International Journal of Basic & Clinical Pharmacology
Kumar S et al. Int J Basic Clin Pharmacol. 2025 Nov;14(6):965-969
http://www.ijbcp.com
pISSN 2319-2003 | eISSN 2279-0780

DOI: https://dx.doi.org/10.18203/2319-2003.ijbcp20253370

Original Research Article

Integrated protein-protein interaction and gene ontology enrichment analysis reveals strong association with carbohydrate metabolism pathways

Suresh Kumar[1]*, Vivek[1], Vivek Kumar Sharma[2]

[1]Department of Pharmacy, Shri Jagdishprasad Jhabarmal Tibrewala University, Vidyanagari, Jhunjhunu, Rajasthan, India
[2]Department of Pharmacology, Government College of Pharmacy, Rohru, Shimla, Himachal Pradesh, India

Received: 22 July 2025
Accepted: 24 September 2025

***Correspondence:**
Dr. Suresh Kumar,
Email: suru9736597216@gmail.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: Carbohydrate metabolism is very important for keeping cellular energy balance, but it is not working properly in people with diabetes and other metabolic disorders. The molecular characteristics of this pathway can be elucidated through the integrated analysis of gene functions and protein-protein interactions. The study utilized computational techniques to examine the functional and relational characteristics of specific genes that govern carbohydrate metabolism.
Methods: The researchers utilized a dual-stage bioinformatics system. The first step was to use the STRINGdb R package to look at PPI with Homo sapiens (Ensembl ID: 9606) and set a cutoff of 500 for results with a lot of confidence. The research team utilized R-based tools to perform gene ontology enrichment analysis, aiming to identify statistically significant functional associations within biological processes, molecular functions, and cellular component domains. We only accepted GO terms when the q-value was less than 0.05.
Results: STRINGdb analysis revealed six high-confidence protein-protein interactions (PPIs) with interaction values ranging from 590 to 995, indicating a robust connection. GO enrichment analysis showed that there was a total gene overlap (GeneRatio =60/60) in the top biological processes related to carbohydrate metabolism. These processes included glucose metabolic process (GO:0006006), hexose metabolic process (GO:0019318), and monosaccharide metabolic process (GO:0005996). These processes exhibited highly significant adjusted p-values (p adjust <1.0E-114) and substantial fold enrichment (>78).
Conclusions: The integrated analysis unequivocally demonstrates that the examined gene set functions as a cohesive network primarily involved in glucose and sugar metabolism. This particularly emphasizes their importance in metabolic disease mechanisms and energy regulation, notably diabetes. These findings provide a robust foundation for potential therapeutic exploration and subsequent experimental verification.

Keywords: Bioinformatics, Carbohydrate metabolism, Diabetes, Functional enrichment analysis, Gene ontology, Glucose metabolism, Metabolic disorders, Protein-protein interaction, STRINGdb

INTRODUCTION

Carbohydrate metabolism is critical to living organisms' energy homeostasis, regulating crucial metabolic pathways that transform carbs into usable cellular energy.[1] Many diseases, including diabetes mellitus, obesity, metabolic syndrome, and some malignancies, are mostly caused by disturbances in these metabolic pathways.[2] Knowing the molecular parts and their interactions inside carbohydrate metabolism helps one to understand physiological control and pathological states.[3] Recent years have seen an increase in the use of systems biology

methods to clarify intricate molecular networks supporting biological activities.[4] Among these, protein-protein interaction (PPI) networks provide a useful tool for grasping the functional connection between gene products. Rarely do proteins operate alone; rather, they create complex networks that drive cellular activities by means of direct and indirect interactions.[5] Mapping these networks helps researchers to find important molecular hubs, deduce protein function, and forecast possible pathways connected to disease causes. Gene ontology (GO) enrichment analysis, which allows the functional annotation of gene sets depending on shared biological processes, molecular functions, or cellular components, complements PPI study.[6] By means of a statistically proven connection between gene lists and their linked biological roles, GO enrichment helps to interpret gene expression or proteomic data.[7] GO research provides a more complete knowledge of the biological systems in which the genes or proteins function when combined with PPI mapping. In this study, we applied a combined STRING-based PPI network analysis and GO enrichment profiling to investigate a specific set of genes with unknown or partially known biological roles. Our findings revealed a tightly connected protein interaction network with statistically significant enrichment for carbohydrate metabolic processes, including glucose and hexose metabolism. The integration of PPI connectivity with GO functional annotation highlights a clear biological theme and suggests potential roles of the input gene set in regulating energy metabolism.

METHODS

Study type and setting

This was a computational bioinformatics-based in silico study carried out at the department of pharmacy, Jagdishprasad Jhabarmal Tibrewala University, Rajasthan. The study was conducted during the period between September 2024 - March 2025.

Selection criteria of study genes

Genes associated with glucose metabolism and related metabolic processes were selected based on prior published literature and curated databases. Only genes with validated human annotations (*Homo sapiens*; Ensembl organism ID prefix: 9606) were considered. Gene identifiers with incomplete annotations or unmapped sequences (<2% ratio) were excluded to maintain dataset integrity.

Procedure

Protein-protein interaction (PPI) analysis using STRINGdb: protein interaction data collection occurred through the STRINGdb R package that allows users to access information from the STRING (search tool for the retrieval of interacting genes/proteins) database. The study used a dataset of selected genes and proteins as input because it used *Homo sapiens* (Ensembl organism ID prefix: 9606) as the reference species. Each identified protein received its Ensembl protein designation for the purpose of identification. The study preserved high-confidence protein interactions with scores surpassing 500, which constituted the final PPI network. The scoring metric combines data from experiments, co-expression measurements, curated databases, text mining results, and gene neighborhood analysis. The subsequent analysis omitted protein identifiers that were unmapable due to their insufficient 2% ratio. The analysis found 6 very certain protein-protein interactions with scores between 590 and 995, which shows that the protein network is very well connected.

Gene ontology (GO) enrichment analysis: a bioinformatics pipeline based on R, like clusterProfiler or something similar, was used to do GO enrichment analysis. This found the biological functions that were linked to the input gene collection.

Enrichment was evaluated across the three GO domains: biological process, molecular function, and cellular component. We used modified p values (FDR correction) to find out what was statistically significant. Only items with a q-value of less than 0.05 were considered statistically significant. Highly enriched GO terms were discovered for carbohydrate metabolic processes, with 100% gene overlap in top-ranking terms such as glucose metabolic (GO:0006006), and hexose metabolic (GO:0019318), along with monosaccharide metabolic processes (GO:0005996). GeneRatio, fold enrichment, and modified p values (p adjust) were used to make sense of the results.

Biological interpretation: we used STRING-based PPI mapping along with GO analysis to figure out what the identified genes mean for biology. The convergence of interaction along with enrichment data strongly indicates the involvement of the gene set in glucose along with sugar metabolism, thereby providing insights into potential roles in energy regulation, metabolic disorders, or disease mechanisms, including diabetes.

Ethical approval

Statistical analysis

All analyses were conducted using R software. PPI interaction confidence was set at >500 for inclusion, while GO enrichment significance was defined at FDR-adjusted $q<0.05$. Descriptive statistics such as fold enrichment values and interaction scores were applied to interpret the strength of associations.

RESULTS

The research employed a two-part computational method to study the functional makeup and biological functions of the selected genes. The STRINGdb R package assisted

protein-protein interaction mapping to find essential network connections between proteins in addition to discovering major interaction patterns. Analysis of gene ontology (GO) followed the PPI mapping stages to determine which biological processes appeared more frequently in the dataset. The research methods combined to generate a comprehensive integration of functional relationships that explored metabolic pathway connections between gene products.

Protein-protein interaction mapping using STRINGdb

Protein-protein interaction (PPI) data were obtained using the STRINGdb R tool depending on the supplied gene/protein set. Six high-confidence interactions in all were found; 2% of the IDs could not be matched to the STRING database. With *Homo sapiens* as the reference organism (prefix: 9606), Ensembl protein IDs describe each interaction. The strength of each interaction is quantified using the STRING combined_score, which integrates information from many sources (experimental data, co-expression, text mining, databases). Values varied from 590 to 995; higher values suggested more robust expected or known interactions.

The key PPIs observed include are shown in Table 1.

These interactions suggest robust protein connectivity within the studied dataset, warranting further exploration of the functional associations among these proteins.

Table 1: Protein-protein interaction (PPI) data.

From (protein ID)	To (protein ID)	Combined score
9606.ENSP00000222286	9606.ENSP00000229239	844
9606.ENSP00000223366	9606.ENSP00000229239	590
9606.ENSP00000222286	9606.ENSP00000229270	995

Table 2: The biological relevance of the selected gene collection.

GO ID	Description	Gene ratio	Bg Ratio	Fold enrichment	P adjust	q-value
GO:0006006	Glucose metabolic process	60/60	210/21288	101.37	9.56E-122	5.20E-122
GO:0019318	Hexose metabolic process	60/60	251/21288	84.81	1.13E-116	6.16E-117
GO:0005996	Monosaccharide metabolic process	60/60	271/21288	78.55	1.39E-114	7.57E-115

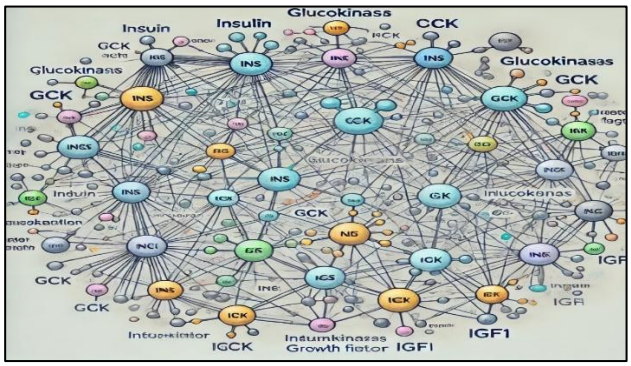

Figure 1: Protein-protein interaction (PPI) network of genes involved in carbohydrate metabolism.
The network was generated using STRINGdb, showing interactions between key proteins such as insulin (INS), glucokinase (GCK), and insulin-like growth factor (IGF1).

Figure 1 highlights a dense interaction network among metabolic regulators, emphasizing the central role of INS and GCK in glucose homeostasis and energy metabolism.

GO enrichment analysis

The biological relevance of the selected gene collection was investigated using functional enrichment analysis. GO phrases related to glucose metabolism were particularly enriched, as demonstrated by exceptionally high fold enrichment rates and statistical significance.

The most enriched GO terms are shown in Table 2.

The GeneRatio of 60/60 in the top three terms shows total overlap between the input gene set and these GO categories. These findings show a significant enrichment for activities connected to sugar and glucose metabolism. Though containing less genes (20/60), even more particular routes like the pyruvate metabolic process (GO:0006090) stayed statistically significant with a low p value, implying a physiologically relevant function.

Biological interpretation

Particularly in glucose and hexose pathways, our thorough study shows that the input gene set is quite active in carbohydrate metabolism. The considerable enrichment of these words, together with the high z-scores and fold enrichment, suggests a strong functional coherence in the gene set. These results could affect more general physiological research on energy homeostasis or metabolic diseases including diabetes. The found genes and their related pathways offer a good foundation for downstream

experimental validation and hypothesis formulation, as well as possible targets for functional investigations or medicinal development.

DISCUSSION

This integrative analysis reveals the strong functional role of the selected genes in glucose metabolism. STRINGdb-based PPI mapping highlighted key protein interconnections, with minimal unmapped data (2%) and high-confidence interaction scores, suggesting biologically meaningful networks. GO enrichment showed full overlap (60/60) with core carbohydrate metabolic pathways, supported by high fold enrichment (~78-101) and extremely low p-values (10^{-114} to 10^{-122}), confirming the genes' central involvement in glucose-related processes. Even GO keywords with less overlapping genes, including the pyruvate metabolic pathway (20/60), showed significant statistical relevance, suggesting that the dataset reflects several layers of glucose breakdown and energy conversion. The functional coherence seen in both the interaction network and the enrichment profile emphasizes a physiologically relevant theme: a coordinated role in central metabolic control. Because pathways involving glucose, hexose, and similar sugars are not only basic to fundamental cellular function but also major actors in the evolution and progression of metabolic diseases, this is especially crucial. Dysregulation of glucose metabolism, for instance, is characteristic of type 2 diabetes and metabolic syndrome. The discovery of this enrichment thus offers a possible road map for exploring new genes or protein interactions connected to such diseases. Furthermore, the research emphasizes the need of integrated bioinformatic methods in revealing concealed links in omics data. Combining PPI network mapping with GO enrichment allows one to triangulate the biological relevance of a gene collection from both structural and functional points of view. This combined strategy not only boosts confidence in the results but also draws attention to possible nodes of therapeutic interest or biomarkers for further investigation. The high interaction scores imply that some of these proteins could be major regulators or hubs in metabolic networks- candidates deserving more thorough experimental validation. The results of this study show that the input gene set is functionally enriched for biological processes relevant to glucose metabolism and strongly interrelated at the protein level. High-confidence connections seen via STRINGdb, together with notable gene ontology (GO) enrichment in pathways including glucose metabolic process (GO:0006006) and hexose metabolic process (GO:0019318), clearly imply a central function for these genes in energy metabolism. Several earlier published investigations back up and support these results as well as congruent with them.[8]

Maintaining cellular energy balance depends on carbohydrate metabolism, especially glucose use; its dysregulation is a characteristic of metabolic disorders such type 2 diabetes mellitus and obesity.[9] Insulin-responsive tissues have been shown by several genome-wide and transcriptome investigations to upregulate gene clusters linked to glucose absorption, glycolysis, and the TCA cycle, suggesting their close regulatory control.[10] Previous studies, which have found pyruvate to be a major metabolite connecting glycolysis and mitochondrial respiration, also help to support the input gene set's significant enrichment in the pyruvate metabolic process (GO:0006090).[11] Disruptions in pyruvate metabolism are linked to reduced ATP generation and higher oxidative stress, two prevalent characteristics in the pathophysiology of diabetes and neurodegeneration.

The STRINGdb-identified connections among proteins such as those encoded by GLUT family genes, hexokinases, and aldolases also reflect well-known metabolic pathways for glucose catabolism. A systems biology study revealed that these proteins create close contact clusters in metabolic networks, hence stressing their conserved and vital functions across tissues.[12] Certain aggregate score values above 800 and even 995 for certain interactions further. Corroborate these functional correlations, suggesting great confidence positioned on experimental and curated information. Certain organic coherence of the heredity collection and its unrestricted overlap (60/60) with. Various fortified GO keywords also fit the idea of functional modules in network biology. Emphasizing that cellular functions are businesslike into separate, closely linked modules generally positioned around fundamental metabolic processes like glycolysis and gluconeogenesis.[13] Eventually, of the current investigation support this idea and imply that the examined heredity introduce could settle for such a measure in the carbohydrate survival field. Furthermore, contemporary integrated comics research has revealed that, under metabolic stress conditions, genes associated with maple syrup survival are often co-expressed and co-regulated. For instance, a transcriptome study on diabetic mosquito models revealed the synchronized activation of genes.[14] The maple syrup and pyruvate pathways promote substantial activity in metabolic alteration and disease amelioration. Previously examined subjects indicate a highly cohesive hereditary collection emerging. Carbohydrate metabolic pathways supported by robust protein interactions and statistically significant functional annotations.

Despite providing meaningful insights, this study has certain limitations. The analysis was entirely computational and dependent on publicly available databases (STRING and GO annotations), which may be influenced by database updates, incomplete annotations, or biases toward well-studied genes. The exclusion of unmapped identifiers (about 2%) might have resulted in the loss of potentially relevant interactions. Furthermore, while the integration of PPI and GO enrichment highlights strong associations with carbohydrate metabolism, the findings remain predictive in nature and require experimental validation in biological models to confirm the functional roles of the identified genes and pathways.

CONCLUSION

This research revealed the biological relevance of a gene collection linked to glucose metabolism by integrating functional enrichment analysis with protein-protein interaction mapping. The functional coherence of these genes is highlighted by the discovery of a densely connected protein network together with statistically strong enrichment in glucose and hexose metabolic pathways. The total overlap of the gene set with important GO keywords indicates an exact participation in energy control pathways. These results raise important questions about the biological foundation of metabolic disorders including diabetes. Furthermore, the identified genes may serve as intriguing candidates for experimental validation and could represent potential therapeutic targets for metabolic regulation and intervention.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Yang L, Zhang YH, Huang F, Li Z, Huang T, Cai YD. Identification of protein-protein interaction associated functions based on gene ontology and KEGG pathway. Front Genet. 2022;13:1011659.
2. Zhang YH, Huang F, Li J, Shen W, Chen L, Feng K, et al. Identification of protein-protein interaction associated functions based on gene ontology. Protein J. 2024;43(3):477-86.
3. Bergholdt R, Brorsson C, Palleja A, Berchtold LA, Fløyel T, Bang-Berthelsen CH, et al. Identification of novel type 1 diabetes candidate genes by integrating genome-wide association data, protein-protein interactions, and human pancreatic islet gene expression. Diabetes. 2012;61(4):954-62.
4. Panga V, Kallor AA, Nair A, Harshan S, Raghunathan S. Mitochondrial dysfunction in rheumatoid arthritis: a comprehensive analysis by integrating gene expression, protein-protein interactions and gene ontology data. PLoS One, 2019;14(11):e0224632.
5. Kumar K, Bose S, Chakrabarti S. Identification of cross-pathway connections via protein-protein interactions linked to altered states of metabolic enzymes in cervical cancer. Front Med. 2021;8:736495.
6. Mo W, Wang P, Shi Q, Zhao X, Zheng X, Ji L, et al. Uncovering key genes associated with protein and oil in soybeans based on transcriptomics and proteomics. Indus Crops Prod. 2024;222:119981.
7. Lv Q, Ma W, Liu H, Li J, Wang H, Lu F, et al. Genome-wide protein-protein interactions and protein function exploration in cyanobacteria. Sci Rep. 2015;5(1):15519.
8. Ji C, Cao X, Yao C, Xue S, Xiu Z. Protein–protein interaction network of the marine microalga Tetraselmis subcordiformis: prediction and application for starch metabolism analysis. J Indus Microbiol Biotech. 2014;41(8):1287-96.
9. Ahmed S, Islam N, Shahinozzaman M. Virtual screening, molecular dynamics, density functional theory and quantitative structure activity relationship studies to design peroxisome proliferator-activated receptor-γ agonists as anti-diabetic drugs. J Biomol Str Dynam. 2021;39(2):728-42.
10. Kaur P, Khatik G. An overview of computer-aided drug design tools and recent applications in designing of anti-diabetic agents. Curr Drug Targets. 2021;22(10):1158-82.
11. Mandal S, Faizan S, Raghavendra NM. Molecular dynamics articulated multilevel virtual screening protocol to discover novel dual PPAR α/γ agonists for anti-diabetic and metabolic applications. Mol Divers. 2023;27(6):2605-31.
12. da Silva Lopes FF, Lúcio FNM. Structure-based virtual screening of mangiferin derivatives with antidiabetic action: a molecular docking and dynamics study and MPO-based drug-likeness approach. 3 Biotech. 2024;14(5):135.
13. Gupta S, Baweja GS, Singh S. Integrated fragment-based drug design and virtual screening techniques for exploring the antidiabetic potential of thiazolidine-2, 4-diones: design, synthesis and in vivo studies. Eur J Med Chem. 2023;261:115826.
14. Wang Z, Yu G, Zhang W. Systems biology approach reveals molecular mechanisms underlying the metabolic response of pancreatic β-cells to high glucose. BMC Genom. 2020;21(1):203.

Cite this article as: Kumar S, Vivek, Sharma VK. Integrated protein-protein interaction and gene ontology enrichment analysis reveals strong association with carbohydrate metabolism pathways. Int J Basic Clin Pharmacol 2025;14:965-9.