ONLINE JOURNAL OF BIOLOGICAL SCIENCES

RESEARCH ARTICLE

Computational Prediction of Diabetic Nephropathy-Associated Comorbidities and Their Key Genes by Analyzing RNA-Seq Data

Chaity Paul, Md. Ashiq Uddin, Rupa Dey, Adrita Alam, Md. Humayun Kabir*

Department of Computer Science and Engineering, Faculty of Engineering, University of Rajshahi, Bangladesh

*Corresponding Author: hkj_cse@ru.ac.bd

Abstract: Diabetic Nephropathy (DN) is a complex health condition that leads to various kidney diseases. According to the World Health Organization, over 2 million deaths have been reported due to diabetes and diabetes related kidney diseases in 2021. And at the recent 78th World Health Assembly, chronic kidney disease was reported to be one of the most common diabetes-related complications, and projected that about 30 to 40% of people living with diabetes will develop the disease. The purpose of this study is to predict the comorbidities of DN and their associated key genes by analyzing publicly available gene expression data of advanced and early stages of diabetes affected kidney samples. First, we collected the gene expression data from the NCBI GEO repository. After preprocessing the data, we identified the highly expressed sample-specific genes for all the samples in the data. Then, we obtained diabetes affected kidney samples' specific genes and investigated their biological processes, pathway activities, disease analyses, and PPI networks. Based on the results, comorbidities and their associated key genes have been predicted for the advanced and early stages of DN. In advanced DN, four diseases, i.e., glomerulonephritis, Ehlers-Danlos syndrome, kidney failure, and collagen disease, and their three key genes, i.e., COL1A1, COL5A1, and ITGB3, have been observed. COL1A1 is known as a potential biomarker of type 2 diabetes (T2D), COL5A1 has the capability to diagnose T2D, and ITGB3 is known as a predictor of T2D. In early DN, two diseases, i.e., amino acid metabolic disorder and proteinuria disease, and their three key genes, i.e., MAOA, MAOB, and DDC, have been examined. MAOA and MAOB expressions are required for insulin secretion, and their inhibition causes T2D. We didn't find any relation of DDC with T2D; based on our analyses, it might be associated with T2D. The findings of this study might be helpful to predict, diagnose, and make treatment plans for diabetes-related kidney diseases and their complexities.

Keywords: Diabetic Nephropathy, Comorbidities, RNA-Seq Gene Expression, Highly Expressed Sample Specific Gene, Gene Enrichment Analysis

Received: 24-03-2025 | **Revised:** 19-09-2025 | **Accepted:** 08-12-2025 | **DOI:** 10.3844/ojbsci.2026.26.02.026

Introduction

Diabetes is one of the major health problems and becoming an epidemic globally [1]. According to the World Health Organization, over 2 million deaths have been reported due to diabetes and diabetes related kidney diseases in 2021. Generally, there are two types of chronic diabetic conditions - type 1 diabetes and type 2 diabetes (T2D). T2D is the most

common type of diabetes for people of age greater than 40. T2D is a complex condition where the pancreas doesn't produce enough insulin and the body becomes resistant to insulin's effects [2]. Deficiency of insulin disrupts the metabolism of lipids, carbohydrates, proteins and it leads to high blood sugar (hyperglycemia), which can cause damage to organs such as the kidneys, eyes, nerves, blood vessels and the immune system [3].

When diabetes affects the kidney, it destroys the kidney's normal function that makes diabetes more complicated which is known as diabetic nephropathy (DN). It is also known as diabetic kidney disease which increases morbidity and mortality in diabetic patients [4]. At the recent 78th World Health Assembly, chronic kidney disease is reported to be one of the most common diabetes-related complications and it is projected that about 30 to 40% of people living with diabetes will develop the disease (https://idf.org/news/who-adopts-first-global-resolution-on-kidney-health/, Accessed on 22 August 2025). Based on the Autoregressive Integrated Moving Average model, the global burden of diabetic nephropathy from 2022 to 2050 has been reported to continuously increase annually resulting in more pressure on the global health system in the future [5]. DN gradually damages the kidney's filtering system. End-stage renal failure and chronic kidney disease may occur due to DN [6]. In the case of T2D, DN is the leading cause of end-stage renal failure, although the rate of kidney function decline differs between people [7]. Early treatment can help prevent or slow down the condition and reduce the risk of complications.

Recently, several bioinformatics research studies have been conducted to identify the key genes regulating DN in association with different conditions and/or diseases. In one study, a microarray dataset of DN has been analyzed for better understanding of DN pathogenesis and finding key genes in disease progression by using weighted gene co-expression network analysis [8]. In another study, potential diagnostic genes have been identified for DN based on hypoxia and immune status by using computational algorithms [9]. Single-cell analysis has been used in another study to investigate the role of PLEKHA1 in oxidative phosphorylation, which is involved in DN and identified as a potential biological marker of DN [10]. One study also analyzed microarray data and identified PDK4 as key gene for DN using co-expression network analysis [11]. Another study identified two key genes, SAMD8 and CYP51A1, in DN based on lipid metabolism [12]. Bioinformatics analyses have been performed in another study and identified CSF1R, CXCL6, VCAM1, JUN, and IL1B as the key genes for common mechanisms between periodontitis and DN in aging [13]. Most of the above studies focused on identifying the key genes of DN in association with only one specific condition or disease. This is the first report to the best of our knowledge regarding the investigation of all the potential comorbidities of DN in different stages and detection of their associated key genes by analyzing RNA-seq gene expression data.

The aim of this study is to predict the comorbidities of DN and their associated genes detection by analyzing the publicly available RNA-seq gene expression datasets collected from diabetes affected kidney samples. The findings of this study may lead to early and better diagnoses of the diseases and their potential treatments.

Materials and Methods

The overall workflow of the analysis is depicted in Fig. 1.

Data Collection and Preprocessing

Publicly available gene expression datasets are collected from the National Center for Biotechnology Information Gene Expression Omnibus repository. The key words to perform the search on the repository are "Type-2 Diabetes", " Dengue", and "RNA-Seq". Two different disease types of RNA-Seq datasets were collected to make variations in the analyses steps to obtain unbiased results. The datasets collected were not all in the same format. We transformed the datasets into log2 scale to observe the data distribution, reduce variability, improve comparability, and enhance visualization. Then, quality checking like bi-modal distribution, box-plot analyses and principal component analyses have been performed to obtain only the good quality RNA-seq datasets. In total, there are 13 datasets and all of them are from human samples except only one from mouse sample. For the mouse dataset we converted the mouse gene symbols to human gene symbols by using the biomaRt R package. The gene symbols of the datasets were in different formats, and hence we transformed the gene symbols of all datasets in ensembl gene id by using the biomaRt R package. We also combined the biological replicate samples by taking averages of them where available in the datasets. After that all the samples of 13 datasets are combined to make a final dataset for our analysis. The final dataset is in a matrix format where the rows represent the genes, and the columns represent the samples names. The details of the datasets are given in Table 1.

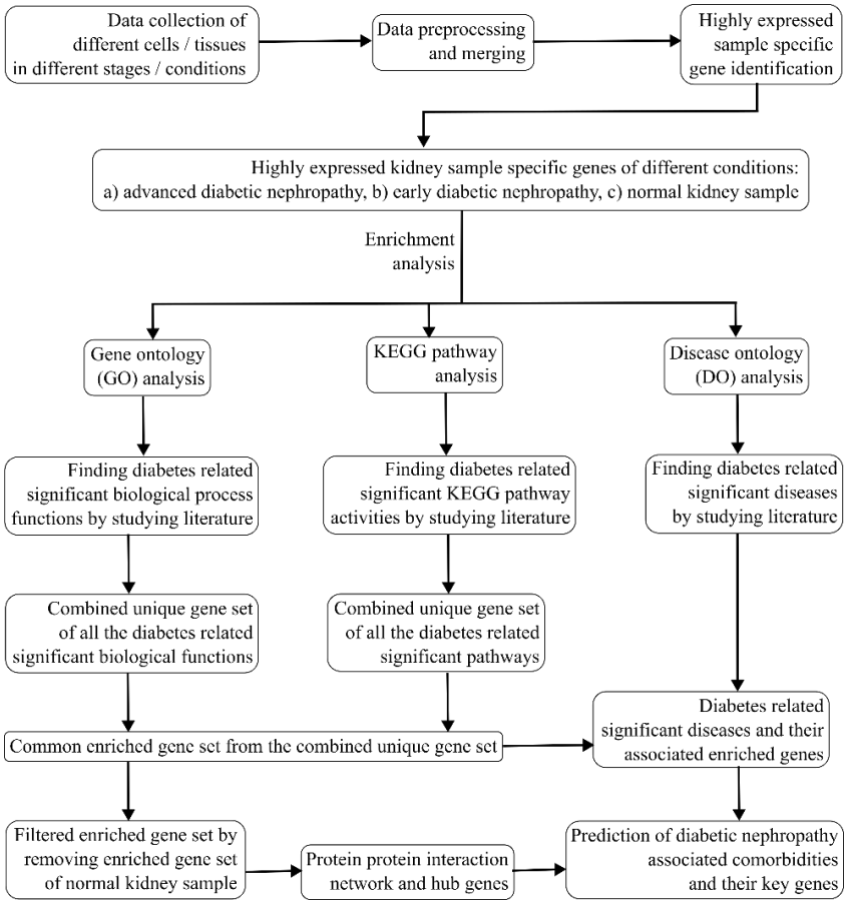


Fig. 1: Overall workflow of the analysis

Table 1: All the collected RNA-Seq datasets analyzed in the study

GSE ID	# of samples	# of genes	Cell type
GSE142025 [14]	36	25727	Kidney
GSE193978 [15]	276	28966	Whole blood cell
GSE154881 [16]	15	60671	Blood cell
GSE153792 [17]	34	58302	Venous blood
GSE240206 [18]	24	55536	Distal part of small intestine/Ileum
GSE137317 [19]	74	58302	Venous blood
GSE215835 [20]	11	60708	PBMC
GSE171487 [21]	16	39376	Plasma blast, Naive B cell
GSE132367 [22]	57	39376	PBMC
GSE178240 [23]	412	39376	PBMC, CD3, CD4, CD4CD8, CD8
GSE155672 [24]	12	39376	Innate lymphoid
GSE176079[25]	26	39376	Classical monocyte, Intermediate monocyte, Nonclassical monocyte
GSE212034 [26]	17	39376	HLADR+CD38+CD69+IFNg+CD8Tcells, HLADR+CD38+CD69+IFNg-CD8Tcells, HLADR+CD38+CD69-IFNg-CD8Tcells, UnstimulatedHLADR+CD38+CD8Tcells, HLADR-CD38-CD8Tcells

Highly Expressed Sample Specific Genes Identification

Previously compendium-based highly expressed sample specific genes have been applied for cell-type identification [27]. Motivated by the study we have also applied the methodology of identifying highly expressed sample specific genes here. First a specific number of highly expressed genes for all samples are detected from the merged dataset. Then housekeeping genes are identified from these highly expressed genes which are expressed in almost all samples of the dataset. Finally, the highly expressed sample specific genes are identified by discarding the housekeeping genes from the samples' highly expressed genes.

Obtaining kidney Sample Specific Genes

Diabetes affected kidney biopsy samples are selected for our analysis here. There are three DN groups of kidney samples in our collected datasets: a) advanced stage DN, b) early stage DN and c) normal samples. There are 21 samples in advanced stage, 6 samples in early stage, and 9 samples in normal group. For each group, highly expressed sample specific genes are collected for all samples and then combined to make the group specific unique set of genes.

Gene Ontology (GO) Analysis

GO provides functional information on genes. GO analysis is performed separately for the specific set of genes collected from the three groups of kidney samples by using clusterProfiler, AnnotationDbi, and org.Hs.eg.db R packages. Statistically significant top ten GO biological process terms have been collected for each case. The clusterProfiler R package is used to visualize enrichment results. From these top ten GO terms, diabetes related terms have been identified by studying literature. Then the combined unique set of genes of the diabetes related GO terms for each group are collected.

Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Analysis

KEGG is a collection of databases which deals with genomes, biological pathways, diseases, drugs, and chemical substances. By using clusterProfiler and org.Hs.eg.db R packages, KEGG pathway analysis is performed on the three groups of gene sets. Like GO analyses, statistically significant top ten KEGG terms have been collected for each group. Diabetes associated terms have been recognized from these top ten KEGG terms by studying literature and then their combined set of unique genes are collected.

Disease Ontology (DO) Analysis

DO has been developed for human disease to provide consistent, reusable and sustainable descriptions of human disease terms, phenotype characteristics and related medical vocabulary disease concepts to the biomedical community. By using clusterProfiler, org.Hs.eg.db and DOSE R packages, DO analysis is performed on the gene sets of advanced and early stages of DN samples. In this case, the top five significant DO terms, i.e., diseases are selected and then identified the diabetes related diseases by searching literature. Finally, the gene set of each of the diabetes related diseases for a stage are predicted by matching every gene of a disease term to the combined set of genes collected from that stage's GO and KEGG analyses.

Protein-Protein Interaction (PPI) Network Construction

PPIs are physical connection between two or more proteins which represent complex biological functions of cells and disease mechanisms. To construct the PPI networks, filtered set of specific genes for advanced and early stages of DN have been identified by deducting the combined set of genes of the normal group's diabetes related to GO and KEGG terms from the combined set of genes of the diabetes related GO and KEGG terms of advanced and early stages of DN respectively. The STRING (https://string-db.org/) database is used here to construct the PPI networks for the specific set of genes for advanced and early stages of DN. The constructed PPI networks are visualized by Cytoscape software. Then hub genes are identified by using the cyto-hubba plug-in of Cytoscape software. Hub genes mean those which have higher connectivity in PPI network.

Results

Highly Expressed Sample Specific Genes Collection for Each DN kidney Sample Group

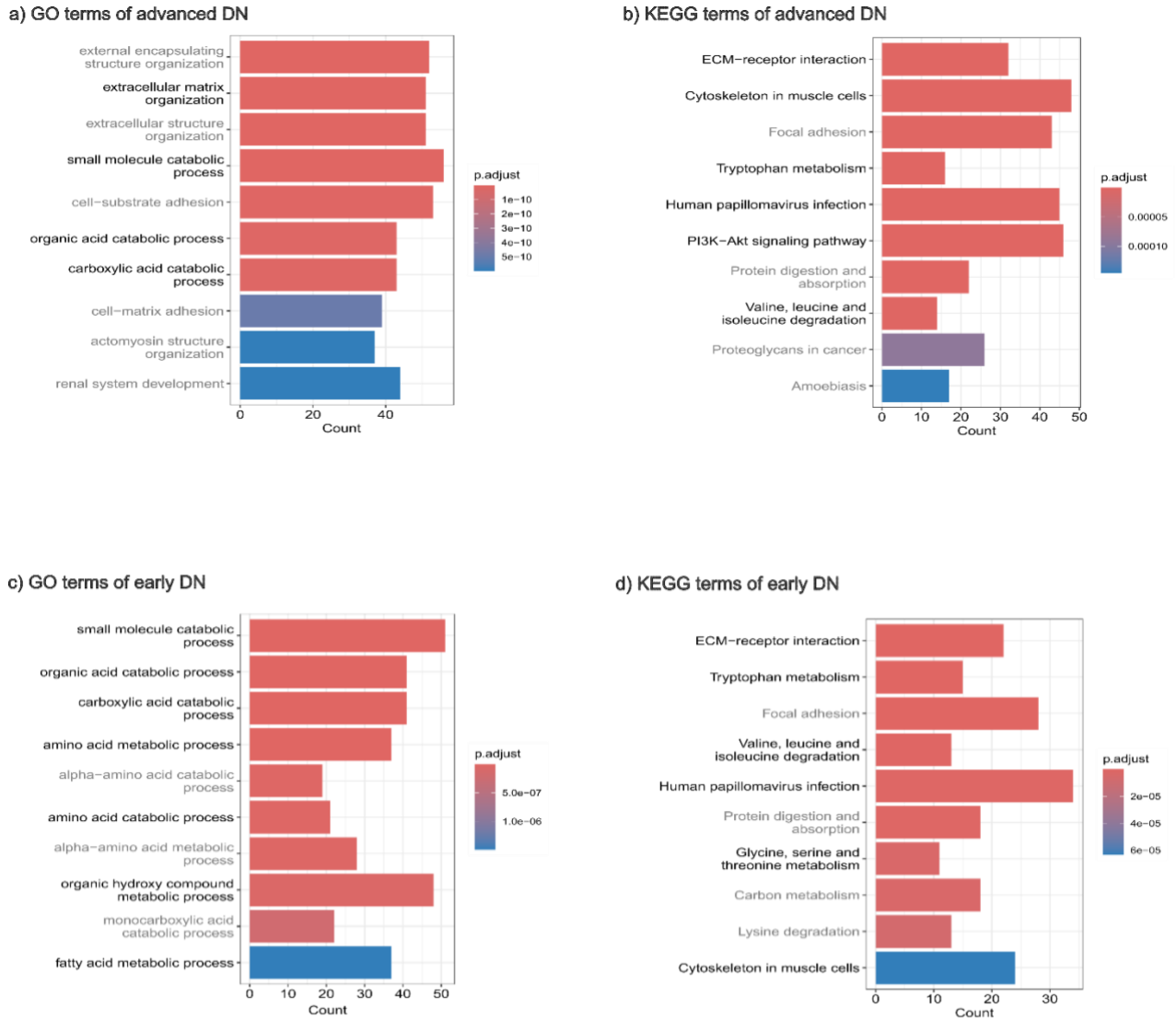
There are 998 samples and 14331 genes in the merged matrix. To identify the highly expressed sample specific genes, firstly top 1000 highly expressed genes for each sample are selected. Then from the highly expressed genes, housekeeping genes (i.e., the genes that are expressed in almost all the samples) are identified by their expression in more than 5% of the total number of samples. Finally, by excluding these housekeeping genes from the top 1000 highly expressed genes of each

DOI: 10.3844/ojbsci.2026.26.02.026

sample, highly expressed sample specific genes are identified. For each DN kidney sample group, all the highly expressed sample specific genes are collected and then combined to make a list of group specific genes. There are 715, 503 and 547 genes in the advanced DN, early DN and normal kidney sample groups respectively.

GO, KEGG and DO Analyses Results

GO and KEGG pathway analyses are performed on the gene sets of advanced DN, early DN and normal kidney samples groups whereas DO analysis is made on for gene sets collected only from advanced and early DN groups. Top 10 GO and KEGG terms whereas top 5 DO terms are selected based on the adjusted p-value for our analyses. From the selected terms which are related to diabetes, they are identified. The bar plot of the GO, KEGG and DO analyses are shown in Fig. 2. The diabetes related known terms are denoted with normal color text, while un-known terms are denoted with light color text in the figure.



DOI: 10.3844/ojbsci.2026.26.02.026

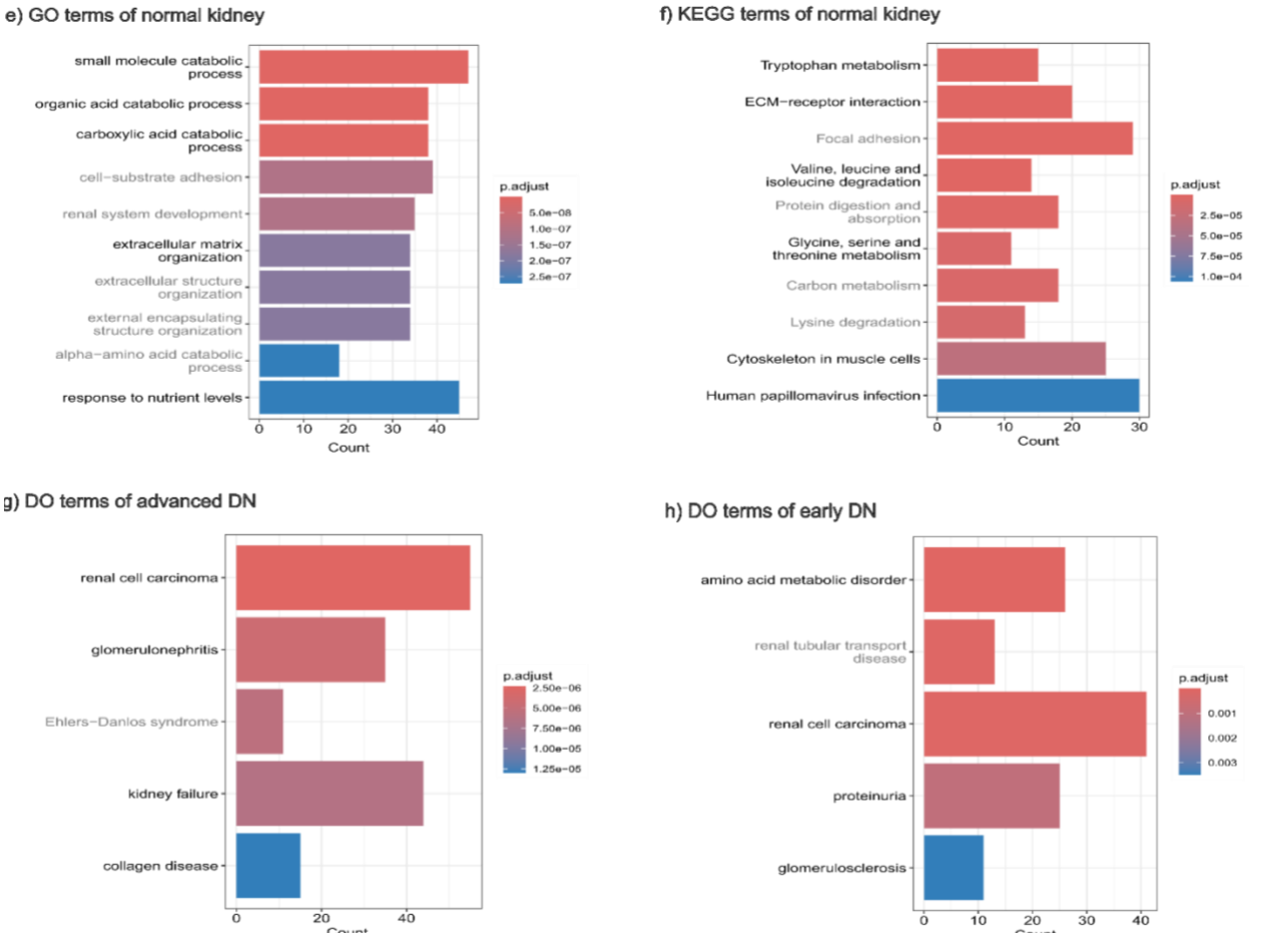


Fig. 2: Top 10 GO and KEGG terms of advanced DN, early DN, and control sample groups; Top 5 DO terms of advanced DN and early DN groups. Normal color texts denote diabetes-associated known terms while light color texts denote unknown terms. The known GO terms are - extracellular matrix organization [28], small molecule catabolic process [29], organic acid catabolic process [29], carboxylic acid catabolic process [30], amino acid metabolic process [31], amino acid catabolic process [29], organic hydroxy compound metabolic process [32], fatty acid metabolic process [33], response to nutrient levels [34]. The known KEGG terms are - ECM-receptor interaction [28], cytoskeleton in muscle cells [35], tryptophan metabolism [36], human papillomavirus infection [37], PI3K-Akt signaling pathway [38], valine, leucine and isoleucine degradation [39], glycine, serine and threonine metabolism [39]. The known DO terms are - renal cell carcinoma [40], glomerulonephritis [41], kidney failure [42], collagen disease [43], amino acid metabolic disorder [39], proteinuria [44], glomerulosclerosis [45].

In the case of advanced DN group, the number of collected genes for related GO terms and KEGG terms are 107 and 106 respectively. The number of common genes between them is 35. From DO analyses of this group, four diabetes related known diseases have been observed among the top five significant diseases. The significant diseases are renal cell carcinoma, glomerulonephritis, Ehlers-Danlos syndrome, kidney failure and collagen diseases, and their associated number of genes are 55, 35, 11, 44 and 15 respectively. When compared these diseases genes with the 35 common genes obtained from GO and KEGG analyses, there are 1, 5, 5, 7 and 8 genes matched with the diseases respectively.

Similarly, for the early DN group, the number of collected genes for related GO terms and KEGG terms are 104 and 76 respectively. The number of common genes between them is 24. From DO analyses, four diabetes related known diseases have been observed among the top five significant diseases. The significant diseases are amino acid metabolic disorder, renal tubular transport disease, renal cell carcinoma, proteinuria, and glomerulosclerosis, and their associated number of genes are 26, 13, 41, 25 and 11 respectively. When compared these diseases genes with the 24 common genes obtained from GO and KEGG analyses, there are 9, 2, 0, 2, and 0 genes matched with the diseases respectively. The diseases and their related genes are shown in Table 2. Finally, diabetes associations of these disease genes are identified by studying literature. The genes related to diabetes are marked by making the text bold in the table.

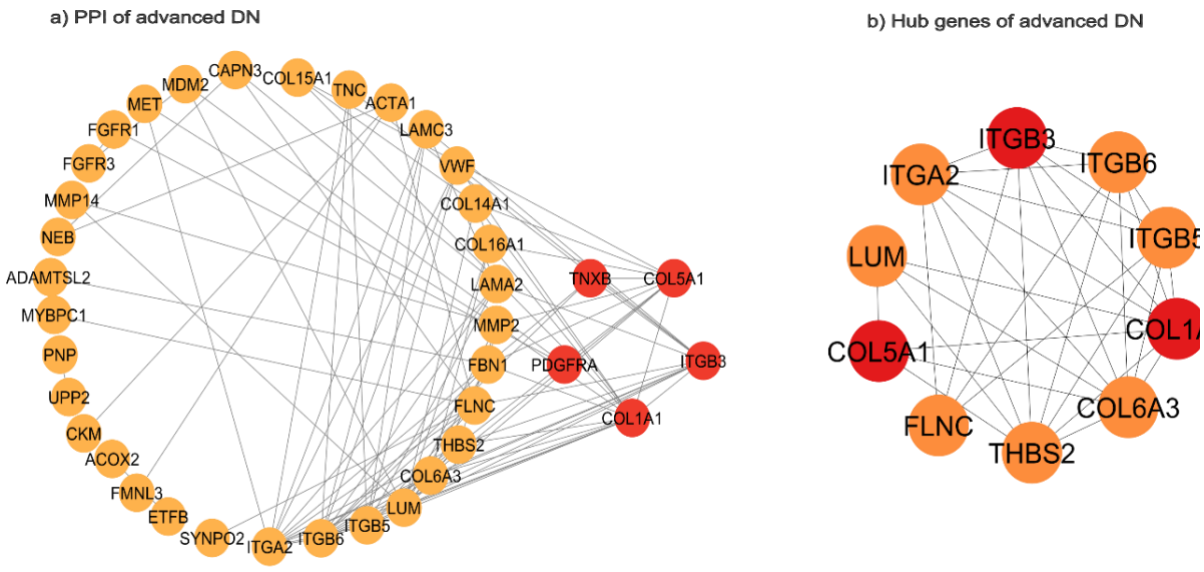
Table 2: Top five diseases and their associated genes

	Disease	Related genes
Advanced diabetic nephropathy	Renal cell carcinoma	PDGFRA
	Glomerulonephritis	COL4A4, **COL1A1, COL3A1, COL4A3, ITGB3**
	Ehlers-Danlos syndrome	**COL1A1, COL1A2,** TNXB**, COL3A1, COL5A1**
	Kidney failure	COL4A4, **COL1A1, COL3A1, COL4A3, ITGB3,** ACAT1, ACMSD
	Collagen Disease	COL4A4, **COL1A1, COL1A2,** TNXB, **COL3A1, COL4A3, COL4A1, COL5A1**
	Amino acid metabolic disorder	ACAT1, **SARDH**, IVD, **GLDC**, MCCC2, PCCA, **PSAT1, MAOB**, **MAOA**
	Renal tubular transport disease	**EHHADH**, ATP6V1B1
Early diabetic nephropathy	Renal cell carcinoma	N/A
	Proteinuria	ACAT1, DDC
	Glomerulosclerosis	N/A

Note: Light color in disease name denotes the disease has no known relation with diabetes; Bold text in gene names denote to have already known associations with diabetes

Generation of PPI Networks and their Hub Genes

Two PPI networks are generated using two sets of filtered specific genes for advanced and early stages of DN as shown in Fig. 3. Only the PPIs that have a weighting threshold value of more than 0.7 are considered for analyses here. For advanced DN, the PPI network is constructed using 63 genes, among them 38 genes are visualized in the PPI network for their above weighting threshold value as shown in Fig. 3(a). Among the 12 disease genes predicted in the previous section for this stage, 5 genes i.e., PDGFRA, ITGB3, COL1A1, TNXB and COL5A1, are found common in the filtered PPI network and are separately displayed in Fig. 3(a). From the PPI network, the top 10 hub genes are identified based on degree of connectivity as shown in Fig. 3(b). Among the hub genes, 3 disease related genes, COL5A1, COL1A1 and ITGB3, are found common and marked in red color in Fig. 3(b). In the case of Early DN stage, 56 genes are used for creating the PPI network and only 25 genes show their connectivity for their PPI weighting threshold value as shown in Fig. 3(c). Among them 4 genes, DDC, PAST1, MAOA and MAOB, are found common in the predicted list of 12 disease genes and are separately displayed in Fig. 3(c). The top 10 hub genes are identified as shown in Fig. 3(d) and among them 3 genes, DDC, MAOA and MAOB, are disease related genes and are marked in red color in Fig. 3(d).



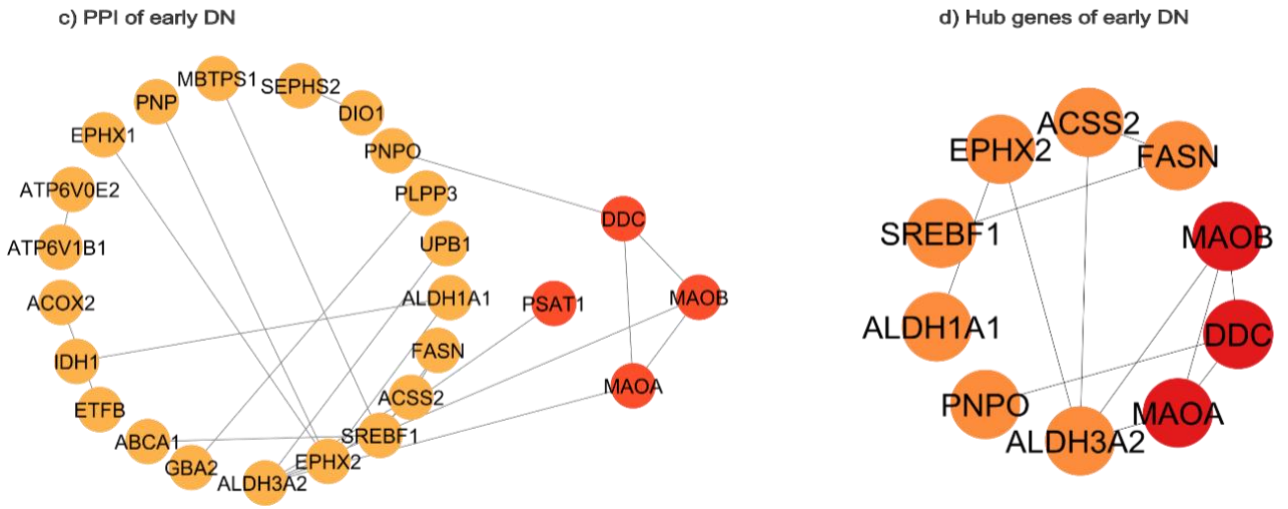


Fig. 3: PPI network and hub genes of advanced and early stages of DN. The disease related genes are marked in red color

Prediction of DN Comorbidities and their Associated Genes

From the PPI analysis, we have found three hub genes in both stages of DN that have several diseases associated as shown in Fig. 4. We did not find any common genes among the hub genes in advanced and early stages of DN. In the case of advanced stage of DN, the three genes are associated with four diseases. COL1A1 is related to all four diseases, COL5A1 is related to two diseases, and ITGB3 is related to two diseases. On the other hand, for early stages of DN, the three genes are associated with only two diseases. Both MAOA and MAOB are related to one disease, and DDC are related to another disease. Like the hub genes, there are also no common diseases associated with hub genes in the advanced and early stages of DN.

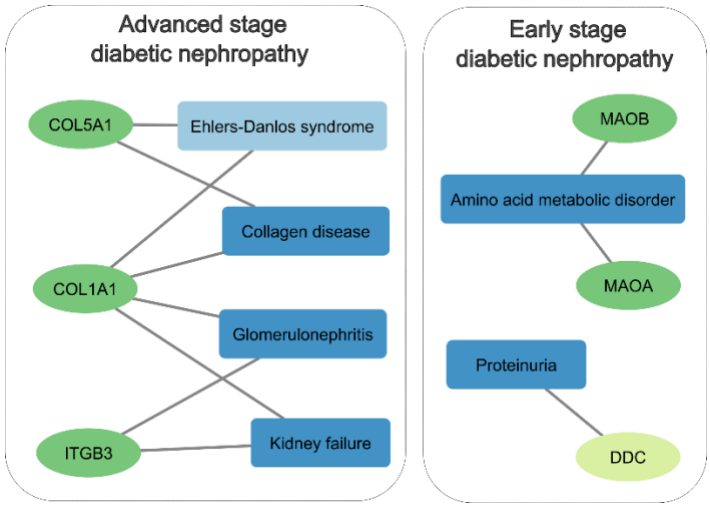


Fig. 4: Comorbidities of advanced and early stages of DN and their associated genes. Dark color denotes known associations with diabetes, while light color denotes unknown associations with diabetes

Discussion

The purpose of this study is to identify the comorbidities for advanced and early stages of DN and their associated genes by analyzing RNA-seq gene expression datasets of diabetes affected kidney samples. First, we have identified highly expressed sample specific genes. Then identified those genes among the highly expressed sample specific genes that are both enriched for diabetes related biological processes and pathway activities. We have also performed disease analyses with the highly expressed sample specific genes. We have identified diabetes associated diseases and their related genes

by matching the disease genes with the diabetes enriched genes. In this position, we predict that diabetes associated diseases and their related genes, both might also have DN associations as the genes are highly expressed specific set of DN genes. We have further filtered the disease genes and their associated diseases from our prediction by establishing the PPI network and the hub genes. Based on the hub genes in the PPI network, we have made our final prediction for disease genes and their associated diseases.

For advanced stage DN, renal cell carcinoma, glomerulonephritis, Ehlers-Danlos syndrome, kidney failure and collagen diseases are identified as comorbidities for diabetes affected kidney sample. In a study of 473 cases with renal cell carcinoma, 120 patients (25.4%) are identified with a history of diabetes [40]. The related identified gene for renal cell carcinoma is PDGFRA, although one study shows that PDGFRA is not significantly different between DN and normal kidneys of human [46]. Diabetes patients are ten times more likely to develop end-stage kidney failure [47]. The associated identified genes for kidney failure are COL4A4, COL1A1, COL3A1, COL4A3, ITGB3, ACAT1 and ACMSD. Among these identified genes four genes, COL1A1, COL3A1, COL4A3 and ITGB3, are already known for diabetes. COL1A1 is a potential biomarker of T2D [48]. Type 2 diabetes has higher expression of COL3A1 [49]. COL4A3 is associated with T2D [50]. ITGB3 is a predictor of the disease at concomitant development of coronary artery disease and T2D [51]. So, the remaining genes, COL4A4, ACAT1 and ACMSD, may also be associated with kidney failure because of diabetes. DN can lead to collagen disease [43], and the identified related genes of the disease are COL4A4, COL1A1, COL1A2, TNXB, COL3A1, COL4A3, COL4A1 and COL5A1. Among these genes, COL1A1, COL1A2, COL3A1, COL4A3, COL4A1 and COL5A1 are known to be associated with diabetes. COL1A2 is a potential gene for diabetes management [52]. COL4A1 is associated with T2D [53]. COL5A1 has the capability to diagnose T2D [54]. The remaining genes, COL4A4 and TNXB, might also be associated with collagen disease due to diabetes. One study reported a case of DN with features of autoimmune-related glomerulonephritis, suggesting a possible connection between the two conditions [41]. The associated identified genes for glomerulonephritis are COL4A4, COL1A1, COL3A1, COL4A3 and ITGB3. Among these identified genes four genes, COL1A1, COL3A1, COL4A3 and ITGB3, are already known for diabetes. So, COL4A4 might also be associated with glomerulonephritis for diabetes. While Ehlers-Danlos syndrome is predicted as a comorbidity of advanced DN in our analyses, there is no established connection between the diseases, and more research is needed to fully understand their relationship.

For early stage of DN, amino acid metabolic disorder, renal tubular transport disease, renal cell carcinoma, proteinuria and glomerulosclerosis are identified as comorbidities of DN. Impaired amino acid metabolism is associated with diabetic kidney disease [39]. The related genes for amino acid metabolic disorder are ACAT1, SARDH, IVD, GLDC, MCCC2, PCCA, PSAT1, MAOB and MAOA. Known diabetes-related genes for amino acid metabolic disease are SARDH, GLDC, PSAT1, MAOA and MAOB. SARDH is a related gene with diabetes [55]. GLDC may affect the development of diabetes [56]. PSAT1 can be used in treating T2D [57]. MAOA and MAOB inhibition may contribute in T2D [58]. The remaining genes, ACAT1, IVD, MCCC2 and PCCA, might also be associated with amino acid metabolic disorder caused by diabetes. For renal cell carcinoma, we did not find any related enrichment genes that are associated with early stage of DN. Proteinuria is a significant factor in DN [44]. The identified associated genes for this disease are ACAT1 and DDC; both genes are currently unknown to have any association with diabetes. Based on our analysis, both genes might be associated with proteinuria disease due to diabetes. Although glomerulosclerosis disease is predicted as a comorbidity of early DN, we did not find any associated enrichment gene of the disease in our analyses. Although we did not find any known association of renal tubular transport disease with diabetes, we identified two enriched genes associated with the disease. EHHADH has a known association with the development and severity of DN in T2D patients [59], while ATP6V1B1 has no known association with diabetes.

On the other hand, when the PPI networks are generated for filtered specific set of genes of advanced and early stages of DN, we have also found some interesting results predicted in our above analyses. In PPI network for advanced stage of DN, 5 genes, PDGFRA, ITGB3, COL1A1, TNXB and COL5A1, among 12 predicted genes in previous section are observed and among these 5 genes 3 genes, COL5A1, COL1A1 and ITGB3, are also observed in hub genes. These 3 genes have already known associations for diabetes. For early stage of DN, 4 genes, DDC, PAST1, MAOA, MAOB, among 12 predicted genes are observed and in the hub genes 3 genes, DDC, MAOA, MAOB, out of 4 genes are also seen. In these 3 genes, MAOA and MAOB are already known gene for diabetes association. We did not find any association of the remaining gene, DDC, with diabetes.

In this study, we emphasized predicting the comorbidities with their key genes of early and advanced stages of DN based on the highly expressed sample specific genes. To do this, first we have considered 1000 highly expressed genes for all samples. Then, from these highly expressed genes we have identified sample specific genes by considering a gene is sample specific if it exists only at less than 5% of total number of samples' highly expressed genes. Then, we used the highly

expressed sample specific genes for each DN stage for GO, KEGG and DO analyses. For GO and KEGG analyses, we have considered top 10 terms based on their adjusted p-values, whereas for DO analysis we have taken only top 5 DO terms based on their adjusted p-values. Based on our analysis procedure we did not observe common key genes as well as the predicted comorbidities of DN stages. This might be a limitation of our findings for disease progression mechanisms from early to advanced stages of DN. To overcome this limitation, we could simply consider more highly expressed genes rather than 1000 and the threshold for specific genes to a greater value than 5% of total number of samples to capture common highly expressed genes between the DN stages. In this case, the specificity of the genes might be less specific than the present analysis results. Also, we could identify the differentially expressed genes among all the expressed genes of the DN stages to capture the common genes of DN stages. In this case, the common genes might be different (i.e., upregulated / downregulated) in different DN stages (i.e., the common genes might be upregulated in one stage and at the same time downregulated in another stage).

The main limitation of our study is that we did not test any of the predicted genes associated with the comorbidities of DN in in vitro or in vivo experimental models. Also, we have only advanced and early stages of DN associated RNA-Seq gene expression data from kidney samples. We could include more variety in gene expression datasets collected from different organs and in different time periods to strength our findings as well as to provide a more holistic view of DN. However, our findings in this study might be important resources for biologists and medical researchers for further studies to develop pathogenesis and targeted therapy of diabetes associated with kidney diseases. In future studies, we hope that we will be able to include all the expressed genes for all the samples in our analysis to capture all the potential key genes associated with the comorbidities related to different DN stages including T2D. Also, we have plans to conduct laboratory experiments with the help of our collaborators to validate our findings for their clinical implications.

Conclusion

Diabetes increases the risk of kidney related diseases significantly through the development of DN. The progression of kidney damage for diabetes contributes to morbidity and mortality. So, it is necessary to focus on prevention, early detection and effective management of kidney diseases in diabetic patients. The findings of our study may contribute to early diagnoses and improved management of DN associated kidney diseases that might ultimately benefit individuals affected by this prevalent metabolic disorder.

Acknowledgment

The authors thank Most. Afia Akhtar for reviewing the manuscript and her necessary suggestions.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author’s Contributions

Chaity Paul: Data collection, data curation, data analysis, methodology, validation, visualization, original draft preparation, review and editing.
Md. Ashiq Uddin: Data collection, data curation, validation, review and editing.
Rupa Dey: Data collection, data curation, validation, review and editing.
Adrita Alam: Data collection, data curation, validation, review and editing.
Md. Humayun Kabir: Conceptualization, Methodology, validation, original draft preparation, review and editing, supervision, project administration.

Ethics

The authors declare that there are no ethical issues regarding the research study.

References

1. Tabish SA. Is diabetes becoming the biggest epidemic of the twenty-first century? Int J Health Sci (Qassim). 2007;1(2):V-VIII.
2. Roden M, Shulman GI. The integrative biology of type 2 diabetes. Nature. 2019;576(7785):51-60. doi:10.1038/s41586-019-1797-8
3. Bascones-Martinez A, Criado-Camara E, Bascones-Ilundain C, Arias S, Bascones-Ilundain J. Periodontal disease and diabetes. In: Singh VK, editor. Diabetes. New York: Springer; 2011. p. 76-87. doi:10.1007/978-1-4614-5441-0_9
4. Valencia WM, Florez H. How to prevent the microvascular complications of type 2 diabetes beyond glucose control. BMJ. 2017;356:i6505. doi:10.1136/bmj.i6505
5. Ma X, Liu R, Xi X, Zhuo H, Gu Y. Global burden of chronic kidney disease due to diabetes mellitus, 1990-2021, and projections to 2050. Front Endocrinol (Lausanne). 2025;16:1513008. doi:10.3389/fendo.2025.1513008
6. Lim A. Diabetic nephropathy-complications and treatment. Int J Nephrol Renovasc Dis. 2014;7:361-381.
7. Rossing K, Christensen PK, Hovind P, Tarnow L, Rossing P, Parving HH. Progression of nephropathy in type 2 diabetic patients. Kidney Int. 2004;66(4):1596-1605. doi:10.1111/j.1523-1755.2004.00925.x
8. Gholaminejad A, Fathalipour M, Roointan A. Comprehensive analysis of diabetic nephropathy expression profile based on weighted gene co-expression network analysis. BMC Nephrol. 2021;22:245. doi:10.1186/s12882-021-02447-2
9. Li C, Su F, Zhang L, Liu F, Fan W, Li Z, et al. Identifying potential diagnostic genes for diabetic nephropathy based on hypoxia and immune status. J Inflamm Res. 2021;14:6871-6891. doi:10.2147/JIR.S341032
10. Zhang C, Li H, Wang S. Common gene signatures and molecular mechanisms of diabetic nephropathy and metabolic syndrome. Front Public Health. 2023;11:1150122. doi:10.3389/fpubh.2023.1150122
11. Han Y, Jin L, Wang L, Wei L, Tu C. Identification of PDK4 as a hub gene for diabetic nephropathy using co-expression network analysis. Kidney Blood Press Res. 2023;48(1):522-534. doi:10.1159/000531288
12. Yang M, Wang J, Meng H, Xu J, Xie Y, Kong W. Identification of key genes in diabetic nephropathy based on lipid metabolism. Exp Ther Med. 2024;28(5):406. doi:10.3892/etm.2024.12695
13. Yan P, Ke B, Fang X. Bioinformatics reveals the relationship between diabetic nephropathy and periodontitis in aging. Heliyon. 2024;10(2):e24872. doi:10.1016/j.heliyon.2024.e24872
14. Fan Y, Yi Z, D'Agati VD, Sun Z, Zhong F, Zhang W, et al. Comparison of kidney transcriptomic profiles of early and advanced diabetic nephropathy. Diabetes. 2019;68(12):2301-2314. doi:10.2337/db19-0204
15. Eckold C, van Doorn CL, Ruslami R, Ronacher K, Riza AL, van Veen S, et al. Impaired resolution of blood transcriptomes in tuberculosis with diabetes. Clin Transl Med. 2023;13(9):e1375. doi:10.1002/ctm2.1375
16. Sun QS, Guo MG, Sha QS, Chen X. Comparison of blood transcriptomic profiles of healthy, T2D and DN. Gene Expression Omnibus (GEO); 2024. (Provide accession number).
17. Briviba M, Ansone L, Silamikelis I, Rovite V, Elbere I, Silamikele L, et al. Whole-blood transcriptome profiling reveals signatures of metformin response. PLoS One. 2020;15(8):e0237400. doi:10.1371/journal.pone.0237400
18. Brīvība M, Silamiķele L, Kalniņa I, Silamiķelis I, Birzniece L, Ansone L, et al. Metformin targets intestinal immune signaling pathways. Front Endocrinol (Lausanne). 2023;14:1232143. doi:10.3389/fendo.2023.1232143
19. Briviba M, Silamikelis I, Kalnina I, Ansone L, Rovite V, Elbere I, et al. Metformin affects transcriptome of peripheral blood cells. PLoS One. 2019;14(11):e0224835. doi:10.1371/journal.pone.0224835
20. Ioannidis LJ, Studniberg SI, Eriksson EM, Suwarto S, Denis D, Liao Y, et al. Impaired effector T cell memory in dengue progression. J Biomed Sci. 2023;30:24. doi:10.1186/s12929-023-00916-4
21. Aggarwal C, Saini K, Reddy ES, Singla M, Nayak K, Chawla YM, et al. Immunophenotyping of plasmablasts in dengue. J Virol. 2021;95(23):e00610-21. doi:10.1128/JVI.00610-21
22. Tian Y, Seumois G, De-Oliveira-Pinto LM, Mateus J, Herrera-de la Mata S, Kim C, et al. Molecular signatures of dengue virus-specific CD4 T cells. Cell Rep. 2019;29(13):4482-4495.e4. doi:10.1016/j.celrep.2019.11.098
23. Poonpanichakul T, Chan-In W, Opasawatchai A, Loison F, Matangkasombut O, Charoensawan V, et al. Innate lymphoid cells in dengue infection. Front Immunol. 2021;12:599805. doi:10.3389/fimmu.2021.599805
24. Maheshwari D, Saini K, Singh P, Singla M, Nayak K, Aggarwal C, et al. Monocyte subsets in dengue pathophysiology. iScience. 2022;25:104384. doi:10.1016/j.isci.2022.104384
25. de Silva Antunes R, Wang H, Tian Y, Tippalagama R, Alahakoon SU, Premawansa G, et al. CD4+CD8+ T cells in dengue infection. Viruses. 2022;14(1):90. doi:10.3390/v14010090
26. Singh P, Bajpai P, Maheshwari D, Chawla YM, Saini K, Reddy ES, et al. Transcriptional heterogeneity of CD8 T cells in dengue. J Virol. 2023;97(11):e00746-23. doi:10.1128/jvi.00746-23
27. Kabir MH, Djordjevic D, O'Connor MD, Ho JWK. C3: R package for cell-type identification. Comput Biol Chem. 2018;77:187-192. doi:10.1016/j.compbiolchem.2018.10.003
28. Kolset SO, Reinholt FP, Jenssen T. Diabetic nephropathy and extracellular matrix. J Histochem Cytochem. 2012;60(12):976-986. doi:10.1369/0022155412465073
29. Zhou LT, Zhang ZJ, Cao JY, Chen H, Zhu YS, Wu X, et al. The molecular mechanism of diabetic nephropathy: a bioinformatics analysis of microarray datasets. Clin Kidney J. 2021;14(3):855-866. doi:10.1093/ckj/sfaa190
30. Collins KS, Eadon MT, Cheng YH, Barwinska D, Melo Ferreira R, McCarthy TW, et al. Alterations in protein translation and carboxylic acid catabolic processes in diabetic kidney disease. Cells. 2022;11(7):1166. doi:10.3390/cells11071166
31. Liu L, Xu J, Zhang Z, Ren D, Wu Y, Wang D, et al. Metabolic homeostasis of amino acids and diabetic kidney disease. Nutrients. 2023;15(1):184. doi:10.3390/nu15010184
32. Zhao J, He K, Du H, Wei G, Wen Y, Wang J, et al. Bioinformatics prediction and experimental verification of biomarkers for diabetic kidney disease. PeerJ. 2022;10:e13932. doi:10.7717/peerj.13932

33. Liu J, Liu S, Yu Z, Qiu X, Jiang R, Li W. Gene regulatory network of type 2 diabetes through multi-omic integration. J Transl Med. 2022;20 doi:10.1186/s12967-022-03826-5
34. Kume S, Koya D, Uzu T, Maegawa H. Role of nutrient-sensing signals in diabetic nephropathy. Biomed Res Int. 2014;2014:315494. doi:10.1155/2014/315494
35. Sanai T, Sobka T, Johnson TS, El-Essawy M, Muchaneta-Kubara EC, El Nahas AM. Expression of cytoskeletal proteins in diabetic nephropathy. Diabetologia. 2000;43(1):91-100. doi:10.1007/s001250050012
36. Debnath S, Velagapudi C, Redus L, Thameem F, Kasinath B, Hura CE, et al. Tryptophan metabolism in diabetic kidney disease. Int J Tryptophan Res. 2017;10:1178646917694600. doi:10.1177/1178646917694600
37. Atiase Y, Effah K, Wormenor CM, Tekpor E, Morkli AC, Boafo E, et al. Prevalence of high-risk human papillomavirus infection among women with diabetes. BMC Womens Health. 2024;24:260. doi:10.1186/s12905-024-03078-z
38. Wang H, Gao L, Zhao C, Fang F, Liu J, Wang Z, et al. Role of PI3K/Akt signaling pathway in chronic kidney disease. Int Urol Nephrol. 2024;56(8):2623-2633. doi:10.1007/s11255-024-03989-8
39. Zhu H, Bai M, Xie X, Wang J, Weng C, Dai H, et al. Impaired amino acid metabolism in diabetic kidney disease. Nutrients. 2022;14(16):3345. doi:10.3390/nu14163345
40. Habib SL, Prihoda TJ, Luna M, Werner SA. Diabetes and risk of renal cell carcinoma. J Cancer. 2012;3:42-48. doi:10.7150/jca.3718
41. Wu HS, Hsieh SC. Diabetic nephropathy simulating autoimmune glomerulonephritis: a case report. Nephrol Renal Ther. 2021. doi:10.24966/nrt-7313/100056
42. Hill NR, Fatoba ST, Oke JL, Hirst JA, O'Callaghan CA, Lasserson DS, et al. Global prevalence of chronic kidney disease. PLoS One. 2016;11(7):e0158765. doi:10.1371/journal.pone.0158765
43. Toro D, Khan R, Mohammed A. Diabetic nephropathy and type III collagen glomerulopathy. Cureus. 2023;15(8):e44284. doi:10.7759/cureus.44284
44. Campbell RC, Ruggenenti P, Remuzzi G. Proteinuria in diabetic nephropathy. Curr Diab Rep. 2003;3(6):497-504. doi:10.1007/s11892-003-0014-0
45. Qian Y, Feldman E, Pennathur S, Kretzler M, Brosius FC. Mechanisms of glomerulosclerosis in diabetic nephropathy. Diabetes. 2008;57(6):1439-1445. doi:10.2337/db08-0061
46. Uehara G, Suzuki D, Toyoda M, Umezono T, Sakai H. PDGF expression in diabetic nephropathy. Clin Exp Nephrol. 2004;8(1):36-42. doi:10.1007/s10157-003-0265-8
47. Natesan V, Kim SJ. Diabetic nephropathy: mechanisms and management. Biomol Ther (Seoul). 2021;29(4):365-372. doi:10.4062/biomolther.2020.204
48. Lin G, Wan X, Liu D, Wen Y, Yang C, Zhao C. COL1A1 as biomarker in type 2 diabetes. Pharmacol Res. 2021;165:105436. doi:10.1016/j.phrs.2021.105436
49. Blencowe M, Furterer A, Wang Q, Gao F, Rosenberger M, Pei L, et al. IAPP-induced beta cell stress in type 2 diabetes. Diabetologia. 2022;65(1):173-187. doi:10.1007/s00125-021-05569-2
50. Saravani S, Yari D, Saravani R, Azadi Ahmadabadi C. Association of COL4A3 polymorphisms with type 2 diabetes. Biomed Rep. 2017;6(3):329-334. doi:10.3892/br.2017.856
51. Muslimova EF, Rebrova TY, Afanasiev SA, Sergienko TN, Repin AN. ITGB3 gene association with coronary disease in diabetes. Diabetes Mellit. 2016;19(5) doi:10.14341/dm7875
52. Zhuang L, Jin G, Wang Q, Ge X, Pei X. Circular RNA COL1A2 in diabetic nephropathy. Appl Biochem Biotechnol. 2023;195(12):7652-7667. doi:10.1007/s12010-023-04501-1
53. Shukla N, Kumari S, Verma P, Kushwah AS, Banerjee M, Sankhwar SN, et al. Genotypic analysis of COL4A1 in diabetic nephropathy. DNA Cell Biol. 2023;42(9):541-547. doi:10.1089/dna.2023.0125
54. Liu Y, Li Y, Shen H, Li Y, Xu Y, Zhou M, et al. Serum fatty acids and diabetic nephropathy. Ther Adv Endocrinol Metab. 2022;13:20420188221118750. doi:10.1177/20420188221118750
55. Gong J, Du X, Li Z, Li X, Guo M, Lu J, et al. Differential gene expression in diabetes. PLoS One. 2018;13(2):e0191212. doi:10.1371/journal.pone.0191212
56. Jog R, Chen G, Wang J, Leff T. Regulation of glycine decarboxylase and oxidative stress. Physiol Rep. 2021;9(15):e14991. doi:10.14814/phy2.14991
57. Yu J, Xiao F, Guo Y, Deng J, Liu B, Zhang Q, et al. PSAT1 regulates insulin sensitivity. Diabetes. 2015;64(5):1591-1602. doi:10.2337/db14-1368
58. Ganic E, Johansson JK, Bennet H, Fex M, Artner I. Monoamine oxidase in type 2 diabetes. Biochem Biophys Res Commun. 2015;468(4):629-635. doi:10.1016/j.bbrc.2015.11.002
59. Kan S, Hou Q, Shi J, Zhang M, Xu F, Liu Z, et al. EHHADH deficiency in diabetic kidney disease. Cell Death Discov. 2024;10:289. doi:10.1038/s41420-024-02066-4