

doi 10.18699/vjgb-24-107

# База данных о генах и белках, ассоциированных с нарушениями метаболизма глюкозы (GlucoGenes®): описание и возможности применения в биоинформатических исследованиях

В.В. Климонтов , К.С. Шишин , Р.А. Иванов , М.П. Пономаренко , К.А. Золотарева , С.А. Лашин <sup>1,2</sup>

<sup>1</sup> Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал Федерального исследовательского центра Института цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

<sup>2</sup> Федеральный исследовательский центр Института цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

 klimontov@mail.ru

**Аннотация.** Данные в области генетики и молекулярной биологии сахарного диабета стремительно накапливаются. Это ставит задачу создания исследовательских инструментов для быстрого поиска, структурирования и анализа информации в этой области. Мы разработали базу данных о генах и белках человека, ассоциированных с высоким уровнем глюкозы (гипергликемией), низким уровнем глюкозы (гипогликемией) и обоими нарушениями. Сведения были собраны с помощью текст-майнинга научных публикаций, проиндексированных в PubMed и PubMed Central, и анализа генных сетей гипергликемии, гипогликемии и вариабельности гликемии, выполненного с помощью биоинформатической системы ANDSystems. Созданный ресурс (GlucoGenes®) доступен по адресу: <https://glucogenes.sysbio.ru/genes/main>. Ресурс предоставляет информацию о генах и белках, связанных с риском развития гипергликемии и гипогликемии; регуляторных молекулах с гипергликемической и антигипергликемической активностью; генах, экспрессия которых повышается при высоком и/или низком уровне глюкозы; генах, экспрессия которых снижается при высоком и/или низком уровне глюкозы, а также о молекулах, связанных с нарушениями метаболизма глюкозы иным образом. На основе ресурса проведен эволюционный анализ генов, ассоциированных с нарушениями метаболизма глюкозы. Результаты анализа выявили значительное увеличение (до 40 %) доли генов, имеющих филостратиграфический индекс (phylostratigraphy age index, PAI), соответствующий времени происхождения многоклеточных организмов. Анализ консервативности последовательностей белков по индексу дивергенции (divergency index, DI) показал, что большинство соответствующих генов высококонсервативны (DI < 0.6) или консервативны (DI < 1). При анализе однонуклеотидного полиморфизма (SNP) в проксимальных районах промоторов, влияющих на сродство TATA-связывающего белка, в базе данных GlucoGenes® найден 181 SNP-маркер, который может снижать (45 SNP-маркеров) или повышать (136 SNP-маркеров) экспрессию 52 генов. Мы полагаем, что разработанный ресурс станет полезным инструментом для дальнейших исследований в области молекулярной биологии диабета.

**Ключевые слова:** ген; белок; сахарный диабет; гипергликемия; гипогликемия; вариабельность глюкозы; база данных; филостратиграфический индекс; однонуклеотидный полиморфизм.

**Для цитирования:** Климонтов В.В., Шишин К.С., Иванов Р.А., Пономаренко М.П., Золотарева К.А., Лашин С.А. База данных о генах и белках, ассоциированных с нарушениями метаболизма глюкозы (GlucoGenes®): описание и возможности применения в биоинформатических исследованиях. *Вавиловский журнал генетики и селекции*. 2024;28(8):1008-1017. doi 10.18699/vjgb-24-107

**Финансирование.** Создание базы данных и веб-ресурса GlucoGenes® выполнено за счет гранта Российского научно-го фонда (№ 20-15-00057-П).

**Благодарности.** Фрагменты по эволюционному анализу генов и анализу SNP реализованы за счет государственного задания ИЦиГ СО РАН (бюджетный проект № FWNR-2022-0006). Авторы выражают искреннюю признательность О.В. Сайк (НИИКЭЛ – филиал ИЦиГ СО РАН) за огромный вклад в сбор данных и ценные советы по формированию базы данных. Мы также благодарим А.М. Мухина (ИЦиГ СО РАН) за техническую помощь в создании веб-ресурса.

## GlucoGenes®, a database of genes and proteins associated with glucose metabolism disorders, its description and applications in bioinformatics research

V.V. Klimontov , K.S. Shishin , R.A. Ivanov , M.P. Ponomarenko , K.A. Zolotareva , S.A. Lashin <sup>1,2</sup>

<sup>1</sup> Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

 klimontov@mail.ru

**Abstract.** Data on the genetics and molecular biology of diabetes are accumulating rapidly. This poses the challenge of creating research tools for a rapid search for, structuring and analysis of information in this field. We have developed a web resource, GlucoGenes®, which includes a database and an Internet portal of genes and proteins associated with high glucose (hyperglycemia), low glucose (hypoglycemia), and both metabolic disorders. The data were collected using text mining of the publications indexed in PubMed and PubMed Central and analysis of gene networks associated with hyperglycemia, hypoglycemia and glucose variability performed with ANDSystems, a bioinformatics tool. GlucoGenes® is freely available at: <https://glucogenes.sysbio.ru/genes/main>. GlucoGenes® enables users to access and download information about genes and proteins associated with the risk of hyperglycemia and hypoglycemia, molecular regulators with hyperglycemic and antihyperglycemic activity, genes up-regulated by high glucose and/or low glucose, genes down-regulated by high glucose and/or low glucose, and molecules otherwise associated with the glucose metabolism disorders. With GlucoGenes®, an evolutionary analysis of genes associated with glucose metabolism disorders was performed. The results of the analysis revealed a significant increase (up to 40 %) in the proportion of genes with Phylostratigraphic Age Index (PAI) values corresponding to the time of origin of multicellular organisms. Analysis of sequence conservation using the Divergence Index (DI) showed that most of the corresponding genes are highly conserved (DI < 0.6) or conservative (DI < 1). When analyzing single nucleotide polymorphism (SNP) in the proximal regions of promoters affecting the affinity of the TATA-binding protein, 181 SNP markers were found in the GlucoGenes® database, which can reduce (45 SNP markers) or increase (136 SNP markers) the expression of 52 genes. We believe that this resource will be a useful tool for further research in the field of molecular biology of diabetes.

**Key words:** gene; protein; diabetes mellitus; hyperglycemia; hypoglycemia; glucose variability; database; phylostratigraphic index; single nucleotide polymorphism.

**For citation:** Klimontov V.V., Shishin K.S., Ivanov R.A., Ponomarenko M.P., Zolotareva K.A., Lashin S.A. GlucoGenes®, a database of genes and proteins associated with glucose metabolism disorders, its description and applications in bioinformatics research. *Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding*. 2024;28(8): 1008-1017. doi 10.18699/vjgb-24-107

## Введение

Сахарный диабет (СД) – одно из самых распространенных и социально значимых заболеваний человека. Согласно оценкам экспертов Международной федерации диабета, в 2021 г. число людей, живущих с СД, в мире достигло 537 млн, к 2045 г. ожидается увеличение этого показателя до 783 млн. Кроме того, более 540 млн человек имеют нарушенную толерантность к глюкозе (International Diabetes Federation, 2021).

В последние годы достигнут значительный прогресс в понимании молекулярных механизмов развития СД и его осложнений. Полногеномные исследования выявили ряд новых генетических локусов, модулирующих риск развития СД и диабетических осложнений в европейских и азиатских популяциях. Исследования в области протеомики, метаболомики и мультиомики пролили свет на молекулярные основы патогенеза заболевания (Langenberg, Lotta, 2018; Lyssenko, Vaag, 2023; Shojima et al., 2023).

В то же время обнаружено влияние воздействия высокой концентрации глюкозы на регуляцию экспрессии генов в различных тканях (Vaulont et al., 2000; Hall et al., 2018; Vega et al., 2020; Zhang S. et al., 2021). Установлено, что влияние высокого уровня глюкозы на экспрессию генов может быть пролонгировано и усугублено эпигенетическими модификациями. Этот механизм считается важным для феномена метаболической памяти и развития диабетических осложнений (Dhawan et al., 2022). Аномально низкий уровень глюкозы также сопровождается рядом биохимических сдвигов. Эти сдвиги связаны прежде всего с реакцией сердечно-сосудистой и нервной систем на гипогликемию (Hanefeld et al., 2016; Rehni, Dave, 2018).

Все большее внимание привлекают молекулярные эффекты повторяющихся эпизодов высокого и низкого уровня глюкозы, которые характерны для феномена высокой

вариабельности гликемии (ВГ). Установлено, что повышенная ВГ увеличивает риск развития микро- и макрососудистых осложнений СД и ассоциирована с повышенной смертностью больных от всех причин и от сердечно-сосудистых заболеваний (Ceriello et al., 2019; Wilmot et al., 2019). На молекулярном уровне патофизиологические изменения, ассоциированные с высокой ВГ, реализуются через повышение или понижение экспрессии большого количества генов и изменение активности сигнальных путей, таких как PI3K/Akt, NF-κB, MAPK (ERK), JNK и TGF-β/Smad (Klimontov et al., 2021).

С учетом огромного количества и разнообразия молекулярных изменений при СД для формирования всестороннего и целостного представления о патогенезе заболевания необходимы передовые аналитические инструменты. Искусственный интеллект, биоинформатика и интегративная системная биология предоставляют новые возможности для изучения сложных заболеваний, таких как СД (Nielsen, 2017; Климонтов и др., 2021; Orlov et al., 2021; Putra et al., 2024).

Перспективным подходом в этой области является анализ генных сетей, т. е. групп координированно функционирующих генов, взаимодействующих друг с другом и определяющих тот или иной фенотипический признак организма (Колчанов и др., 2013). Ранее, используя текстмайнинг научных публикаций, индексируемых в базах данных Pubmed и Pubmed Central (PMC), а также биоинформационный анализ, мы реконструировали генные сети, ассоциированные с нарушениями метаболизма глюкозы (НМГ): гипергликемией, гипогликемией и ВГ, и собрали большое количество материалов о молекулах и белках, так или иначе связанных с этими метаболическими нарушениями (Saik, Klimontov, 2020–2022).

На основе полученных результатов была создана база данных, содержащая информацию о генах и белках, свя-

занных с НМГ: гипергликемией, гипогликемией и обоими нарушениями. В этой статье мы представляем описание возможностей базы данных, а также приводим результаты двух биоинформатических исследований, проведенных с ее использованием. В первом исследовании был выполнен эволюционный анализ генов НМГ, во втором – анализ однонуклеотидного полиморфизма (SNP) проксимальных районов длиной 90 п. о. промоторов генов человека, связанных с нарушениями метаболизма глюкозы.

## Материалы и методы

**Формирование веб-ресурса.** Материал для ресурса GlucoGenes® был получен ранее в ходе реконструкции и анализа генных сетей гипергликемии, гипогликемии и ВГ, проведенных с использованием биоинформатической системы ANDSysTEM (разработана в ИЦиГ СО РАН; доступ: <https://anddigest.sysbio.ru/index.php>). ANDSysTEM строит ассоциативные генные сети на основе автоматического анализа текстов (текст-майнинга) научных публикаций, проиндексированных в базах данных PubMed и PMC (Ivanisenko et al., 2015, 2019). Детали анализа полученных с помощью ANDSysTEM генных сетей НМГ описаны ранее (Saik, Klimontov, 2020–2022).

Для программной реализации базы данных были выбраны реляционная модель данных и система управления базой данных PostgreSQL (<https://www.postgresql.org/>).

При проектировании веб-ресурса была выбрана клиент-серверная архитектура, состоящая из трех основных компонентов: клиент, сервер и база данных. Для разработки использованы фреймворки Vue.js и Flask; управление доступом к базе данных осуществляется через программный доступ на основе технологии REST.

**Филостратиграфический анализ и анализ дивергенции генов, ассоциированных с НМГ.** Филостратиграфический анализ – это метод, направленный на определение эволюционного происхождения генов путем анализа присутствия их ортологов – генов, кодирующих гомологичные белки, разошедшиеся в результате видообразования в геномах различных видов. Такой подход выявляет ключевые моменты в эволюции генома, когда наблюдался резкий рост числа новых генов, а также помогает идентифицировать гены, уникальные для отдельных таксонов (Domazet-Lošo, Tautz, 2010).

Нами проведен эволюционный анализ генов, включенных в базу данных GlucoGenes®, с применением филостратиграфического возрастного индекса (phylostratigraphy age index, PAI) и индекса дивергенции (divergency index, DI). Значение PAI показывает, насколько таксон, отражающий возраст гена, удален от корня филогенетического дерева (Мустафин и др., 2021). Под таксоном, отражающим возраст гена, понимается таксон, в котором произошло расхождение изучаемого вида с наиболее удаленным родственным таксоном, в котором был найден ортолог рассматриваемого гена. Чем больше значение PAI исследуемого гена, тем моложе этот ген. Для филостратиграфического анализа использовали программный комплекс OrthoWeb (<https://orthoweb.sysbio.cytogen.ru/run.html>). При расчете PAI применяли метод расчета по группам ортологии KEGG (Kanehisa et al., 2016).

Индекс DI – это показатель эволюционной изменчивости гена. Он рассчитывается на основе соотношения  $dN/dS$ , где  $dN$  – доля несинонимичных замен в последовательностях ДНК исследуемого гена и его ортолога;  $dS$  – доля синонимичных замен. Значение этого показателя было рассчитано на основе сравнения генов человека с генами близкородственных организмов из семейства Hominidae, а именно ортологов, найденных у западной гориллы *Gorilla gorilla gorilla*, суматранского orangutanга *Pongo abelii* и шимпанзе обыкновенного *Pan troglodytes*. Для расчета значений  $dN/dS$  применяли модель LPB93 (Yang, 2007). Значение DI в диапазоне от 0 до 1 показывает, что ген подвергается стабилизирующему отбору, 1 – нейтральной эволюции, а более 1 – движущему отбору.

**Анализ SNP в проксимальных районах длиной 90 п. о. промоторов генов человека, связанных с НМГ.** В базе знаний Human\_SNP\_TATAdb (Филонов и др., 2023) осуществили поиск вариантов SNP в проксимальных районах длиной 90 п. о. промоторов генов человека, связанных с НМГ, которые способны статистически значимо уменьшать или увеличивать сродство TATA-связывающего белка (ТВП) к этим промоторам и, соответственно, уровни экспрессии генов. Для дальнейшей работы оставили SNP, для которых описаны клинические проявления в свободно доступной базе данных ClinVar (Landrum et al., 2014). Наконец, с использованием базы данных PubMed для всех отобранных таким путем клинических SNP-маркеров заболеваний в промоторах генов человека, ассоциированных с нарушениями метаболизма глюкозы, осуществили курируемую аннотацию влияния полиморфного изменения экспрессии этих генов на уровни глюкозы у пациентов-носителей соответствующих SNP.

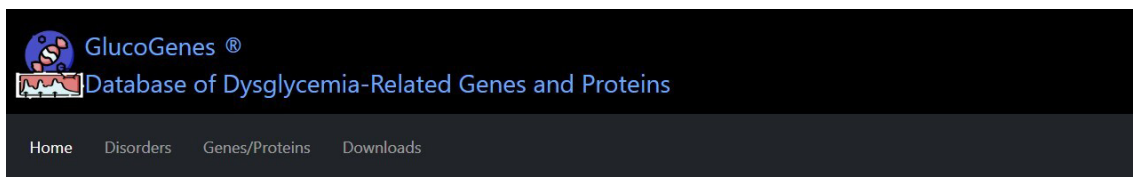
## Результаты

### Ресурс GlucoGenes®

Ресурс GlucoGenes® находится в свободном доступе по адресу: <https://glucogenes.sysbio.ru/genes>. Интерфейс ресурса показан на рис. 1.

База данных GlucoGenes® включает шесть компонентов (таблиц). Таблица Genes содержит названия генов, их описание и идентификаторы NCBI. Таблица Proteins включает названия белков, их описание и идентификаторы в базе данных UniProt, а также ссылки на соответствующие гены в базе данных GlucoGenes®. Таблица Glycemia\_related\_conditions представляет информацию о НМГ (гипергликемия, гипогликемия). Таблица Types\_of\_glycemia\_gene\_association содержит информацию о типах ассоциаций между молекулами и НМГ. Таблица References включает идентификаторы статей PubMed или PubMed Central с кратким извлечением данных. Таблица Glycemia\_gene\_association является сводной, аккумулирует информацию из всех вышеперечисленных таблиц. Структура базы данных представлена на рис. 2. В настоящее время база данных включает 561 ген, связанный с НМГ, и 2115 ссылок на литературные источники.

Веб-ресурс GlucoGenes® включает четыре функциональных раздела.



Gene	Association with Low Glucose	Reference
<a href="#">ABCC8</a>	SNPs associated with the risk of Low Glucose	1. <a href="#">PMCID: 1291378</a> ; 2. <a href="#">PMID: 20042013</a> ; 3. <a href="#">PMID: 21142918</a> ;
<a href="#">ACE</a>	SNPs associated with the risk of Low Glucose	1. <a href="#">PMID: 16981144</a> ; 2. <a href="#">PMCID: 3814531</a> ; 3. <a href="#">PMID: 17333108</a> ; 4. <a href="#">PMID: 20546161</a> ; 5. <a href="#">PMID: 12547848</a> ; 6. <a href="#">PMID: 12637987</a> ; 7. <a href="#">PMID: 18404608</a> ;
<a href="#">ADCYAP1</a>	Other relations with Low Glucose	1. <a href="#">PMCID: 4909203</a> ;

Рис. 1. Фрагмент интерфейса интернет-портала GlucoGenes®.

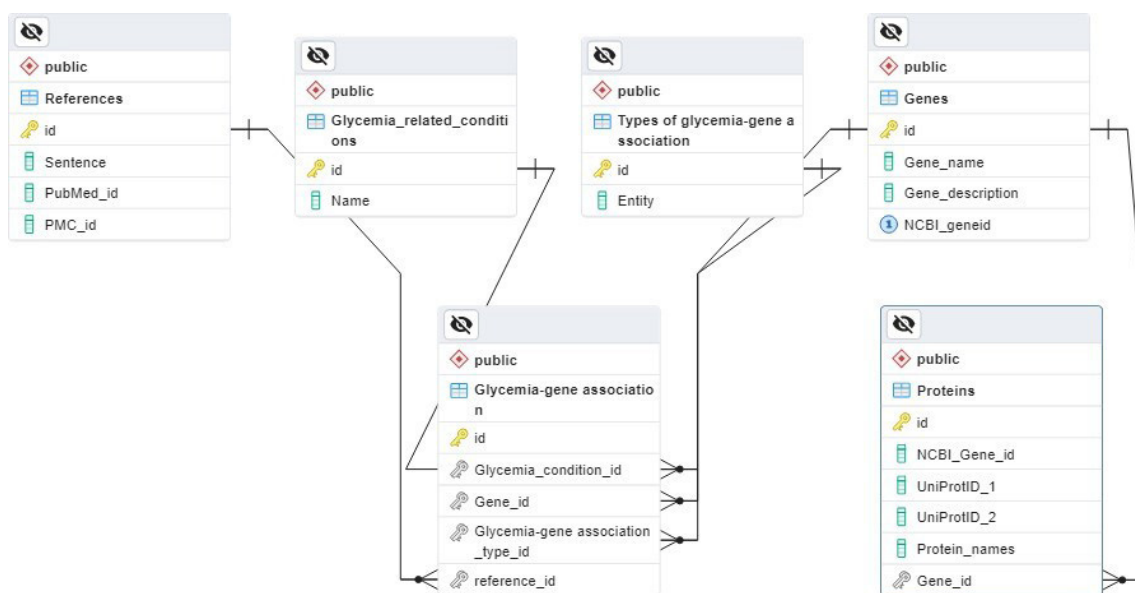


Рис. 2. Диаграмма «сущность-связь» базы данных GlucoGenes®.

1. На главной странице (Номере) представлена общая информация о ресурсе и используемых терминах. Мы обозначили гипергликемию, или высокий уровень глюкозы в культуральной среде, как High Glucose (HG), а гипогликемию, или низкий уровень глюкозы в культуральной среде, – как Low Glucose (LG).
2. Страница НМГ (Disorders). Здесь представлены списки генов, ассоциированных с HG, LG, а также с высоким и низким уровнем глюкозы (HLG). Для каждого гена указан тип ассоциации с НМГ. Среди ассоциаций с НМГ мы выделили следующие категории: однонуклеотидные полиморфизмы (SNP), ассоциированные с риском HG, LG или HLG; белки с гипергликемической активностью; белки с антигипергликемическим эффектом; гены, экспрессия которых повышается в условиях HG; гены, экспрессия которых повышается в условиях LG; гены, экспрессия которых понижается в условиях HG; гены, экспрессия которых понижается в условиях LG;

другие ассоциации с HG, LG или HLG. Для каждого гена и ассоциации приведены ссылки на соответствующие публикации в PubMed.

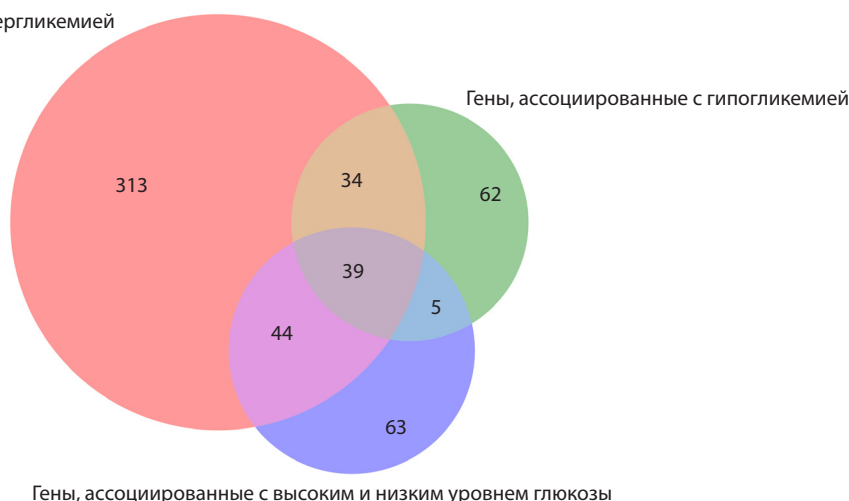
3. Каталог генов/белков, связанных с НМГ (Genes/Proteins). Здесь можно найти название гена и NCBI-идентификатор гена, название(я) белка(ов), кодируемого(ых) геном, а также типы ассоциации с НМГ.
4. Страница загрузки (Downloads). С этой страницы можно скачать списки генов, ассоциированных с HG, LG и HLG, а также все ассоциированные гены в формате Excel. Поиск в системе может осуществляться по имени гена, идентификатору гена NCBI или типу НМГ.

Доступ к данным портала возможен и без использования графического интерфейса пользователя, через программный интерфейс (API) REST. Данный интерфейс позволяет получить необходимую информацию при отправке запроса на веб-сервер в виде строки URL. В ответ на такой запрос сервер возвращает результат в виде тек-

**Таблица 1.** Списки белок-кодирующих генов человека, для которых был проведен филогенетический анализ

Группа генов	Описание	Число генов
Все гены в геноме <i>Homo sapiens</i>	Все белок-кодирующие гены человека, для которых были посчитаны значения PAI и DI	19 566
Гены, ассоциированные только с гипергликемией	Список генов из GlucoGenes, ассоциированных с высоким уровнем глюкозы	430
Гены, ассоциированные только с гипогликемией	Список генов из GlucoGenes, ассоциированных с низким уровнем глюкозы	140
Гены, ассоциированные с высоким и низким уровнем глюкозы	Список генов из GlucoGenes, ассоциированных с высоким и низким уровнем глюкозы	151

Гены, ассоциированные с гипергликемией



**Рис. 3.** Диаграмма Венна пересечений групп генов выборки.

стовой страницы или файла, информация в котором структурирована в соответствии с форматом JavaScript Object Notation (JSON) (<http://json.org/>). Полученный текстовый файл можно открыть любым текстовым редактором. Файл также может быть обработан с помощью различных программных средств, в том числе программ, написанных пользователем в универсальных средах моделирования (например, Matlab, Scilab и др.) или на языках программирования высокого уровня (например, Python, R, C++, Java и др.). Пример REST-запроса приведен ниже (результатом является текстовый файл в структурированном формате JSON):

<https://glucogenes.sysbio.ru/api/genes/<geneid>> – возвращает карточку с описанием гена <geneid>.

**Эволюционные характеристики генов, связанных с НМГ.** Мы рассчитали индексы PAI и построили их распределение как для списка белок-кодирующих генов в геноме *Homo sapiens*, так и для генов, представленных в базе данных GlucoGenes®: для генов, ассоциированных с гипергликемией, гипогликемией и колебаниями уровня глюкозы (табл. 1). Примечательно, что часть генов имела ассоциацию более чем с одной патологией (рис. 3).

Распределение значений PAI для всех генов в геноме человека является мультимодальным и имеет два ярко выраженных пика на уровне Cellular Organisms: Metazoa и Vertebrata-Euteleostomi (рис. 4). Первый пик – самый

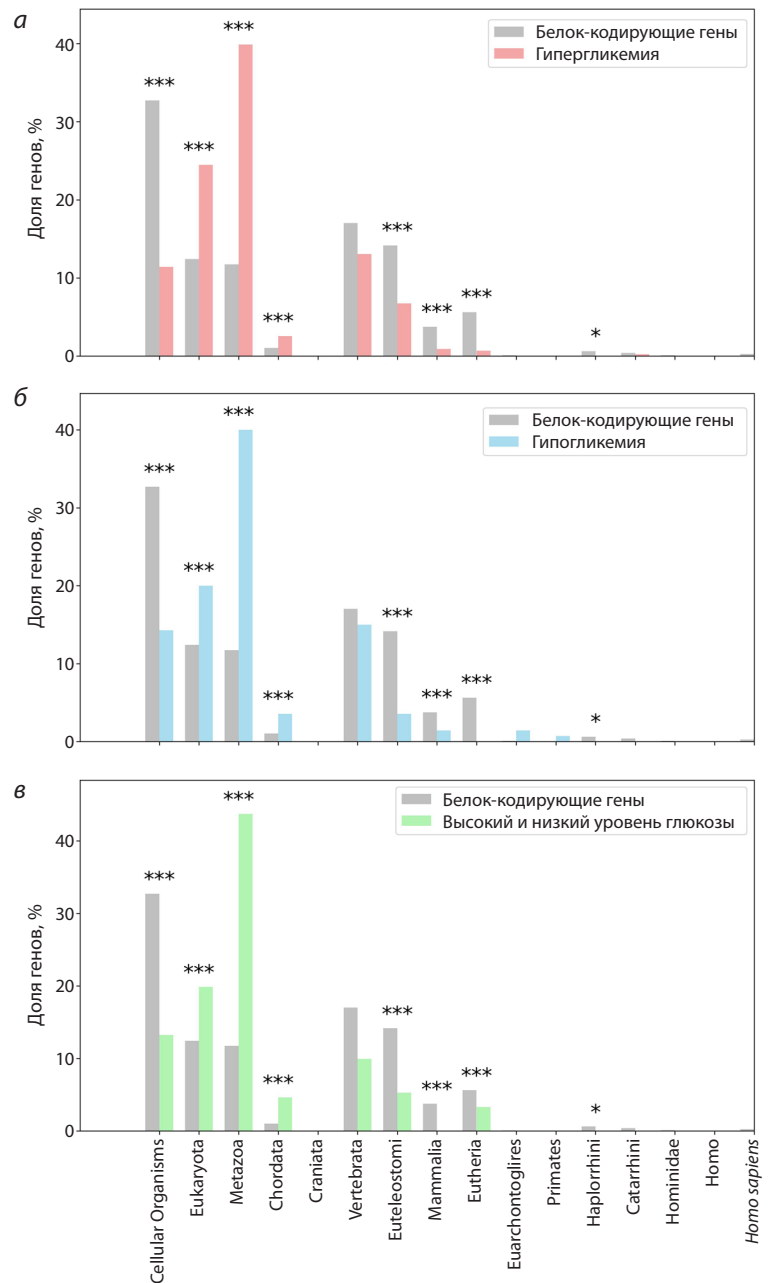
большой; почти 55 % генов в геноме *H. sapiens* имеют PAI от 1 до 3. Второй пик охватывает 32 % генов.

Результаты анализа выявили значительное увеличение доли генов (до 40 %), связанных с регуляцией уровня глюкозы, имеющих индекс PAI = 3, во всех трех категориях (рис. 4). В частности, к этой группе генов относятся *TCF7L2*, *PPARG*, *GCGR*, *IRS1* и *MTNR1B*, продукты которых – важные регуляторы метаболизма глюкозы.

Анализ консервативности последовательностей для тех же списков генов (рис. 5) показал, что большинство из изученных генов высококонсервативны (DI < 0.6) или консервативны (DI < 1). Это свидетельствует о сохранении их функций в процессе эволюции и подчеркивает их критическую роль в биологических процессах, связанных с регуляцией уровня глюкозы. Однако был обнаружен ряд генов с DI выше 1, что указывает на недавнее воздействие на них движущего отбора. К таким генам относятся *SPPI*, *CALCA*, *CD33*, *SULT2A1*, *TNF*, *ECM1*, *CYP3A4* и *EDN1*.

**Анализ SNP в проксимальных районах длиной 90 п. о. промоторов генов человека, связанных с НМГ.** Всего в базе данных GlucoGenes® нами найден 181 SNP-маркер, который может снижать (45 SNP-маркеров) или повышать (136 SNP-маркеров) экспрессию 52 генов человека, изменяя таким путем уровни глюкозы у пациентов с минорными аллелями этих SNP. В табл. 2 в качестве при-

PAI	Таксон
1	Cellular Organisms
2	Eukaryota
3	Metazoa
4	Chordata
5	Craniata
6	Vertebrata
7	Euteleostomi
8	Mammalia
9	Eutheria
10	Euarchontoglires
11	Primates
12	Haplorrhini
13	Catarrhini
14	Hominidae
15	Homo
16	<i>Homo sapiens</i>



**Рис. 4.** Распределение белок-кодирующих генов, ассоциированных с НМГ, по значениям PAI.

Здесь и на рис. 5: *a* – все белок-кодирующие гены человека (All\_CDS) в качестве контрольного списка и гены, ассоциированные с высоким уровнем глюкозы (Hyperglycemia); *b* – все белок-кодирующие гены человека (All\_CDS) в качестве контрольного списка и гены, ассоциированные с низким уровнем глюкозы (Hypoglycemia); *v* – все белок-кодирующие гены человека (All\_CDS) в качестве контрольного списка и гены, ассоциированные с высоким и низким уровнем глюкозы (High, and Low Glucose-related genes). Звездочками отмечены столбцы со статистически достоверно различающимися значениями между выборками генов из базы данных и выборкой всех белок-кодирующих генов человека: \*  $p$ -value < 0.05, \*\*\*  $p$ -value < 0.001. Статистическая проверка проводилась при помощи критерия chi-square.

мера представлено 10 SNP в промоторах генов *ABCC8*, *INSR* и *PGM1* человека, доступных в базе данных ClinVar (Landrum et al., 2014), для которых в базе данных GlucoGenes® предлагается механизм ассоциаций с нарушениями метаболизма глюкозы.

Согласно данным, представленным в табл. 2, минорные аллели промоторов генов *ABCC8*, *INSR* и *PGM1* характеризуются измененным сродством к ТВР, что может влиять на уровень экспрессии этих генов и объяснять их связь с

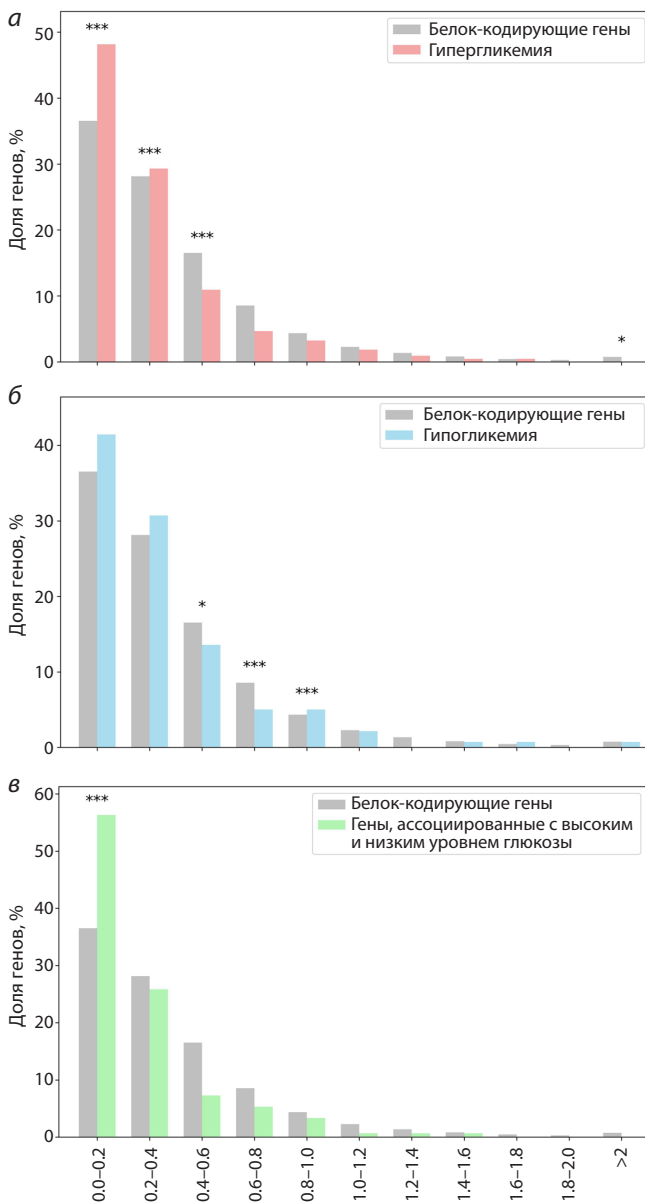
НМГ. Более полная информация о найденных SNP-маркерах приведена в Приложении<sup>1</sup>.

### Обсуждение

Достижения в изучении молекулярной биологии СД представляют широкие возможности для внедрения технологий высокоточной медицины в лечение заболевания.

<sup>1</sup> Приложение см. по адресу:

<https://vavilovj-icg.ru/download/pict-2024-28/appx36.pdf>



**Рис. 5.** Распределение белок-кодирующих генов, ассоциированных с НМГ, по значениям DI.

В частности, выявление специфических для заболевания биомаркеров дает новые перспективы в диагностике, мониторинге, прогнозировании заболевания и его исходов, в фармакогенетике современных сахароснижающих препаратов, а также в поиске новых терапевтических средств (Chung et al., 2020). Быстрое накопление данных о молекулярных основах генетической предрасположенности к СД и молекулярных механизмах его осложнений определяет потребность в исследовательских инструментах для структурированного поиска информации в этой области.

Мы создали базу данных генов и белков, для которых были продемонстрированы ассоциации с НМГ: гипергликемией, гипогликемией и обоими нарушениями. Разработанный веб-ресурс под названием GlucoGenes® (<https://glucogenes.sysbio.ru/genes/main>) может быть использован

для сбора, поиска и визуализации информации о генах и белках, так или иначе связанных с НМГ. Доступ к базе данных, интегрированной в GlucoGenes®, осуществляется через программный интерфейс на основе REST для просмотра записей. Графический пользовательский интерфейс позволяет просматривать записи и экспортировать их содержимое в формат Excel. База данных содержит каталоги генов и белков, ассоциированных с НМГ, с информацией о типах ассоциаций и ссылками на рефераты соответствующих публикаций в PubMed или полнотекстовые статьи в PMC. Списки генов и белков доступны для скачивания. Ограничением ресурса является аккумуляция данных только из источников, проиндексированных в Pubmed и PMC. Очевидна необходимость регулярного обновления информации.

Созданный ресурс может быть полезен для решения исследовательских задач в области биоинформатики и молекулярной биологии СД. В частности, он может применяться для отбора генов и белков с целью исследования генетической предрасположенности к СД в различных популяциях, в изучении молекулярных аспектов патогенеза СД, в поиске потенциальных биомаркеров осложнений заболевания, молекул-таргетов для терапии и иных задач. В этой статье мы приводим примеры использования разработанного ресурса для решения научных задач в биоинформатике.

Первое исследование посвящено эволюционному происхождению генов, ассоциированных с НМГ. Эволюционный анализ генов при помощи филогенетики является ключевым инструментом в биологии, позволяющим понять фундаментальные механизмы, лежащие в основе разнообразия жизни на Земле. Эволюционная история генов раскрывает информацию о том, как различные функции и структуры развивались и адаптировались к изменениям окружающей среды. Это знание не только помогает в реконструкции филогенетических деревьев, но и способствует идентификации генов, ответственных за адаптивные изменения и специфические физиологические процессы, такие как метаболизм глюкозы. Проведенный филогенетический анализ показал, что среди генов, ассоциированных с метаболизмом глюкозы, значительную долю (до 40 %) составляют гены с PAI = 3, что соответствует времени происхождения многоклеточных организмов (Maloof et al., 2010). Большинство изученных генов оказались высококонсервативными (DI < 0.6) или консервативными (DI < 1). Полученные результаты подчеркивают важность генов, ассоциированных с НМГ, в регуляции специализированных метаболических процессов, характерных для сложных организмов.

Во втором исследовании данные веб-ресурса были использованы для анализа SNP в проксимальных районах промоторов генов человека, влияющих на родство TATA-связывающего белка. Интеграция данных GlucoGenes® с информацией из других баз данных об ассоциациях SNP с рядом заболеваний человека, с одной стороны, и биоинформатических оценок изменения уровня глюкозы у пациентов с этими SNP, с другой стороны, отражает молекулярные механизмы возможного влияния НМГ на течение этих заболеваний.

**Таблица 2.** SNP, влияющие на сходство TBP, в промоторах генов *ABCC8*, *INSR* и *PGM1*, ассоциированных с НМГ

Интернет-портал GlucosGenes®	База данных ClinVar (Landrum et al., 2014)	База данных dbSNP (Day, 2010)	База данных Human_SNP_TATAdb (Филонов и др., 2023)	База данных PubMed				
Символ гена	Статус гликемии	Ассоциации с SNP	Значимость	Влияние на течение болезни (Ref)				
		dbSNP ID:min	$K_D$ , нМ/л, <i>in silico</i>					
		5'-фланг, 10 bp	WT	$z$				
		3'-фланг, 10 bp	$M_0 \pm SEM$	$p$				
		WT → min	$\Delta$	$\rho$				
		G → T						
<i>PGM1</i>	Low Glucose	rs199928376:T	46.52 ± 3.87	3.89	$10^{-3}$	B	↑	Избыток PGM1 – ответ на дефицит глюкозы при раке легких (Li et al., 2020)
		tgccscaac	gggtaaggga					
<i>INSR</i>	High and Low Glucose & High Glucose	rs32052290:T	62.14 ± 4.83	18.81	$10^{-6}$	A	↑	Избыток <i>Insr</i> улучшал здоровье мышей db/db (Sasaki et al., 2015)
		Insulin-resistant diabetes mellitus & acanthosis nigricans; Rabson-Mendenhall syndrome, leprechaunism syndrome	tcttgggtgg					
<i>ABCC8</i>	High and Low Glucose & Low Glucose & High Glucose	rs1409900082:A	74.04 ± 5.71	14.58	$10^{-6}$	A	↑	Влияние диеты матери на потомство у мышей (Zhang Q. et al., 2024): избыток <i>Abcc8</i> защищает потомство
		ABCC8-related disorder, hypoglycemia, hyperinsulinism, diabetes mellitus, monogenic diabetes	tggtcctcag					
		rs1409900082:T	74.04 ± 5.71	6.99	$10^{-6}$	A		
		rs195706366:A	53.2 ± 4.54	20.58	$10^{-6}$	A		
		rs1957150758:A	74.04 ± 5.71	8.06	$10^{-6}$	A		
		rs541244107:A	59.19 ± 4.91	3.04	$10^{-2}$	C		
		rs541244107:C	59.19 ± 4.91	3.83	$10^{-3}$	B		
		rs565662182:A	53.2 ± 4.54	8.65	$10^{-6}$	A		
		rs58241708:T	53.2 ± 4.54	2.68	$10^{-2}$	C		

Примечание. WT и min – анцестральный (норма) и минорный (патология) аллели SNP;  $K_D$  – равновесная константа диссоциации комплекса TBP-промотор, выраженная в наномолях на литр (нМ/л);  $M_0$  и SEM – контекстно-зависимая оценка *in silico* и ее стандартизованная ошибка;  $z$ ,  $p$  и  $\rho$  – значение z-статистики Фишера и уровень ее значимости, а также эвристическая приоритизация оценок *in silico* от наилучшей (A) до наихудшей (D) в алфавитном порядке;  $\Delta$  – экспрессия выше (↑) или ниже (↓).



## Заклучение

GlucoGenes® – это ресурс, сочетающий в себе графический интерфейс пользователя и базу данных генов и белков, связанных с гипергликемией, гипогликемией и обоими метаболическими нарушениями. Ресурс использован для биоинформатического анализа эволюционных характеристик генов, ассоциированных с указанными нарушениями, а также при анализе SNP в проксимальных районах промоторов генов, влияющих на средство ТАТА-связывающего белка. Показано, что значительная часть генов, ассоциированных с НМГ, являются эволюционно древними и консервативными. Идентифицированы SNP-маркеры, которые могут снижать (45 SNP-маркеров) или повышать (136 SNP-маркеров) экспрессию 52 генов.

## Список литературы / References

- Климонтон В.В., Бериков В.Б., Сайк О.В. Искусственный интеллект в диабетологии. *Сахар. диабет.* 2021;24(2):156-166. doi 10.14341/DM12665
- [Klimontov V.V., Berikov V.B., Saik O.V. Artificial intelligence in diabetology. *Sakharnyi Diabet = Diabetes Mellitus.* 2021;24(2): 156-166. doi 10.14341/DM12665 (in Russian)]
- Колчанов Н.А., Игнатъева Е.В., Подколodная О.А., Лихошвай В.А., Матушкин Ю.Г. Генные сети. *Вавиловский журнал генетики и селекции* 2013;17(4/2):833-850
- [Kolchanov N.A., Ignatieva E.V., Podkolodnaya O.A., Likhoshvai V.A., Matushkin Yu.G. Gene networks. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding.* 2013;17(4/2):833-850 (in Russian)]
- Мустафин З.С., Лашин С.А., Матушкин Ю.Г. Филостратиграфический анализ генных сетей заболеваний человека. *Вавиловский журнал генетики и селекции.* 2021;25(1):46-56. doi 10.18699/VJ21.006
- [Mustafin Z.S., Lashin S.A., Matushkin Yu.G. Phylostratigraphic analysis of gene networks of human diseases. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding.* 2021;25(1):46-56. doi 10.18699/VJ21.006]
- Филонов С.В., Подколodный Н.Л., Подколodная О.А., Твердохлеб Н.Н., Пономаренко П.М., Рассказов Д.А., Богомолов А.Г., Пономаренко М.П. Human\_SNP\_TATAdb – база данных о SNP, статистически достоверно изменяющих средство ТАТА-связывающего белка к промоторам генов человека: полногеномный анализ и варианты использования. *Вавиловский журнал генетики и селекции.* 2023;27(7):728-736. DOI 10.18699/VJGB-23-85
- [Filonov S.V., Podkolodnyy N.L., Podkolodnaya O.A., Tverdokhlebl N.N., Ponomarenko P.M., Rasskazov D.A., Bogomolov A.G., Ponomarenko M.P. Human\_SNP\_TATAdb: a database of SNPs that statistically significantly change the affinity of the TATA-binding protein to human gene promoters: genome-wide analysis and use cases. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding.* 2023;27(7):728-736. doi 10.18699/VJGB-23-85]
- Ceriello A., Monnier L., Owens D. Glycaemic variability in diabetes: clinical and therapeutic implications. *Lancet Diabetes Endocrinol.* 2019;7(3):221-230. doi 10.1016/S2213-8587(18)30136-0
- Chung W.K., Erion K., Florez J.C., Hattersley A.T., Hivert E.F., Lee C.G., McCarthy M.I., Nolan J.J., Norris J.M., Pearson E.R., Phillipson L., McElvaine A.T., Cefalu W.T., Rich S.S., Franks P.W. Precision medicine in diabetes: a Consensus Report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia.* 2020;63(9):1671-1693. doi 10.1007/s00125-020-05181-w
- Day I.N. dbSNP in the detail and copy number complexities. *Hum. Mutat.* 2010;31(1):2-4. doi 10.1002/humu.21149
- Dhawan P., Vasishta S., Balakrishnan A., Joshi M.B. Mechanistic insights into glucose induced vascular epigenetic reprogramming in type 2 diabetes. *Life Sci.* 2022;298:120490. doi 10.1016/j.lfs.2022.120490
- Domazet-Lošo T., Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature.* 2010; 468(7325):815-819. doi 10.1038/nature09632
- Hall E., Dekker Nitert M., Volkov P., Malmgren S., Mulder H., Bacos K., Ling C. The effects of high glucose exposure on global gene expression and DNA methylation in human pancreatic islets. *Mol. Cell. Endocrinol.* 2018;472:57-67. doi 10.1016/j.mce.2017.11.019
- Hanefeld M., Frier B.M., Pistrosch F. Hypoglycemia and cardiovascular risk: is there a major link? *Diabetes Care.* 2016;39(S.2):S205-S209. doi 10.2337/dcS15-3014
- International Diabetes Federation. IDF Diabetes Atlas, 10th ed. Brussels, 2021
- Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSySystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst. Biol.* 2015;9(S2):S2. doi 10.1186/1752-0509-9-S2-S2
- Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSySystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics.* 2019;20(1):34. doi 10.1186/s12859-018-2567-6
- Kanehisa M., Sato Y., Kawashima M., Furumichi M., Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44(D1):D457-D462. doi 10.1093/nar/gkv1070
- Klimontov V.V., Saik O.V., Korbut A.I. Glucose variability: How does it work? *Int. J. Mol. Sci.* 2021;22(15):7783. doi 10.3390/ijms22157783
- Landrum M.J., Lee J.M., Riley G.R., Jang W., Rubinstein W.S., Church D.M., Maglott D.R. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014;42:D980-D985. doi 10.1093/nar/gkt1113
- Langenberg C., Lotta L.A. Genomic insights into the causes of type 2 diabetes. *Lancet.* 2018;391(10138):2463-2474. doi 10.1016/S0140-6736(18)31132-2
- Li Y., Liang R., Sun M., Li Z., Sheng H., Wang J., Xu P., Liu S., Yang W., Lu B., Zhang S., Shan C. AMPK-dependent phosphorylation of HDAC8 triggers PGM1 expression to promote lung cancer cell survival under glucose starvation. *Cancer Lett.* 2020;478:82-92. doi 10.1016/j.canlet.2020.03.007
- Lyssenko V., Vaag A. Genetics of diabetes-associated microvascular complications. *Diabetologia.* 2023;66(9):1601-1613. doi 10.1007/s00125-023-05964-x
- Maloof A.C., Porter S.M., Moore J.L., Dudás F.Ö., Bowring S.A., Higgins J.A., Fike D.A., Eddy M.P. The earliest Cambrian record of animals and ocean geochemical change. *Geol. Soc. Am. Bull.* 2010; 122(11-12):1731-1774. doi 10.1130/B30346.1
- Nielsen J. Systems biology of metabolism: A driver for developing personalized and precision medicine. *Cell Metab.* 2017;25(3):572-579. doi 10.1016/j.cmet.2017.02.002
- Orlov Y.L., Anashkina A.A., Klimontov V.V., Baranova A.V. Medical genetics, genomics and bioinformatics aid in understanding molecular mechanisms of human diseases. *Int. J. Mol. Sci.* 2021;22(18): 9962. doi 10.3390/ijms22189962
- Putra S.E.D., Martriano Humardani F., Antonius Y., Jonathan J., Thalia Mulyanata L. Epigenetics of Diabetes: A bioinformatic approach. *Clin. Chim. Acta.* 2024;557:117856. doi 10.1016/j.cca.2024.117856
- Rehni A.K., Dave K.R. Impact of hypoglycemia on brain metabolism during diabetes. *Mol. Neurobiol.* 2018;55(12):9075-9088. doi 10.1007/s12035-018-1044-6
- Saik O.V., Klimontov V.V. Bioinformatic reconstruction and analysis of gene networks related to glucose variability in diabetes and its complications. *Int. J. Mol. Sci.* 2020;21(22):8691. doi 10.3390/ijms21228691

- Saik O.V., Klimontov V.V. Hypoglycemia, vascular disease and cognitive dysfunction in diabetes: insights from text mining-based reconstruction and bioinformatics analysis of the gene networks. *Int. J. Mol. Sci.* 2021;22(22):12419. doi 10.3390/ijms2222212419
- Saik O.V., Klimontov V.V. Gene networks of hyperglycemia, diabetic complications, and human proteins targeted by SARS-CoV-2: what is the molecular basis for comorbidity? *Int. J. Mol. Sci.* 2022;23:7247. doi 10.3390/ijms23137247
- Sasaki T., Kuroko M., Sekine S., Matsui S., Kikuchi O., Susanti V.Y., Kobayashi M., Tanaka Y., Yuasa T., Kitamura T. Overexpression of insulin receptor partially improves obese and diabetic phenotypes in db/db mice. *Endocr. J.* 2015;62(9):787-796. doi 10.1507/endocrj.ej15-0255
- Shojima N., Yamauchi T. Progress in genetics of type 2 diabetes and diabetic complications. *J. Diabetes Investig.* 2023;14(4):503-515. doi 10.1111/jdi.13970
- Vaulont S., Vasseur-Cognet M., Kahn A. Glucose regulation of gene transcription. *J. Biol. Chem.* 2000;275(41):31555-31558. doi 10.1074/jbc.R000016200
- Vega M.E., Finlay J., Vasishtha M., Schwarzbauer J.E. Elevated glucose alters global gene expression and tenascin-C alternative splicing in mesangial cells. *Matrix Biol. Plus.* 2020;8:100048. doi 10.1016/j.mbplus.2020.100048
- Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 2007;24(8):1586-1591. doi 10.1093/molbev/msm088
- Wilmot E.G., Choudhary P., Leelarithna L., Baxter M. Glycaemic variability: The under-recognized therapeutic target in type 1 diabetes care. *Diabetes Obes. Metab.* 2019;21(12):2599-2608. doi 10.1111/dom.13842
- Zhang S., Ke Z., Yang C., Zhou P., Jiang H., Chen L., Li Y., Li Q. High glucose causes distinct expression patterns of primary human skin cells by RNA sequencing. *Front. Endocrinol.* 2021;12:603645. doi 10.3389/fendo.2021.603645
- Zhang Q., Xiao X., Zheng J., Li M., Yu M., Ping F., Wang T., Wang X. DNA methylation regulates pancreatic gene expression and links maternal high-fat diet to the offspring glucose metabolism. *J. Nutr. Biochem.* 2024;123:109490. doi 10.1016/j.jnutbio.2023.109490

---

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 02.10.2024. После доработки 08.11.2024. Принята к публикации 12.11.2024.