Research Article

Investigation of the Mechanism of Complement System in Diabetic Nephropathy via Bioinformatics Analysis

Bojun Xu,[1] Lei Wang,[2] Huakui Zhan ◉**,[1] Liangbin Zhao,[1] Yuehan Wang,[1] Meng Shen,[3] Keyang Xu,[4] Li Li,[1] Xu Luo,[5] Shasha Zhou,[1] Anqi Tang,[1] Gang Liu,[1] Lu Song,[1] and Yan Li** ◉**[1]**

[1]*Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, 610072 Sichuan, China*
[2]*Key Laboratory of Chinese Internal Medicine of Ministry of Education and Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China*
[3]*Chengdu Seventh People's Hospital, Chengdu, 610213 Sichuan, China*
[4]*Centre for Cancer and Inflammation Research, School of Chinese Medicine, Hong Kong Baptist University, Hong Kong*
[5]*The First Affiliated Hospital of Guangxi University of Chinese Medicine, Nanning 530023, China*

Correspondence should be addressed to Yan Li; yanlieu@qq.com

Objectives. Diabetic nephropathy (DN) is a major cause of end-stage renal disease (ESRD) throughout the world, and the identification of novel biomarkers via bioinformatics analysis could provide research foundation for future experimental verification and large-group cohort in DN models and patients. *Methods.* GSE30528, GSE47183, and GSE104948 were downloaded from Gene Expression Omnibus (GEO) database to find differentially expressed genes (DEGs). The difference of gene expression between normal renal tissues and DN renal tissues was firstly screened by GEO2R. Then, the protein-protein interactions (PPIs) of DEGs were performed by STRING database, the result was integrated and visualized via applying Cytoscape software, and the hub genes in this PPI network were selected by MCODE and topological analysis. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were carried out to determine the molecular mechanisms of DEGs involved in the progression of DN. Finally, the Nephroseq v5 online platform was used to explore the correlation between hub genes and clinical features of DN. *Results.* There were 64 DEGs, and 32 hub genes were identified, enriched pathways of hub genes involved in several functions and expression pathways, such as complement binding, extracellular matrix structural constituent, complement cascade related pathways, and ECM proteoglycans. The correlation analysis and subgroup analysis of 7 complement cascade-related hub genes and the clinical characteristics of DN showed that C1QA, C1QB, C3, CFB, ITGB2, VSIG4, and CLU may participate in the development of DN. *Conclusions.* We confirmed that the complement cascade-related hub genes may be the novel biomarkers for DN early diagnosis and targeted treatment.

1. Introduction

Diabetic nephropathy (DN), the most common microvascular complication of diabetes, is becoming the leading cause of ESRD worldwide [1]. The typical clinical features of DN consist of the decreased glomerular filtration rate and persistent proteinuria [2]. Currently, there are several pathways reported to be involved in the pathogenesis of DN, including the activation of polyol and protein kinase C (PKC) pathway [3], the generation of advanced glycation end products (AGE) [4], and intraglomerular hypertension caused by glo-

merular hyperfiltration [5]. In addition, microinflammation and subsequent extracellular matrix (ECM) pathways are also involved in the progression of DN [6]. Furthermore, the complement system is reported to involve in the development of DN but mostly focused on the lectin pathway [7, 8]. There are few studies aimed to explore the potential significance of the classic and alternative complement pathways in the pathogenesis of DN at the transcriptional level [9]. However, the reported studies of these complement pathways are not comprehensive and detailed and other pathways of the complement system as well as their clinicopathological
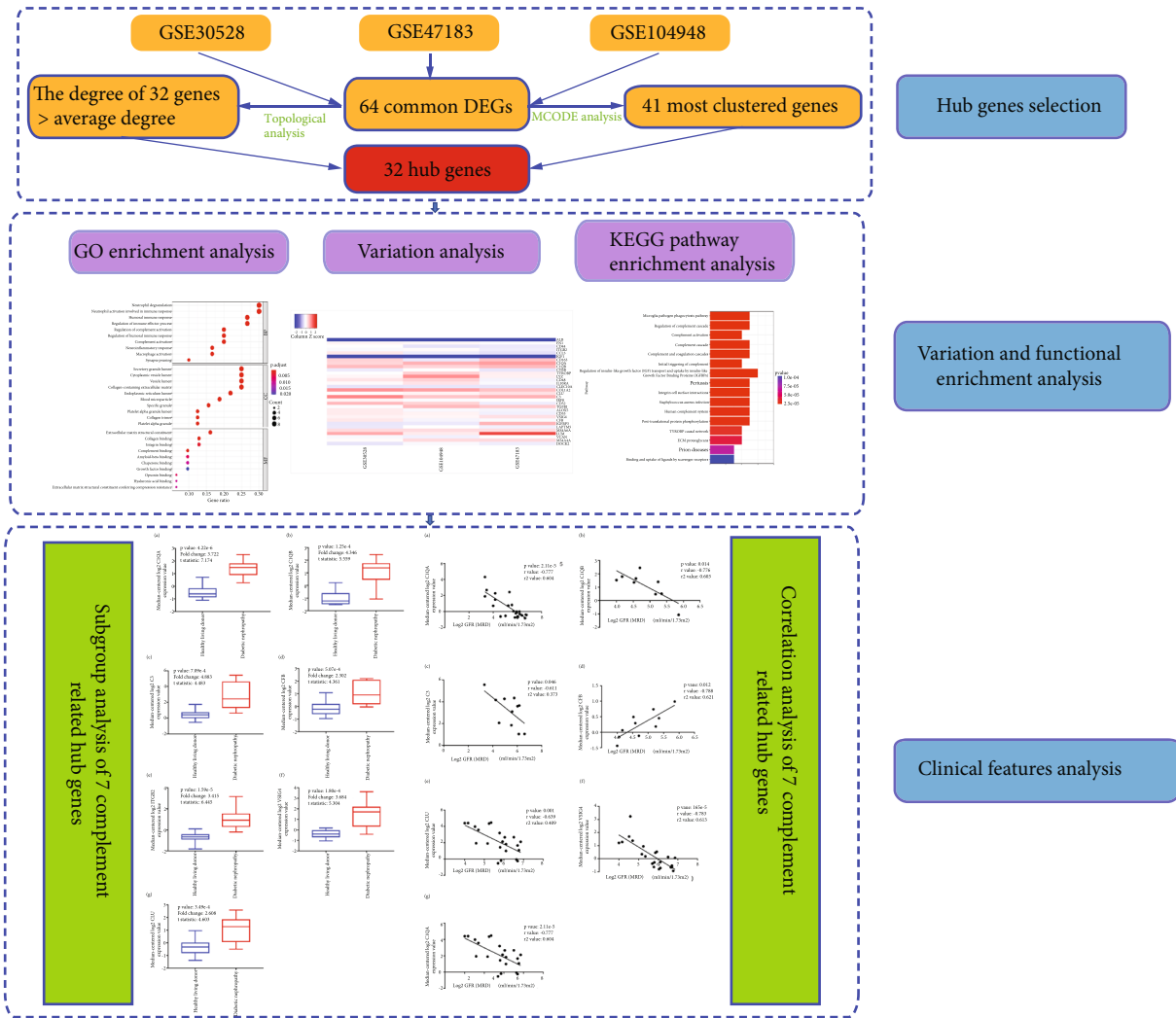
FIGURE 1: A flow chart based on an integration strategy of multiple-microarray analysis.

relevance with DN are still poorly understood. Taking together, the pathogenic and molecular mechanisms of DN have not been elucidated comprehensively yet, the prevalence rate of DN is high, treatment is difficult, and prognosis is poor [10, 11]. Therefore, new diagnostic biomarkers and novel therapies of DN should be further studied that will be beneficial for improving the clinical prognostic of DN.

Genome-wide transcriptome analysis using microarray and bioinformatics technology enable the identification of biomarkers for disease progression and gain insights into the disease pathogenesis and molecular classification [12, 13]. With the wide application of genome transcriptome analysis, a large amount of core slice data has been produced and stored in a public database, including the GEO database [14]. Recently, some microarray data analyses about DN have been carried out, and numerous differentially expressed genes (DEGs) have been identified. Researchers have used microarray data from DN models of different species to determine molecular mechanisms and genetic factors involved in DN [15, 16]. Previous bioinformatics analyses using human DN gene chip data (GSE47183) from the GEO database found that the VSIG4, CD163, C1QA,

C1QB, MS4A6A, COL6A3, COL1A2, CD44, FN1, NPHS1, WT1, PLCE1, TNNT2, TNNI1, and TNNC1 genes played important roles in DN progression through ECM-receptor interaction, PI3K-Akt signaling pathway, focal adhesion, proteoglycans in cancer, and complement and coagulation cascades [17]. Yang et al. [18] identified hub genes associated with DN using GSE30528 chip data, including VEGFA, ITGA3, ITGB5, COL4A3, COL4A5, CBLB, and CCL19. In addition, the key miRNAs related to DN were predicted based on the hub genes, including miR-200b/c, miR-29a/b/c, miR-25, miR-27, miR-23, miR-181, miR-17, miR-506, and miR-124a. All these genes were mainly enriched in the ECM-receptor interaction and PI3K/Akt signaling pathways to initiate the pathogenesis of DN [18]. Liu et al. [19] carried out a weighted gene coexpression network analysis of GSE104948 chip data and discovered that FCER1G played a crucial role in the pathogenesis of DN.

In this study, we reanalyzed three microarray datasets, GSE30528, GSE47183, and GSE104948; key biomarkers were identified by selecting the significant differentially expressed genes (DEG) between DN and normal glomerular samples. Then, biological processes and signaling pathways that

participated in DN will be explored on basis of DEGs. The pathogenesis of DN was studied by GO and KEGG pathway enrichment as well as PPI network analysis. Additionally, the Nephroseq v5 online platform was used to analyze correlations and to perform subgroup analysis among the hub genes and clinical features of DN to further explore the pathogenesis, pathophysiological and molecular mechanisms involved in DN. In conclusion, a total of 64 DEGs and 32 hub genes were identified, which may be potential diagnostic biomarkers and therapeutic targets for preventing the occurrence and development of DN; the flow chart of this study is shown in Figure 1.

2. Methods and Materials

2.1. Microarray Data Analysis. We downloaded the datasets including GSE30528 [9], GSE47183 [20], and GSE104948 [16] from GEO (http://www.ncbi.nlm.nih.gov/geo) [21], an international public functional genomics data repository of next-generation sequence, chips, and microarrays. The DEGs in DN and normal renal tissues were firstly selected by the GEO2R online tool [22] with |LogFC| >1 and $P < 0.05$. Subsequently, the overlap DEGs among the three datasets were shown in the Venn diagram via using FunRich, a software widely used for gene and protein functional enrichment and interaction network analysis [23].

2.2. Protein-Protein Interaction (PPI) Network Construction, Topological Analysis, and Hub Gene Identification. A PPI network of overlapping DEGs was established based on the STRING platform (https://string-db.org/) [24] to retrieve nearly all functional interactions among the expressed proteins. Protein interaction information derived from the STRING database was imported into the Cytoscape software where the interaction information was integrated and visualized [25].

Then, the most significant modules in the PPI network were selected by molecular complex detection (MCODE) [26]; the parameters of clustering and scoring were MCODE score ≥ 2, degree cutoff = 2, node score cutoff = 0.2, max depth = 100, and k-score = 2. At the same time, topological analysis was performed by using NetworkAnalyzer in Cytoscape software and four topological features (degree, betweenness centrality, average shortest path length, and closeness centrality) were analyzed [27]. Subsequently, selecting nodes with degrees higher than the average number, taking the intersection of the topological analysis, and clustering analysis result as the hub genes. If the LogFC < 0, the expression of the hub gene was deemed to be downregulated, while LogFC > 0, gene expression was deemed to be upregulated. Additionally, a network of the hub genes and their coexpression network was performed by using FunRich software.

2.3. GO and KEGG Pathway Enrichment Analysis. GO is a comprehensive and widely used database for the classification of gene functions, consisting of biological process (BP), molecular function (MF), and cell component (CC) [28]. KEGG (http://www.kegg.jp/) is an encyclopedia of genomes, which links genomic information with higher-order func-
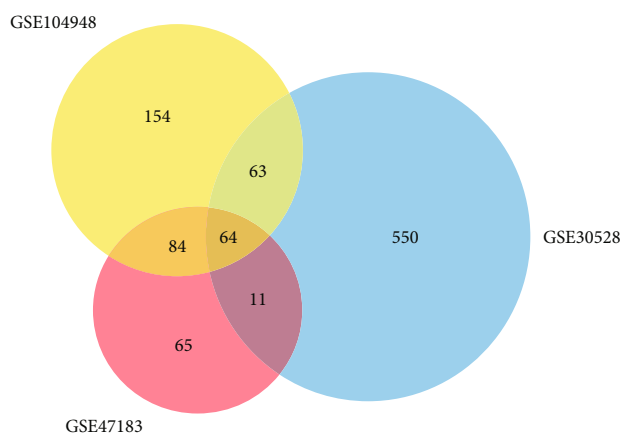

Figure 2: Overlapping DEGs of multiple microarrays.

tional information to capture significantly enriched biological pathways [29]. In our study, GO functional annotation and KEGG pathway enrichment analysis were performed through applying the clusterProfiler package of R software, when $P < 0.05$ was deemed as a screening threshold.

2.4. The Association between Hub Genes and Clinical Features of DN. Correlation analysis and subgroup analysis between hub genes and clinical features were carried out via the Nephroseq v5 online tool (http://v5.nephroseq.org) [30] to confirm the potential functions of hub genes participated in DN.

2.5. Statistical Analysis. Pearson's correlation analysis between hub genes and glomerular filtration rate (GFR) in patients with DN was conducted via applying Nephroseq v5. Comparisons between two subgroups were carried out via an unpaired Student's t-test. All tests were two-tailed, with a P value < 0.05 considered statistically significant. The statistical analyses were performed by using GraphPad Prism (version 7.0; GraphPad Software, La Jolla, California).

3. Results

3.1. Identification of DEGs in DN. We used the GEO database to search the gene expression profiles of GSE30528, GSE47183, and GSE104948 in DN and normal renal samples. These three datasets contain 51 normally functioning human glomerulus samples and 35 diabetic human glomerulus samples. These study samples were derived from healthy human transplant donors, diagnostic renal biopsies, and tumor-nephrectomy specimens. Then, the gene information obtained from the database was analyzed by GEO2R. The cutoff criteria were |LogFC| > 1 and $P < 0.05$. We found 224, 687, and 365 DEGs in GSE47183, GSE30528, and GSE104948 datasets, respectively. Subsequently, these DEGs from three datasets were imported into FunRich to identify the common DEGs; a total of 64 DEGs were detected (Figure 2).

3.2. PPI Network Analysis and Hub Gene Selection. A total of 64 common DEGs were imported into STRING online database to construct the PPI network, the interaction network was based on the selected targets with a medium confidence
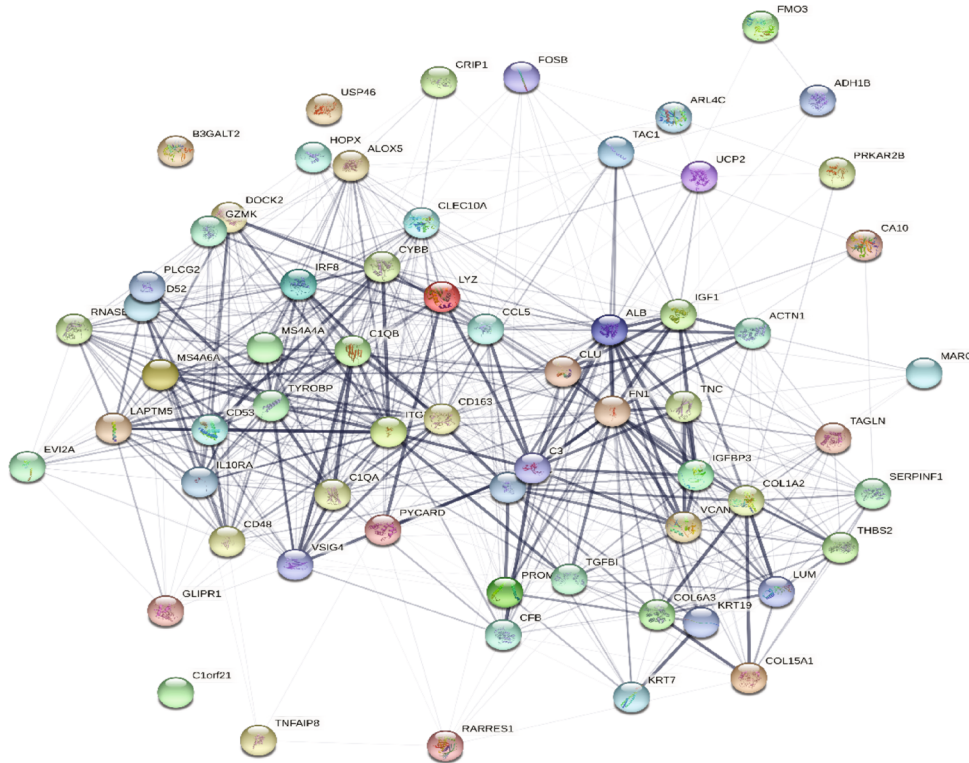
Figure 3: The PPI network of overlapping DEGs of three microarray datasets. Each circular node represents a protein target, and the 3D structure in the circular nodes shows the protein spatial structure. The lines among different nodes represent the association among potential protein targets, while the width of lines was according to the action intensity.
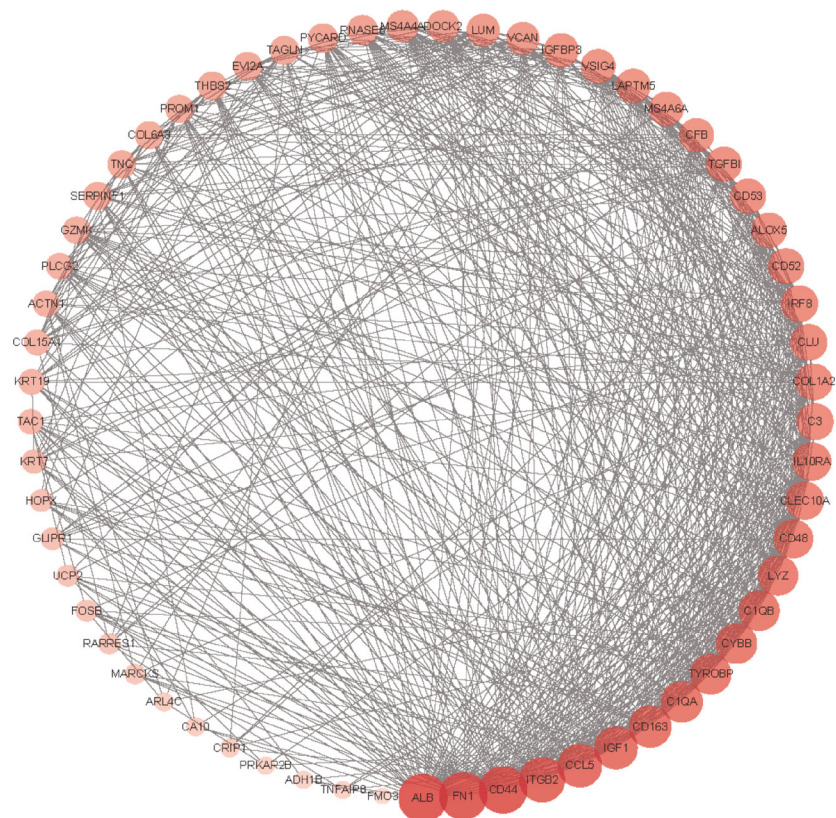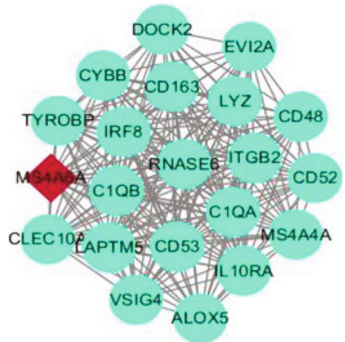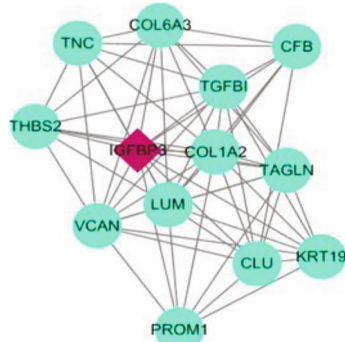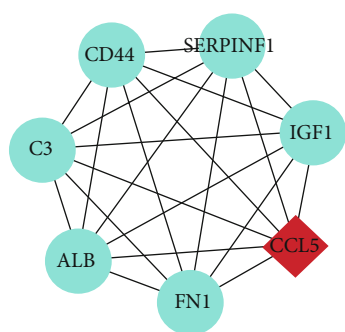

Figure 4: PPI network visualization and analysis.

TABLE 1: The most significant clusters of common DEGs.

Cluster	Network	Nodes	Edges	Score	Cluster genes
1		21	203	20.3	EVI2A, RNASE6, LAPTM5, LYZ, CD53, CLEC10A, TYROBP, DOCK2, ALOX5, IL10RA, C1QA, MS4A6A, CD52, C1QB, ITGB2, MS4A4A, CD48, CD163, VSIG4, CYBB, IRF8
2		13	62	10.3	TGFBI, VCAN, THBS2, CFB, CLU, COL1A2, IGFBP3, COL6A3, LUM, KRT19, TNC, PROM1, TAGLN
3		7	21	7	CD44, ALB, FN1, CCL5, SERPINF1, IGF1, C3

The red nodes display the core genes of the cluster; the green nodes represent the normal cluster genes. Furthermore, the density of lines among different nodes shows the interaction relationship between cluster genes.

score of 0.15, and finally, 652 edges and 64 key nodes were embodied (Figure 3). PPI network of common DEGs was visualized by Cytoscape (Figure 4), and the three most significant modules were recognized by the MCODE plug-in of Cytoscape.

Among these three modules, a total of 41 DEGs were identified (Table 1). In addition, the average node degree is 21.377 after the topological analysis of common DEGs, we selected the genes higher than the average degree, taking the intersection of the genes higher than the average degree, and most significant clustering genes result as the hub genes. The hub genes included 2 downregulated genes (ALB and IGF1) and 30 upregulated genes (FN1, CD44, ITGB2, CCL5, CD163, C1QA, C1QB, CYBB, TYROBP, LYZ, CD48, IL10RA, CLEC10A, COL1A2, CLU, C3, IRF8, CD52, TGFBI, ALOX5, CD53, VSIG4, CFB, IGFBP3, LAPTM5, MS4A6A, LUM, VCAN, MS4A4A, and DOCK2), which are exhibited in Table 2. Furthermore, the coexpression network

analysis of hub genes is displayed in Figure 5 and the heat map of the hub gene expression in the three GEO datasets is shown in Figure 6.

3.3. GO and KEGG Pathway Enrichment Analysis of Hub Genes in DN. After applying the clusterProfiler package for hub gene enrichment analysis, we selected the top 10 remarkably enriched BP terms for analysis, including regulation of complement activation, neutrophil degranulation, neutrophil activation involved in immune response, regulation of humoral immune response, humoral immune response, neuroinflammatory response, complement activation, regulation of immune effector process, macrophage activation, and synapse pruning. Besides, the top 10 CC terms were screened, consisting of secretory granule lumen, cytoplasmic vesicle lumen, vesicle lumen, blood microparticle, collagen-containing extracellular matrix, endoplasmic reticulum lumen, platelet alpha granule lumen, specific granule,

TABLE 2: Hub genes of diabetic nephropathy.

Groups	Gene symbol	GSE30528		GSE104948		GSE47183	
		LogFC	P value	LogFC	P value	LogFC	P value
Up	FN1	1.540307	$1.86E-06$	1.688679	$6.53E-14$	1.411328	0.000115
	CD44	1.410913	0.00288	1.154713	$1.03E-09$	1.118944	0.004117
	ITGB2	1.288601	0.000137	1.624558	$1.71E-09$	1.176759	0.015233
	CCL5	1.720493	0.000615	1.281747	$1.58E-08$	1.010635	0.002817
	CD163	1.973348	$2.66E-06$	1.260515	$4.52E-05$	2.25682	0.006126
	C1QA	1.922276	$3.04E-08$	2.469202	$7.39E-08$	2.643819	0.002315
	C1QB	2.204686	$4.22E-07$	2.123409	$8.65E-08$	2.331991	0.003483
	CYBB	1.082376	$2.88E-05$	1.559454	$6.84E-10$	1.081369	0.01398
	TYROBP	1.3831	0.000242	2.191158	$1.95E-08$	1.720981	0.000578
	LYZ	1.529532	0.00258	2.79008	$2.08E-08$	1.693295	0.000448
	CD48	1.736706	0.000071	2.1464	$7.11E-11$	1.271107	0.002955
	IL10RA	1.468163	0.000276	1.928371	$3.65E-09$	1.340032	0.004027
	CLEC10A	1.005234	$6.1E-06$	1.197336	$6.34E-07$	1.050224	0.032825
	COL1A2	2.590987	$9.18E-06$	2.656871	$2.14E-07$	2.275034	0.005026
	CLU	1.472253	0.000508	1.734574	$1.59E-06$	1.912975	0.002851
	C3	2.679961	0.000157	2.281075	$3.84E-06$	2.430408	0.017567
	IRF8	1.583247	0.000131	1.488973	$1.09E-06$	1.493004	0.005593
	CD52	2.232654	$9.78E-05$	2.140349	$6.98E-13$	1.169746	0.007569
	TGFBI	1.32786	0.000642	2.396142	$3.41E-13$	2.168562	$3.43E-05$
	ALOX5	1.673735	0.00568	1.450238	$7.51E-09$	1.277775	0.001606
	CD53	1.672055	$5.65E-05$	1.74099	$5.92E-08$	1.230233	0.012145
	VSIG4	1.576657	$5.22E-06$	1.696336	$1.31E-07$	1.919467	0.003
	CFB	1.720388	$6.19E-05$	1.382782	$5.74E-05$	1.365459	0.021665
	IGFBP3	1.033501	$9.66E-05$	1.23278	0.000966	2.409502	0.002813
	LAPTM5	1.062031	0.00158	1.554621	$9.52E-08$	1.238422	0.001673
	MS4A6A	1.869037	$6.81E-06$	2.054775	$9.74E-09$	1.959976	0.00497
	LUM	2.374619	$4.49E-05$	1.843351	0.00363	3.496266	0.004413
	VCAN	1.273776	0.00177	1.33971	$4.88E-05$	1.440879	0.034958
	MS4A4A	1.446755	$2.65E-05$	2.039072	$1.15E-08$	1.994681	0.00792
	DOCK2	1.009537	0.0028	1.457016	$1.34E-09$	1.181051	0.003642
Down	ALB	-1.78623	0.00896	-2.38803	0.000092	-2.63068	0.007517
	IGF1	-2.57456	$7.11E-05$	-1.6702	$1.25E-05$	-1.63423	$3.29E-05$

collagen trimer, and platelet alpha granule. Furthermore, the top 10 MF terms were selected, containing collagen binding, complement binding, extracellular matrix structural constituent, integrin binding, opsonin binding, amyloid-beta binding, hyaluronic acid binding, extracellular matrix structural constituent conferring compression resistance, chaperone binding, and growth factor binding. These processes are of great significance to further understand the hub genes that participated in the progression of DN. The results of the GO analysis are illustrated in Figure 7.

After the KEGG pathway enrichment analysis, a total of 16 significantly enriched pathways were selected out based on the threshold of $P < 0.05$ (Figure 8). The results indicated that these genes were mainly associated with regulation of insulin-like growth factor (IGF) transport and uptake by insulin-like growth factor binding proteins (IGFBPs), post-translational protein phosphorylation, ECM proteoglycans, complement and coagulation cascades, complement cascade, initial triggering of complement, regulation of complement cascade, human complement system, and complement activation. The results prove that hub genes derived from these three datasets may participate in the progression and development of DN by regulating complement cascade, insulin resistance, and inflammatory reaction. Among them, complement cascade-related pathways were most enriched.
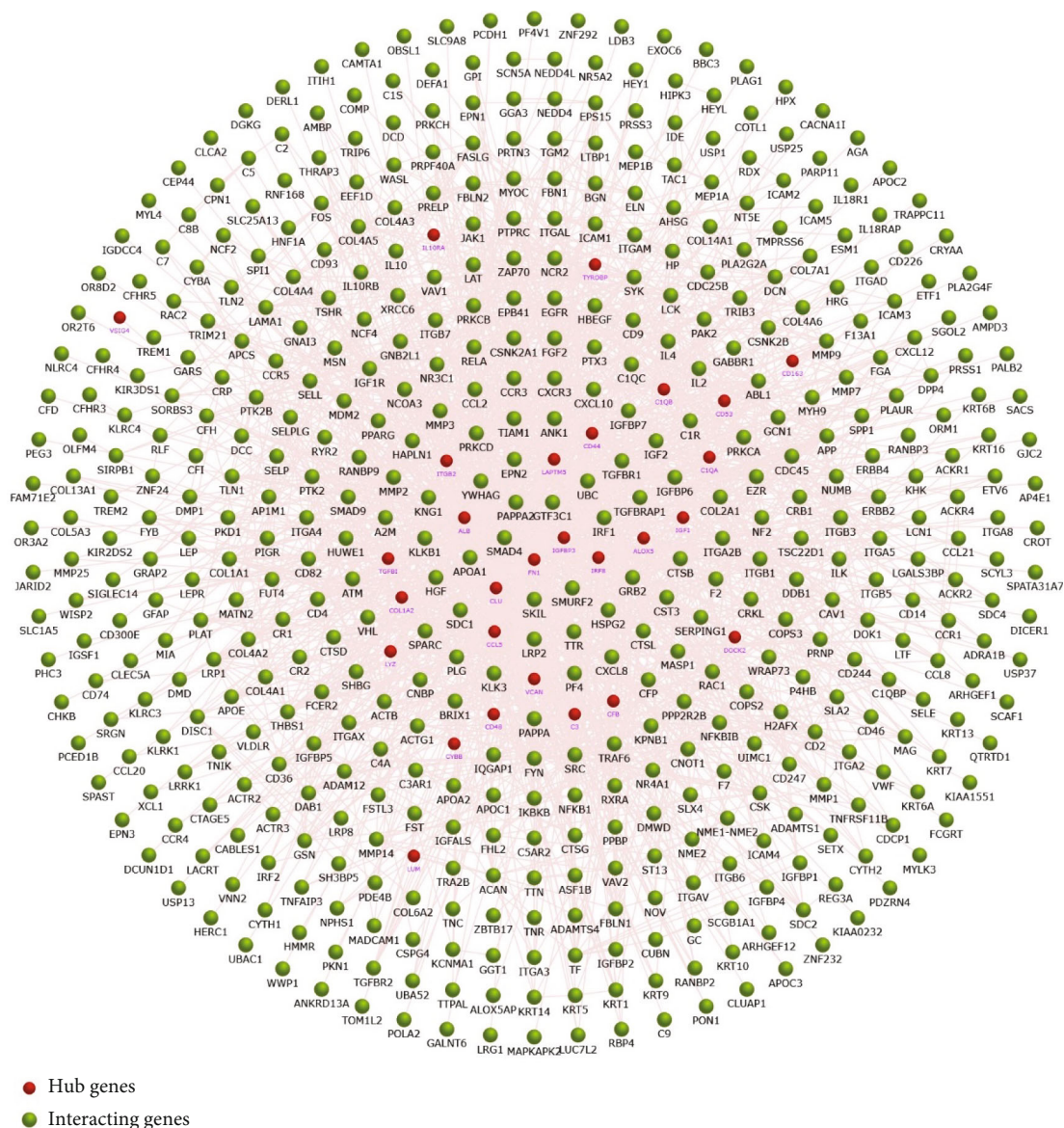
FIGURE 5: The coexpression network analysis of hub genes.

- ● Hub genes
- ● Interacting genes

3.4. Association between the Hub Genes and Clinical Features of DN. Of the overlapping DEGs identified in this study, 32 were recognized as hub genes. With the use of Nephroseq v5, the expression of complement cascade-related hub genes (C1QA, C1QB, C3, CFB, ITGB2, VSIG4, and CLU) showed the difference between DN patients and healthy living donors (Figure 9); we found all complement cascade-related hub genes were downregulated in the DN renal tissues compared with healthy kidney samples. In addition, the correlation between the complement cascade-related hub genes and GFR of DN patients was determined (Figure 10). The expression of CFB in DN renal tissue samples was positively correlated with GFR. Thus, the expression of CFB may contribute to the maintenance and improvement of renal function. The expression of the C1QA, C1QB, C3, ITGB2, VSIG4, and CLU was negatively correlated with GFR. Therefore, the expression changes of these six genes may result in the occurrence and development of DN.

4. Discussion

The prevalence of type 2 diabetes has risen dramatically worldwide, and DN is one of the most common complications of type 2 diabetes which has become the main cause of ESRD [31]. DN is featured as glomerular injury, glomerular hypertrophy, and glomerular basement membrane thickening [32]. Many cases of DN have a delayed diagnosis and are complicated to treat, which may contribute to the poor renal prognosis of patients with DN [33]. However, DN is the result of multiple gene interactions and the molecular mechanisms of DN remain poorly understood because of the complexity of the etiology differences [34]. Therefore, potential biomarkers for early diagnosis and targeted treatments are urgently needed. Currently, collagen binding and ECM-receptor interaction have already been verified to make a considerable contribution to the development of DN [35, 36]. In addition, there are two main complement cascade-
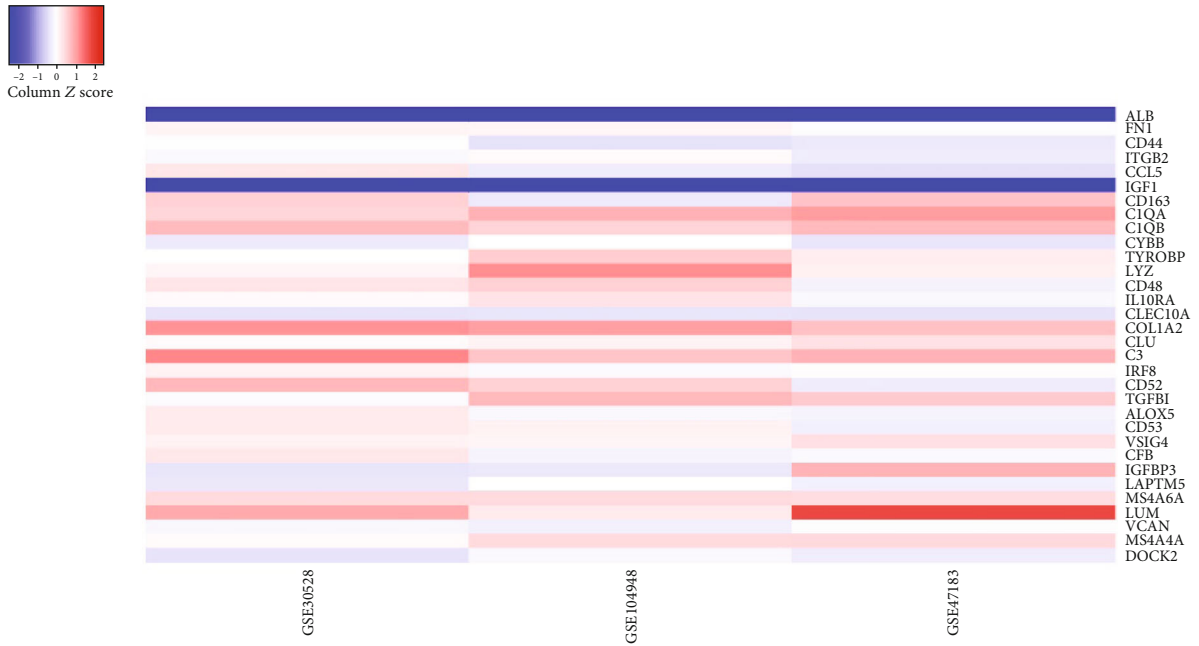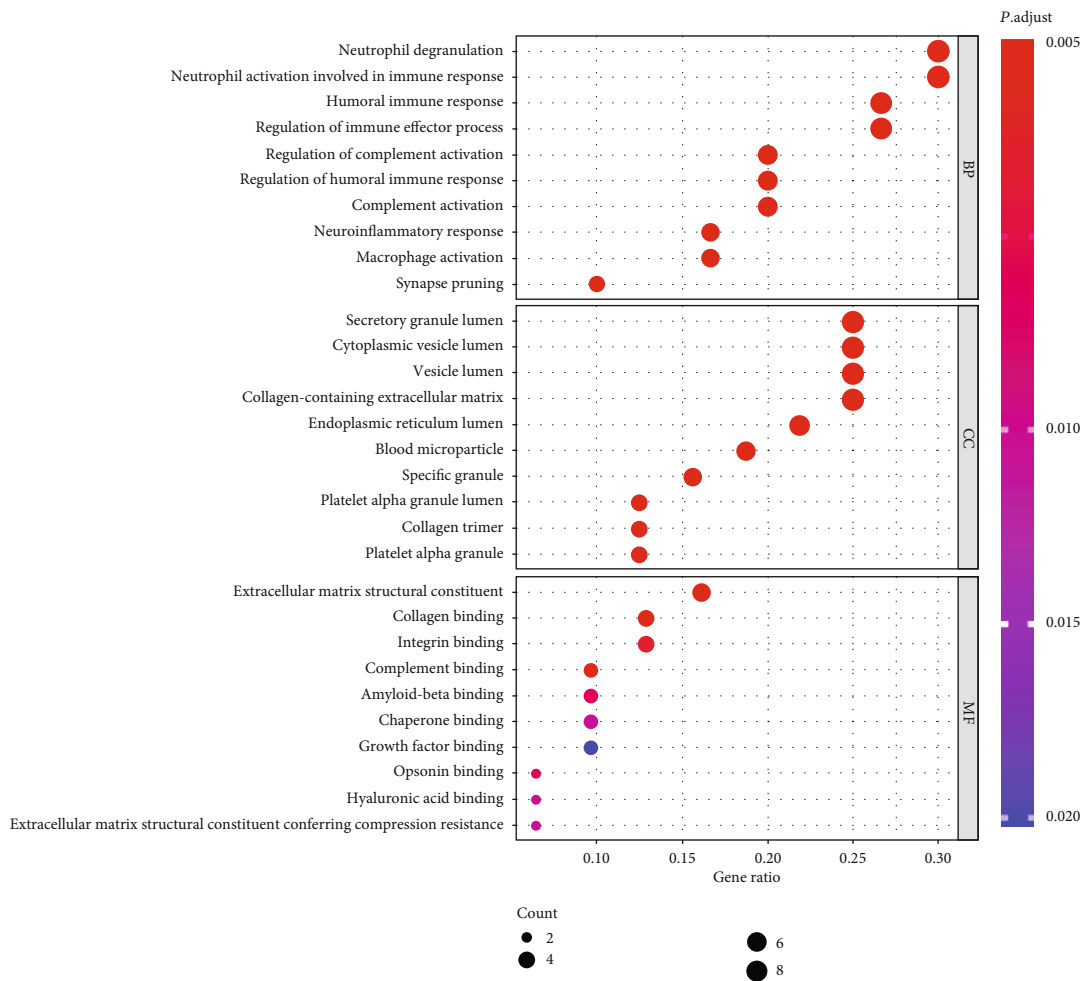
FIGURE 6: The heat map of the hub genes.

FIGURE 7: Enriched GO terms for BP, MF, and CC of hub genes. The color of the node is displayed in a gradient from red to blue according to the ascending order of the P value, while the size of the node is showed according to the ascending order of the number of gene counts.
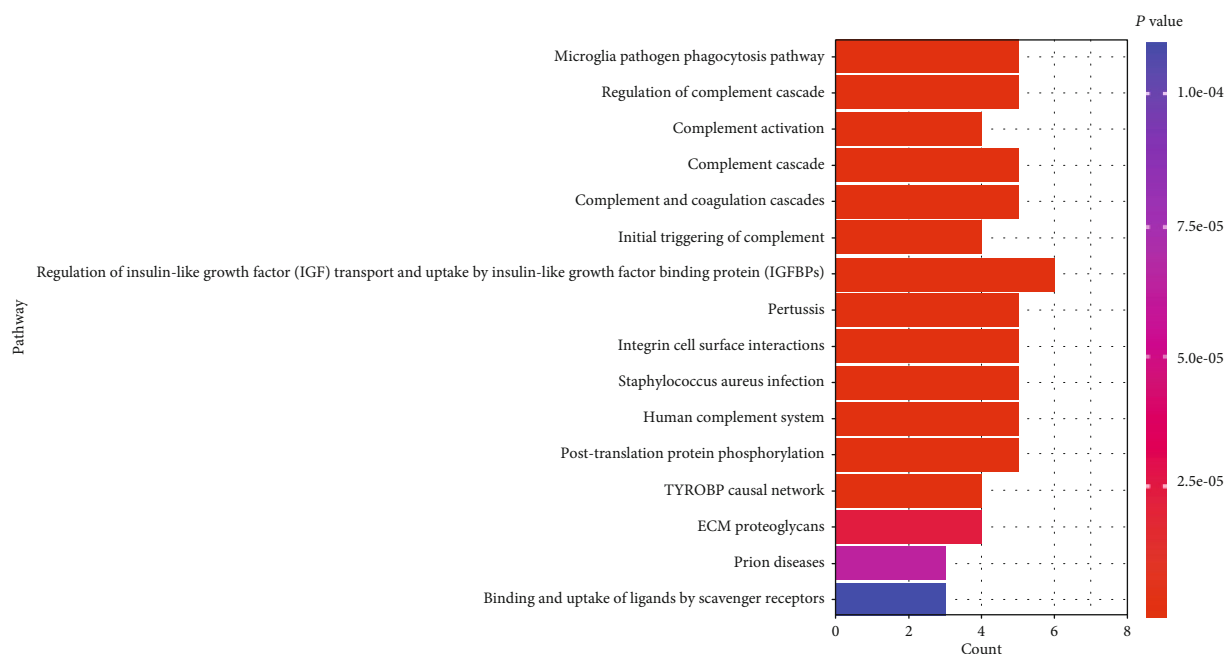
FIGURE 8: KEGG pathway analysis of hub genes. A total of 16 pathways are screened out according to the criteria of $P < 0.05$. The longitudinal axis displays the name of different pathways, and the transverse axis shows the count of enriched genes. In addition, the length of the bar is showed according to the ascending order of the number of gene counts, while the color of the bar is displayed in a gradient from blue to red according to the descending order of the P value.

associated mechanisms that have been reported to be involved in the development and progression of DN [37]. First, hyperglycaemia is considered to cause glycation of complement regulatory proteins which can lead to dysfunction of their regulatory capacity [38]. Second, the activation of the lectin pathway in response to glycated proteins is expressed on the surface of cells due to overexposure to glucose [39].

The development of microarray technology enables us to explore genetic alterations in DN and have a better understanding of the molecular mechanisms, eventually identifying the novel markers in DN [40–43]. In our study, a set of 64 overlapping DEGs from the GSE30528, GSE47183, and GSE104948 datasets were identified. Among them, the expression of 32 hub genes (30 upregulated genes and 2 downregulated genes) was selected for performing GO and KEGG enrichment analyses to explore the molecular mechanisms of hub genes involved in the development of DN. The GO enrichment analysis revealed that hub genes were involved in multiple biological processes, including regulation of complement activation, regulation of humoral immune response, humoral immune response, complement activation, regulation of immune effector process, and macrophage activation. The hub genes, such as C1QA, C1QB, C3, CFB, ITGB2, VSIG4, and CLU, can participate in complement cascade [44–47]. CD163, CD44, CD48, CD52, and CD53 may involve in humoral immune response and macrophage activation [48, 49]. Thus, the biological processes of hub genes are relatively consistent with the pathogenesis and mechanism of DN. Additionally, cellular components constitute secretory granule lumen, cytoplasmic vesicle lumen, vesicle lumen, blood

microparticle, collagen-containing extracellular matrix, endoplasmic reticulum lumen, platelet alpha granule lumen, specific granule, collagen trimer, and platelet alpha granule. It indirectly elucidates the complexity of the pathogenesis of DN and its damage to several cellular components [5]. Moreover, molecular functions are mostly enriched in collagen binding, complement binding, extracellular matrix structural constituent, amyloid-beta binding, and extracellular matrix structural constituent conferring compression resistance. It reveals that hub genes may target these molecular functions to affect DN progression and it is consistent with previous studies [50].

After performing KEGG enrichment analysis, we found that hub genes were mainly enriched in the regulation of IGF transport and uptake by IGFBPs, ECM proteoglycans, and complement cascade-related pathways. ECM proteoglycans, one important ECM component, show a more complex pattern of changes in DN, which could be mediated by TGF-β [51]. IGF-IGFBP signaling components play an important role in the maintenance of normal renal function and the development of DN. IGF-I expression increases in the diabetic kidney in the autocrine/paracrine manner that promotes matrix production, mesangial cell proliferation, and migration, but this process can be opposed by IGFBPs [52].

Moreover, experimental and clinical evidence has showed that multiple components of the complement system involved in the pathogenesis of DN [37]. And we also found hub genes were mostly enriched in complement cascade-associated pathways, consisting of complement and coagulation cascades, complement cascade, initial triggering of complement, regulation of complement cascade, human complement system, and complement activation. In these 32 hub genes, we selected 7 complement cascade-
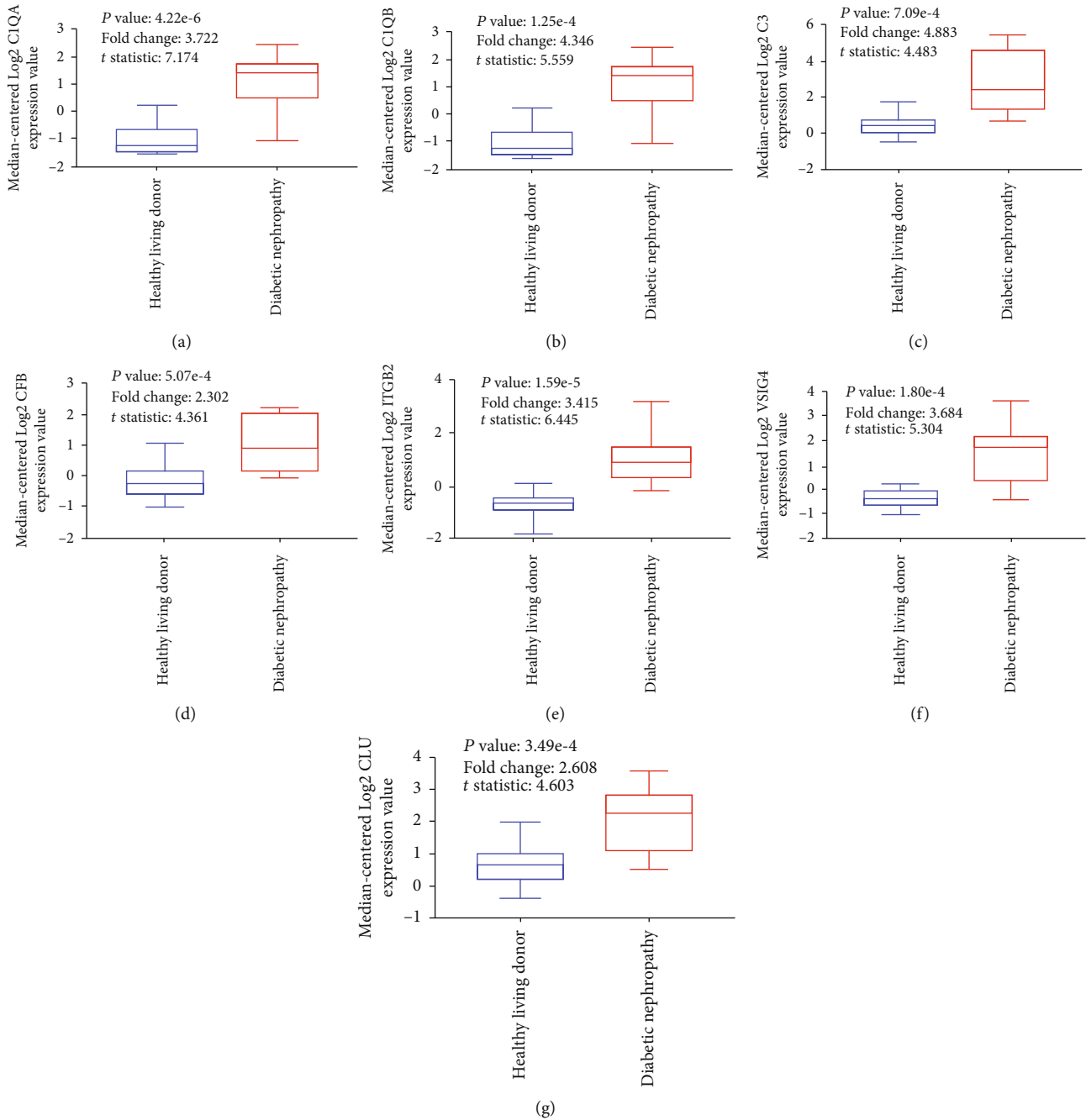
FIGURE 9: The different expressions of complement cascade-related hub genes (C1QA, C1QB, C3, CFB, ITGB2, VSIG4, and CLU) in DN renal tissues and healthy kidney tissues. (a) The expression of C1QA in DN renal tissues is higher than in healthy kidney tissues. (b) The expression of C1QB in DN renal tissues is higher than in healthy kidney tissues. (c) The expression of C3 in DN renal tissues is higher than in healthy kidney tissues. (d) The expression of CFB in DN renal tissues is higher than in healthy kidney tissues. (e) The expression of ITGB2 in DN renal tissues is higher than in healthy kidney tissues. (f) The expression of VSIG4 in DN renal tissues is higher than in healthy kidney tissues. (g) The expression of CLU in DN renal tissues is higher than in healthy kidney tissues.

related genes, including C1QA, C1QB, C3, CFB, ITGB2, VSIG4, and CLU. Among them, C3 is the central component in the complementary system which plays an important role in the classical complement pathway and alternative complement pathway [47, 53]. We found that C3 expression was higher in DN renal samples compared with healthy renal samples and was negatively correlated with GFR ($P = 0.046$, $r = -0.611$). In addition, the classic complement pathway

was activated in DN renal tissues with significant increases of C1QA and C1QB mRNAs [47]. We discovered that C1QA and C1QB expressions were higher in DN renal samples and were negatively correlated with GFR (($P < 0.001$, $r = -0.777$) and ($P = 0.014$, $r = -0.776$), respectively). CFB is a crucial factor for activating the alternative complement pathway [54], which is elevated in adipose tissue and serum from patients with type 2 diabetes [55], but the correlation
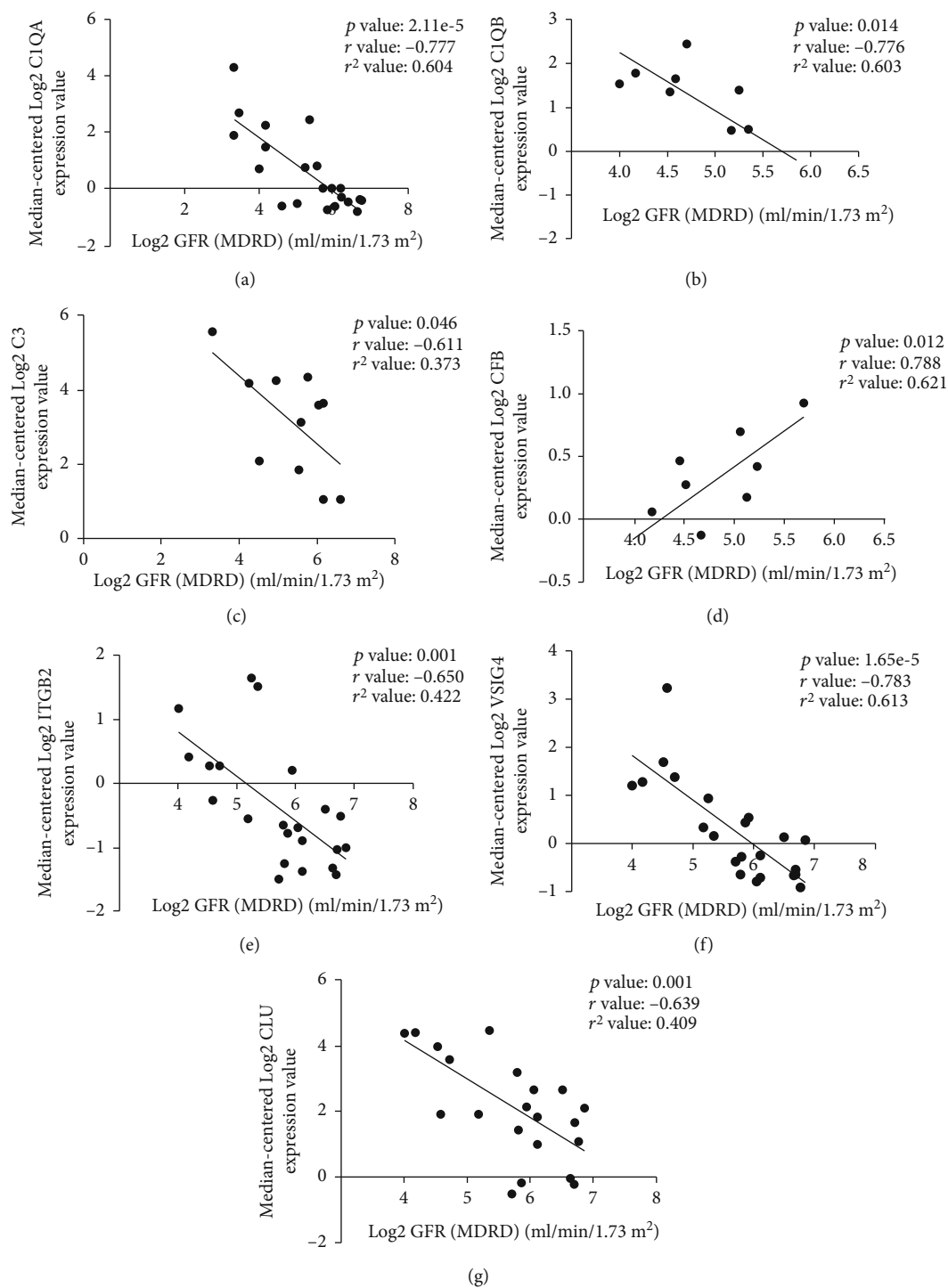
FIGURE 10: The correlation analysis between complement cascade-related hub genes (C1QA, C1QB, C3, CFB, ITGB2, VSIG4, and CLU) and the expression of GFR in DN patients. (a) The expression of C1QA was negatively correlated with GFR ($P < 0.001$, $r = -0.777$). (b) The expression of C1QB was negatively correlated with GFR ($P = 0.014$, $r = -0.776$). (c) The expression of C3 was negatively correlated with GFR ($P = 0.046$, $r = -0.611$). (d) The expression of CFB was positively correlated with GFR ($P = 0.012$, $r = 0.788$). (e) The expression of ITGB2 was negatively correlated with GFR ($P = 0.001$, $r = -0.650$). (f) The expression of VSIG4 was negatively correlated with GFR ($P < 0.001$, $r = -0.783$). (g) The expression of CLU was negatively correlated with GFR ($P = 0.001$, $r = -0.639$).

between CFB and DN has not been illustrated yet. We found that CFB expression was higher in DN renal samples compared with healthy renal samples and was positively correlated with GFR ($P = 0.012$, $r = 0.788$). ITGB2 is implicated in the network linking complement system to T cell activation [56], but it also remains poorly understood about the relationship with DN. We found that ITGB2 expression was higher in DN renal samples and was negatively

correlated with GFR ($P = 0.001$, $r = -0.650$). Besides, VSIG4 is a B7 family-related protein and acts as a complement receptor for C3 [57]; it has been reported that VSIG4 can induce the epithelial-mesenchymal transition of renal tubular cells when exposing to high-glucose condition [58], but how VSIG4 affect DN through the complement system is still unknown. We found that VSIG4 expression was higher in DN renal samples and was negatively correlated with GFR ($P < 0.001$, $r = -0.783$). CLU is a kind of complement regulatory protein which can bind to C5b-7 and inhibit the generation of membrane attack complex [59] and was found upregulated in the glomerular of both DN patients and streptozotocin-induced diabetic mice [46]. We found that CLU expression was higher in DN renal samples and was negatively correlated with GFR ($P = 0.001$, $r = -0.639$). Taking these findings together will offer novel potential targets in future DN research.

Independent microarray analysis always induces false-positive rates [60]. Due to the heterogeneity of the tissues or samples in independent studies or a single cohort study, gene array results are always either limited or inconsistent. In this study, we reanalyzed three microarray datasets (GSE30528, GSE47183, and GSE104948) and combined various bioinformatics methods and expression profiling techniques to identify candidate DEGs and predict hub genes in DN. In addition, we also used a clinical database (Nephroseq v5 platform) to analyze correlations and perform subgroup analysis among the hub genes and clinical manifestations of DN.

However, there are some limitations of this study. Firstly, the predictions above were not confirmed by experiments and are rather preliminary in silico studies; these are only useful for an initial screening. It is required to be validated by in vitro/vivo experiments and large-group cohort before any valid conclusion could be taken. Secondly, because of the heterogeneity of detailed demographic data, it is difficult to obtain the more convincing association between hub genes and the severity of diabetic glomerular injury using the samples from different datasets. Thirdly, this study had a relatively small sample size; in future studies, we also need to combine and analyze more clinical samples based on similar demographic data to validate these results.

5. Conclusions

The microarray and bioinformatics technology have provided new perspectives for researchers to study the potential molecular mechanisms and regulatory targets of DN. A total of 64 DEGs and 32 hub genes were identified as biomarkers for the clinical diagnosis of DN and as potential targets for novel treatments. However, predictions were not verified by experiments, and the number of samples used for analysis was limited. Thus, the specific molecular mechanisms and biological functions of hub genes need further exploration. In conclusion, our study found hub genes which might be involved in the pathogenesis of DN especially in complement cascade-related signaling pathways. At the same time, we also linked the expression of hub genes with the clinical manifes-

tation of DN, which may provide the novel methodologies for DN early diagnosis and targeted therapies.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declared that they have no competing interests.

Authors' Contributions

Bojun Xu and Lei Wang contributed equally to this article. All authors have contributed to this study and approved its submission.

Acknowledgments

References

[1] L. Zhang, J. Long, W. Jiang et al., "Trends in chronic kidney disease in China," *The New England Journal of Medicine*, vol. 375, no. 9, pp. 905-906, 2016.

[2] K. Azushima, S. B. Gurley, and T. M. Coffman, "Modelling diabetic nephropathy in mice," *Nature Reviews. Nephrology*, vol. 14, no. 1, pp. 48–56, 2018.

[3] D. Kawanami, K. Matoba, and K. Utsunomiya, "Signaling pathways in diabetic nephropathy," *Histology and Histopathology*, vol. 31, no. 10, pp. 1059–1067, 2016.

[4] D. Sanajou, A. Ghorbani Haghjo, H. Argani, and S. Aslani, "AGE-RAGE axis blockade in diabetic nephropathy: current status and future directions," *European Journal of Pharmacology*, vol. 833, pp. 158–164, 2018.

[5] J. Wada and H. Makino, "Inflammation and the pathogenesis of diabetic nephropathy," *Clinical Science (London, England)*, vol. 124, no. 3, pp. 139–152, 2013.

[6] L. F. Zeng, Y. Xiao, and L. Sun, "A glimpse of the mechanisms related to renal fibrosis in diabetic nephropathy," *Advances in Experimental Medicine and Biology*, vol. 1165, pp. 49–79, 2019.

[7] L. Z. Guan, Q. Tong, and J. Xu, "Elevated serum levels of mannose-binding lectin and diabetic nephropathy in type 2 diabetes," *PLoS One*, vol. 10, no. 3, article e0119699, 2015.

[8] T. K. Hansen, on behalf of the FinnDiane Study Group, C. Forsblom et al., "Association between mannose-binding lectin, high-sensitivity C-reactive protein and the progression of diabetic nephropathy in type 1 diabetes," *Diabetologia*, vol. 53, no. 7, pp. 1517–1524, 2010.

[9] K. I. Woroniecka, A. S. D. Park, D. Mohtat, D. B. Thomas, J. M. Pullman, and K. Susztak, "Transcriptome analysis of human diabetic kidney disease," *Diabetes*, vol. 60, no. 9, pp. 2354–2369, 2011.

[10] B. Thomas, "The Global Burden of Diabetic Kidney Disease: Time Trends and Gender Gaps," *Current Diabetes Reports*, vol. 19, no. 4, p. 18, 2019.

[11] C. Qi, X. Mao, Z. Zhang, and H. Wu, "Classification and Differential Diagnosis of Diabetic Nephropathy," *Journal of Diabetes Research*, vol. 2017, Article ID 8637138, 7 pages, 2017.

[12] C. S. Pareek, R. Smoczynski, and A. Tretyn, "Sequencing technologies and genome sequencing," *Journal of Applied Genetics*, vol. 52, no. 4, pp. 413–435, 2011.

[13] S. Udhaya Kumar, B. Rajan, D. Thirumal Kumar et al., "Involvement of Essential Signaling Cascades and Analysis of Gene Networks in Diabesity," *Genes*, vol. 11, no. 11, p. 1256, 2020.

[14] S. Udhaya Kumar, D. Thirumal Kumar, R. Bithia et al., "Analysis of differentially expressed genes and molecular pathways in familial hypercholesterolemia involved in atherosclerosis: a systematic and bioinformatics approach," *Frontiers in Genetics*, vol. 11, p. 734, 2020.

[15] T. Wu, Q. Li, T. Wu, and H. Y. Liu, "Identification of biological targets of therapeutic intervention for diabetic nephropathy with bioinformatics approach," *Experimental and Clinical Endocrinology & Diabetes*, vol. 122, no. 10, pp. 587–591, 2014.

[16] P. C. Grayson, S. Eddy, J. N. Taroni et al., "Metabolic pathways and immunometabolism in rare kidney diseases," *Annals of the Rheumatic Diseases*, vol. 77, no. 8, pp. 1226–1233, 2018.

[17] M. Zeng, J. Liu, W. Yang et al., "Identification of key biomarkers in diabetic nephropathy via bioinformatic analysis," *Journal of Cellular Biochemistry*, vol. 120, no. 5, pp. 8676–8688, 2019.

[18] F. Yang, Z. Cui, H. Deng et al., "Identification of miRNAs-genes regulatory network in diabetic nephropathy based on bioinformatics analysis," *Medicine (Baltimore)*, vol. 98, no. 27, article e16225, 2019.

[19] S. Liu, C. Wang, H. Yang, T. Zhu, H. Jiang, and J. Chen, "Weighted gene co-expression network analysis identifies FCER1G as a key gene associated with diabetic kidney disease," *Ann Transl Med*, vol. 8, no. 21, p. 1427, 2020.

[20] W. Ju, C. S. Greene, F. Eichinger et al., "Defining cell-type specificity at the transcriptional level in human disease," *Genome Research*, vol. 23, no. 11, pp. 1862–1873, 2013.

[21] R. Edgar, M. Domrachev, and A. E. Lash, "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository," *Nucleic Acids Research*, vol. 30, no. 1, pp. 207–210, 2002.

[22] S. Davis and P. S. Meltzer, "GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor," *Bioinformatics*, vol. 23, no. 14, pp. 1846-1847, 2007.

[23] M. Pathan, S. Keerthikumar, C. S. Ang et al., "FunRich: an open access standalone functional enrichment and interaction network analysis tool," *Proteomics*, vol. 15, no. 15, pp. 2597–2601, 2015.

[24] A. Franceschini, D. Szklarczyk, S. Frankild et al., "STRING v9.1: protein-protein interaction networks, with increased coverage and integration," *Nucleic Acids Research*, vol. 41, no. D1, pp. D808–D815, 2013.

[25] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.

[26] W. P. Bandettini, P. Kellman, C. Mancini et al., "MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study," *Journal of Cardiovascular Magnetic Resonance*, vol. 14, no. 1, p. 83, 2012.

[27] Y. Assenov, F. Ramírez, S. E. Schelhorn, T. Lengauer, and M. Albrecht, "Computing topological parameters of biological networks," *Bioinformatics*, vol. 24, no. 2, pp. 282–284, 2008.

[28] The Gene Ontology Resource, "The Gene Ontology Resource: 20 years and still GOing strong," *Nucleic Acids Research*, vol. 47, no. D1, pp. D330–d338, 2019.

[29] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: new perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Research*, vol. 45, no. D1, pp. D353–d361, 2017.

[30] S. Eddy, L. H. Mariani, and M. Kretzler, "Integrated multiomics approaches to improve classification of chronic kidney disease," *Nature Reviews. Nephrology*, vol. 16, no. 11, pp. 657–668, 2020.

[31] Y. Zhang, J. Niu, and H. K. Choi, "Excess mortality among persons with type 2 diabetes," *The New England Journal of Medicine*, vol. 374, no. 8, p. 788, 2016.

[32] C. W. Tung, Y. C. Hsu, Y. H. Shih, P. J. Chang, and C. L. Lin, "Glomerular mesangial cell and podocyte injuries in diabetic nephropathy," *Nephrology (Carlton)*, vol. 23, Suppl 4, pp. 32–37, 2018.

[33] D. Fineberg, K. A. Jandeleit-Dahm, and M. E. Cooper, "Diabetic nephropathy: diagnosis and treatment," *Nature Reviews Endocrinology*, vol. 9, no. 12, pp. 713–723, 2013.

[34] J. M. Forbes and M. E. Cooper, "Mechanisms of diabetic complications," *Physiological Reviews*, vol. 93, no. 1, pp. 137–188, 2013.

[35] H. Ha and H. B. Lee, "Reactive Oxygen Species and Matrix Remodeling in Diabetic Kidney," *Journal of the American Society of Nephrology*, vol. 14, no. 90003, pp. 246S–2249, 2003.

[36] J. I. Jun and L. F. Lau, "Taking aim at the extracellular matrix: CCN proteins as emerging therapeutic targets," *Nature Reviews. Drug Discovery*, vol. 10, no. 12, pp. 945–963, 2011.

[37] A. Flyvbjerg, "The role of the complement system in diabetic nephropathy," *Nature Reviews. Nephrology*, vol. 13, no. 5, pp. 311–318, 2017.

[38] X. Qin, A. Goldfine, N. Krumrei et al., "Glycation inactivation of the complement regulatory protein CD59: a possible role in the pathogenesis of the vascular complications of human diabetes," *Diabetes*, vol. 53, no. 10, pp. 2653–2661, 2004.

[39] J. Fortpied, D. Vertommen, and E. Van Schaftingen, "Binding of mannose-binding lectin to fructosamines: a potential link between hyperglycaemia and complement activation in diabetes," *Diabetes/Metabolism Research and Reviews*, vol. 26, no. 4, pp. 254–260, 2010.

[40] S. Mishra, M. I. Shah, S. Udhaya Kumar et al., "Network analysis of transcriptomics data for the prediction and prioritization of membrane-associated biomarkers for idiopathic pulmonary fibrosis (IPF) by bioinformatics approach," *Advances in Protein Chemistry and Structural Biology*, vol. 123, pp. 241–273, 2021.

[41] S. Udhaya Kumar, D. Thirumal Kumar, R. Siva et al., "Dysregulation of signaling pathways due to differentially

expressed genes from the B-cell transcriptomes of systemic lupus erythematosus patients - a bioinformatics approach," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 276, 2020.

[42] H. Yan, G. Zheng, J. Qu et al., "Identification of key candidate genes and pathways in multiple myeloma by integrated bioinformatics analysis," *Journal of Cellular Physiology*, vol. 234, no. 12, pp. 23785–23797, 2019.

[43] S. U. Kumar, D. T. Kumar, R. Siva, C. G. P. Doss, and H. Zayed, "Integrative bioinformatics approaches to map potential novel genes and pathways involved in ovarian cancer," *Frontiers in Bioengineering and Biotechnology*, vol. 7, p. 391, 2019.

[44] H. Huang, D. Li, X. Huang et al., "Association of complement and inflammatory biomarkers with diabetic nephropathy," *Annals of Clinical and Laboratory Science*, vol. 49, no. 4, pp. 488–495, 2019.

[45] X. Q. Li, D. Y. Chang, M. Chen, and M. H. Zhao, "Complement activation in patients with diabetic nephropathy," *Diabetes & Metabolism*, vol. 45, no. 3, pp. 248–253, 2019.

[46] J. He, K. L. Dijkstra, K. Bakker et al., "Glomerular clusterin expression is increased in diabetic nephropathy and protects against oxidative stress-induced apoptosis in podocytes," *Scientific Reports*, vol. 10, no. 1, p. 14888, 2020.

[47] K. J. Kelly, Y. Liu, J. Zhang, and J. H. Dominguez, "Renal C3 complement component: feed forward to diabetic kidney disease," *American Journal of Nephrology*, vol. 41, no. 1, pp. 48–56, 2015.

[48] C. Q. F. Klessens, M. Zandbergen, R. Wolterbeek et al., "Macrophages in diabetic nephropathy in patients with type 2 diabetes," *Nephrology, Dialysis, Transplantation*, vol. 32, no. 8, pp. 1322–1329, 2017.

[49] J. Melin, O. Hellberg, K. Funa, R. Hällgren, E. Larsson, and B. C. Fellström, "Ischemia-induced renal expression of hyaluronan and CD44 in diabetic rats," *Nephron. Experimental Nephrology*, vol. 103, no. 3, pp. e86–e94, 2006.

[50] T. W. Cohen Tervaert, A. L. Mooyaart, K. Amann et al., "Pathologic classification of diabetic nephropathy," *Journal of the American Society of Nephrology*, vol. 21, no. 4, pp. 556–563, 2010.

[51] H. Yokoyama and T. Deckert, "Central role of TGF-beta in the pathogenesis of diabetic nephropathy and macrovascular complications: a hypothesis," *Diabetic Medicine*, vol. 13, no. 4, pp. 313–320, 1996.

[52] T. L. Vasylyeva and R. J. Ferry Jr., "Novel roles of the IGF-IGFBP axis in etiopathophysiology of diabetic nephropathy," *Diabetes Research and Clinical Practice*, vol. 76, no. 2, pp. 177–186, 2007.

[53] H. J. Müller-Eberhard, "Molecular organization and function of the complement system," *Annual Review of Biochemistry*, vol. 57, no. 1, pp. 321–347, 1988.

[54] P. J. Lachmann, "Looking back on the alternative complement pathway," *Immunobiology*, vol. 223, no. 8-9, pp. 519–523, 2018.

[55] P. M. Coan, M. Barrier, N. Alfazema et al., "Complement factor B is a determinant of both metabolic and cardiovascular features of metabolic syndrome," *Hypertension*, vol. 70, no. 3, pp. 624–633, 2017.

[56] A. Couto Alves, S. Bruhn, A. Ramasamy et al., "Dysregulation of complement system and CD4+ T cell activation pathways implicated in allergic response," *PLoS One*, vol. 8, no. 10, article e74821, 2013.

[57] L. Vogt, N. Schmitz, M. O. Kurrer et al., "VSIG4, a B7 family-related protein, is a negative regulator of T cell activation," *The Journal of Clinical Investigation*, vol. 116, no. 10, pp. 2817–2826, 2006.

[58] E.-Y. Gong, H. A. Jo, S. H. Park, D. R. Cha, D. Y. Hur, and S. Y. Han, "VSIG4 Induces Epithelial-Mesenchymal Transition of Renal Tubular Cells under High-Glucose Conditions," *Life*, vol. 10, no. 12, p. 354, 2020.

[59] K. Yamada, Y. Hori, N. Hanafusa et al., "Clusterin is up-regulated in glomerular mesangial cells in complement-mediated injury," *Kidney International*, vol. 59, no. 1, pp. 137–146, 2001.

[60] C. Wei, J. Li, and R. E. Bumgarner, "Sample size for detecting differentially expressed genes in microarray experiments," *BMC Genomics*, vol. 5, no. 1, p. 87, 2004.