Identification of Hub Genes Associated With Clinical Characteristics and Potential Therapy of Diabetic Nephropathy by Bioinformatics Analysis.

Si Xu

Wuhan University Second Clinical Hospital: Wuhan University Zhongnan Hospital

Sha Wu

Wuhan University Zhongnan Hospital

Min Yang

Wuhan University Zhongnan Hospital

Xiaoning Li (✉ xiliusa2000@126.com)

Wuhan University Zhongnan Hospital <https://orcid.org/0000-0003-4596-798X>

Research

Keywords: Bioinformatics analysis, Diabetic nephropathy, differentially expressed genes, Hub genes, Diagnostic biomarkers, Therapy

Posted Date: December 31st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-136161/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: To provide molecular markers and potential targeted molecular therapy for diabetic nephropathy by screening hubgenes based on bioinformatic analysis.

Results: We found 91 differentially expressed genes (DEGs) between diabetic nephropathy tissues and normal kidney tissues. Majority DEGs were significantly enriched in the extracellular matrix structural constituent, collagen-containing extracellular matrix. KEGG pathway analysis showed that most of DEGs participated in PI3K-Akt signaling pathway, AGE-RAGE signaling pathway in diabetic complications. Five high relevant sub-networks and the top 16 genes according to 12 topological algorithms were screened out and also five co-expressed gene modules were identified by WGCNA. Eventually, 5 hub genes were identified by taking the intersection which might be involved in the progression of DN. And 11 microRNAs were associated with related genes in WebGestalt.

Conclusions: We identified five hub genes, namely COL1A2, COL6A3, COL15A1, CLU and LUM, and their related microRNAs(especially miR29 and miR196), which might be used as diagnostic biomarkers and therapeutic targets for diabetic nephropathy.

1. Background

Diabetic nephropathy (DN) was one of the most common microvascular complications of diabetes mellitus and also the main cause of end-stage renal disease (ESRD) [1, 2]. In terms of pathophysiology, diabetic nephropathy can be divided into glomerular lesion, tubules and interstitial atrophy and fibrosis. The clinical features of diabetic nephropathy were characterized by hyperglycemia, continuous decrease in GFR, accompanied by sustained increase in proteinuria and SCR [3]. Therefore, these four characteristics were usually used in clinical diagnosis, but there was no specific molecular diagnosis of diabetic nephropathy. Meanwhile, the treatment of diabetic nephropathy could only provide symptomatic supportive treatment which was limited for preventing DN progression such as blood glucose control and blood pressure lowering, etc [4, 5]. Thus, it was an urgent task to identify molecular biomarkers and explore underlying mechanisms involved in the process of DN, which could contribute to find new diagnostic markers and therapeutic approaches.

Genome-wide transcriptome analysis using expression arrays could be used to identify biomarkers for disease progression and to gain insights into the disease pathogenesis and molecular therapy [6]. With the wide application of genome transcriptome analysis, a large amount of data was stored in a public database. Reanalyzing these data could provide significant clues for other researches [6, 7]. In recent years, numerous differentially expressed genes (DEGs) had been found by microarray data analysis on DN; a few such genes were promising biomarkers in terms of DN diagnosis and prognosis[8]. Moreover, most previous studies focused on DEG screening which was limited to single microarray data, rather than the interactions among genes. Weighted gene co-expression network analysis (WGCNA) could be used to identify correlations among genes. Also, WGCNA could identify highly correlated gene modules to confirm

hubgenes [9]. In this study, we integrated five microarray datasets from the Gene Expression Omnibus database (GEO) database which compared DN with normal samples for analysis to reduce false positive rate. In addition, we also compared and intersected the DEGs with the hub genes screened by WGCNA, so that the common genes were more reliable as biomarkers or therapeutic targets.

MicroRNA(miRNA) was a kind of highly conservative single small non-coding RNA that could regulate the expression level and function of their target genes in the process of protein synthesis[10, 11]. In recent years, studies had found that the miRNA could participate in the development regulation, cell growth, differentiation and apoptosis, and oncogenesis[12]. They interacted with the complementary strand of mRNAs and lead to the degradation of the corresponding mRNAs; they also interfered with protein production by suppressing protein synthesis[13]. Generally, miRNAs were involved in a wide range of diseases, including neurological disease, heart disease, and cancer and so on[14]. Therefore, miRNAs could achieve a novel targeted molecular therapy by regulating the expression of targeted hub genes.

2. Result

2.1. Identification of DEGs and enrichment analysis of GO and KEGG

After pre-processing and normalizing, expression values of 7564 genes in 140 kidney samples were obtained from merged five microarray data sets. Overall, 91 DEGs were involved in DN as shown in the heatmap (Figure 1a), of which 49 were down-regulated and 42 were up-regulated (Figure 1b).

GO functional enrichment analysis revealed that most significant enrichment of molecular function (MF) included extracellular matrix structural constituent, SMAD binding, heparin binding; the enrichment of cell composition (CC) included collagen-containing extracellular matrix, Collagen Trimer and blood microparticle, while biological process (BP) enrichment included kidney development, renal system development and nephron development, etc(Figure 1c). KEGG pathway analysis showed the top 10 pathways were mainly enriched in PI3K-Akt signaling pathway, AGE-RAGE signaling pathway in diabetic complications, Protein digestion and absorption, Complement and coagulation cascades, Rap1 signaling pathway, Cytokine-cytokine receptor interaction and ECM-receptor interaction, and so on(Figure 1d).

2.2. PPI network of DEGs and screening of hub genes of DN with Cytoscape.

Search tool for the retrieval of interacting genes (STRING 11.0; <https://string-db.org>) was applied to establish a PPI network of DEGs. Interaction with a combined score >0.7 was set as the cut-off point (Figure 2a). Cytoscape software was used to visualize module analysis for PPI network of gene signatures and hub genes by MCODE. After MCODE detection, 30 genes were found to participate in 5 clusters (Figure 2 b-f). 16 genes were identified as hub genes, VEGFA, COL1A2, ALB, FOS, EGR1, EGF,

COL3A1, CLU, CTGF, SST, EGR2, LUM, FOSB, COL6A3, PROM1, COL15A1 through integration in 12 algorithms of Cytohubba (TABLE 1). Ultimately, these 16 genes were intersected with 30 genes in the modules of MCODE to obtain 14 key genes (Figure 3b), including VEGFA, COL1A2, ALB, FOS, EGR1, EGF, COL3A1, CLU, SST, EGR2, LUM, FOSB, COL6A3 and COL15A1. Meanwhile, the common hub genes between renal tubules and renal glomeruli were obtained in the Venn diagram: COL1A2, COL6A3, COL15A1, LUM, CLU, EGF, ALB, FOS, FOSB (Figure 4d1).

2.3. The gene modules and hub genes confirmed by WGCNA.

In the WGCNA, according to the results of the scale-free fit index and average connectivity, the soft-threshold was determined as 9 (Figure 3a). Based on hierarchical clustering, the genes were divided into 5 modules (Figure 3c). The yellow module was the most significant negative correlation between modules and normal samples ($r = -0.64$, $p = 3e-17$), and displayed the most significant positive correlation between modules and DN samples ($r = 0.64$, $p = 3e-17$) (Figure 3d). There are 323 genes in this module in DN and normal kidney. Eventually, the four methods intersected to get five hub genes (Figure 3b), which were COL1A2, COL6A3, COL15A1, CLU and LUM, respectively.

2.4. Relationship between key genes and targeted microRNAs of DN.

The top ten of key miRNAs that targeted the 14 key genes, were predicted: the miR29 family (miR29a/b/c), miR196a/b, miR199A, miR377, miR380-5P, miR191, miR186, miR1, miR206, miR-383 and miR-134. A total of eight groups were identified as $FDR < 0.05$ (TABLE 2). In particular, miR29a/b/c and miR196a/b, which were related to COL1A2/COL6A3/COL15A1, also had a high enrichment ratio.

2.5. Association between interested genes and clinical features of DN.

To verify potential roles of hub genes in DN, correlation analysis and subgroup analysis between hub genes and clinical features were performed using Nephroseq v5 online tool. First, the results showed that mRNA expression of VEGFA, ALB, EGR1, EGF, EGR2, FOS, FOSB, CTGF and SST positively correlated with GFR in tubulointerstitium of DN patients (Figure 4a-j), while mRNA expression of VEGFA, ALB and SST positively correlated with GFR in glomerulus of DN patients (Figure 4a1-c1), suggesting most of the down-regulated hub genes may delay or inhibit the development of DN. In addition, it also suggested that the effects of the genes in different sites may be opposite. Meanwhile, the mRNA expression of COL1A2, COL3A1, COL6A3, COL15A1, CLU, LUM and PROM1 reversely correlated with GFR in DN patients (Figure 5), indicating that those up-regulated hub genes may contribute to the progression of DN. Second, the up-

regulation of hub genes COL1A2, COL3A1, CLU and LUM mRNA expression was positively correlated with SCR in DN (Figure 6a–d), suggesting that those up-regulated hub genes may contribute to the progression of DN. The down-regulation of VEGFA and EGF mRNA expression was negatively correlated with SCR (Figure 6e, h, i), while EGR1 and EGR2 mRNA expression were positively correlated with SCR in DN (Figure 6f, g), indicating that VEGFA and EGF may have protective roles in DN, while EGR1 and EGR2 may promote the development of DN. Third, the mRNA expression of VEGFA, LUM and CTGF reversely correlated with proteinuria (Figure 7a–c). In addition, the expression of ALB in nephrotic proteinuria group was lower than that of subnephrotic proteinuria group (Figure 7d). However, there was no difference in COL3A1 and PROM1 expression between nephrotic proteinuria group and subnephrotic proteinuria group (Figure 9f, g).

Especially, five Hub genes that were integrated in four methods which expression were up-regulated in DN compared with normal control, the expression of LUM negatively correlated with GFR ($p < 0.0001$, $r = -0.757$) (Figure 5f) in renal tubules, and negatively correlated in renal glomerulus ($p < 0.0001$, $r = -0.744$) (Figure 5e1); LUM expression was positively correlated with SCR ($p = 0.009$, $r = 0.613$) (Figure 6d), but negatively correlated with proteinuria ($p = 0.024$, $r = -0.774$) (Figure 7c). The expression of COL1A2 negatively correlated with GFR ($p < 0.0001$, $r = -0.810$) (Figure 5a) in renal tubules, and negatively correlated in renal glomerulus ($p = 0.0004$, $r = -0.693$) (Figure 5a1); COL1A2 expression was positively correlated with SCR ($p = 0.010$, $r = 0.607$) (Figure 6a), and COL1A2 expression in the nephroproteinuria group was significantly higher than that in the subnephrotic proteinuria group (Figure 7e). The expression of CLU negatively correlated with GFR ($p < 0.0001$, $r = -0.813$) in renal tubules, and negatively correlated in renal glomerulus ($p = 0.001$, $r = -0.639$); CLU expression was positively correlated with SCR ($p = 0.010$, $r = 0.607$). In the available data, only COL6A3 and COL15A1 were confirmed to be negatively correlated with GFR, respectively.

3. Discussion

DN was the main cause of renal failure worldwide, it was quite clear that glomerulopathy and tubulopathy played a central role in the progression of DN. Recently, diabetic nephropathy had been pay more attention to acquire earlier diagnosis and additional targeted invention [16]. Increasing evidence indicated that autophagy[17], inflammation[18], oxidative stress, advanced glycation end products (AGEs) and apoptosis[19], mitochondrial dysfunction[20], and abnormal activation of the renin angiotensin system[21] were closely involved in diabetic nephropathy. Of course, the occurrence of diabetic nephropathy was influenced by genetic factors [22]. In this study, we successfully screened 91 DEGs between DN and normal populations from the GEO database. Bioinformatics tools such as GO and KEGG contributed to untangle the interaction between DEGs. The results showed that DEGs, especially hub genes, were mainly involved in cell differentiation, basement membrane, extracellular matrix components, collagen fibril organization, and other processes. KEGG enrichment analysis of DEGs showed that many of these genes were related to ECM-receptor interaction, Rap1 signaling pathway, AGE-RAGE signaling pathway in diabetic complications, protein digestion and absorption and PI3K-Akt signaling pathway. In terms of pathological changes, proliferation of the renal basement membrane, mesangial matrix

hyperplasia and fiber hyperplasia were showed with DN. Therefore, the increase of extracellular matrix (ECM) protein which included laminin, perlecan and collagen [23] could be used as a marker of DN. Meanwhile, abnormal deposition of large amounts of proteins in the basement membrane and extracellular matrix eventually lead to glomerulosclerosis. In addition, AGE/RAGE signaling elicited activation of multiple intracellular signal pathways (also contained several of the above pathways) promoted the expression of pro-inflammatory cytokines, thereby accelerating the progression of DN [24].

Fourteen differential expression genes, COL1A2, COL3A1, COL6A3, COL15A1, LUM, CLU, EGF, VEGFA, ALB, SST, FOS, FOSB, EGR1 and EGR2, were identified as key genes. Among these genes, the down-regulation of key genes such as Epidermal Growth Factor (EGF) was an active polypeptide used to promote the repair and regeneration of the damaged epidermis [25]. EGF mainly combined with EGFR to perform its functions, including activation of PI³K/Akt pathway and extracellular regulated protein kinases(ERK) pathway[26, 27].Report showed that the reduced expression of EGF particularly in the proximal tubular cells was confirmed in DN[28].And blocking EGF and EGFR activity significantly reduced diabetes-related renal enlargement and glomerular hypertrophy[29].Vascular endothelial growth factor A (VEGFA), an important endogenous angiogenic factor, was synthesized and secreted by tubular cells. Previous studies identified that vascular endothelial growth factor (VEGF) A is thought to be associated with glomerular endothelial cell dysfunction and tubular cell damage in diabetic nephropathy. And the renal angiopenia, cortical atrophy and urinary protein were further aggravated with decreased expression of the VEGFA [30, 31]. Albumin (ALB) encodes serum albumin, which regulates the plasma colloid osmotic pressure. Of course, albumin was a basic indicator of renal disease. Serum albumin also had a variety of physiological functions such as stabilizing colloid osmotic pressure, anti-inflammation and increasing blood volume. Somatostatin (SST) was a natural cyclic peptide inhibitor, and exogenous supplementation of SST attenuated the renal hypertrophy caused by diabetes and also decreased albuminuria [32]. The results of this study showed that the increase in SST expression led to a significant increase in GFR expression. Members of the Fos family, including c-Fos and FosB, dimerized with Jun family members to form the activator protein-1 (AP-1) transcription factor complex. In acute kidney injury, phosphorylation levels of c-FOS and FOSB in proximal renal tubules were increased, but C-FOS and FOSB protein were extremely unstable and rapidly degraded [33]. Meanwhile, c-FOS expression was increased in glomerular mesangial cells in high glucose environment [34]. However, the results showed that the decrease in FOS and FOSB expression lead to a significant decrease in GFR. Over-expression of miR29b could inhibit the activation of hepatic stellate cells by inhibiting the expression of C-FOS mRNA in liver fibrosis [35]. FOSB was an ap-1 transcription factor. In small intestinal neuroendocrine tumors, down-regulated expression of miRNA1 upregulated the expression of miRNA1 targeted oncogene FOSB, leading to disease progression [36]. Early growth response proteins (EGRs), as a transcriptional regulatory family, were involved in the process of cell growth, differentiation, apoptosis, and even carcinogenesis [37, 38]. EGR1 and EGR2 were members of the EGRs family genes [39]. EGR1 was proved as a transcriptional activator of oxidative stress and the levels of EGR1 mRNA and EGR1 protein in the kidneys of diabetic Kidney Disease mice were significantly increased [40]. Meanwhile, the thickness of glomerular basement membrane and renal fibrosis were alleviated after EGR1 knockdown in diabetic kidney disease mice [41].

EGR2 silencing reduced COL1A1 and COL3A1 mRNA levels, and diminished accumulation of myofibroblasts and macrophages. Liver fibrosis was also alleviated because collagen accumulation was reduced by half [42]. The results showed that the decrease in EGR1 and EGR2 expression lead to a significant decrease in GFR and SCR.

Next, we will highlight these hub genes. Clusterin (CLU), an extracellular chaperone, was a ubiquitously expressed protein that could be identified in various body fluids and tissues. Expression of CLU could lead to various processes including apoptosis, inflammation, lipid transportation, cell adhesion, morphologic transformation, membrane recycling, and suppression of complement system [43, 44]. Increased expression of CLU was found in several renal disease models such as angiotensin-induced fibrosis and unilateral ureteral obstruction (UUO) [45]. Meanwhile, under-expression of CLU was shown to accelerate renal fibrosis induced by UUO, whereas its over-expression was shown to prevent UUO-induced renal fibrosis[46].However, the results of this study showed that the increase in CLU expression lead to a significant decrease in GFR and a significant increase in SCR in diabetic nephropathy patients. The enrichment analysis showed that CLU was mapped to complement and coagulation cascades. There was also considerable evidence that activated complement systems and coagulant events contributed to the progression of DN, such as complement C3, C1, and C7,but did not mention CLU[47–49]. Together with biglycan, fibromodulin, and decorin, lumican (LUM) belonged to small leucine-rich extracellular matrix proteoglycans (SLRPs) [50]. Members of the proteoglycan family could form complexes with cytokines [51]. They were primarily considered to play a role as organizers of the extracellular matrix. Targeted disruption of the genes for decorin, fibromodulin, and LUM all led to abnormal collagen fibril morphology [52]. Their glomerular localization was confirmed, although they were abundant in the tubulointerstitium.Decorin and Biglycan are shown clustered in areas of tubulointerstitial fibrosis in some people with kidney disease, and decorin was considered to be the best predictor of progressive renal failure [53]. LUM was significantly upregulated in the tubulointerstitium and glomeruli at all stages of diabetic nephropathy [52]. In our work, the mRNA expression of LUM in DN patients was negatively correlated with GFR and proteinuria and positively correlated with SCR. Four key genes including collagen type III α 1 chain (COL3A1), collagen type I α 2 chain (COL1A2), collagen type VI α 3 chain (COL6A3) and collagen type XV α 1 chain (COL15A1) were up-regulated in diabetic nephropathy tissues. The mRNA expression of COL3A1 increased in the kidneys under hyperglycaemic/hyperinsulinaemic conditions [54]. Previous report showed Collagen constituted a microfibrillar network composed of several different types of collagen fibrils, including type I and type VI alpha 1–3 (encoded by the genes COL1 and COL6A1-3, respectively)[55].The level of collagen I (α 2) was elevated in the renal cortex of diabetic mice. And the elevation of collagen I (α 2) could contribute to tubulointerstitial fibrosis in diabetic nephropathy [56]. In our research, COL1A2 and COL3A1 were in targeted correlation with the miRNA196a/b. Little research had been reported on miR196, one study had shown that miR196 was associated with transcription factors and affects cancer development and progression [57]. Later, it might be possible to explore the effect of miR196 on diabetic nephropathy.The interaction among DN and hub genes COL6A3 and COL15A1 had not been reported yet. Collagen type VI α 3chain (COL6A3) was a beaded filament-forming collagen involved in cell-binding [58]. In human adipose tissue, up-regulation of COL6A3 associated

positively with obesity-related inflammation, insulin resistance and metabolic dysregulation. COL6A3 might partly promote insulin resistance by restricting adipogenesis [59]. We had observed knockdown of COL6A3 suppressed macrophage chemoattractant protein MCP1, supporting pro-inflammatory effects of COL6A3 [60]. Insulin resistance and inflammation were the risk factors for diabetic nephropathy. Moreover, the results of this study showed that the increase in COL6A3 expression lead to a decrease in GFR in diabetic nephropathy patients. Collagen type XV α 1 chain (COL15A1) encoded cytoskeletal protein restin, which stabilized microvessels and cells as well as suppresses angiogenesis [61]. Abnormal angiogenesis was an evolving mechanism proposed to contribute to the early phase of DN [62]. Results indicated that the increase in COL15A1 expression lead to a significant decrease in GFR. And correlation analysis of key genes with microRNAs showed both COL6A3 and COL15A1 were in targeted correlation with the miRNA29a/b/c. The miR29 family consisted of miR29a, miR29b, miR29c, and the difference between these subtypes was based on their genomic locus [63]. Reported that the higher expression of miR29 family in T2DM patients and susceptible to T2DM [64]. MiR29 family played an imperative role in chronic conditions of kidney diseases. Various targets of miR29 family in signal transduction pathways, including Wnt/ β and TGF- β 1 signaling pathways involved in fibrosis, were regulated by targeting profibrotic genes in DN[65]. Meanwhile, miR29 family was involved in maintenance of genes expression that encoded for ECM proteins such as collagen I, III and IV, not VI and XV [66]. And miR29 family also regulated inflammatory genes to slow the progression of DN. Finally, multiple studies had shown that miR29a/b could act as an anti-fibrosis factor, providing a treatment for glomerulosclerosis, while miR29c promoted inflammatory response by inhibiting the expression of TTP, leading to the further development of DN [65, 67].

Of course, there were many shortcomings of this study. First, although 5 original microarray data sets of diabetic nephropathy were integrated and the total number of samples was sufficient, diabetic nephropathy did not account for the majority of the samples. Second, the lack of detailed demographic information made it difficult to obtain a more convincing correlation between the selected genes and the severity of diabetic nephropathy using the same samples. Third, due to the lack of online data, the information of selected genes and related clinical indicators was insufficient, resulting in the unclear specific role of some genes. Therefore, further clinical and basic studies are needed in the future to verify our results and clarify the biological role of these genes and related microRNA in diabetic nephropathy.

4. Conclusions

Through transcriptome data analysis, 91 DEGs, 14 key genes especially 5 hub genes and 11 micrnas(especially miR29 and miR196) were identified as biomarkers for the clinical diagnosis of DN and as potential targets for new therapies. However, the specific molecular mechanism and biological functions of these genes still require further research.

5. Methods

5.1. Acquisition and processing of microarray data

Gene expression data related to diabetic nephropathy was downloaded from GEO online database (<https://www.ncbi.nlm.nih.gov/geo/>). By searching the following search terms: “diabetic nephropathy,” “Homo sapiens,” and “Expression profiling by array,” five microarray data including GSE30122, GSE1009, GSE47185-GPL11670, GSE47185-GPL14663 and GSE104954-GPL22945 were screened out by searching the following search terms: “diabetic nephropathy,” “Homo sapiens,” and “Expression profiling by array”. There were a total of 140 samples in these five microarray data, including 61 diabetic nephropathy samples and 79 normal kidney samples, which from different batches renal tissue samples. The probe data in the matrix file was transformed into gene symbol data through affymetrix platform file, if a single gene corresponds to multiple probes, the average expression value is calculated. After the integration of five matrix data through Perl, batch normalization was conducted through the SVA (surrogate variable analysis) package of R language (version 3.6.1).

5.2. Screening of differential genes and enrichment analysis of GO and KEGG.

DEGs were calculated and screened by Limma (Linear Models for Microarray Data) package of R language. Only when the FDR (False Discovery Rate) < 0.05 and $|\log FC| > 1$, can it be identified as the DEGs for subsequent analysis. Gene Ontology (GO), which consists of three independent ontology, namely biological processes (BP), molecular function (MF) and cellular component (CC), was also used for disease and functional analysis [6]. KEGG is an integrated database resource for the biological interpretation of genome sequences and other high-throughput data[15]. GO function and KEGG pathway enrichment analysis were performed by “ClusterProfiler” package of R. Cut-off values for GO and KEGG enrichment analysis: $P = 0.05$, $q = 0.05$.

5.3. Screening of key differential genes with Cytoscape.

To analyze the interactions between gene-encoded proteins, the DEGs were introduced into the searching tool of the interaction genes database (STRING11.0) to construct a protein-protein interaction (PPI) network, and then the results were imported into CytoHubba used to calculate the value of each DEG according to 12 topological algorithms in Cytoscape. Download the data and sorted it, selecting the top 16 hub nodes. The public hub nodes calculated by ≥ 5 topological algorithms were finally designated as key genes. Besides, sub-networks were screened and the node degrees were calculated using Molecular Complex Detection (MCODE) of Cytoscape software. The parameters were set as follows: MCODE score ≥ 3 , degree cut-off = 2, node score cut-off = 0.2, max depth = 100 and k-score = 2, and then the high relevant modules were obtained. Meanwhile, all the samples of different batches were divided into DN tubular tissue and DN glomerular tissue, which were compared with healthy kidney tubular tissue and healthy kidney glomerular tissue respectively for analysis, and relevant key genes were obtained according to the above analysis method. The intersections between key genes associated with glomeruli and key genes associated with tubules were taken.

5.4. Screening of hub genes with weighted gene co-expression network analysis.

Weighted gene co-expression network analysis (WGCNA) was suitable for identifying gene modules with the certain phenotype [9]. The risk of DN was considered the result of multi-molecule interaction. WGCNA was used to identify the related gene modules (each module contained multiple molecules). A scale-free co-expression network was constructed for all genes in five microarray data using the WGCNA package [9] in R version 4.0.3 (<https://www.r-project.org/>) and gene modules were identified, and the gene module with the highest correlation was found in the module-trait diagram. Based on gene significance and module membership, the “hclust” function in the WGCNA package was used to screen the hub genes of the targeted- module. Finally, DEGs, genes of highly relevant MCODE module, genes derived from 12 algorithms and genes of highest correlation modules in WGCNA were intersected to obtain the hub genes.

5.5. Correlation analysis of key genes and microRNAs.

WEB-based GENE SeT AnaLysis Toolkit (WebGestalt) was a functional enrichment analysis web tool, which supported multiple enrichment analysis algorithms, and covered a comprehensive database of functional annotations. In my research, WebGestalt was used to predict miRNAs that targeted the key genes associated with DN. The genes in the modules with high correlation analyzed by MCODE were imported into WebGestalt, the miRNAs targeted by WebGestalt were predicted. Differentially expressed microRNAs were selected according to the top ten P values of Significance Level.

5.6. Acquisition and processing of data related to gene and clinical characteristics.

Nephroseq was a free platform to the academic community for integrative data mining of genotype/phenotype data, which was supported by the University of Michigan O'Brien Renal Center. The primary goal of the Applied Systems Biology Core was to provide to the renal research community a platform for integrative data mining of comprehensive renal disease gene expression data sets. By inputting interested genes, Relevant data between relevant genes and GFR, SCR and proteinuria were obtained in Nephroseq (<https://www.nephroseq.org>), and then the data was imported into GraphPad Prism 6 to draw correlation analysis diagram.

Abbreviations

DN : diabetic nephropathy; ECM : extracellular matrix; GEO : Gene Expression Omnibus; GO : gene ontology; KEGG : kyoto encyclopedia of genes and genomes; miRs : microRNAs; PPI : protein–protein interaction; WebGestalt : WEB-based GENE SeT AnaLysis Toolkit. CG: Cockcroft Gault; GFR: glomerular filtration rate; MDRD: modification of diet in renal disease; SCR: serum creatine.

Declarations

Acknowledgements

The authors thank the various free online databases for their support. We also thank our colleagues for their technical support in data generation.

Authors' contributions

XS conceived and designed the study. XS and WS performed the bioinformatics pipeline and processed data for the analysis. XS performed correlation analysis. XS wrote the manuscript. WS and YM put forward reasonable Suggestions for the discussion section of the study. LXN revised the manuscript. All authors read and approved the final manuscript.

Funding

There is no funding to support for this study.

Availability of data and materials

The Human Diabetic nephropathy transcriptome data set is downloaded from The GEO DataSets browser of The National Center for Biotechnology Information (NCBI) [<https://www.ncbi.nlm.nih.gov/gds>]. Data supporting the correlation between clinical features and hub-genes in this conclusion can be found in nephroseq[<https://www.nephroseq.org>].

Ethics approval and consent to participate

This study does not require ethical approval and consent due to the use of data in an online database for analysis.

Consent for publication

Not applicable.

Competing interests

The authors declare that there is no conflict of interests.

References

1. Gilbertson DT, Liu JN, Xue JL et al. Projecting the number of patients with end-stage renal disease in the United States to the year 2015. *J. Am. Soc. Nephrol.* 2005; 16:3736-41.
2. Thomas MC, Cooper ME, Zimmet P. Changing epidemiology of type 2 diabetes mellitus and associated chronic kidney disease. *Nat Rev Nephrol.* 2016; 12:73-81.
3. Fineberg D, Jandeleit DK, Cooper ME. Diabetic nephropathy: diagnosis and treatment. *Nat Rev Endocrinol.* 2013; 9(12):713-723.
4. Jian X, Yang QL, Xiao S et al. The effects of a sodium-glucose cotransporter 2 inhibitor on diabetic nephropathy and serum oxidized low-density lipoprotein levels. *Eur Rev Med Pharmacol Sci.* 2018; 22: 3994-3999.

5. Johnson SA, Spurney RF. Twenty years after ACEIs and ARBs: emerging treatment strategies for diabetic nephropathy. *Am. J. Physiol. Renal Physiol.*2015; 309: F807-20.
6. Mao YD, Zhao QH, Yin S et al. Genome-wide expression profiling and bioinformatics analysis of deregulated genes in human gastric cancer tissue after gastroscopy. *Asia Pac J Clin Oncol.*2018; 14: e29-e36.
7. Jiang ZH, Zhou X, Li R et al. Whole transcriptome analysis with sequencing: methods, challenges and potential solutions. *Cell.Mol.LifeSci.*2015; 72:3425-39.
8. Li ZT, Liu JX, Wang WW et al. Investigation of hub genes involved in diabetic nephropathy using biological informatics methods. *Ann Transl Med.*2020; 8:1087.
9. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis.*BMC Bioinformatics.*2008; 9:559.
10. Ludwig N, Leidinger P, Becker K et al. Distribution of miRNA expression across human tissues. *Nucleic Acids Res.* 2016; 44: 3865-77.
11. Kulcheski FR, Christoff AP, Margis R. Circular RNAs are miRNA sponges and can be used as a new class of biomarker. *J. Biotechnol.*2016; 238: 42-51.
12. Chen X, Yan CG, Clarence, Zhang X et al. HGIMDA: Heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget.*2016; 7: 65257-65269.
13. Valencia S, Liu J, Hannon G J et al. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* 2006; 20, 515–524.
14. Paul P, Chakraborty A, Sarkar D et al. Interplay between miRNAs and human diseases. *J. Cell. Physiol.* 2018; 233, 2007–2018.
15. Zhang LJ, Lu DH, Liu MQ, et al. Identification and interaction analysis of key miRNAs in medullary thyroid carcinoma by bioinformatics analysis. *Mol Med Rep.*2019 20: 2316-2324.
16. Zeni L, Norden Anthony G W, Cancarini G et al.A more tubulocentric view of diabetic kidney disease. *J. Nephrol.*2017; 30: 701-717.
17. Woo CY, Kc Rj, Kim M et al. Autophagic flux defect in diabetic kidney disease results in megamitochondria formation in podocytes. *Biochem. Biophys. Res. Commun.*2020; 521: 660-667.
18. Xu BH, Sheng JY, You Yk et al. Deletion of Smad3 prevents renal fibrosis and inflammation in type 2 diabetic nephropathy. *Metab. Clin. Exp.* 2020; 103: 154013.
19. Tong YN, Liu S, Gong R et al. Ethyl Vanillin Protects against Kidney Injury in Diabetic Nephropathy by Inhibiting Oxidative Stress and Apoptosis. *Oxid Med Cell Longev.*2019; 2019: 2129350.
20. Zhou D, Zhou M, Wang ZY et al. PGRN acts as a novel regulator of mitochondrial homeostasis by facilitating mitophagy and mitochondrial biogenesis to prevent podocyte injury in diabetic nephropathy. *Cell Death Dis.* 2019; 10: 524.
21. Wang Y, Fang QL, Jin YY et al. Blockade of myeloid differentiation 2 attenuates diabetic nephropathy by reducing activation of the renin-angiotensin system in mouse kidneys. *Br. J. Pharmacol.*2019; 176: 2642-2657.

22. Soler MJ, Batlle D. Single-cell RNA profiling of glomerular cells in diabetic kidney: a step forward for understanding diabetic nephropathy. *Ann Transl Med.* 2019; 7:S340.
23. Chan GC, Eng DG, Miner JH, et al. Differential expression of parietal epithelial cell and podocyte extracellular matrix proteins in focal segmental glomerulosclerosis and diabetic nephropathy. *Am J Physiol Renal Physiol.* 2019; 317(6):F1680-F1694.
24. Wang ZW, Zhang JQ, Chen L et al. Glycine Suppresses AGE/RAGE Signaling Pathway and Subsequent Oxidative Stress by Restoring Glo1 Function in the Aorta of Diabetic Rats and in HUVECs. *Oxid Med Cell Longev.*2019; 2019: 4628962.
25. Yang FM, Qin XM, Zhang T et al. Pinctada MartensiiEffect of Oral Administration of Active Peptides of on the Repair of Skin Wounds. *Mar Drugs.*2019; 17: undefined.
26. Overstreet JM, Wang YQ, Wang X et al. Selective activation of epidermal growth factor receptor in renal proximal tubule induces tubulointerstitial fibrosis.*FASEB J.* 2017; 31: 4407-4421.
27. Chen JC, Harris RC. Interaction of the EGF Receptor and the Hippo Pathway intheDiabeticKidney. *J.Am.Soc.Nephrol.*2016; 27:1689-700.
28. Rudnicki M, Beckers A, Neuwirt H et al. RNA expression signatures and posttranscriptional regulation in diabetic nephropathy. *Nephrol Dial Transplant.*2015, null: iv35-42.
29. Tang JH, Liu N, Zhuang SG.Role of epidermal growth factor receptor in acute and chronic kidney injury. *Kidney Int.*2013; 83: 804-10.
30. Onions KL, Gamez M, Buckner NR et al. VEGFC Reduces Glomerular Albumin Permeability and Protects Against Alterations in VEGF Receptor Expression in Diabetic Nephropathy. *Diabetes.*2019; 68: 172-187.
31. Li XY, Yang SS, Yan MH et al.. Interstitial HIF1A induces an estimated glomerular filtration rate decline through potentiating renal fibrosis in diabetic nephropathy. *Life Sci,* 2020; 41: 117109.
32. Segev Y, Eshet R, Rivkis I et al. Comparison between somatostatin analogues and ACE inhibitor in the NOD mouse model of diabetic kidney disease.*Nephrol. Dial. Transplant.*2004; 19: 3021-8.
33. Iwatsuki M, Inageda K, Matsuoka M. Cadmium induces phosphorylation and stabilization of c-Fos in HK-2 renal proximal tubular cells. *Toxicol. Appl. Pharmacol.*2011; 251: 209-16.
34. Huang K, Huang J, Chen C, et al. AP-1 regulates sphingosine kinase 1 expression in a positive feedback manner in glomerular mesangial cells exposed to high glucose.*Cell Signal.*2014; 26: 629-638.
35. Sekiya Y, Ogawa T, Yoshizato K et al. Suppression of hepatic stellate cell activation by microRNA-29b. *Biochem. Biophys. Res. Commun.*2011; 412: 74-9.
36. Miller Helen C,Frampton Adam E,Malczewska Anna et al. MicroRNAs associated with small bowel neuroendocrine tumours and their metastases. *Endocr. Relat. Cancer.*2016; 23: 711-26.
37. Collins S, Wolfraim LA, Drake CG et al. Cutting edge: TCR-induced NAB2 enhances T cell function by coactivating IL-2 transcription. *J. Immunol.*2006; 177:8301–8305.

38. Skerka C, Decker EL, Zipfel PF. Coordinate expression and distinct DNA-binding characteristics of the four EGR-zinc finger proteins in Jurkat T lymphocytes. *Immunobiology*.1997; 198:179–191.
39. Fei YC, Yu H, Huang S et al. Expression and prognostic analyses of early growth response proteins (EGRs) in human breast carcinoma based on database analysis. *PeerJ*.2019; 7: e8183.
40. Hu F, Xue M, Li Y et al. Early Growth Response 1 (Egr1) Is a Transcriptional Activator of NOX4 in Oxidative Stress of Diabetic Kidney Disease. *J Diabetes Res*. 2018; 2018: 3405695.
41. Yang YL, Hu F, Xue M et al. Early growth response protein-1 upregulates long noncoding RNA Arid2-IR to promote extracellular matrix production in diabetic kidney disease. *Am. J. Physiol*. 2019; 316: C340-C352.
42. Vollmann EH1, Cao L2, Amatucci A, et al. Identification of Novel Fibrosis Modifiers by In Vivo siRNA Silencing. *Mol Ther Nucleic Acids*.2017; 16(1); 7:314323.
43. Uddin MS, Kabir MT, Begum MM et al. Exploring the Role of CLU in the Pathogenesis of Alzheimer's Disease. *Neurotox Res*. 2020; undefined: undefined.
44. Trougakos IP, Gonos ES. Chapter 9: Oxidative stress in malignant progression: The role of Clusterin, a sensitive cellular biosensor of free radicals. *Adv Cancer Res*.2009; 104: 171–210.
45. Jung GS, Jeon JH, Jung YA et al. Clusterin/apolipoprotein J attenuates angiotensin II-induced renal fibrosis. *PLoS One*.2014; 22; 9(8):e105635.
46. Jung GS, Kim MK, Jung YA et al. Clusterin attenuates the development of renal fibrosis. *J. Am. Soc. Nephrol*.2012; 23: 73-85.
47. Tang Sydney CW, Yiu WH. Innate immunity in diabetic kidney disease. *Nat Rev Nephrol*.2020; 16: 206-222.
48. Tang SM, Wang XF, Deng TC et al. Identification of C3 as a therapeutic target for diabetic nephropathy by bioinformatics analysis. *Sci Rep*.2020; 10: 13468.
49. Sun ZJ, Li XQ, Chang DY, et al. Complement deposition on renal histopathology of patients with diabetic nephropathy. *Diabetes Metab*.2019; 45: 363-368.
50. Schaefer L, Raslik I, Grone H J et al. Small proteoglycans in human diabetic nephropathy: discrepancy between glomerular expression and protein accumulation of decorin, biglycan, lumican, and fibromodulin. *FASEB J*.2001; 15: 559-61.
51. Moss J, Shore I, Woodrow D: An ultrastructural study of the colocalization of biglycan and decorin with AA amyloid fibrils in human renal glomeruli. *Amyloid*.1998; 5:43–48.
52. Neame P J, Kay C J, McQuillan D J et al. Independent modulation of collagen fibrillogenesis by decorin and lumican. *Cell. Mol. Life Sci*.2000; 57: 859-63.
53. Takeuchi Y, Kodama Y, Matsumoto T: Bone matrix decorin binds transforming growth factor- β and enhances its bioactivity. *J Biol Chem*.1994; 269:32634–32638.
54. Gaikwad AB, Gupta J, Tikoo KH. Epigenetic changes and alteration of Fbn1 and Col3A1 gene expression under hyperglycaemic and hyperinsulinaemic conditions. *Biochem. J*. 2010; 432: 333-41.

55. Dankel SN, Grytten E, Bjune JI et al. COL6A3 expression in adipose tissue cells is associated with levels of the homeobox transcription factor PRRX1. *Sci Rep.* 2020, 10: 20164.
56. Das F, Ghosh CN, Venkatesan BC et al. PDGF receptor- β uses Akt/mTORC1 signaling node to promote high glucose-induced renal proximal tubular cell collagen I ($\alpha 2$) expression. *Am. J. Physiol. Renal Physiol.* 2017; 313: F291-F307.
57. Li Y, Zhang M, Chen H et al. Ratio of miR-196s to HOXC8 messenger RNA correlates with breast cancer cell migration and metastasis. *Cancer Res.* 2010; 70: 7894–7904.
58. Specks U, Mayer U, Nischt R, et al. Structure of recombinant N-terminal globule of type VI collagen alpha 3 chain and its binding to heparin and hyaluronan. *EMBO Journal.* 1992; 11(12), 4281–4290.
59. Dankel, S. N. et al. COL6A3 expression in adipocytes associates with insulin resistance and depends on PPAR γ and adipocyte size. *Obesity* .2014; 22, 1807–1813.
60. Gesta, S. et al. Reduced expression of collagen VI alpha 3 (COL6A3) confers resistance to inflammation-induced MCP1 expression in adipocytes. *Obesity.* 2016; 24, 1695–1703.
61. John H, Radtke K, Ständker L et al. Identification and characterization of novel endogenous proteolytic forms of the human angiogenesis inhibitors restin and endostatin. *Biochimica et Biophysica Acta.* 2005; 1747(2), 161–170.
62. NAKAGAWA T, KOSUGI T, HANEDA M et al. Abnormal angiogenesis in diabetic nephropathy. *Diabetes* .2009; 58: 1471-1478.
63. Jiang H, Zhang G, Wu J-H et al. Diverse roles of miR-29 in cancer. *Oncol Rep.* 2014; 31(4):1509–16.
64. Kong L, Zhu J, Han W et al. Significance of serum microRNAs in pre-diabetes and newly diagnosed type 2 diabetes: a clinical study. *Acta Diabetol.* 2011; 48(1):61–9.
65. Dasare AP, Gondaliya P, Srivastava A et al. A therapeutic approach towards microRNA29 family in vascular diabetic complications: A boon or curse? *J Diabetes Metab Disord.* 2019; 18: 243-254.
66. Wang B, Komers R, Carew R et al. Suppression of microRNA-29 expression by TGF- $\beta 1$ promotes collagen expression and renal fibrosis. *J Am Soc Nephrol.* 2012; 23(2):252–65.
67. Guo J, Li J, Zhao J et al. MiRNA-29c regulates the expression of inflammatory cytokines in diabetic nephropathy by targeting tristetraprolin. *Sci Rep.* 2017; 7.

Tables

TABLE 1

Hub genes of diabetic nephropathy provided by 12 algorithms of CytoHubba

Groups	Genes	Gene title	Number of crosses
Up	COL1A2	Collagen type I alpha2	11
	COL3A1	Collagen type III alpha1	9
	CLU	Clusterin	9
	LUM	Lumican	6
	COL6A3	Collagen type VI alpha3	6
	COL15A1	Collagen type XV alpha1	5
	PROM1	Prominin 1	5
Down	VEGFA	Vascular endothelial growth factor B	11
	ALB	Albumin	11
	EGR1	Early growth response 1	10
	FOS	Fos proto-oncogene	10
	EGF	Epidermal growth factor	9
	CTGF	Connective tissue growth factor	8
	SST	Somatostatin	7
	EGR2	Early growth response 2	7
	FOSB	FosB proto-oncogene	6

Number of crosses—Each number represents a number of intersection algorithm of the 12 algorithms.

TABLE 2

Hub genes-related MicroRNAs

MicroRNA	Genes	EnrichmentRatio	FDR
MIR-29A,MIR-29B,MIR-29C	COL15A1,COL3A1,COL6A3,VEGFA,FOS	8.836705	0.000081293
MIR-196A,MIR-196B	COL1A2,COL3A1	12.91901	0.009646
MIR-199A	VEGFA,FOS	10.07967	0.015526
MIR-377	VEGFA,EGR1	9.31218	0.018049
MIR-380-5P	COL1A2	41.69318	0.023746
MIR-191	EGR1	31.62931	0.031197
MIR-186	COL3A1,VEGFA	6.769373	0.032840
MIR-1,MIR-206	VEGFA,FOSB	6.156040	0.039137
MIR-383	VEGFA	18.34500	0.053253
MIR-134	VEGFA	17.63942	0.055331

Figures

Figure 1

Heatmap and Volcano plot of DEGs and enrichment analysis of GO and KEGG. (a) Heatmap of 91 DEGs screened. Red areas represent highly expressed genes and green areas represent lowly expressed genes in DN patients compared with normal controls. (b) Volcano plot analysis identifies DEGs. Red dots represent up-regulated genes, green dots represent down-regulated genes, and black dots represent normally expressed genes in DN patients compared with normal controls. (c) GO enrichment barplot of DEGs. The x-axis represented gene count and left y-axis represented GO terms. Different color represented different p value. (d) KEGG cnetplot of DEGs. Fold change represented Threshold of multiple changes in gene expression. Different size of circles represented the number of genes in the pathway.

Figure 2

PPI network and five significant modules of DEGs. Red circles represent up-regulated DEGs and green circles represent down-regulated DEGs. (a) PPI network of DEGs created by STRING. The nodes represent proteins and the edges represent the interactions of proteins. (b) The most significant module identified by MCODE. (c) The second significant module identified by MCODE. (d) The third significant module

identified by MCODE. (e) The forth significant module identified by MCODE. (f) The fifth significant module identified by MCODE.

Figure 3

Weighted gene co-expression network analysis to obtain the modules associated with the development of DN and normal kidney samples. (a) Graphical representation showing the selection of soft-thresholding powers. (b) The Venn diagram identified five intersecting genes. (c) Hierarchical clustering dendrogram. (d) Histogram showing module-trait correlations. Yellow module was the module which had the strongest correlation with DN.

Figure 4

Positive correlations between down-regulated hub genes mRNA expression and GFR in tubules of DN patients, P value was a significant difference, r was the correlation coefficient. (a) The expression of VEGFA. (b) The expression of ALB. (c) The expression of EGR1. (d) The expression of EGR2. (e) The expression of SST. (f) The expression of EGF. (g) The expression of EGF. (h) The expression of FOS. (i) The expression of FOSB. (j) The expression of CTGF. Positive correlation between down-regulated hub gene mRNA expression and GFR in glomerulus of DN. (a1) The expression of VEGFA. (b1) The expression of ALB. (c1) The expression of SST. $P < 0.05$ was considered statistically significant. CG: Cockcroft Gault; GFR: glomerular filtration rate; MDRD: modification of diet in renal disease. (d1) The Venn diagram of hub genes between glomerulus and tubules.

Figure 5

Negative correlations between up-regulated hub gene mRNA expression and GFR in renal tubulointerstitium of DN patients, P value was a significant difference, r was the correlation coefficient. (a) The expression of COL1A2. (b) The expression of COL3A1. (c) The expression of COL6A3. (d) The expression of COL15A1. (e) The expression of CLU. (f) The expression of LUM. (g) The expression of PROM1 Negative correlation between up-regulated hub gene mRNA expression and GFR in glomerulus of

DN patients. (a1)The expression of COL1A2.(b1) The expression of COL6A3. (c1)The expression of COL15A1.(d1)The expression of CLU.(e1) The expression of LUM. P < 0.05 was considered statistically significant. CG: Cockcroft Gault; GFR: glomerular filtration rate; MDRD: modification of diet in renal disease.

Figure 6

Correlation between mRNA expression of up-regulated hub genes and SCR in DN patients, P value was a significant difference, r was the correlation coefficient. (a) The expression of COL1A2 positively correlated with SCR. (b) The expression of COL3A1 positively correlated with SCR. (c) The expression of CLU

positively correlated with SCR. (d) The expression of LUM positively correlated with SCR. Correlation between mRNA expression of down-regulated hub genes and SCR in DN patients. (e) The expression of VEGFA negatively correlated with SCR. (f) The expression of EGR1 positively correlated with SCR. (g) The expression of EGR2 positively correlated with SCR. (h) The expression of EGF negatively correlated with SCR in tubulointerstitium of DN patients. (i) The expression of EGF negatively correlated with SCR in glomerulus of DN patients. $P < 0.05$ was considered statistically significant. SCR: serum creatine.

Figure 7

Association between mRNA expression of hub genes and proteinuria in DN patients. (a) The expression of VEGFA negatively correlated with proteinuria. (b) The expression of CTGF negatively correlated with proteinuria. (c) The expression of LUM negatively correlated with proteinuria. (d) The expression of ALB in nephrotic proteinuria group was lower than that of subnephrotic proteinuria group. (e) The expression of COL1A2 in nephrotic proteinuria group was higher than that of subnephrotic proteinuria group. (f) There was no difference in COL3A1 expression between nephrotic proteinuria group and subnephrotic proteinuria group. (g) There was no difference in PROM1 expression between nephrotic proteinuria group and subnephrotic proteinuria group. $P < 0.05$ was considered statistically significant. * $P < 0.05$, ** $P < 0.01$, n meant no significant difference .