


doi 10.18699/vjgb-24-107

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 ^{1, 2}

¹ 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

² 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 Selektzii = Vavilov Journal of Genetics and Breeding*. 2024;28(8): 1008-1017. doi 10.18699/vjgb-24-107

Funding. The GlucoGenes® database and web resource was created with the support of the Russian Science Foundation (grant No. 20-15-00057-П).


Acknowledgements. The sections on evolutionary analysis of genes and SNP analysis were performed using the Bioinformatics Sharing Centre supported by Budget Project No. FWNR-2022-0006. The authors express their sincere gratitude to O.V. Saik (RICEL – branch of ICG SB RAS) for her significant contribution to data collection and valuable advice on the development of the database. We also thank A.M. Mukhin (ICG SB RAS) for technical assistance in creating the web resource.

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 ^{1, 2}

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

² Федеральный исследовательский центр Институт 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 родство 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.

Introduction

Diabetes is one of the most common and socially significant human diseases. According to experts from the International Diabetes Federation, the number of people living with diabetes worldwide reached 537 million in 2021 and is expected to rise to 783 million by 2045. In addition, more than 540 million people have impaired glucose tolerance (International Diabetes Federation, 2021).

In recent years, significant progress has been made in understanding the molecular mechanisms underlying the development of diabetes and its complications. Genome-wide association studies have identified a number of novel genetic loci that modulate the risk of diabetes and diabetic complications in European and Asian populations. Proteomics, metabolomics and multiomics studies have shed light on the molecular basis of disease pathogenesis (Langenberg, Lotta, 2018; Lyssenko, Vaag, 2023; Shojima et al., 2023).

At the same time, the effects of exposure to high glucose concentrations on the regulation of gene expression in different tissues have been identified (Vaulont et al., 2000; Hall et al., 2018; Vega et al., 2020; Zhang S. et al., 2021). It has been shown that the effects of high glucose levels on gene expression can be prolonged and exacerbated by epigenetic modifications. This mechanism is considered to be important for the phenomenon of metabolic memory and the development of diabetic complications (Dhawan et al., 2022). Abnormally low glucose levels are also associated with a number of biochemical shifts. These shifts are primarily related to the response of the cardiovascular and nervous systems to hypoglycemia (Hanefeld et al., 2016; Rehni, Dave, 2018).

The molecular effects of repeated episodes of high and low glucose levels, which characterize the phenomenon of

high glycemic variability (GV), attract increasing attention. Elevated GV has been found to increase the risk of microvascular and macrovascular diabetic complications and is associated with increased all-cause and cardiovascular mortality (Ceriello et al., 2019; Wilmot et al., 2019). At the molecular level, the pathophysiological changes associated with high GV are realized through increased or decreased expression of a large number of genes and altered activity of signaling pathways such as PI3K/Akt, NF- κ B, MAPK (ERK), JNK and TGF- β /Smad (Klimontov et al., 2021b).

Given the vast number and diversity of molecular changes in diabetes, advanced analytical tools are necessary to form a comprehensive and holistic understanding of the disease's pathogenesis. Artificial intelligence, bioinformatics, and integrative systems biology provide new opportunities for studying complex diseases such as diabetes (Nielsen, 2017; Klimontov et al., 2021a; Orlov et al., 2021; Putra et al., 2024).

A promising approach in this field is the analysis of gene networks, i.e. groups of genes that function in a coordinated manner, interact with each other, and determine specific phenotypic traits of an organism (Kolchanov et al., 2013). Previously, using text mining of scientific publications indexed in the PubMed and PubMed Central (PMC) databases, along with bioinformatic analysis, we reconstructed gene networks associated with glucose metabolism disorders (GMDs): hyperglycemia, hypoglycemia, and GV, gathering a large amount of data on molecules and proteins related to these metabolic disorders in some way (Saik, Klimontov, 2020–2022).

Based on the obtained data, a database was created containing information about genes and proteins associated with GMDs: hyperglycemia, hypoglycemia, and both conditions.

In this article, we present a description of the database's capabilities and provide the results of two bioinformatics studies conducted using it. The first study involved an evolutionary analysis of GMD genes, and the second involved an analysis of single nucleotide polymorphisms (SNPs) in 90 base-pair proximal regions of human gene promoters associated with GMDs.

Materials and methods

Development of the web resource. The material for creating the GlucoGenes® web resource was previously accumulated during the reconstruction and analysis of gene networks for hyperglycemia, hypoglycemia, and GV, conducted using ANDSystem, a bioinformatics tool (developed at Institute of Cytology and Genetics SB RAS; access: <https://anddigest.sysbio.ru/index.php>). ANDSystem constructs associative gene networks based on text mining of scientific publications indexed in the PubMed and PMC databases (Ivanisenko et al., 2015, 2019). The details of the analysis of GMD gene networks generated using ANDSystem were described previously (Saik, Klimontov, 2020–2022).

A relational data model and the PostgreSQL database management system (<https://www.postgresql.org/>) were chosen for the software implementation of the database.

For the design of the web resource, a client-server architecture was chosen, consisting of three main components: client, server and database. The Vue.js and Flask frameworks were used for development, and access management to the database is implemented through programmatic access based on REST technology.

Phylostratigraphic analysis and divergence analysis of genes associated with GMDs. Phylostratigraphic analysis is a method aimed at determining the evolutionary origin of genes by analyzing the presence of their orthologs, which are genes encoding homologous proteins that have diverged through speciation in the genomes of different species. This approach identifies key points in genome evolution, where a sharp increase occurred in the number of new genes, and helps to identify genes unique to specific taxa (Domazet-Lošo, Tautz, 2010).

We performed evolutionary analysis of genes included in the GlucoGenes® database using the phylostratigraphy age index (PAI) and divergency index (DI). The PAI value indicates how far the taxon reflecting the gene's age is from the root of the phylogenetic tree (Mustafin et al., 2021). The taxon reflecting the gene's age is considered as the taxon where the studied species diverged from the most distantly related taxon in which an ortholog of the gene has been found. The higher the PAI value of a gene, the younger it is. The Orthoweb software package (<https://orthoweb.sysbio.cytogen.ru/run.html>) was used for phylostratigraphic analysis. For PAI calculation, the method based on KEGG orthology groups was used (Kanehisa et al., 2016).

DI is an indicator of a gene's evolutionary variability. DI is calculated based on the dN/dS ratio, where dN is the proportion of nonsynonymous substitutions in the DNA sequences of the studied gene and its ortholog; dS is the proportion of

synonymous substitutions. This index was calculated by comparing human genes with genes from closely related organisms in the Hominidae family, specifically orthologs found in the western lowland gorilla *Gorilla gorilla gorilla*, Sumatran orangutan *Pongo abelii*, and common chimpanzee *Pan troglodytes*. The LPB93 model (Yang, 2007) was used to calculate dN/dS values. A DI value ranging from 0 to 1 indicates that a gene is undergoing stabilizing selection, a value of 1 indicates neutral evolution, and a value greater than 1 indicates positive selection.

Analysis of SNPs in 90-bp proximal regions of human gene promoters associated with GMD. In the Human_SNP_TATAdb knowledge base (Filonov et al., 2023), we searched for SNP variants in 90-bp proximal regions of human gene promoters associated with GMD that could statistically significantly decrease or increase the affinity of TATA-binding protein (TBP) to these promoters and consequently affect gene expression levels. Among all these SNPs, only those with clinical manifestations described in the publicly available ClinVar database (Landrum et al., 2014) were selected for further work. Finally, using the PubMed database, we performed curated annotations of how polymorphic changes in gene expression affected glucose levels in patients carrying these SNPs for all clinically relevant SNP markers located in promoters of genes associated with glucose metabolism disorders.

Results

GlucoGenes® web resource

The GlucoGenes® web resource is freely available at: <https://glucogenes.sysbio.ru/genes>. The interface of the resource is shown in Fig. 1.

The GlucoGenes® database consists of six components (tables). The Genes table contains gene names, descriptions and NCBI identifiers. The Proteins table includes protein names, descriptions, UniProt database identifiers, and links to the corresponding genes in the GlucoGenes® database. The Glycemia_related_conditions table provides information on glycemic disorders (hyperglycemia and hypoglycemia). The Types_of_glycemia_gene_association table contains information on the types of associations between molecules and glycemic disorders. The References table contains article identifiers in PubMed or PubMed Central with brief data extracts. The Glycemia_gene_association table is a summary table that aggregates information from all of the above-mentioned tables. The structure of the database is shown in Fig. 2. The database currently includes 561 genes associated with GMDs and 2,115 references to literature sources.

The GlucoGenes® web portal consists of four functional sections.

1. Homepage: The homepage provides general information about the resource and the terms used. We define hyperglycemia or high glucose levels in the culture medium as High Glucose (HG) and hypoglycemia or low glucose levels in the culture medium as Low Glucose (LG).

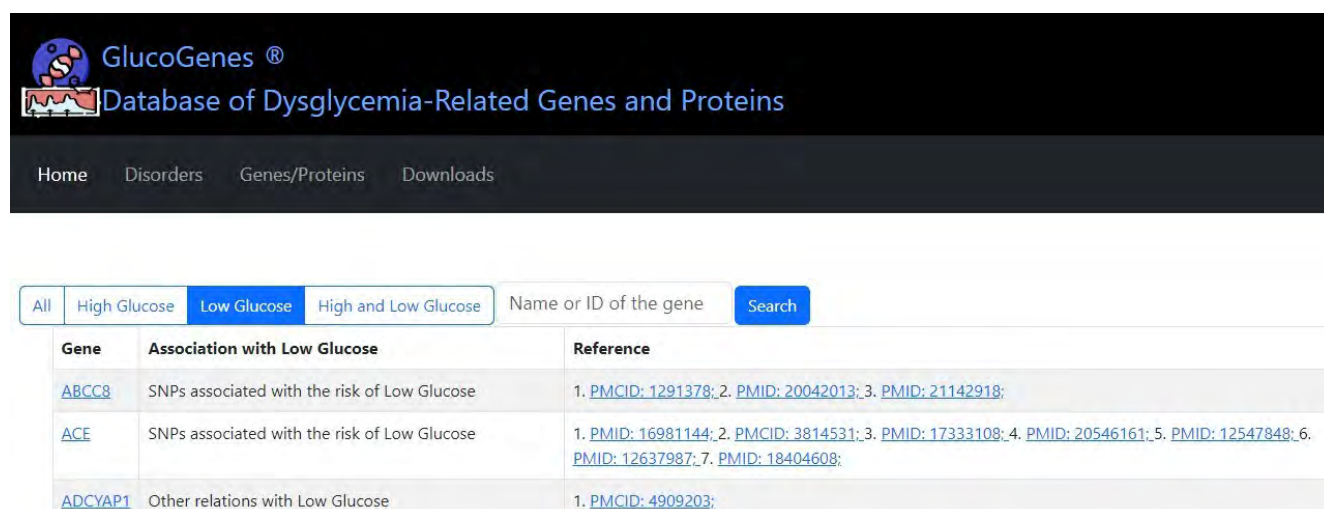


Fig. 1. GlucoGenes® website interface.

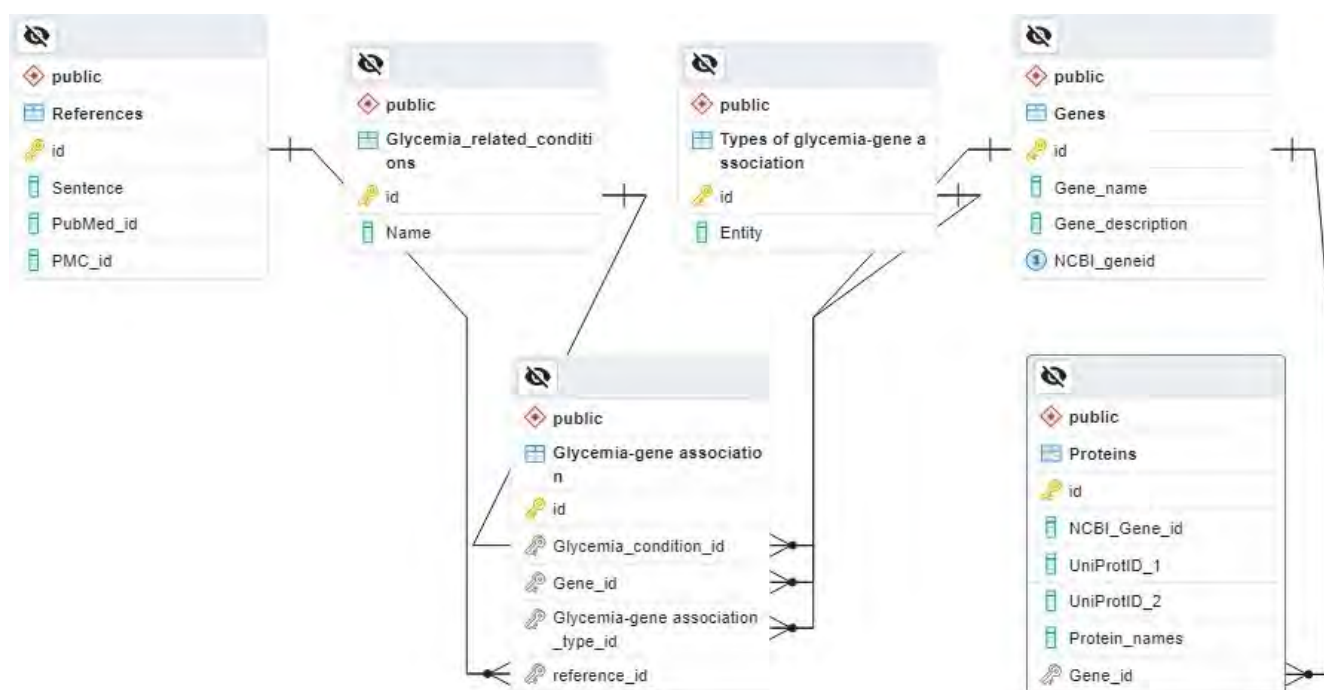


Fig. 2. Entity-relationship diagram of the GlucoGenes® database.

2. Disorders page: This page presents lists of genes associated with HG, LG, as well as with high and low glucose levels (HLG). For each gene, the type of association with glycemic disorders is indicated. The following categories of associations with glycemic disorders are highlighted: SNPs associated with HG, LG, or HLG; proteins with hyperglycemic activity; proteins with antihyperglycemic effects; genes up-regulated by HG; genes up-regulated by LG; genes down-regulated by HG; genes down-regulated by LG; and other associations with HG, LG, or HLG. For each gene and its association, references to relevant publications in PubMed are provided.

3. Genes/Proteins catalog: This section allows users to find gene names and NCBI gene identifiers, names of protein(s) encoded by the gene, and types of associations with GMD.

4. Downloads page: From this page, users can download lists of genes associated with HG, LG, and HLG, as well as all associated genes in Excel format. Search within the system is available by gene name, NCBI gene identifier, or type of GMD.

Data from the portal can also be accessed without using the graphical user interface via a REST application programming interface (API). This interface allows users to retrieve

Table 1. Lists of human protein-coding genes analyzed through phylostratigraphic analysis

| Gene group | Description | Number of genes |
|--|--|-----------------|
| All genes in the <i>Homo sapiens</i> genome | All human protein-coding genes for which PAI and DI values were calculated | 19,566 |
| Genes associated only with hyperglycemia | List of genes from GlucoGenes associated with high glucose levels | 430 |
| Genes associated only with hypoglycemia | List of genes from GlucoGenes associated with low glucose levels | 140 |
| Genes associated with both high and low glucose levels | List of genes from GlucoGenes associated with both high and low glucose levels | 151 |

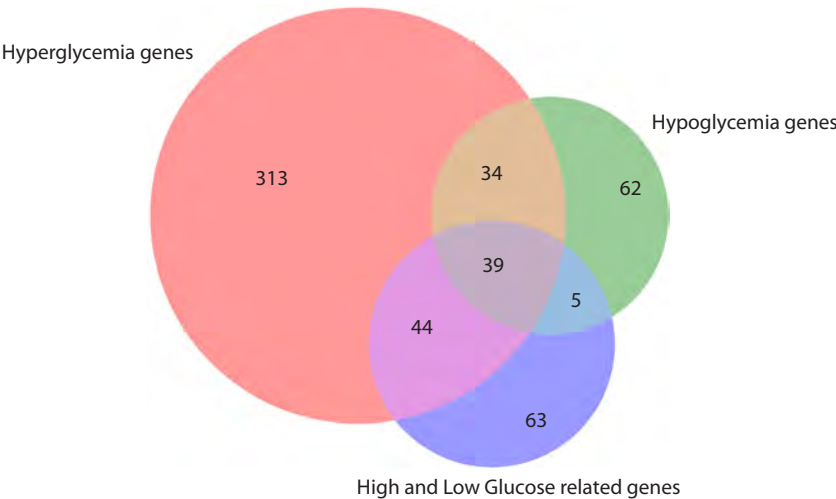


Fig. 3. Venn diagram showing intersections of gene groups.

required information by sending a request to the web server in the form of a URL string. In response to such a request, the server returns results as a text page or file, where the information is structured according to the JavaScript Object Notation (JSON) format (<http://json.org/>). The resulting text file can be opened with any text editor. It can also be processed using various software tools, including user-written programs in general-purpose modeling environments (e.g., Matlab, Scilab) or high-level programming languages (e.g., Python, R, C++, Java).

An example of a REST request is given below (the result is a text file in structured JSON format): <https://glucogenes.sysbio.ru/api/genes/<geneid>> – returns a card with a description of the gene <geneid>.

Evolutionary characteristics of genes associated with GMDs. We calculated PAI indices and plotted their distribution both for the list of protein-coding genes in the *Homo sapiens* genome and for genes represented in the GlucoGenes® database including those associated with hyperglycemia, hypoglycemia and glucose fluctuations (Table 1). Please, note that some genes were associated with more than one GMD (Fig. 3).

The distribution of PAI values for all genes in the human genome is multimodal, with two pronounced peaks at the levels of Cellular Organisms, Metazoa and Vertebrata-Euteleostomi (Fig. 4). The first peak is the largest; almost

55 % of genes in the *H. sapiens* genome have a PAI between 1 and 3. The second peak covers 32 % of the genes.

The results of the analysis showed a significant increase (up to 40 %) in the proportion of genes involved in glucose regulation with a PAI index = 3 in all three categories (Fig. 4). In particular, this group includes the *TCF7L2*, *PPARG*, *GCGR*, *IRS1* and *MTNR1B* genes, the products of which are important regulators of glucose metabolism.

Sequence conservation analysis for the same gene lists (Fig. 5) showed that most of the genes studied are highly conserved (DI < 0.6) or conserved (DI < 1). This indicates the conservation of their functions during evolution and highlights their critical role in biological processes related to glucose regulation. However, several genes with a DI greater than 1 were identified, indicating recent exposure to positive selection. These genes include *SPPI*, *CALCA*, *CD33*, *SULT2A1*, *TNF*, *ECM1*, *CYP3A4* and *EDN1*.

Analysis of SNPs in 90-bp proximal regions of human gene promoters associated with GMD. A total of 181 SNP markers were identified in the GlucoGenes® database, which may either decrease (45 SNP markers) or increase (136 SNP markers) the expression of 52 human genes, thereby altering glucose levels in patients carrying minor alleles of these SNPs. Table 2 provides an example of 10 SNPs located in the promoters of the human *ABCC8*, *INSR*, and *PGM1* genes, available in the ClinVar database (Landrum et al., 2014).

| PAI | Taxon |
|-----|---------------------|
| 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> |

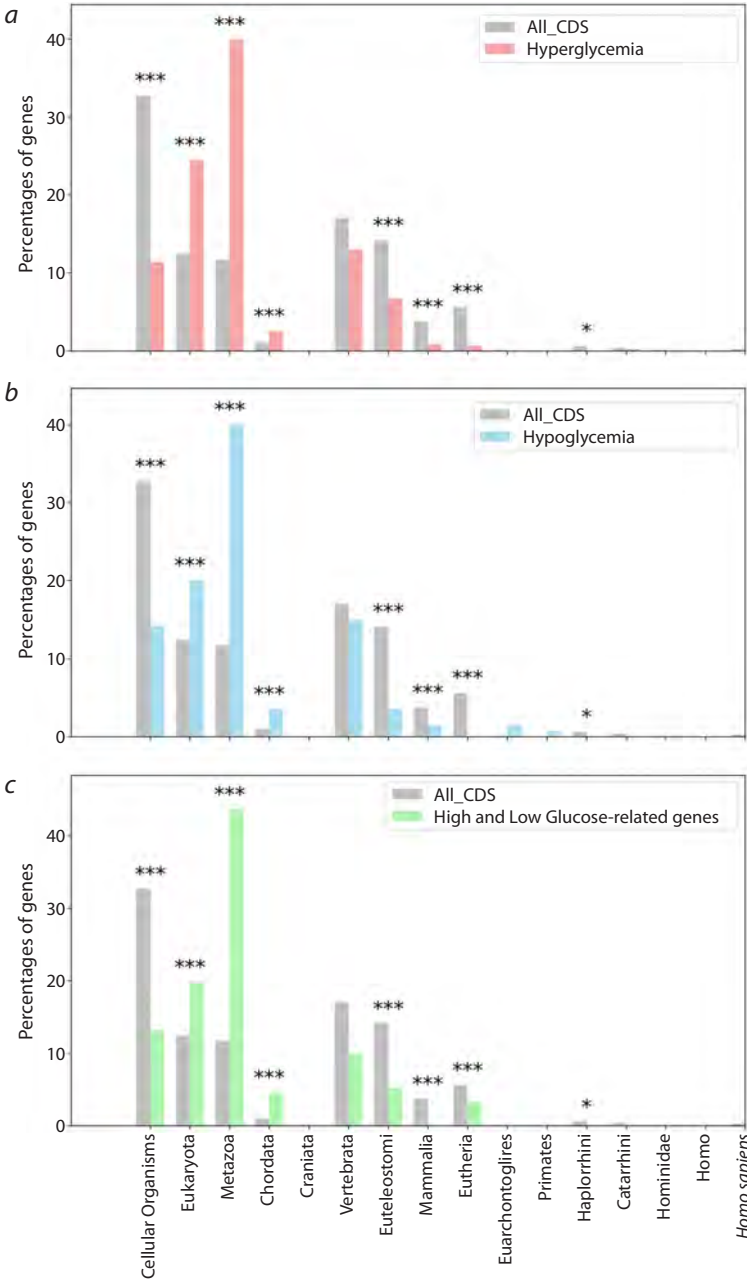


Fig. 4. Distribution of protein-coding genes associated with GMDs by PAI values.

Here and in Fig. 5: *a* – all human protein-coding genes (All_CDS) as a control group compared to genes associated with high glucose levels (Hyperglycemia); *b* – all human protein-coding genes (All_CDS) as a control group compared to genes associated with low glucose levels (Hypoglycemia); *c* – all human protein-coding genes (All_CDS) as a control group compared to genes associated with both high and low glucose levels (High and Low Glucose-related genes). Columns marked with asterisks indicate statistically significant differences between gene samples from the database and the sample of all human protein-coding genes: **p*-value < 0.05, ****p*-value < 0.001. Statistical testing was performed using the chi-square test.

According to the data presented in Table 2, minor alleles of the *ABCC8*, *INSR*, and *PGM1* gene promoters exhibit altered affinity for TBP, which may affect the expression levels of these genes and explain their association with GMDs. More detailed information on the identified SNP markers can be found in Supplementary Material¹.

¹ Supplementary Material is available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Klimontov_Engl_28_8.pdf

Discussion

Advances in the study of the molecular biology of diabetes open up broad opportunities for the implementation of precision medicine technologies in the treatment of this disease. In particular, the identification of disease-specific biomarkers offers new prospects for diagnosis, monitoring, prognosis of the disease and its outcomes, pharmacogenetics of modern glucose-lowering drugs, as well as the search for new thera-

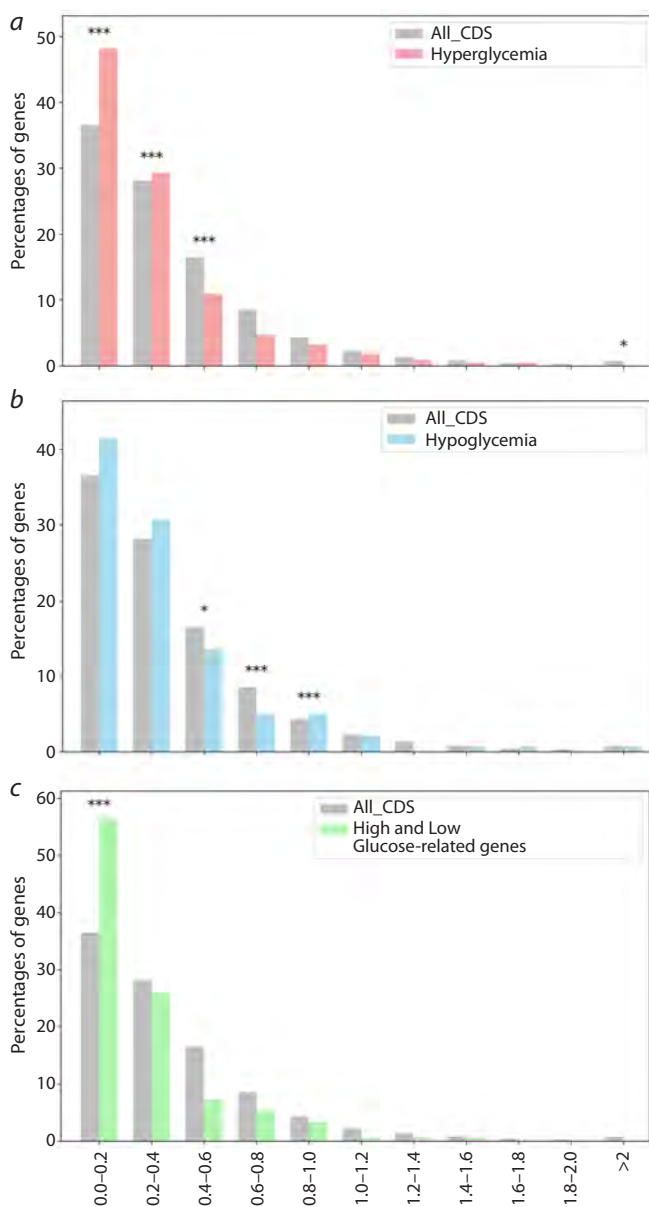


Fig. 5. Distribution of protein-coding genes associated with GMD by DI values.

peutic agents (Chung et al., 2020). The rapid accumulation of data on the molecular basis of genetic predisposition to diabetes and the molecular mechanisms of its complications underscores the need for research tools to facilitate structured information retrieval in this field.

We have developed a database of genes and proteins that have demonstrated associations with GMDs, including hyperglycemia, hypoglycemia, or both. The web-based resource, named GlucoGenes® (<https://glucogenes.sysbio.ru/genes/main>), can be utilized to collect, search, and visualize information on genes and proteins associated with GMDs. Access to the database integrated into GlucoGenes® is provided via a REST-based API for record browsing. A graphical user interface allows users to view records and

export their content in Excel format. The database contains catalogs of genes and proteins associated with GMDs, including information on the types of associations and links to abstracts of relevant publications in PubMed or full-text articles in PMC. Gene and protein lists are available for download. A limitation of this resource is that it accumulates data only from articles indexed in the PubMed and PMC. Regular information updates are evidently required.

The developed resource may prove useful for addressing research challenges in bioinformatics and the molecular biology of diabetes. Specifically, it can be applied to select genes and proteins for studying genetic predisposition to diabetes in various populations, investigating the molecular aspects of pathogenesis, searching for potential biomarkers of diabetic complications, identifying potential therapeutic targets, and other tasks. In this study, we present examples of using the developed resource to solve research tasks in bioinformatics studies.

The first task focuses on the evolutionary origin of genes associated with GMDs. Evolutionary analysis of genes using phylostratigraphy is a key tool in biology, enabling an understanding of the fundamental mechanisms underlying the diversity of life on Earth. The evolutionary history of genes provides insights into how various functions and structures have evolved and adapted to environmental changes. This knowledge not only aids in reconstructing phylogenetic trees but also helps to identify genes responsible for adaptive changes and specific physiological processes, such as glucose metabolism. The conducted phylostratigraphic analysis revealed that among genes associated with glucose metabolism, a significant proportion (up to 40 %) are genes with PAI = 3, corresponding to the origin of multicellular organisms (Maloof et al., 2010). Most of the studied genes were found to be highly conserved ($DI < 0.6$) or conserved ($DI < 1$). The obtained results emphasize the importance of GMD-associated genes in regulating specialized metabolic processes characteristic of complex organisms.

During the second task, data from the web resource were used to analyze SNPs in proximal regions of human gene promoters that affect the affinity for TBP. The integration of GlucoGenes® data with information on SNP associations with various human diseases from other databases, on the one hand, and bioinformatic assessments of changes in glucose levels in patients carrying these SNPs, on the other hand, reflects the molecular mechanisms through which GMDs may influence the progression of these diseases.

Conclusion

GlucoGenes® is a resource that combines a graphical user interface with a database of genes and proteins associated with hyperglycemia, hypoglycemia, and both metabolic disorders. The resource has been utilized for bioinformatic analysis of the evolutionary characteristics of genes associated with these disorders, as well as for the analysis of SNPs in proximal promoter regions of genes that affect the

Table 2. SNPs affecting TBP affinity in the promoters of the *ABCC8*, *INSR* and *PGM1* genes associated with GMDs

| GlucoGenes® web resource | | ClinVar database (Landrum et al., 2014) | | | Human_SNP_TATAdb database (Filonov et al., 2023) | | | | PubMed database | | |
|---|---|--|----------------|----------------------------|--|--------------------------------|---------------|---------------|-----------------|--------------------------------|-----|
| | | Association with SNP | | dbSNP database (Day, 2010) | | K_D , nM/l, <i>in silico</i> | | Significance | | Effect on disease course (Ref) | |
| Gene symbol | Glycaemia status | Disease | dbSNP ID:min | 5' flank, 10 bp | WT → min | 3' flank, 10 bp | WT | min | z | p | Δ |
| | | | | | | | $M_0 \pm SEM$ | $M_0 \pm SEM$ | | | |
| <i>PGM1</i> | Low Glucose | PGM1-congenital disorder of glycosylation, inborn genetic diseases | rs199928376:T | tgccgccaac | G→T | gggtaaggga | 46.52±3.87 | 36.68±3.28 | 3.89 | 10 ⁻³ | B ↑ |
| Excess PGM1: A response to glucose deficiency in lung cancer (Li et al., 2020) | | | | | | | | | | | |
| <i>INSR</i> | High and Low Glucose & High Glucose | Insulin-resistant diabetes mellitus & acanthosis nigricans; Rabson-Mendenhall syndrome, leprechaunism syndrome | rs532052290:T | cgctcggag | C→T | tcttggtgg | 62.14±4.83 | 17.87±1.92 | 18.81 | 10 ⁻⁶ | A ↑ |
| Excess <i>Insr</i> : Improved health outcomes in db/db mice (Sasaki et al., 2015) | | | | | | | | | | | |
| <i>ABCC8</i> | High and Low Glucose & Low Glucose & High Glucose | ABCC8-related disorder, hypoglycemia, hyperinsulinism, diabetes mellitus, monogenic diabetes | rs1409900082:A | agcatttcc | G→A | tggtcctcag | 74.04±5.71 | 30.75±2.85 | 14.58 | 10 ⁻⁶ | A ↑ |
| Maternal diet effects on offspring in mice (Zhang Q. et al., 2024): Excess Abcc8 protects offspring | | | | | | | | | | | |
| | | | rs1409900082:T | agcatttcc | G→T | tggtcctcag | 74.04±5.71 | 52.37±3.26 | 6.99 | 10 ⁻⁶ | A |
| | | | rs1955706366:A | cctgtccagt | G→A | ctaagcgagt | 53.2±4.54 | 12.08±1.4 | 20.58 | 10 ⁻⁶ | A |
| | | | rs1957150758:A | cttgccgac | C→A | actttccga | 74.04±5.71 | 47.43±3.75 | 8.06 | 10 ⁻⁶ | A |
| | | | rs541244107:A | agctgcaagg | G→A | ccgggcgcg | 59.19±4.91 | 50.3±3.4 | 3.04 | 10 ⁻² | C |
| | | | rs541244107:C | agctgcaagg | G→C | ccgggcgcg | 59.19±4.91 | 47.44±3.82 | 3.83 | 10 ⁻³ | B |
| | | | rs565662182:A | tgcccccat | G→A | gaggagcagt | 53.2±4.54 | 31.21±2.78 | 8.65 | 10 ⁻⁶ | A |
| | | | rs58241708:T | tgcacagatc | C→T | ttcgtgtcca | 53.2±4.54 | 45.33±3.8 | 2.68 | 10 ⁻² | C |

Note. WT and min – ancestral (wild-type) and minor (pathological) alleles of SNPs; K_D – equilibrium dissociation constant of the TBP-promoter complex, expressed in nanomoles per liter, nM/l; M_0 and SEM – context-dependent *in silico* assessment and its standardized error; z, p and p – Fisher's z-statistic, its significance level, and heuristic prioritization of *in silico* evaluations from best (A) to worst (D) in alphabetical order; Δ – expression increased (↑) or decreased (↓).

affinity for TBP. It has been demonstrated that a significant proportion of genes associated with GMDs are evolutionarily ancient and conserved. SNP markers that can decrease (45 SNP markers) or increase (136 SNP markers) the expression of 52 genes have been identified.

References

- 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 M.F., Lee C.G., McCarthy M.I., Nolan J.J., Norris J.M., Pearson E.R., Philipson 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
- Filonov S.V., Podkolodny N.L., Podkolodnaya O.A., Tverdokhlebov 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 Selektii = Vavilov Journal of Genetics and Breeding.* 2023;27(7):728-736. doi 10.18699/VJGB-23-85
- 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. ANDSystem: 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 ANDSystem 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., Berikov V.B., Saik O.V. Artificial intelligence in diabetology. *Sakhamyi Diabet = Diabetes Mellitus.* 2021a;24(2):156-166. doi 10.14341/DM12665 (in Russian)
- Klimontov V.V., Saik O.V., Korbut A.I. Glucose variability: How does it work? *Int. J. Mol. Sci.* 2021b;22(15):7783. doi 10.3390/ijms22157783
- Kolchanov N.A., Ignatieva E.V., Podkolodnaya O.A., Likhoshvai V.A., Matushkin Yu.G. Gene networks. *Vavilovskii Zhurnal Genetiki i Selektii = Vavilov Journal of Genetics and Breeding.* 2013;17(4/2):833-850 (in Russian)
- 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
- Mustafin Z.S., Lashin S.A., Matushkin Yu.G. Phylostratigraphic analysis of gene networks of human diseases. *Vavilovskii Zhurnal Genetiki i Selektii = Vavilov Journal of Genetics and Breeding.* 2021;25(1):46-56. doi 10.18699/VJ21.006
- 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/ijms222212419
- 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., Leelarathna 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

Conflict of interest. The authors declare no conflict of interest.

Received October 2, 2024. Revised November 8, 2024. Accepted November 12, 2024.