Доступность репозитория генетической информации обеспечивается комбинацией технологических (стандартизированные форматы данных, техническая инфраструктура), организационных (метаданные и аннотации) и правовых механизмов (персональные данные и этические аспекты). Самый распространенный в настоящее время вид доступа к данным — веб-интерфейсы. Такой доступ может быть бесплатным или платным (лицензионным). Например, данные из IGSR общедоступны и могут свободно распространяться [4].

Описание характеристик ведущих генетических репозитория

Проведен обзор открытых источников с целью сравнения репозитория генетических и иммунологических данных человека. В соответствии с предварительно определенными критериями оценки были идентифицированы 15 ведущих репозитория, обладающих значительными массивами данных. Сформирован перечень сводных данных для каждого отобранного репозитория, который приведен в таблице. В перечень также включены ссылки на основополагающие статьи, описывающие структуру, назначение и особенности функционирования репозитория, если такие ссылки были указаны на их официальных сайтах.

Среди 15 указанных репозитория 1 (6,25%) разработан в Нидерландах, 2 (12,5%) — в Испании, 6 (37,5%) — в Великобритании (1 из которых совместно с Испанией), 7 (43,75%) — в США. В перечне, указанном в таблице 1, 12 (80%) составляют репозитории с открытым доступом к данным и 3 (20%) с закрытым доступом, но с подробным описанием сведений, содержащихся в репозитории. Чаще всего данные для загрузки представлены в формате txt. Однако следует отметить, что в перечень внесены не все существующие в мире репозитории, а только наиболее крупные. Эти репозитории выделяются не только объемом данных, но и открытостью, что способствует международному сотрудничеству и повышению воспроизводимости исследований.

Отдельно следует отметить, что некоторые репозитории предоставляют доступ только к части данных. Однако закрытые информационные ресурсы, специализирующиеся на хранении конфиденциальных данных, также находят свое применение в исследованиях. Содержащиеся в этих базах сведения, по согласованию с их владельцем, обеспечивают доступ к клиническим, фенотипическим и генетическим показателям, что делает их незаменимыми для медицинских исследований, где важна защита частной информации. Также существуют закрытые репозитории, которые не вошли в перечень.

Общие принципы организации генетических репозитория

В ходе исследования установлено, что важными условиями существования качественного репозитория являются:

Таблица. Сводные данные о репозиториях

№	Название репозитория	Направление	Описание	Страна	Объем данных	Доступность данных	Форматы данных для загрузки
1	The International Genome Sample Resource (IGSR) ²	Генетика	Каталог общих генетических вариаций человека, включающий образцы, взятые с согласия людей	Великобритания	2504 образцов из 26 популяций	Открытый доступ	VCF, FASTQ, BAM
2	Allele Frequency Net Database (AFND) ³	Генетика, иммунология	База данных, содержащая частоты аллелей иммунных генов и соответствующие им аллели в различных популяциях	Великобритания	Количество частот — 155 685 (HLA), 6731 (KIR), 4376 (цитокин), 877 (MIC) от 14 264 290 чел. Исследования населения — 1802 человека, данные генов/аллелей — 1786 человек, данные гаплотипа — 684 человек, данные о генотипе — 192 человека	Открытый доступ	CSV
3	The NHGRI-EBI Catalog of human genome-wide association studies (GWAS) ⁴	Генетика	Содержит информацию об ассоциациях генетических маркеров с фенотипами, полезными для фармакогенетики	Великобритания	7083 публикации, 692 444 основные ассоциации и 96 947 полных сводных статистик. Данные сопоставлены с Genome Assembly GRCh38.p14 и dbSNP Build 15	Открытый доступ	TSV, OWL/RDF, EFO
4	The Genome Aggregation Database (gnomAD) ⁵	Генетика	Представляет экзомные последовательности и полные геномы для изучения редких генетических вариантов	США	730 947 экзомных последовательностей и 76 215 полных геномных последовательностей от неродственных индивидуумов с различным происхождением	Открытый доступ	VCF, TSV
5	Gene Expression Omnibus (NCBI GEO) ⁶	Генетика	База данных для профилирования экспрессии генов и профилирования метилирования РНК	США	4348 наборов данных	Открытый доступ	TXT, XML, SOFT, MIAME.
6	Biobanking Netherlands (BBMRI-NL) ⁷	Генетика	Набор данных и «омик»-сигнатур заболеваний	Нидерланды	Генетические, эпигенетические, транскриптомные и метаболомные данные 35 000 образцов из 29 когорт	Закрытый доступ	TXT

² IGSR: The International Genome Sample Resource (проект 1000 Genomes). <http://www.internationalgenome.org/> (дата обращения: 27.11.2024).

³ Allele Frequency Net Database. <http://www.allelefrequenciest.net> (дата обращения: 27.11.2024).

⁴ GWAS: The NHGRI-EBI Catalog of human genome-wide association studies. <https://www.ebi.ac.uk/gwas/> (дата обращения: 27.11.2024).

⁵ gnomAD: The Genome Aggregation Database. <https://gnomad.broadinstitute.org/> (дата обращения: 27.11.2024).

⁶ NCBI GEO: Gene Expression Omnibus. <https://www.ncbi.nlm.nih.gov/geo/> (дата обращения: 27.11.2024).

⁷ BBMRI: Biobanking Netherlands. <https://www.bbMRI.nl/> (дата обращения: 28.11.2024).

Продолжение таблицы

№	Название репозитория	Направление	Описание	Страна	Объем данных	Доступность данных	Форматы данных для загрузки
7	Database on the Molecular Basis of Familial Hemophagocytic Lymphohistiocytosis (FHLdb) ⁸	Генетика, иммунология	База данных вариантов семейного гемофагоцитарного лимфогистиоцитоза	Испания	Информация о зарегистрированных вариантах в 4 генах, связанных с FHL (<i>PRF1</i> , <i>UNC13D</i> , <i>STXBP2</i> , <i>STX11</i>), включающих 579 вариантов (в т.ч. миссенс, нонсенс, индел, сплайсинг и т. д.)	Открытый доступ	JSON
8	AIRR Data Commons ⁹	Генетика, иммунология	Данные использования технологий секвенирования для изучения репертуаров антител/рецепторов В-клеток и Т-клеток	США	5,2 млрд аннотированных последовательностей, 67 000 клонов, 133 000 отсортированных, одиночных клеток В/Т	Открытый доступ	MIAIRR, YAML/JSON
9	European Genome-phenome Archive (EGA) ¹⁰	Генетика, иммунология	Архивирование и распространение персональных идентифицируемых генетических и фенотипических данных	Великобритания, Испания	11 775 наборов генетических, фенотипических и клинических данных	Закрытый доступ	TAR.GZ
10	The Immune Epitope Database (IEDB) ¹¹	Генетика, иммунология	Ресурс для поиска и экспорта иммунных эпитопов	США	Пептидные эпитопы 1 621 303; непептидные эпитопы 3189; анализ Т-клеток 541 542; анализ В-клеток 1 414 095; анализ лигандов МНС 4 881 627; организмы — источники эпитопа 4579; ссылки 25 400	Открытый доступ	XLSX, CSV, TCB, JSON
11	Online Mendelian Inheritance in Man (OMIM) ¹²	Генетика	Постоянно обновляемый каталог генов, генетических нарушений и фенотипических признаков у человека и их взаимосвязи	США	Аутосомных генов 26 080; Х-связанных 1382; Y-связанных 63; митохондриальных 72	Открытый доступ	TXT
12	Central Repository — National Institutes of Health (NIDDK) ¹³	Генетика	Централизованный исследовательский ресурс данных о диабете, заболеваниях пищеварительной системы и почек	США	5767 наборов данных	Закрытый доступ	CSV

⁸ FHLdb: Database of variants of Familial Hemophagocytic Lymphohistiocytosis syndrome. <https://www.biotoclin.org/FHLdb/> (дата обращения: 28.11.2024).⁹ AIRR Data Commons. <https://docs.airr-community.org/en/stable/index.html> (дата обращения: 28.11.2024).¹⁰ EGA: European Genome-phenome Archive. <https://www.ebi.ac.uk/ega> (дата обращения: 28.11.2024).¹¹ IEDB: The Immune Epitope Database. <https://www.iedb.org/> (дата обращения: 28.11.2024).¹² OMIM: Online Mendelian Inheritance in Man. <http://www.omim.org/> (дата обращения: 28.11.2024).¹³ NIDDK Central Repository — National Institutes of Health. <https://repository.niddk.nih.gov/home/> (дата обращения: 29.11.2024).

Продолжение таблицы

№	Название репозитория	Направление	Описание	Страна	Объем данных	Доступность данных	Форматы данных для выгрузки
13	Functional genomics data (Array-Express) ¹⁴	Генетика	Коллекция данных функциональной геномики	Великобритания	78 511 наборов данных	Открытый доступ	CEL, TXT, XML
14	European Nucleotide Archive (ENA) ¹⁵	Генетика	Ресурс биоданных, в том числе нуклеотидов. Хранилище, обеспечивающее доступ к аннотированным последовательностям ДНК и РНК, к информации об экспериментальных процедурах и т.д.	Великобритания	Сборка 2 046 549; последовательность 23 430 609; кодирующая ДНК 38 244 291; не кодирующая ДНК 1 265 026; прочтение 2 990 747; анализ 998 067	Открытый доступ	CSV
15	Database of Genotypes and Phenotypes (dbGaP) ¹⁶	Генетика	База данных генотипов и фенотипов	США	Генотип 4 039 007; анализ экспрессии 422 847; соматические мутации 100 614; геном 683 996; эпигеном 88 733	Открытый доступ	TXT

Таблица составлена авторами

- стандартизация, так как размещаемые в репозиториях данные должны соответствовать определенным установленным стандартам с целью возможности их корректной обработки, сравнения, проведения автоматизированного анализа и долгосрочного хранения;
- воспроизводимость, в том числе важна унификация форматов файлов;
- наличие описания экспериментальных условий и метаданных, описывающих, как данные были получены, какие образцы использовали и какие анализы были проведены.

Форматы хранения данных

В работе нами был выявлен высокий уровень гетерогенности экспериментальных платформ и методов анализа, что приводит к разнообразию форматов данных. Это связано в первую очередь с человеческим фактором: для решения одинаковой биоинформатической задачи разные специалисты используют разные методологические подходы, что приводит к отсутствию возможности сопоставления полученных результатов в будущем.

Вместе с тем для обеспечения совместимости данных при обмене информацией между различными репозиториями используются стандартизированные форматы и онтологии. Так, формат FASTQ используется для хранения последовательностей прочтений, полученных методом секвенирования; SAM — текстовый

формат для хранения выровненных последовательностей; BAM — бинарный формат для хранения выровненных последовательностей; VCF — для вариаций генома, более компактный, чем SAM, используется для хранения и анализа больших объемов данных секвенирования; TXT (текстовый формат, tab-delimited) — наиболее простой и распространенный формат, содержащий значения экспрессии для каждого гена в каждом образце, подходит для импорта в большинство программ анализа данных; CSV (comma-separated values) — похож на формат TXT, но разделение значений запятыми, с широкой поддержкой; SOFT (simple omnibus format in text) — более структурированный формат, чем TXT/CSV, содержащий метаданные о платформе, образцах, а также данные экспрессии генов, что позволяет автоматизировать обработку и анализ большого количества данных, рекомендуется, как правило, для комплексного анализа; MIAME (minimum information about a microarray experiment) — формат на основе XML, содержит больше метаданных, чем SOFT, обеспечивая максимальную воспроизводимость и интерпретируемость данных, часто используется для обмена информацией между базами данных и применяется для анализа; TCB и EFO — файлы для работы в Binary Data или другом программном обеспечении; RDF — формат для представления взаимосвязанных данных; XML — размеченный файл, который позволяет определять и хранить данные; MIAIRR — наборы данных, определяющие, какая информация должна сопровождать данные репертуара

¹⁴ ArrayExpress: Functional genomics data. <https://www.ebi.ac.uk/biostudies/arrayexpress> (дата обращения: 27.11.2024).

¹⁵ ENA: European Nucleotide Archive. <https://www.ebi.ac.uk/ena/browser/home> (дата обращения: 29.11.2024).

¹⁶ dbGaP (Database of Genotypes and Phenotypes). <https://www.ncbi.nlm.nih.gov/gap/> (дата обращения: 29.11.2024).

TCR/BCR, чтобы их можно было корректно интерпретировать; YAML — структурированная запись информации; JSON — стандартный текстовый формат для хранения и передачи структурированных данных; TAR.GZ — архив данных; XLSX — данные в виде таблицы; CEL — содержит все типы данных, которые получены; TSV — текстовый формат для представления таблиц баз данных; OWL — язык описания онтологий.

Компоненты и структура репозитория

Обычно репозиторий состоит из нескольких компонентов: база данных — центральное хранилище информации, где хранятся данные, например последовательности ДНК, профили экспрессии генов, фенотипические данные, а также метаданные и онтологии; веб-интерфейс, позволяющий пользователям искать, фильтровать, выгружать, визуализировать и анализировать данные; система контроля версий, которая позволяет отслеживать изменения данных и метаданных, а также восстанавливать предыдущие версии при необходимости; система управления доступом, которая определяет, какие пользователи имеют доступ к каким данным, что особенно важно для защиты конфиденциальности персональных данных.

Отмечается разнообразие метаданных: в некоторых репозиториях содержатся данные только по стране проживания людей, участвующих в исследовании; в других могут содержаться более подробные данные, например о возрасте, поле и этнической принадлежности людей, участвующих в исследовании. Отдельно следует обозначить ресурсы, направленные на визуализацию генетических данных, так называемые геномные браузеры.

Структура баз данных бывает: централизованной (доступной для широкого круга исследователей; как правило, такой подход применяется в крупных международных проектах); децентрализованной (доступной только для ограниченного круга пользователей; такой подход применяется во многих исследовательских институтах и университетах). Помимо этого следует отметить, что могут создаваться консорциумы, объединяющие усилия нескольких исследовательских групп для создания и поддержания крупных баз данных. При любой структуре репозитория важным элементом системы является сохранность данных на протяжении длительного времени, что особенно важно для долгосрочных исследований. В настоящий момент встречается применение облачных технологий для хранения генетических и иммунологических данных.

Генетические базы данных являются фундаментом современных геномных исследований и представляют собой обширные коллекции информации о геноме человека, других организмов, включая вариации последовательностей ДНК, их частоту в различных популяциях и связь с фенотипическими признаками. Эти базы данных можно разделить на две основные подкатегории: общие геномные базы данных (например, проект 1000 Genomes) и базы данных генотипирования и фенотипирования (например, NHGRI GWAS) [6].

Иммунологические базы данных (например, IEDB) играют ключевую роль в понимании механизмов функционирования иммунной системы, которая является сложной сетью клеток, тканей и молекул, защищающих

организм от инфекций и других угроз. Эти базы данных можно разделить на следующие категории: базы данных иммунных рецепторов, базы данных иммунофенотипирования [13].

Мультиомиксные базы данных представляют собой интегративные базы данных, полученных с помощью различных «омик»-технологий, включая геномику, транскриптомику, протеомику и метаболомику. Применение таких баз обеспечивает более полное понимание биологических процессов и сложных взаимодействий между разными уровнями биологической организации (например, применяется в GEO) [8].

Репозитории обычно предоставляют открытый доступ к данным, хотя иногда могут быть ограничения, связанные с конфиденциальностью или авторскими правами. Также встречаются репозитории смешанного типа, предоставляющие часть данных в открытом доступе, а другую часть в закрытом или платном доступе.

В некоторых репозиториях пользователи могут проводить поиск по различным критериям (например, тип данных, организм, заболевание) и выгружать их для дальнейшего анализа.

Конфиденциальность данных

Конфиденциальность данных в генетических и иммунологических репозиториях является критически важным аспектом. Для ее обеспечения используется целый комплекс мер, направленных на защиту персональной информации участников исследований. Основными мерами обеспечения конфиденциальности являются: анонимизация данных; агрегация данных, когда данные нескольких индивидов объединяются для создания более крупных групп, что затрудняет идентификацию отдельных участников; шифрование данных при хранении на серверах и при передаче между системами; многоуровневая ролевая модель доступа; многофакторная аутентификация; логирование действий; детекция признаков несанкционированного доступа; соглашения о конфиденциальности; регулярные проверки безопасности на уязвимости; дополнительные меры могут применяться меры по ограничению срока хранения данных.

Риски функционирования репозитория

Основными рисками при работе репозитория являются этические, правовые, биологические [17]. Следует отметить этические риски, связанные с получением генетических и иммунологических данных в исследованиях. Так, например, при проведении исследования геномов коренных народов прибрежного Эквадора [18], американских индейцев [19], учитывая, что исследуемые когорты являются узкораспространенными и обладают уникальными характерными отличиями, даже при использовании самых современных методов анонимизации существует риск утечки данных. Неправильное использование генетических данных также может привести к ложным выводам и дискриминации. А открытие генетических маркеров, связанных с определенными заболеваниями, может способствовать социальной стигматизации. При этом монополизация генетических и иммунологических данных может

ограничить доступ к важным медицинским исследованиям и разработкам.

Помимо прочего, в зоне риска находится вопрос защиты интеллектуальной собственности на методы получения данных и на сами генетические и иммунологические данные. Генетические данные часто рассматриваются как открытие естественного явления, а не как изобретение, что затрудняет их патентование. Вместе с тем генетические данные регулярно обновляются и дополняются, что создает сложности в определении границ интеллектуальной собственности. Также стоит отметить, что в настоящий момент существуют значительные отличия в законодательстве разных стран, что затрудняет защиту интеллектуальной собственности генетических и иммунологических данных на международном уровне. Защита интеллектуальной собственности, основанной на генетических и иммунологических данных, может вступать в противоречие с другими правами, такими как право на неприкосновенность частной жизни и право на информацию. Существуют также и технические сложности, такие как идентификация и отслеживание данных, защита от несанкционированного использования. Возможными решениями обозначенных проблем могут стать:

- создание независимых этических комитетов для оценки проектов, связанных с использованием генетических данных;
- повышение осведомленности общественности о важности защиты генетических и иммунологических данных и связанных с этим этических проблемах;
- регламентация открытого доступа к генетическим данным для ускорения научных исследований и предотвращения монополизации;
- разработка гибких лицензионных соглашений, которые бы обеспечивали баланс между защитой интеллектуальной собственности и общественным доступом;
- создание международных соглашений и стандартов для регулирования защиты интеллектуальной собственности на генетические данные.

Основные биологические риски связаны с неконтролируемым выпуском генно-модифицированной пищевой продукции; редактированием генома человека; созданием биологического оружия [17]. В России процедуры выпуска генно-модифицированной пищевой продукции носят разрешительный порядок на основе лицензирования, сертификации и регистрации генно-модифицированных организмов, а также контроля. В Российской Федерации в Уголовном кодексе определены положения, регулирующие сферу ответственности за создание и использование биологического оружия.

Генетические репозитории на территории Российской Федерации

Российские репозитории генетических и иммунологических данных не были включены в исследование ввиду несоответствия критериям отбора, однако их

упоминание необходимо для формирования целостного представления о развитии направления на территории Российской Федерации.

В Российской Федерации активно развивается направление генетических исследований, в связи с чем возникает ряд важных вопросов, связанных с хранением, использованием и защитой генетических данных. В 2019 году была утверждена Федеральная научно-техническая программа «Развитие генетических технологий»¹⁷. Эта программа направлена на развитие генетических технологий в России. Так, в рамках программы были образованы три центра геномных исследований мирового уровня.

В России существует несколько открытых генетических репозиторий, доступных для исследователей. В 2021 году Центр Genetico и лаборатория Сербалаб в сотрудничестве с Институтом биоинформатики создали и разместили в открытом доступе первую российскую базу данных генетических вариантов и их встречаемости в российской популяции — RUSeq. Эта база содержит информацию о генетических вариантах, выявленных у более 6000 человек. Следует отметить, что, как и в большинстве зарубежных репозиторий, в RUSeq хранятся деперсонализированные данные [20].

Также существует Национальный агрегатор открытых репозиторий российских университетов (НОРА)¹⁸, который объединяет результаты исследований российских ученых и предоставляет доступ к материалам, опубликованным в открытом доступе. На базе Всероссийского института генетических ресурсов растений имени Н.И. Вавилова (ВИР) также осуществляется сбор, хранение и изучение генетических ресурсов растений и предоставляется доступ к коллекциям института для научных исследований.

В 2024 году вступил в силу закон о создании Национальной базы генетической информации¹⁹ — государственной информационной системы для обеспечения национальной безопасности, охраны жизни и здоровья граждан. Она гарантирует суверенитет в сфере хранения и использования генетических данных, а также обмен информацией между государственными органами и обладателями соответствующей информации. Это позволит проводить масштабные генетические исследования и разрабатывать новые методы диагностики и терапии заболеваний, развивать фармацевтическую промышленность и повышать качество медицинской помощи.

В период 2020–2024 гг. в ФМБА России была разработана одна из крупнейших в мире «База данных популяционных частот генетических вариантов»²⁰, в которой содержатся данные 120 000 условно здоровых людей, а также информация о более чем 550 000 000 уникальных генетических вариантах и их распространенности в российской популяции. Следует отметить, что структура этой базы данных централизованная и аналогична международным репозиториям.

Существует Национальная генетическая инициатива «100 000 + Я» — уникальный российский проект, направленный на улучшение методов диагностики

¹⁷ Федеральная научно-техническая программа развития генетических технологий на 2019–2027 годы. <http://government.ru/docs/36457/> (дата обращения: 29.11.2024).

¹⁸ «Национальный агрегатор открытых репозиторий» (НОРА). <https://www.openrepository.ru/> (дата обращения: 29.11.2024).

¹⁹ Национальная база генетической информации. <http://nrcki.ru/product/mic-izvestiya/-48080.shtml> (дата обращения: 29.11.2024).

²⁰ База данных популяционных частот генетических вариантов. <https://nir.cspfmba.ru/info> (дата обращения: 29.11.2024).

и терапии наследственных и онкологических заболеваний путем определения генотипов 100 000 россиян из разнообразных географических регионов и разных популяций для поиска генетических вариантов, которые встречаются на территории России, обобщения их сходства и выявления различий²¹. Инициатива «100 000 + Я» реализуется на базе ООО «Биотек кампус» и была разработана совместно ПАО «НК «Роснефть» и МГУ им. М.В. Ломоносова.

Отдельно следует отметить российский ресурс, внедренный ФМБА России, — Национальный информационный ресурс, содержащий сведения о популяционных иммунологических и генетических исследованиях, проводимых в Российской Федерации. Сведения о таких исследованиях содержат результаты научно-исследовательских работ, направленных на получение информации, неотъемлемо связанной с молекулярно-генетическими особенностями человека, способствующих изучению показателей состояния здоровья, формированию прогноза риска развития хронических заболеваний, а также оценке функциональных особенностей иммунной системы человека в норме и при патологических состояниях. В мире существует достаточно много разрозненных репозиториях, содержащих генетические и иммунологические данные, однако ресурс ФМБА России является уникальным решением, предоставляющим возможность комплексного анализа сведений о большом количестве исследований в едином цифровом интерфейсе. Функционирование этого ресурса в будущем может позволить ускорить формирование коллабораций научных коллективов на основе анализа их компетенций и опыта в применении передовых методов проведения генетических и иммунологических популяционных исследований, что может способствовать развитию науки в области здравоохранения на национальном уровне. В настоящий момент аналоги подобного ресурса на территории Российской Федерации отсутствуют.

Помимо прочего, в России активно развивается сеть биобанков. Биобанк — это репозиторий, который в том числе содержит биологические образцы человека (кровь, слюну, ткани) и связанные с ними генетические данные. Для каждого биобанка необходим контроль качества. Например, такие биобанки созданы на базах: ФГБУ «НМИЦ ТПМ» Минздрава России²², ФГБУ «НМИЦ им. В.А. Алмазова» Минздрава России²³ и Первого МГМУ им. И.М. Сеченова Минздрава России²⁴. Биообразцы из этих биобанков используются для проведения различных исследований, включая изучение генетических биомаркеров заболеваний и разработку новых лекарственных препаратов.

В работе анализировали информацию исключительно из открытых баз данных без учета репозитория с закрытым для учреждений из Российской Федерации и платным доступом, что является основным ограничением проведенного исследования. Также можно отметить, что в задачи представленной работы не входило проведение анализа и установление особенностей всех существующих репозиториях, в связи

с чем в исследование был включен только ряд репозиториях с большим количеством хорошо документированных данных.

Стандартизация репозиториях генетических данных играет ключевую роль в обеспечении совместимости, воспроизводимости и эффективного использования информации, так как унифицированные форматы хранения, такие как FASTQ для последовательностей, VCF для геномных вариантов и BAM/SAM для выравниваний, минимизируют риски ошибок при анализе, а поддержка облачных интеграций и открытых API (Application Programming Interface, интерфейс прикладного программирования) позволяет масштабировать исследования, избегая ручного скачивания больших объемов данных. Стандарты метаданных гарантируют полноту описания образцов (источник, методы секвенирования), что важно для корректного сравнения результатов между различными исследованиями. Единые биоинформатические протоколы снижают вариативность обработки данных. Без такой стандартизации интеграция масштабных проектов, таких как IGSR, с клиническими базами была бы невозможна, что существенно замедлило бы идентификацию патогенных мутаций и развитие персонализированной медицины. В настоящее время проводится ряд исследований с целью оптимизации форматов хранения генетических данных [21] и протоколов их биоинформатической обработки [22].

Следует отметить, что совместное хранение генетических и иммунологических данных успешно реализуется в следующих репозиториях: Allele Frequency Net Database, FHLdb, AIRR Data Commons, EGA, IEDB. Необходимо подчеркнуть, что разработка единого мирового репозитория с возможностью интеграции модулей, содержащих стандартизованные результаты генетических и популяционных иммунологических исследований, может оказать значимое положительное влияние на развитие медицины и привести к ускорению технологического прогресса.

Этическое регулирование генетических исследований направлено на баланс между научным прогрессом и защитой прав людей, включая обязательное получение информированного согласия, гарантию анонимности данных и предотвращение возможной дискриминации. При этом особое внимание уделяется уязвимым группам (коренным народам, пациентам с редкими заболеваниями), чьи данные могут быть использованы только при условии их прямого участия в принятии решений. Современные вызовы, такие как риск реидентификации анонимных геномов или применение искусственного интеллекта в анализе ДНК, требуют постоянного обновления правовых норм и усиления кибербезопасности для сохранения доверия общества к генетическим исследованиям. Существует мнение о том, что этические комитеты должны иметь более значимый вес, чем в настоящий момент [23]. Важно отметить, что защита от этических рисков является актуальной темой обсуждений как в России, так и за рубежом.

На территории Российской Федерации генетическая информация и определение ее правового режима

²¹ Национальная генетическая инициатива «100 000 + Я». <https://www.biotechcampus.ru/> (дата обращения: 29.11.2024).

²² Биобанк на базе ФГБУ «НМИЦ ТПМ» Минздрава России. <https://gnicpm.ru/scientific-directions/biobank.html> (дата обращения: 29.11.2024).

²³ Биобанк на базе ФГБУ «НМИЦ им. В.А. Алмазова» Минздрава России. http://www.almazovcentre.ru/?page_id=69701 (дата обращения: 29.11.2024).

²⁴ Биобанк на базе Первого МГМУ им. И.М. Сеченова Минздрава России. <https://www.sechenov.ru/pressroom/news/zamorozhennaya-kollektsiya-kak-sechenovskiy-universitet-sozdaet-biobank-dlya-nauchnykh-issledovaniy/> (дата обращения: 29.11.2024).

регулируются Федеральным законом от 3 декабря 2008 г. № 42-ФЗ²⁵, в соответствии с которым геномную информацию определяют в качестве персональных данных, включающих кодированную информацию об определенных фрагментах дезоксирибонуклеиновой кислоты физического лица или неопознанного трупа, не характеризующие их физиологические особенности, и Федеральным законом от 27 июля 2006 г. № 152-ФЗ²⁶, в соответствии с которым геномная информация тесно связана с категорией биометрических персональных данных (ст. 11) и требует оформления согласия индивида на ее последующую обработку и распространение (ст. 7). Однако существует мнение о том, что некорректно относить генетическую информацию только к персональным данным, поскольку обладателями данной информации являются не только само лицо, принявшее участие в исследовании, но и его генетические родственники [24]. В связи с этим представляется целесообразным на законодательном уровне определение генетической информации в качестве самостоятельного вида данных. После формирования такого определения возникнет необходимость дальнейшего законодательного регулирования обращения с генетической информацией. Стоит подчеркнуть, что многие российские авторы научных публикаций отмечают необходимость усовершенствования нормативно-правовой базы, например в части генетической паспортизации населения [25], защиты генетических данных [26] или развития геномного права [27].

Патентование генетических последовательностей как вещества допустимо, но сопряжено с рисками дублирования прав на идентичные последовательности, что требует ограничения абсолютной защиты. Альтернатива — патентование конкретного применения последовательности (например, терапевтического метода), требующее международной гармонизации из-за конфликтов интересов. Современные нормы фокусируются на промышленной применимости, обязывая раскрывать функциональную аннотацию гена, а не только его структуру. Режим ноу-хау хотя и возможен, но ограничивает доступ к данным, препятствуя научным инновациям, и малопривлекателен из-за отсутствия исключительных прав. Также существует мнение, что необходимо более четкое разграничение понятий «открытие» и «изобретение» в области генетических исследований [28]. В большинстве правовых систем, включая российскую, в настоящее время признается принципиальная возможность патентования гена (части гена), но при определенных условиях, установление которых призвано не допустить монополизации знания о природе.

ВЫВОДЫ

В долгосрочной перспективе развитие генетических и популяционных иммунологических исследований

позволит значительно улучшить здоровье населения и повысить качество жизни. Международные репозитории генетических данных играют ключевую роль в развитии современной медицины и науки, предоставляя исследователям доступ к объемным базам данных для изучения генетической изменчивости, предрасположенности к заболеваниям и реакции на лечение. Однако для эффективного использования этих данных необходимо решить проблемы, связанные с повышением качества и безопасности хранимой информации, обеспечением конфиденциальности, стандартизацией методических решений для удобства анализа и интерпретации данных специалистами широкого профиля, в том числе не обладающими навыками программирования (молекулярные биологи, врачи и др.). Постоянное развитие технологий и появление новых стандартов позволяют репозиториям становиться все более эффективным и удобным инструментом для пользователей. С точки зрения развития репозитория перспективной представляется разработка интеграционного модуля для применения при обработке и аннотации данных искусственного интеллекта со стандартизированными протоколами. Использование такого модуля для аннотации данных и автоматизации анализа позволит ускорить обработку информации и повысить эффективность использования массивов хранимых данных. Также развитие технологий конфиденциального анализа данных, таких как федеративное обучение искусственного интеллекта, обеспечит возможность совместного использования данных без их передачи третьим лицам. Совершенствование репозитория данных способствует внедрению результатов генетических и иммунологических исследований в практическую медицину.

Значительный объем генетических данных, накопленных в России в результате проведения около 15 000 научно-исследовательских работ с 2010 по 2025 год, подчеркивает насущную необходимость создания и развития национальных генетических и иммунологических репозиториях. Эффективное использование этих данных крайне важно, поскольку анализ генетической и иммунологической информации позволяет глубже понять основы заболеваний, усовершенствовать оценку рисков, разработать новые методы диагностики, профилактики, терапии и создать инновационные лекарства. При этом для повышения международной конкурентоспособности и вклада российских репозиториях в глобальные исследования следует предусмотреть механизмы частичного открытого доступа к анонимизированным и агрегированным данным отечественных репозиториях, а также проработать возможности интеграции с международными базами данных.

Литература / References

1. Кочетова ОВ, Авзалетдинова ДШ, Кoryтина ГФ, Моругова ТВ, Бобоедова ОВ. Роль генов иммунного ответа в развитии сахарного диабета 2 типа. *Медицина*. 2022;4:1–9. Kochetova OV, Avzaletdinova DS, Korytina GF, Morugova TV, Boboedova OV. The Role of the Genes of Immune Response

²⁵ Федеральный закон от 3 декабря 2008 г. № 42-ФЗ «О государственной геномной регистрации в Российской Федерации» (с изменениями и дополнениями).

²⁶ Федеральный закон от 27 июля 2006 г. № 152-ФЗ «О персональных данных».

- in the Development of Type 2 Diabetes. *Medicine*. 2022;4:1–9 (In Russ.).
<https://doi.org/10.29234/2308-9113-2022-10-4-1-9>
2. Трошина ЕА, Юкина МЮ, Нуралиева НФ, Мокрышева НГ. Роль генов системы HLA: от аутоиммунных заболеваний до COVID-19. *Проблемы Эндокринологии*. 2020;66(4):9–15.
 Troshina EA, Yukina MYu, Nuralieva NF. The role of HLA genes: from autoimmune diseases to COVID-19. *Problems of Endocrinology*. 2020;66(4):9–15 (In Russ.).
<https://doi.org/10.14341/probl12470>
 3. Deutsch AJ, Udler MS. Phenotypic and genetic diversity in diabetes across populations. *The Journal of Clinical Endocrinology and Metabolism*. 2025;dgaf234.
<https://doi.org/10.1210/clinem/dgaf234>
 4. Fairley S, Lowy-Gallego E, Perry E, Flicek P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Research*. 2020;48(D1):D941–7.
<https://doi.org/10.1093/nar/gkz836>
 5. Gonzalez-Galarza FF, McCabe A, Melo Dos Santos EJ, Ghattaoraya G, Jones AR, Middleton D. Allele Frequency Net Database. *Methods in Molecular Biology*. 2024;2809:19–36.
https://doi.org/10.1007/978-1-0716-3874-3_2
 6. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledge-base and deposition resource. *Nucleic Acids Research*. 2023;51(D1):D977–85.
<https://doi.org/10.1093/nar/gkac1010>
 7. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. Author Correction: A genomic mutational constraint map using variation in 76,156 human genomes. *Nature*. 2024;625(7993):92–100.
<https://doi.org/10.1038/s41586-023-06045-0>
 8. Clough E, Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Research*. 2024;52(D1):D138–44.
<https://doi.org/10.1093/nar/gkad965>
 9. van den Akker EB, Trompet S, Wolf JJHB, Beekman M, Suchiman HED, Deelen J, et al. Metabolic Age Based on the BBMRI-NL 1H-NMR Metabolomics Repository as Biomarker of Age-related Disease. *Circulation: Genomic and Precision Medicine*. 2020;13(5):541–7.
<https://doi.org/10.1161/CIRCGEN.119.002610>
 10. Viñas-Giménez L, Padilla N, Battle-Masó L, Casals F, Riviere JG, Martínez-Gallo M, et al. FHLdb: A Comprehensive Database on the Molecular Basis of Familial Hemophagocytic Lymphohistiocytosis. *Frontiers in Immunology*. 2020;31(11):107.
<https://doi.org/10.3389/fimmu.2020.00107>
 11. Marquez S, Babrak L, Greiff V, Hoehn KB, Lees WD, Prak ETL, et al. AIRR Community. Adaptive Immune Receptor Repertoire (AIRR) Community Guide to Repertoire Analysis. *Methods in Molecular Biology*. 2022;2453:297–316.
https://doi.org/10.1007/978-1-0716-2115-8_17
 12. Fernández-Orth D, Rueda M, Singh B, Moldes M, Jene A, Ferri M, et al. A quality control portal for sequencing data deposited at the European genome-phenome archive. *Briefings in Bioinformatics*. 2022;23(3):bbac136.
<https://doi.org/10.1093/bib/bbac136>
 13. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini Sh, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Research*. 2019;47(D1):D339–43.
<https://doi.org/10.1093/nar/gky1006>
 14. Rasmussen SA, Hamosh A, Amberger J, Arnold C, Bocchini C, O'Neill MJF, et al. What's in a name? Issues to consider when naming Mendelian disorders. *Genetics in Medicine*. 2020;22(10):1573–5.
<https://doi.org/10.1038/s41436-020-0851-0>
 15. Thakur M, Brooksbank C, Finn RD, Firth HV, Foreman J, Freeberg M, et al. EMBL's European Bioinformatics Institute (EMBL-EBI) in 2024. *Nucleic Acids Research*. 2024;gkae1089.
<https://doi.org/10.1093/nar/gkae1089>
 16. Wong KM, Langlais K, Tobias GS, Fletcher-Hoppe C, Krasnewich D, Leeds HS, et al. The dbGaP data browser: a new tool for browsing dbGaP controlled-access genomic data. *Nucleic Acids Research*. 2017;45(D1):D819–26.
<https://doi.org/10.1093/nar/gkw1139>
 17. Каримов ВХ, Казанцев ДА. Потенциальные угрозы использования генетических технологий и правовые пути их разрешения. *Вопросы Безопасности*. 2022;1:48–63.
 Karimov VKh, Kazantsev DA. Potential threats of using genetic technologies and legal ways to resolve them. *Security Issues*. 2022;1:48–63 (In Russ.).
<https://doi.org/10.25136/2049-7543.2022.1.36744>
 18. Brandt DYC, Del Brutto OH, Nielsen R. Signatures of natural selection may indicate a genetic basis for the beneficial effects of oily fish intake in indigenous people from coastal Ecuador. *G3 (Bethesda)*. 2025;15(4):jkaf014.
<https://doi.org/10.1093/g3journal/jkaf014>
 19. Arnaiz-Villena A, Lledo T, Silvera-Redondo C, Juarez I, Vaquero-Yuste Ch, Martin-Villa JM, et al. The Origin of Amerindians: A Case Study of Secluded Colombian Chimila, Wiwa, and Wayúu Ethnic Groups and Their Trans-Pacific Gene Flow. *Genes (Basel)*. 2025;16(3):286.
<https://doi.org/10.3390/genes16030286>
 20. Barbitoff YA, Khmelkova DN, Pomerantseva EA, Slepchenkov AV, Zubashenko NA, Mironova IV, et al. Expanding the Russian allele frequency reference via cross-laboratory data integration: insights from 7452 exome samples. *National Science Review*. 2024;11(10):nwae326.
<https://doi.org/10.1093/nsr/nwae326>
 21. Poterba T, Vittal C, King D, Goldstein D, Goldstein JI, Schultz P, et al. The scalable variant call representation: enabling genetic analysis beyond one million genomes. *Bioinformatics*. 2024;41(1):btae746.
<https://doi.org/10.1093/bioinformatics/btae746>
 22. Patsakis M, Provatas K, Baltoumas FA, Chantzi N, Mouratidis I, Pavlopoulos GA, et al. MAFin: motif detection in multiple alignment files. *Bioinformatics*. 2025;41(4):btaf125.
<https://doi.org/10.1093/bioinformatics/btaf125>
 23. Пржиленский ВИ. Правовое и этическое регулирование генетических исследований. *Вестник Российского Университета Дружбы Народов. Серия: Юридические Науки*. 2021;25(1):214–31.
 Przhilenskiy VI. Legal and ethical regulation of genetic research. *RUDN Journal of Law*. 2021;25(1):214–31 (In Russ.).
<https://doi.org/10.22363/2313-2337-2021-25-1-214-231>
 24. Ахтямова ЕВ, Алсынбаева ЭМ, Масалимова АА. Проблемы правового регулирования защиты генетической информации, полученной методами преимплантационной и пренатальной генетической диагностики, в Российской Федерации. *Правовое Государство: Теория и Практика*. 2022;3(69):20–6.
 Akhtyamova EV, Alsynbayeva EM, Masalimova AA. Problems of legal regulation of the protection of genetic information obtained by preimplantation and prenatal genetic diagnostics in the Russian Federation. *The Rule of Law: Theory and Practice*. 2022;3(69):20–6 (In Russ.).
<https://doi.org/10.33184/pravgos-2022.3.3>
 25. Хусаинова РИ, Миннихметов ИР, Ялаев БИ, Ахтямова ЕВ, Алсынбаева ЭМ. Юридические проблемы защиты прав человека в Российской Федерации при использовании молекулярно-генетических технологий в медицине. *Гены и клетки*. 2021;3:97–103.
 Khusainova RI, Minniakhmetov IR, Yalaev BI, Akhtyamova EV, Alsynbayeva EM. Legal problems in the protection of human rights in the Russian Federation in the use of molecular genetic

- technologies in medicine. *Genes & Cells*. 2021;3:97–103 (In Russ.).
<https://doi.org/10.23868/202110014>
26. Рассолов ИМ, Чубукова СГ. Защита генетических данных при генетическом тестировании и генотерапевтическом лечении: информационно-правовые аспекты. *Актуальные Проблемы Российского Права*. 2020;15(5):65–72.
 Rassolov IM, Chubukova SG. Protection of Genetic Data in Genetic Testing and Gene-Therapy Treatment: IT Law Aspects. *Actual Problems of Russian Law*. 2020;15(5):65–72 (In Russ.).
<https://doi.org/10.17803/1994-1471.2020.114.5.065-072>
 27. Трикоз ЕН, Мустафина-Бредихина ДМ, Гуляева ЕЕ. Правовое регулирование процедуры генного редактирования: опыт США и стран ЕС. *Вестник Российского Университета Дружбы Народов. Серия: Юридические науки*. 2021;25(1):67–86.
 Trikoz EN, Mustafina-Bredikhina DM, Gulyaeva EE. Legal regulation of gene editing procedure: USA and EU experience. *RUDN Journal of Law*. 2021;25(1):67–86 (In Russ.).
<https://doi.org/10.22363/2313-2337-2021-25-1-67-86>
 28. Новоселова ЛА, Кольздорф МА. Генетическая информация как объект интеллектуальных прав. *Вестник Пермского университета. Юридические науки*. 2020;48:290–321.
 Novoselova LA, Kolzendorf MA. Genetic Information as Intellectual Property. *Perm University Herald. Juridical Sciences*. 2020;48:290–321 (In Russ.).
<https://doi.org/10.17072/1995-4190-2020-48-290-321>

Вклад авторов. Все авторы подтверждают соответствие своего авторства критериям ICMJE. Наибольший вклад распределен следующим образом: Титова А.Г. — концептуализация, создание черновика рукописи; Трусов Г.А. — разработка методологии исследования; Баёв А.В. — проведение исследования; Сосин Д.В. — редактирование рукописи; Нечаев Д.Н. — администрирование данных; Ломов А.Н. — предоставление ресурсов; Макаров В.В. — утверждение окончательного варианта рукописи; Юдин В.С. — руководство и координирование исследованием; Юдин С.М. — администрирование проекта.

ОБ АВТОРАХ

Титова Анастасия Германовна

<https://orcid.org/0000-0003-1822-9842>
 ATitova@cspfmmba.ru

Трусов Георгий Александрович

<https://orcid.org/0000-0002-8922-6342>
 GTrusov@cspfmmba.ru

Баёв Артур Владимирович

<https://orcid.org/0000-0002-9046-1737>
 ABayov@cspfmmba.ru

Сосин Дмитрий Витальевич, канд. биол. наук

<https://orcid.org/0000-0003-1471-8905>
 DSosin@cspfmmba.ru

Нечаев Денис Николаевич

<https://orcid.org/0000-0003-0071-4866>
 DNechaev@cspfmmba.ru

Ломов Алексей Николаевич

<https://orcid.org/0000-0003-2036-3514>
 ALomov@cspfmmba.ru

Макаров Валентин Владимирович, канд. биол. наук

<https://orcid.org/0000-0001-9495-0266>
 Makarov@cspfmmba.ru

Юдин Владимир Сергеевич, канд. биол. наук

<https://orcid.org/0000-0002-9199-6258>
 VYudin@cspfmmba.ru

Юдин Сергей Михайлович, д-р мед. наук, профессор

<https://orcid.org/0000-0002-7942-8004>
 info@cspfmmba.ru