# Identification of tubulointerstitial genes and ceRNA networks involved in diabetic nephropathy via integrated bioinformatics approaches

**Haiyan Cao**
Tianjin Medical University General Hospital https://orcid.org/0000-0001-6367-1838
**Xiaosheng Rao**
First Affiliated Hospital of Guangzhou Medical University
**Junya Jia**
Tianjin Medical University General Hospital
**Tiekun Yan**
Tianjin Medical University General Hospital
**Dong Li** ( lidong430@126.com )

**Research**

**Keywords:** Diabetic nephropathy, Integrated bioinformatics approaches, tubulointerstitial, genes, ceRNA networks

**Posted Date:** May 24th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1630303/v1

# Abstract

**Background:** Diabetic nephropathy (DN) is the major cause of end-stage renal disease worldwide. The mechanism of tubulointerstitial lesions in DN is not fully elucidated. This article aims to identify novel genes and clarify the molecular mechanisms for the progression of DN through integrated bioinformatics approaches.

**Method:** We downloaded microarray datasets from Gene Expression Omnibus (GEO) database and identified the differentially expressed genes (DEGs). Enrichment analyses, construction of Protein-protein interaction (PPI) network, and visualization of the co-expressed network between mRNAs and microRNAs (miRNAs) were performed. Additionally, we validated the expression of hub genes and analyzed the Receiver Operating Characteristic (ROC) curve in another GEO dataset. Clinical analysis and ceRNA networks were further analyzed.

**Results:** Totally 463 DEGs were identified, and enrichment analyses demonstrated that extracellular matrix structural constituents, regulation of immune effector process, positive regulation of cytokine production, phagosome, and complement and coagulation cascades were the major enriched pathways in DN. Three hub genes (CD53, CSF2RB, and LAPTM5) were obtained, and their expression levels were validated by GEO datasets. Pearson analysis showed that these genes were negatively correlated with the glomerular filtration rate (GFR). After literature searching, the ceRNA networks among circRNAs/lncRNAs, miRNAs, and mRNAs were constructed. The predicted RNA pathway of NEAT1/XIST-hsa-miR-155-5p/hsa-miR-486-5p-CSF2RB provides an important perspective and insights into the molecular mechanism of DN.

**Conclusion:** In conclusion, we identified three genes, namely CD53, CSF2RB, and LAPTM5, as hub genes of tubulointerstitial lesions in DN. They may be closely related to the pathogenesis of DN and the predicted RNA regulatory pathway of NEAT1/XIST-hsa-miR-155-5p/hsa-miR-486-5p-CSF2RB presents a biomarker axis to the occurrence and development of DN.


# Introduction

Diabetic nephropathy (DN) is a destructive complication of diabetes mellitus and a major cause of end-stage renal disease [1]. As an overwhelming majority of the renal parenchyma, the tubulointerstitial compartment plays an irreplaceable role in the pathogenesis of DN. Representative tubulointerstitial changes of DN include the hypertrophy of the renal tubular, thickening of the tubular basement membrane, and the fibrosis of the renal interstitial, which promote the development of DN [2, 3]. Some emerging biological pathways, such as oxidative stress, transforming growth factor-β activation, proinflammatory cytokines, and chemokines, are also involved in tubulointerstitial lesions of DN [4]. Since its complex pathogenesis, exploring potential genes and mechanisms could provide novel insight into the understanding of DN.

In the last few years, much attention has been paid to the integrated analysis of transcriptomic and microarray due to its widespread use in multiple diseases, including cancers and non-cancers. The previous explorations combined with bioinformatics have revealed several biomarkers in DN, such as Protein S and COL4A3 [5, 6]. The presence of microalbuminuria has been considered a well-known biomarker to indicate the occurrence of the advanced DN [7]. Whereas microalbuminuria cannot sufficiently predict the DN patients with non-proteinuria or with slight pathophysiological changes, exploring extra biomarkers of tubulointerstitial lesions to indicate the early stage of DN seems indispensable. Furthermore, competitive endogenous RNA (ceRNA) networks that based on the interactions among the main non-coding RNAs (ncRNAs), including microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and circular RNAs (circRNAs), are capable of providing a significant opportunity to advance the understanding of biomarkers and mechanisms that closely related to DN.

Our present work identified the difference in the mRNA expression profiles of human tubulointerstitial samples of DN and controls. The limma package was used to recognize differentially expressed genes (DEGs). Gene set enrichment analysis (GSEA), Gene Ontology (GO), and Kyoto encyclopedia of genes and genomes (KEGG) Pathways were performed to explore the mechanisms that impact the pathogenesis of DN. The Protein-protein Interaction (PPI) network was constructed to obtain hub genes. Validation datasets were further used to validate the expression and plot the ROC curve of hub genes. The clinical features of the hub genes were explored. Based on the interaction relationship between the predicted miRNAs, lncRNAs, and circRNAs, we constructed ceRNA networks to elucidate a novel mechanism for accelerating DN progression in transcriptional networks. The identified hub genes and potential ceRNA networks are expected to provide insights into the molecular mechanism of tubulointerstitial lesions in DN.

# Materials And Methods

### Data sources

We screened relevant gene expression datasets from the Gene Expression Omnibus (GEO) genomics data repository (http://www.ncbi.nlm.nih.gov/geo) using diabetic nephropathy and tubulointerstitial as the keywords [8]. The screening criteria are set as follows: the microarray expression profile should include human tubulointerstitial samples and provide complete raw data. Eventually, we selected the microarray dataset (GSE30529) that was downloaded from the Affymetrix GPL571 platform (Affymetrix Human Genome U133A 2.0 Array) as the test dataset. GSE30529 was the subseries of GSE30122, from which we collected human tubulointerstitial tissues of 10 DN samples (GSM757014-GSM757023) and 12 controls (GSM757024-GSM757035).

### Quality control and identification of DEGs

We used GEOquery (version 2.54.1) and limma package (version 3.42.2) in R software (version 3.6.3) to read data and complete the analysis of DEGs, respectively. Data normalization and log2 conversion were included in the pre-processed analysis. Probe sets with no relevant gene symbols or gene symbols with

an excess of one probe set were removed or averaged. To evaluate the quality of the selected dataset, we analyzed the processed data boxplot, principal component analysis (PCA), and uniform manifold approximation and projection (UMAP). We considered these genes with adj. P-value < 0.05 and |logFC (fold change)| ≥ 1 as DEGs. We further used the ggplot2 (version 3.3.3) package and ComplexHeatmap (version 2.2.0) package of R software to draw the volcano and heatmaps of the acquired DEGs [9].

## Enrichment analyses

To evaluate the association between gene sets and biological signals, we performed the GSEA analysis [10]. After downloading GSEA_4.1.0 and c5: GO gene sets (c5.all.v7.1.symbols. gmt), we analyzed it via the clusterProfiler package (version 3.14.3) in R software [11]. The threshold value of |Normalized Enrichment Score (NES)| > 1, False Discovery Rates (FDR) < 0.05, and p-value < 0.05 were considered statistically significant in GSEA. GO and KEGG enrichment analyses of identified DEGs were annotated with clusterProfiler package (version 3.14.3) and visualized by ggplot2 (version 3.3.3). GO analysis can provide scientists with the better understanding of the functional annotations of the gene products [12]. KEGG Pathway is a tool to help biologists understand the systemic functions of the cell and even the organism [13]. Q value ≤ 0.05 and gene count ≥ 2 were considered to indicate statistically significant results.

## PPI network construction and significant module analysis

STRING (version 11.0), an online database that can explore the underlying relationship among proteins, was used to construct the complex PPI network with a combined score ≥ 0.7 [14]. Cytoscape (version 3.8.2) was used to visualize the PPI network [15]. Minimal Common Oncology Data Elements (MCODE) (version 2.0.0), a plugin of Cytoscape, was used to construct significant modules with the filter criteria : degree cutoff = 2, max depth = 100, K-core = 2, and node score cutoff = 0.2.

## Identification of hub genes and prediction of target miRNAs

We used cytoHubba (version 0.1), a plugin of Cytoscape software, to obtain hub genes by acquiring the interacted genes of the top 10 genes identified by the following algorithm modules, namely Maximal Clique Centrality (MCC), Density of Maximum Neighborhood Component (DMNC), Maximum Neighborhood Component (MNC), and Degree [16]. Three online tools (miRmap, microT, miRanda) were applied to predict target miRNAs of hub genes, and we chose the miRNAs that appeared more than once in searching tools as target miRNAs. Given the regulatory relationship between the hub genes (mRNA) and their target miRNAs, we used Cytoscape software to perform the co-expressed network of mRNAs-miRNAs.

## Validation and ROC curve analysis of hub genes by another GEO dataset

The GEO dataset (GSE104954) was recognized as validation dataset with a total of human tubulointerstitial tissues of 7 DN (GSM2811029-GSM2811035) and 18 controls (GSM2811043-GSM2811060). We used the validation dataset to verify the mRNA expression and plot the ROC curve of

the hub genes by the ggplot2 package (version 3.3.3). The statistical method of t-test was used to compare the differences between the DN samples and controls, and we considered P-value < 0.05 as significant.

### Construction of ceRNA networks

StarBase (http://starbase.sysu.edu.cn/index.php) is a novel database developed to accelerate the comprehensive understanding of miRNA-target intersection maps based on the CLIP-Seq and Degradome-Seq data [17]. We used StarBase to explore the interactive relationships between the lncRNAs, circRNAs, and specific miRNAs with the screening criteria of mammalian, human, h19 genome, strict stringency (>=5) of CLIP-Data, with or without degradome data, and with or without data of pan-cancer. Finally, the intersected lncRNAs and circRNAs were recognized as the final target results of certain miRNAs. The ceRNA networks among mRNAs, miRNAs, lncRNAs, and circRNAs were constructed and visualized by the Cytoscape software.

### Clinical and statistical analysis

The data from Nephroseq v5 online platform (http://v5.nephroseq.org) was extracted to explore clinical characteristics of gene expression in human tubulointerstitial tissues [18]. We performed a correlation analysis of glomerular filtration rate (GFR) [19], proteinuria [20], and age in tubulointerstitial samples of DN patients and compared gene expression between the subnephrotic proteinuria group and nephrotic proteinuria group [20]. A statistical method of Student's t-test was used to analyze the data between the two groups. The cutoff criterion was set as p□0.05.

## Results

### Quality control and identification of DEGs

Fig. 1 shows the overall flowchart of this present research. The processed data boxplot displayed that the medians of each sample were almost on the same line (Fig. 2A). The PCA and UMAP plots suggested a good distinction between DN and control samples (Fig. 2B, C). After downloading the tubulointerstitial data from GSE30529, we recognized 463 DEGs that consist of 340 up-regulated DEGs and 123 down-regulated DEGs. Fig. 2D and Fig. 2E showed the volcano map and heatmap of the DEGs.

### Analysis of the functional characteristics

To evaluate GO analysis of all genes, we performed GSEA analysis. We found that the most enriched gene sets were associated with extracellular matrix structural constituent, complement activation, and positive regulation of T cell proliferation in DN samples (Fig. 3). We subsequently performed the GO and KEGG pathway analyses of DEGs. Compared to the controls, DN patients are significantly enriched in the extracellular matrix and inflammatory response, including collagen-containing extracellular matrix, extracellular structure organization, and regulation of immune effector process (Fig. 4A). Based on the KEGG pathways, the top 10 enriched pathways involved Phagosome, Staphylococcus aureus infection,

Pertussis, Leishmaniasis, Tuberculosis, Complement and coagulation cascades, Cell adhesion molecules, Viral myocarditis, Systemic lupus erythematosus, and Toxoplasmosis (Fig. 4B).

### PPI network construction and module analysis

The complex PPI network was constructed with 277 nodes and 803 edges in Fig. 5A. Through plugin MCODE of Cytoscape, we recognized 3 significant modules with an MCODE score ≥ 5. Module one had the highest score with 11 (11 nodes and 55 edges), module two had the second-highest score with 5.625 (17 nodes and 45 edges), followed by module three had the third-highest score with 5.60 (6 nodes and 14 edges) (Fig. 5B-E).

### Identification of hub genes and prediction of target miRNAs

Four algorithms were used to identify the hub genes via the Venn diagram. As a result, 3 intersected genes were obtained (Fig. 6A). Table 1 lists the hub genes together with their full name and descriptions. We then explored 24 target miRNAs of 3 hub genes based on the regulatory relationship between mRNAs and miRNAs. There are 25 mRNA-miRNA pairs in the co-expressed mRNA-miRNA network (Fig. 6B).

### Validation and ROC curve of hub genes

To assess the reliability of these hub gene expressions, we used validation datasets to explore the mRNA levels of the obtained hub genes. The results displayed that 3 hub genes showed up-regulated mRNA levels in DN (Fig. 7A). To further evaluate the sensitivity and specificity of the hub genes, we analyzed the ROC curve with the validation datasets. Among the results, CSF2RB had the highest areas under the curve (AUC) score ( AUC value: 0.897), CD53 had the second-highest AUC score ( AUC value: 0.846), followed by LAPTM5 had the third-highest AUC score ( AUC value: 0.843) (Fig. 7B).

### Clinical analysis

Based on the Nephroseq v5 platform, we investigated the possible impact of the hub genes in tubulointerstitial samples in DN. Correlation analyses showed that the mRNA levels of CSF2RB, CD53, and LAPTM5 in human tubulointerstitial samples were negatively correlated with GFR (Fig. 8A-D), indicating that these hub genes may accelerate the tubulointerstitial lesions of DN. Then, the mRNA levels of CD53 and LAPTM5 in human tubulointerstitial samples had positive correlations with proteinuria (Fig. 8E, F), further suggesting that CD53 and LAPTM5 may be involved in the progression of tubulointerstitial injuries of DN. Additionally, we found that, in contrast with the subnephrotic proteinuria group, the mRNA expression level of LAPTM5 in renal tubulointerstitial samples was higher in the nephrotic proteinuria group (Fig. 8G). Furthermore, the mRNA levels of CSF2RB, CD53, and LAPTM5 were conversely related to their age (Fig. 8H, J).

### Construction of ceRNA networks

It is well known that miRNAs can inhibit mRNA translation or lead to mRNA degradation by binding to target mRNAs, thus achieving the regulation of gene expression. However, the ceRNA hypothesis interpreted that lncRNAs, circRNAs, and pseudogenes that act as competing endogenous RNAs or natural microRNA sponges can increase gene expression by binding microRNA response elements (MREs) [21]. Based on the Starbase platform, we predicted corresponding lncRNAs and circRNAs of target miRNAs. Finally, we selected the lncRNAs, circRNAs that appeared in most of the forecasted results of target miRNAs to be our final lncRNAs, circRNAs. Of the predicted results in the Starbase tool, we regarded the circRNAs that qualified with the highest score or /and the most samples in the circBase database to be the eventual target circRNAs since one transcript corresponds to multiple shear sites of circRNA.

Just as the forecasted results of circRNAs and lncRNAs (Fig. 9A, B, C), we acquired 22 circRNAs and 1 lncRNA of 3 miRNAs of CD53, which were paired into 66 circRNA-miRNA interactions, 3 lncRNA-miRNA interactions, and 3 miRNA-mRNA interactions; 9 circRNAs and 2 lncRNAs of 18 miRNAs of CSF2RB, which were paired into 135 circRNA-miRNA interactions, 18 lncRNA-miRNA interactions, and 29 miRNA-mRNA interactions; and 5 circRNAs and 1 lncRNA of 4 miRNAs of LAPTM5, which were paired into 20 circRNA-miRNA interactions, 4 lncRNA-miRNA interactions, and 4 miRNA-mRNA interactions. Given the ceRNA hypothesis, we subsequently put two down-regulated miRNAs and two up-regulated lncRNAs into further analyses. According to these data, we can infer that the regulatory pathway of NEAT1/XIST-hsa-miR-155-5p/hsa-miR-486-5p-CSF2RB may play an essential role in DN (Fig. 9D).

# Discussion

DN is the major cause of renal failure across the world. The pathogenetic mechanisms of the tubulointerstitium, including the activation of intracellular signals [1], adverse changes secondary to the glomerulus, and hypoxia-induced lesions [22, 23], have emerged as a critical part of the occurrence and progression of DN. Although a great deal of effort has been undertaken so far, we still lack a thorough understanding of the pathogenesis of DN. Fortunately, this emerging combination of high-throughput microarray technology and bioinformatics methods allows us to identify important genes and biological pathways associated with tubulointerstitial lesions in DN and gain insights into the pathogenesis of DN.

Totally 12548 genes are included in the GSEA enrichment analysis, and 463 DEGs are included for GO and KEGG analyses. GSEA and GO enrichment analyses both suggested that extracellular matrix (ECM) structural constituent, ECM organization, collagen-containing ECM, and extracellular structure organization were highly enriched in DN groups. The hallmark of the pathogenesis of DN is an increased ECM accumulation, and the ECM levels are usually regulated by a balance between the deposition and degradation of the ECM components [24]. The deposition of ECM proteins, including collagen I, decorin, and biglycan, are all involved in tubulointerstitial fibrosis of DN [25, 26]. As significant endopeptidases for degrading ECM protein components, matrix metalloproteinases (MMPs) can accelerate diverse physiologic and pathological damages in DN [27]. In human DN, MMPs dysregulation interferes with ECM degradation and hydrolysis and is involved in an established renal injury [28].

Additionally, GO and KEGG pathway analyses were also enriched in regulation of immune effector process, positive regulation of cytokine production, phagosome, complement and coagulation cascades, and cell adhesion molecules. The inflammatory response is a mechanism induced by activated stimuli to provide protection, while non-resolving inflammation triggers the collateral damage of tissues and organisms [29]. Under high glucose conditions, increasing evidence points out the critical role of inflammation response in the development and progression of DN [30]. The activation of the complement system, deposition of immunoglobulin or immunoglobulin complexes, and activation of macrophages are all involved in the pathogenesis of DN [31–33].

After constructing the PPI network, we acquired three hub genes ( CSF2RB, CD53, and LAPTM5 ). Further validation suggested that they showed up-regulated expressions and had high AUC values. Combined with their negative correlations with GFR, we speculated that CSF2RB, CD53, and LAPTM5 might be the key genes in human tubulointerstitial lesions of DN. The subsequent mRNA-miRNA network and ceRNA networks were accomplished to offer the opportunity to understand the mechanisms of DN from the transcriptomic level.

CSF2RB(colony-stimulating factor 2 receptor β subunit ) is the common subunit of receptors for interleukin 3 (IL3), GM-CSF, and IL5 and is responsible for the initiation of signal transduction triggered by ligand binding [34]. Considerable evidence has revealed that IL3, GM-CSF, and IL5 regulate multiple inflammatory responses that contribute to the pathology in chronic inflammation [35, 36]. Since inflammatory responses are the key contributors to the development of DN [37], we speculated that CSR2RB has a critical role in the molecular mechanisms of DN. Our work found that CSF2RB expressed an increased mRNA level in human tubulointerstitial samples of DN and had a high diagnostic value (AUC = 0.897). Furthermore, the level of CSF2RB in human renal tubulointerstitial samples was conversely correlated with GFR, and we thus regarded CSR2RB as a key gene in the progression of tubulointerstitial lesions in DN.

CD53, a component of the tetraspanin family member expressed in the immune compartment, has been reported to regulate integrin-related function by adjusting integrin signaling pathways and changing the general distribution of integrins located on the immune cell surface [38, 39]. As known that relevant biological pathways about integrins have been recognized as an essential factor in the pathogenesis of DN [19], we then supposed that CD53 might play potential roles in DN. In our study, we investigated that CD53 showed an up-regulated level in DN samples. This level had a negative correlation with GFR and a positive correlation with proteinuria in human tubulointerstitial samples of DN, further illustrating the impact of CD53 involved in tubulointerstitial lesions of DN.

LAPTM5 (lysosomal-associated protein transmembrane 5) can exert positive proinflammatory effects in macrophages through activating NF-κB and MAPK signaling pathways, promoting the production of proinflammatory cytokines [40]. The accumulation of macrophages in diabetic kidneys has been proven to associate with DN progression as this accumulation is closely related to interstitial myofibroblast accumulation and interstitial fibrosis scores [41–44]. In our study, we found that LAPTM5 was

augmented in DN, and this level was negatively correlated with GFR and actively correlated with proteinuria in human tubulointerstitial samples of DN. Combined with the results mentioned earlier and its diagnostic value (AUC value:0.843), we regarded LAPTM5 as a hub gene in DN.

Based on the predicted target miRNAs, lncRNAs, and circRNAs, we subsequently visualized the ceRNA networks by Cytoscape software. After a literature search for target miRNAs and lncRNAs, the down-regulated miRNAs and up-regulated lncRNAs were chosen for further analysis. Among them, hsa-miR-155-5p and hsa-miR-486-5p were down-regulated in DN samples [45, 46]. Additionally, lncRNAs of Nuclear Enriched Abundant Transcript-1 (NEAT1) and X-inactive specific transcript (XIST) have emphasized their high expression level in DN. Under hyperglycemia, the expression of NEAT1 was up-regulated in DN and led to renal tubular epithelial-mesenchymal transition (EMT) and renal fibrosis [47]. Another lncRNA of XIST also showed an up-regulated level in human DN, and this increased expression of XIST is closely related to renal fibrosis [48]. Based on the proposed ceRNA hypothesis, these observations may support the possibility that the regulatory pathway of NEAT1/XIST-hsa-miR-155-5p/hsa-miR-486-5p-CSF2RB might be involved in the molecular mechanisms of DN complicated with tubulointerstitial lesions.

Nevertheless, our article still has some limitations. First, this is a retrospective study that may command Vivo experiments with a larger sample size to prove our observations; Second, the definite function of the hub genes and the molecular mechanisms under DN needs to be further verified in future investigations before being applied in clinical practice, which will be the main direction of our future work.

# Conclusion

Our research identified three genes, namely CSF2RB, CD53, and LAPTM5, as the tubulointerstitial genes of DN and helped us further elucidate the molecular mechanisms of DN at the transcriptome level. The hypothetical pathway of NEAT1/XIST-hsa-miR-155-5p/hsa-miR-486-5p-CSF2RB may play a vital role in the progression of tubulointerstitial lesions implicated in early DN.

# Abbreviations

ceRNA:Competitive endogenous RNA; DN: Diabetic nephropathy; GEO: Gene expression omnibus; DEGs: Differentially expressed genes; PPI: Protein-protein Interaction; miRNA: MicroRNA; ROC: Receiver operating characteristic; GFR: Glomerular filtration rate; NEAT1: Nuclear enriched abundant transcript-1; XIST: X-inactive specific transcript; ncRNAs: Non-coding RNAs; lncRNAs: Long non-coding RNAs; CircRNAs: circular RNAs; GSEA: Gene set enrichment analysis; GO: Gene ontology; KEGG: Kyoto encyclopedia of genes and genomes; PCA: Principal component analysis; UMAP: Uniform manifold approximation and projection; NES: Normalized enrichment score; FDR: False discovery rates; MCODE: Minimal common oncology data elements; MCC: Maximal clique centrality; DMNC: Density of maximum neighborhood component; MNC: Maximum neighborhood component; AUC: Areas under the curve; ECM: Extracellular matrix; MMPs: Matrix metalloproteinases.

# Declarations


### Acknowledgments

We appreciate all investigators and staff in this study


### Authors' Contributions

DL conceived and designed the study. HC and XR analyzed the data for GEO datasets and wrote the manuscript. JJ and TY modified the manuscript. The author(s) read and

approved the final manuscript.


### Funding

None


### Availability of data and materials

The datasets presented in this study can be found in Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/). The names of the repository/repositories and accession number(s) can be found in the article.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not Applicable.

### Competing interests

The authors declare no competing interests.


### Author details

[1]Department of Nephrology, Tianjin Medical University General Hospital, Tianjin, China. [2]Department of Urology, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China.


# References


1. Samsu N. Diabetic Nephropathy: Challenges in Pathogenesis, Diagnosis, and Treatment. Biomed Res Int. 2021;2021:1497449.

2. Maezawa Y. Takemoto M, Yokote K. Cell biology of diabetic nephropathy: Roles of endothelial cells, tubulointerstitial cells and podocytes. J Diabetes Investig. 2015;6(1):3–15.
3. Santoro D. Torreggiani M, Pellicanò V. Cernaro V, Messina RM. Longhitano E, *et al.* Kidney Biopsy in Type 2 Diabetic Patients: Critical Reflections on Present Indications and Diagnostic Alternatives. Int J Mol Sci. 2021;22(11):5425.
4. Warren AM. Knudsen ST, Cooper ME. Diabetic nephropathy: an insight into molecular mechanisms and emerging therapies. Expert Opin Ther Targets. 2019;23(7):579–91.
5. Zhong F. Chen H, Xie Y. Azeloglu EU, Wei C. Zhang W, *et al.* Protein S Protects against Podocyte Injury in Diabetic Nephropathy. J Am Soc Nephrol. 2018;29(5):1397–410.
6. Salem RM. Todd JN, Sandholm N. Cole JB, Chen WM. Andrews D, *et al.* Genome-Wide Association Study of Diabetic Kidney Disease Highlights Biology Involved in Glomerular Basement Membrane Collagen. J Am Soc Nephrol. 2019;30(10):2000–16.
7. Papadopoulou-Marketou N. Kanaka-Gantenbein C, Marketos N. Chrousos GP, Papassotiriou I. Biomarkers of diabetic nephropathy: A 2017 update. Crit Rev Clin Lab Sci. 2017;54(5):326–42.
8. Barrett T. Wilhite SE, Ledoux P. Evangelista C, Kim IF. Tomashevsky M, *et al.* NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41:991–5.
9. Gu Z. Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847–9.
10. Joly JH. Lowry WE, Graham NA. Differential Gene Set Enrichment Analysis: a statistical approach to quantify the relative enrichment of two gene sets. Bioinformatics. 2021;36(21):5247–54.
11. Yu G. Wang LG, Han Y. He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.
12. Zheng Q. Wang XJ. GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. Nucleic Acids Res. 2008;36:358–63.
13. Kanehisa M. Furumichi M, Tanabe M. Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1):D353–61.
14. Szklarczyk D. Gable AL, Nastou KC. Lyon D, Kirsch R. Pyysalo S, *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021;49(D1):D605–12.
15. Otasek D. Morris JH, Bouças J. Pico AR, Demchak B. Cytoscape Automation: empowering workflow-based network analysis. Genome Biol. 2019;20(1):185.
16. Chin CH. Chen SH, Wu HH. Ho CW, Ko MT. Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8(Suppl 4):11.
17. Li JH. Liu S, Zhou H. Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res. 2014;42:92–7.
18. Lay AC. Hale LJ, Stowell-Connolly H. Pope RJP, Nair V. Ju W, *et al.* IGFBP-1 expression is reduced in human type 2 diabetic glomeruli and modulates β1-integrin/FAK signalling in human podocytes.

Diabetologia. 2021;64(7):1690–702.
19. Woroniecka KI. Park AS, Mohtat D. Thomas DB, Pullman JM. Susztak K. Transcriptome analysis of human diabetic kidney disease. Diabetes. 2011;60(9):2354–69.
20. Schmid H. Boucherot A, Yasuda Y. Henger A, Brunner B. Eichinger F, *et al.* Modular activation of nuclear factor-kappaB transcriptional programs in human diabetic nephropathy. Diabetes. 2006;55(11):2993–3003.
21. Salmena L. Poliseno L, Tay Y. Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011;146(3):353–8.
22. Bazzi C. Petrini C, Rizza V. Arrigo G, D'Amico G. A modern approach to selectivity of proteinuria and tubulointerstitial damage in nephrotic syndrome. Kidney Int. 2000;58(4):1732–41.
23. Gilbert RE. Proximal Tubulopathy: Prime Mover and Key Therapeutic Target in Diabetic Kidney Disease. Diabetes. 2017;66(4):791–800.
24. Garcia-Fernandez N. Jacobs-Cachá C, Mora-Gutiérrez JM. Vergara A, Orbe J. Soler MJ. Matrix Metalloproteinases in Diabetic Kidney Disease. J Clin Med. 2020;9(2):472.
25. Kolset SO. Reinholt FP, Jenssen T. Diabetic nephropathy and extracellular matrix. J Histochem Cytochem. 2012;60(12):976–86.
26. Stokes MB. Holler S, Cui Y. Hudkins KL, Eitner F. Fogo A, *et al.* Expression of decorin, biglycan, and collagen type I in human renal fibrosing disease. Kidney Int. 2000;57(2):487–98.
27. Zakiyanov O. Kalousová M, Zima T. Tesař V. Matrix metalloproteinases and tissue inhibitors of matrix metalloproteinases in kidney disease. Adv Clin Chem. 2021;105:141–212.
28. Kim SS. Shin N, Bae SS. Lee MY, Rhee H. Kim IY, *et al.* Enhanced expression of two discrete isoforms of matrix metalloproteinase-2 in experimental and human diabetic nephropathy. PLoS ONE. 2017;12(2):e0171625.
29. Goldszmid RS. Trinchieri G. The price of immunity. Nat Immunol. 2012;13(10):932–8.
30. Rayego-Mateos S. Morgado-Pascual JL, Opazo-Ríos L. Guerrero-Hue M, García-Caballero C. Vázquez-Carballo C, *et al.* Pathogenic Pathways and Therapeutic Approaches Targeting Inflammation in Diabetic Nephropathy. Int J Mol Sci. 2020;21(11):3798.
31. Shahzad K. Bock F, Dong W. Wang H, Kopf S. Kohli S, *et al.* Nlrp3-inflammasome activation in non-myeloid-derived cells aggravates diabetic nephropathy. Kidney Int. 2015;87(1):74–84.
32. Flyvbjerg A. The role of the complement system in diabetic nephropathy. Nat Rev Nephrol. 2017;13(5):311–8.
33. Tesch GH. Diabetic nephropathy - is this an immune disorder? Clin Sci (Lond). 2017;131(16):2183–99.
34. Croxford AL. Lanzinger M, Hartmann FJ. Schreiner B, Mair F. Pelczar P, *et al.* The Cytokine GM-CSF Drives the Inflammatory Signature of CCR2 + Monocytes and Licenses Autoimmunity. Immunity. 2015;43(3):502–14.

35. Dougan M. Dranoff G, Dougan SK. GM-CSF. IL-3, and IL-5 Family of Cytokines: Regulators of Inflammation. Immunity. 2019;50(4):796–811.
36. Wicks IP. Roberts AW. Targeting GM-CSF in inflammatory diseases. Nat Rev Rheumatol 2016;12(1):37–48.
37. Wada J. Makino H. Inflammation and the pathogenesis of diabetic nephropathy. Clin Sci (Lond). 2013;124(3):139–52.
38. Wee JL. Schulze KE, Jones EL. Yeung L, Cheng Q. Pereira CF, *et al.* Tetraspanin CD37 Regulates β2 Integrin-Mediated Adhesion and Migration in Neutrophils. J Immunol. 2015;195(12):5770–9.
39. Rocha-Perugini V. González-Granado JM, Tejera E. López-Martín S, Yañez-Mó M. Sánchez-Madrid F. Tetraspanins CD9 and CD151 at the immune synapse support T-cell integrin signaling. Eur J Immunol. 2014;44(7):1967–75.
40. Glowacka WK. Alberts P, Ouchida R. Wang JY, Rotin D. LAPTM5 protein is a positive regulator of proinflammatory signaling pathways in macrophages. J Biol Chem. 2012;287(33):27691–702.
41. Calle P. Hotter G. Macrophage Phenotype and Fibrosis in Diabetic Nephropathy. Int J Mol Sci. 2020;21(8):2806.
42. Awad AS. Kinsey GR, Khutsishvili K. Gao T, Bolton WK. Okusa MD. Monocyte/macrophage chemokine receptor CCR2 mediates diabetic renal injury. Am J Physiol Renal Physiol. 2011;301(6):1358–66.
43. Nguyen TQ. Tarnow L, Andersen S. Hovind P, Parving HH. Goldschmeding R, *et al.* Urinary connective tissue growth factor excretion correlates with clinical markers of renal disease in a large population of type 1 diabetic patients with diabetic nephropathy. Diabetes Care. 2006;29(1):83–8.
44. Yonemoto S. Machiguchi T, Nomura K. Minakata T, Nanno M. Yoshida H. Correlations of tissue macrophages and cytoskeletal protein expression with renal fibrosis in patients with diabetes mellitus. Clin Exp Nephrol. 2006;10(3):186–92.
45. Barutta F. Tricarico M, Corbelli A. Annaratone L, Pinach S. Grimaldi S, *et al.* Urinary exosomal microRNAs in incipient diabetic nephropathy. PLoS ONE. 2013;8(11):e73798.
46. Duan YR. Chen BP, Chen F. Yang SX, Zhu CY. Ma YL, *et al.* LncRNA lnc-ISG20 promotes renal fibrosis in diabetic nephropathy by inducing AKT phosphorylation through miR-486-5p/NFAT5. J Cell Mol Med. 2021;25(11):4922–37.
47. Yang YL. Xue M, Jia YJ. Hu F, Zheng ZJ. Wang L, *et al.* Long noncoding RNA NEAT1 is involved in the protective effect of Klotho on renal tubular epithelial cells in diabetic kidney disease through the ERK1/2 signaling pathway. Exp Mol Med. 2020;52(2):266–80.
48. Yang J. Shen Y, Yang X. Long Y, Chen S. Lin X, *et al.* Silencing of long noncoding RNA XIST protects against renal interstitial fibrosis in diabetic nephropathy via microRNA-93-5p-mediated inhibition of CDKN1A. Am J Physiol Renal Physiol. 2019;317(5):1350–8.

# Tables

**Table 1** The details of the intersected hub genes.

| Gene symbol | Full name | Description |
|---|---|---|
| CD53 | CD53 Molecule | CD53 is a member of the transmembrane 4 superfamilies. It contributes to the transduction of CD2-generated signals in T cells and natural killer cells and has been suggested to play a role in growth regulation. |
| LAPTM5 | Lysosomal Protein Transmembrane 5 | LAPTM5 is a transmembrane receptor that is associated with lysosomes. |
| CSF2RB | Colony Stimulating Factor 2 Receptor Subunit Beta | CSF2RB is a high-affinity receptor for interleukin-3, interleukin-5, and granulocyte-macrophage colony-stimulating factor. |

# Figures

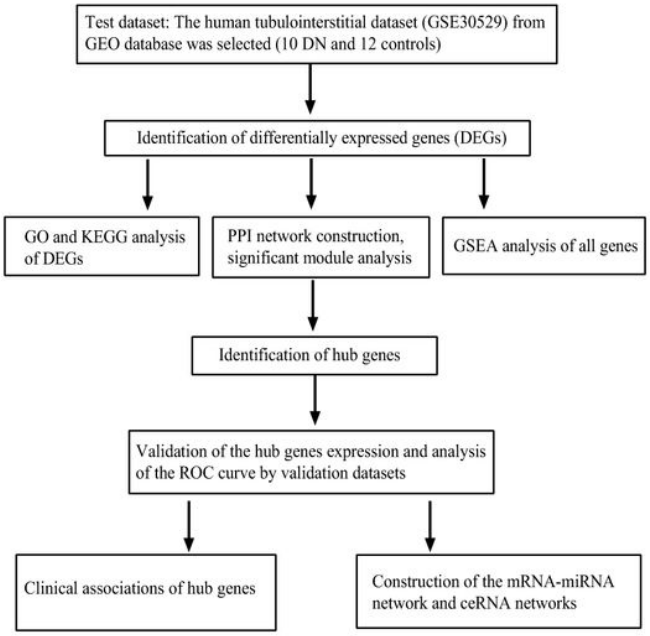
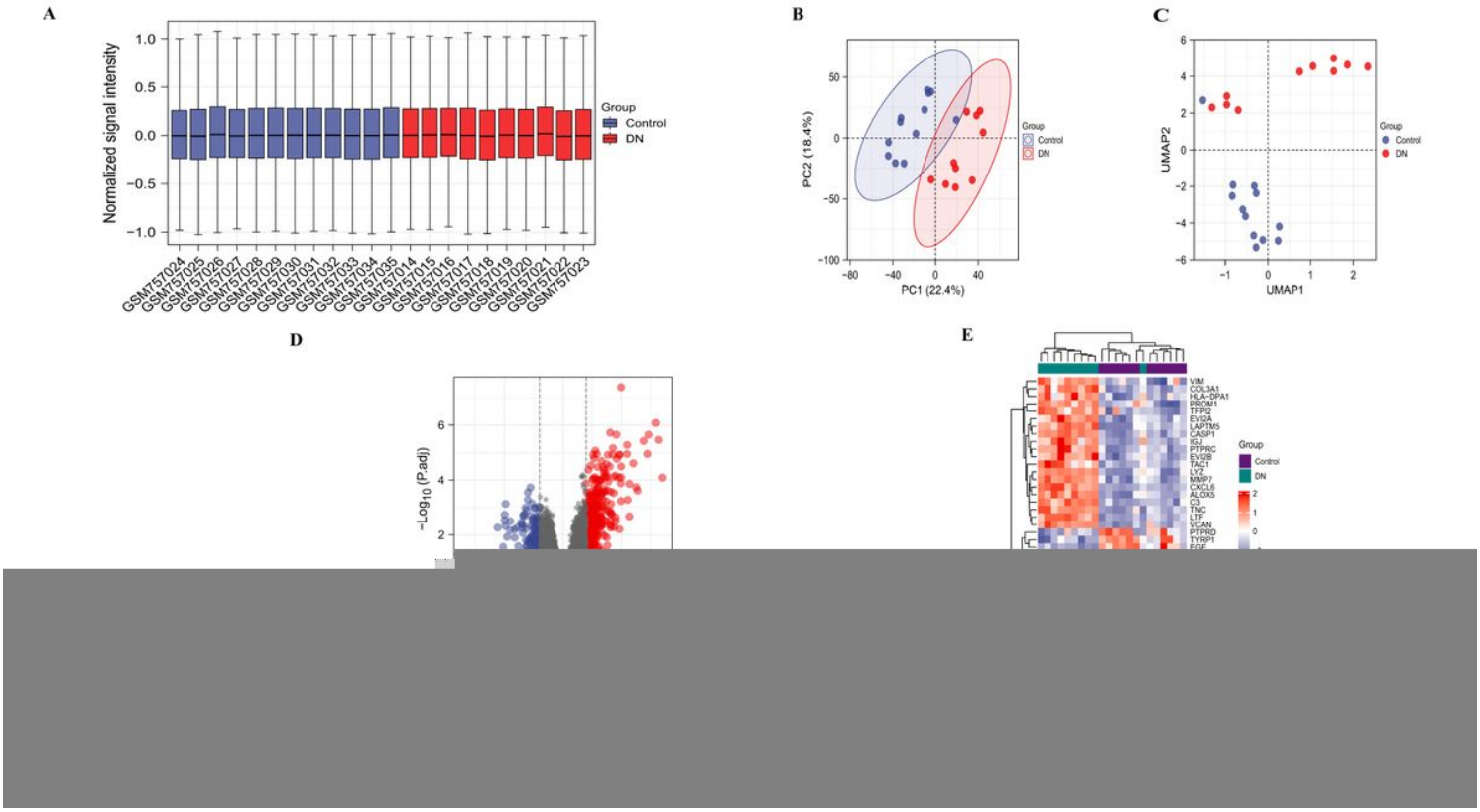



**Fig. 1** The flowchart of this study.

## Figure 1

See image above for figure legend.




## Figure 2

See image above for figure legend.

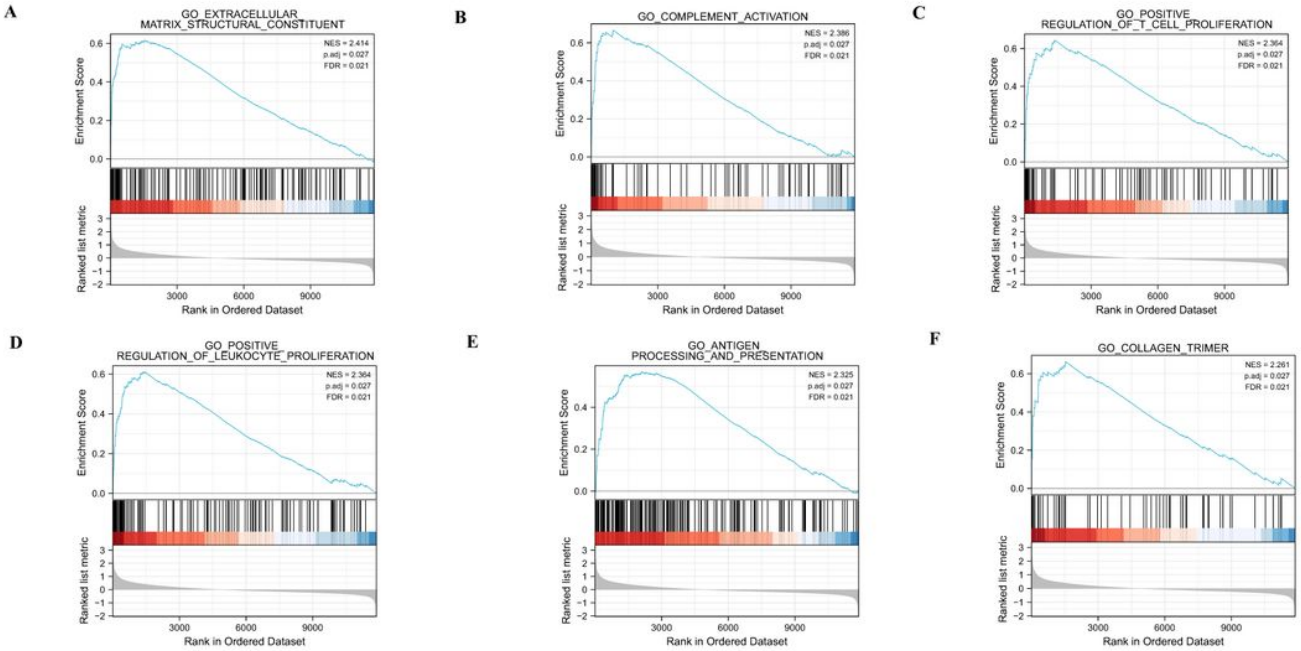



**Fig. 3** GSEA plot showed the most enriched gene sets in DN. **A** The most signifcant enriched gene set was extracellar matrix structural constituent (ES = 0.617, NES = 2.414, $p < 0.05$). **B** The second signifcant gene set was complement activation (ES = 0.666, NES = 2.386, $p < 0.05$). **C** The third signifcant enriched gene set was positive regulation of T cell production (ES = 0.644, NES = 2.364, $p < 0.05$). **D** The fourth signifcant enriched gene set was positive rgulation of leukocyte proliferation (ES = 0.612, NES = 2.364, $p < 0.05$). **E** The fifth signifcant enriched gene set was antigen processing and presentation (ES = 0.570, NES = 2.325, $p < 0.05$). **F** The sixth signifcant enriched gene set was collagen trimer (ES = 0.662, NES = 2.261, $p < 0.05$). The defined criteria for significant gene sets were |Normalized Enrichment Score (NES)| > 1, False Discovery Rates (FDR) < 0.05 and p-value < 0.05. DN: diabetic nephropathy; ES: enrichment score; NES: normalized enrichment score.

## Figure 3

See image above for figure legend.

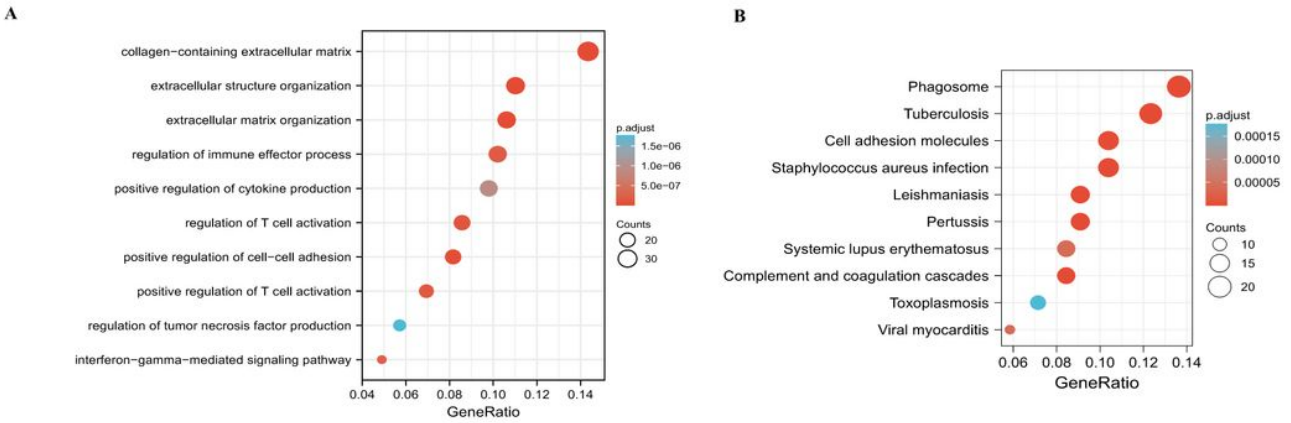



**Fig. 4** Enrichment analysis of DEGs. **A** The bubble plot showing the top 10 enriched biological processes of DEGs. **B** The bubble plot showing the top 10 enriched pathways of KEGG in DN samples. DN: diabetic nephropathy.

## Figure 4

See image above for figure legend.

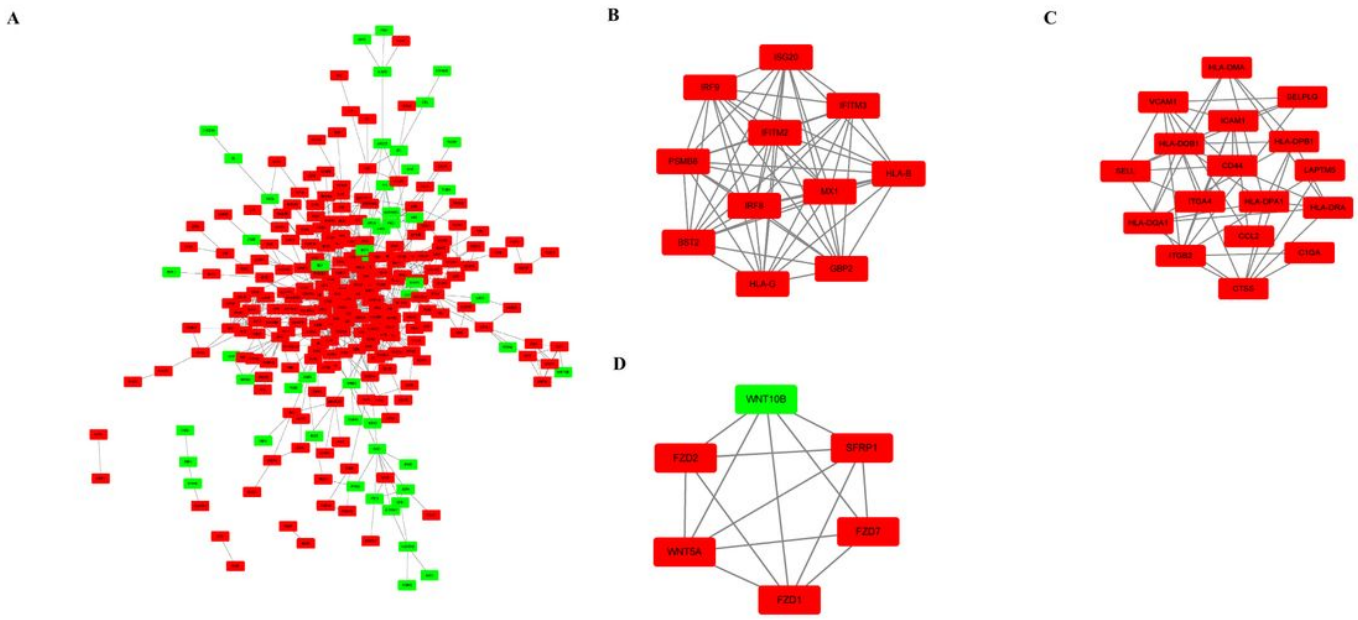



**Fig. 5** PPI network and three significant modules. **A** The PPI network includes 277 nodes and 803 edges. **B-D** three modules are extracted with MCODE score > 5. Red nodes indicate up-regulated DEGs and green nodes indicate down-regulated DEGs. PPI: protein-protein interaction.

## Figure 5

See image above for figure legend.

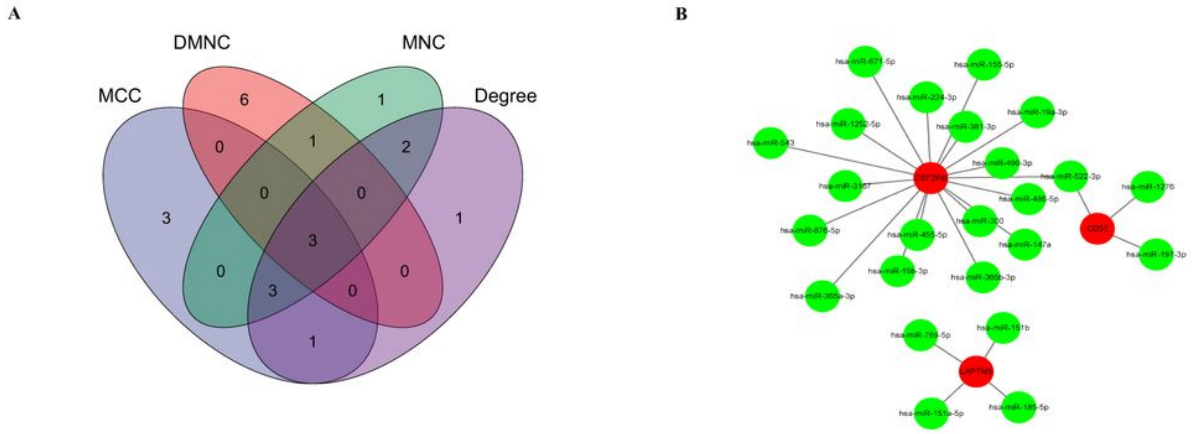



**Fig. 6** Identification of hub genes and the co-expressed network of mRNAs-miRNAs. **A** The Venn diagram of three hub genes screened by four algorithms. **B** The mRNA-miRNA network was visualized by Cytoscape. CSF2RB had 18 target miRNAs, CD53 had 3 target miRNAs, and LAPTM5 had 4 target miRNAs. Red circles represented mRNAs and green circles represented miRNAs.

## Figure 6

See image above for figure legend.

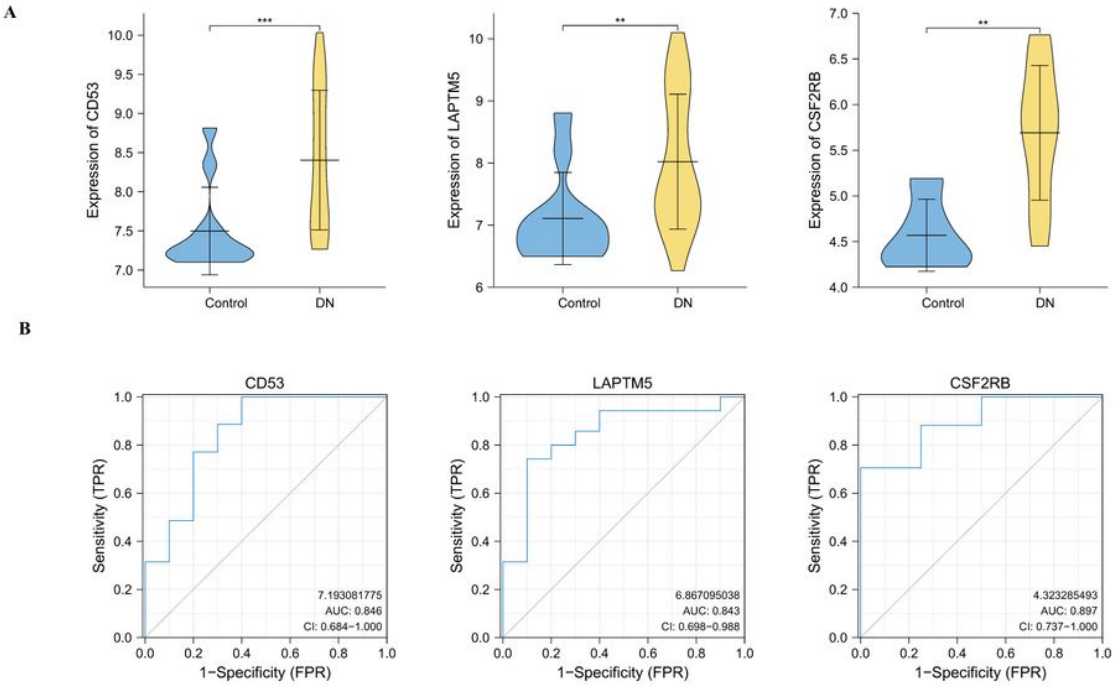



**Fig. 7** Verification and ROC curve analysis of the hub genes by another human tubulointerstitial dataset. **A** CD53, LAPTM5, and CSF2RB are up-regulated in DN samples compared with controls. ***: p<0.001, **: p<0.01, *: p<0.05. **B** CSF2RB has the highest score (AUC value: 0.897), CD53 has the second highest score (AUC value: 0.846) and LAPTM5 has the third highest score (AUC value: 0.843). ROC: Receiver Operating Characteristic; AUC: Area under the curve; DN: diabetic nephropathy.

## Figure 7

See image above for figure legend.

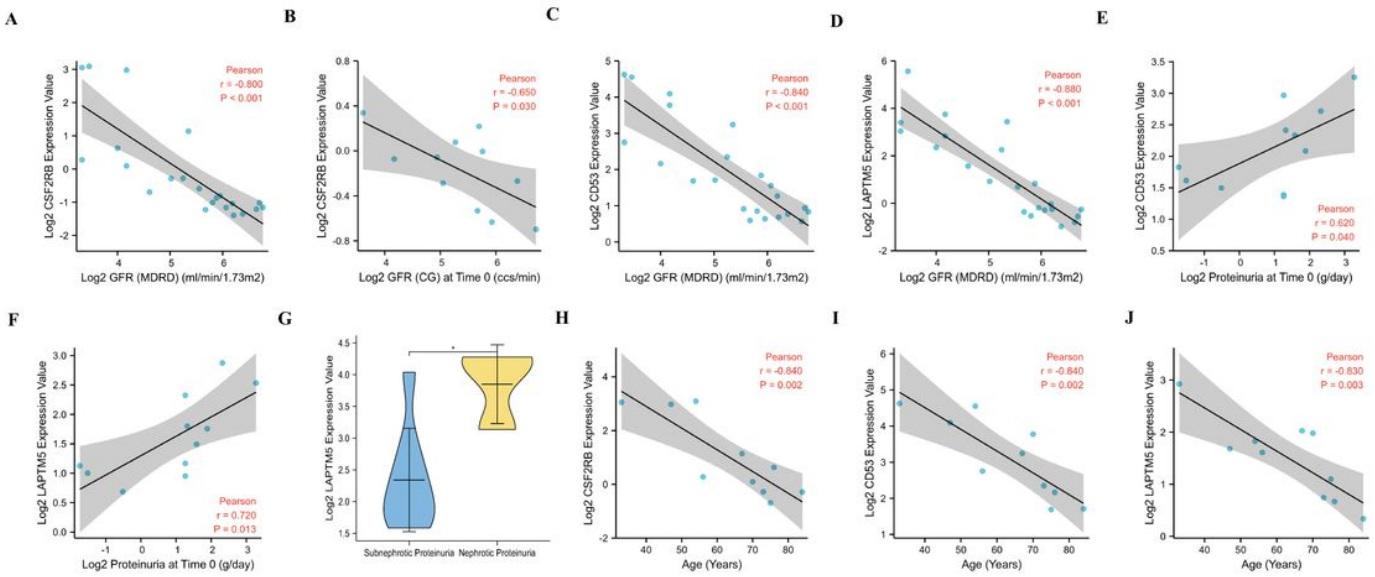



**Fig. 8** Explorations between clinical features and hub genes (CSF2RB, CD53 and LAPTM5) in human tubulointerstitial samples. **A** The mRNA expression of CSF2RB was negatively correlated with GFR (MDRD). **B** The mRNA expression of CSF2RB was negatively correlated with GFR (CG). **C** The mRNA expression of CD53 was negatively correlated with GFR (MDRD). **D** The mRNA expression of LAPTM5 was negatively correlated with GFR (MDRD). **E** The mRNA expression of CD53 was positively correlated with proteinuria. **F** The mRNA expression of LAPTM5 was positively correlated with proteinuria. **G** The mRNA expression of LAPTM5 in nephrotic proteinuria group was higher than that of subnephrotic proteinuria group. **H** The mRNA expression of CSR2RB was negatively correlated with age. **I** The mRNA expression of CD53 was negatively correlated with age. **J** The mRNA expression of LAPTM5 was negatively correlated with age. GFR: glomerular filtration rate; MDRD: modification of diet in renal disease; CG: Cockcroft Gault; *: $p < 0.05$; $P < 0.05$ was considered significant.

## Figure 8

See image above for figure legend.

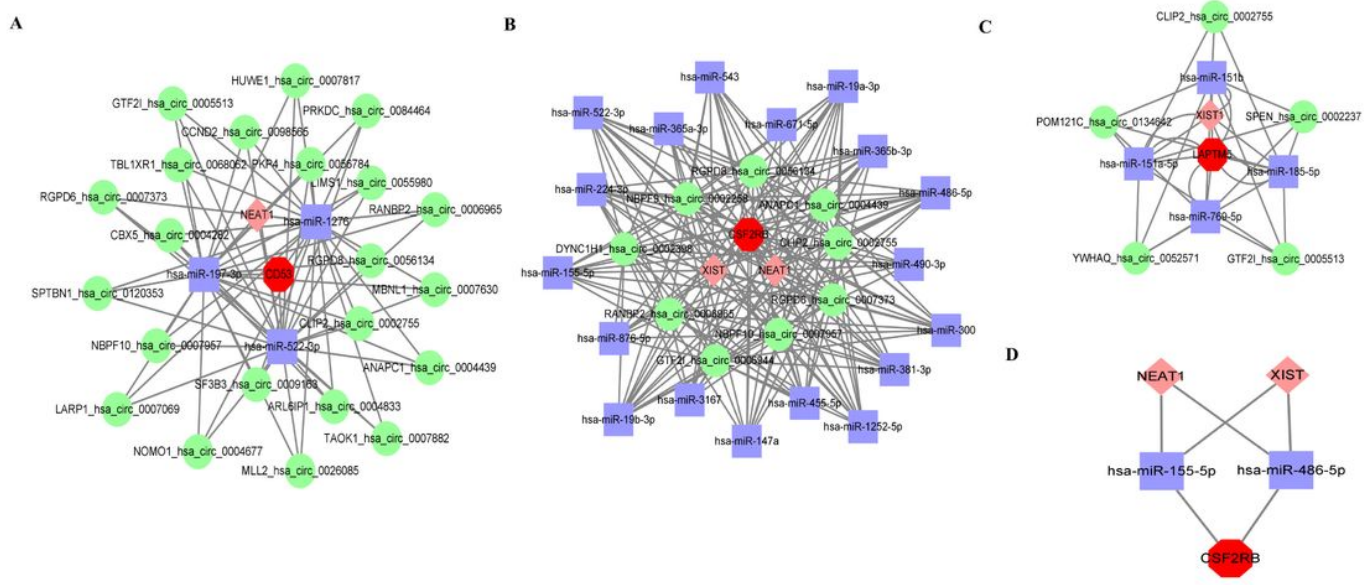



**Fig. 9** Three ceRNA networks of CD53, CSF2RB and LAPTM5 and the potential RNA regulatory pathways. **A** The ceRNA network of CD53. **B** The ceRNA network of CSF2RB. **C** The ceRNA network of LAPTM5. **D** NEAT1/XIST-miR-155-5p/miR-486-5p-CSR2RB. Red octagons note the hub genes, purple rectangles note miRNAs, pink diamonds note lncRNAs, and syan circles note circRNAs.

## Figure 9

See image above for figure legend.