



Identification and Verification of Diagnostic Biomarkers for Glomerular Injury in Diabetic Nephropathy Based on Machine Learning Algorithms

Hongdong Han^{1†}, Yanrong Chen^{1†}, Hao Yang², Wei Cheng¹, Sijing Zhang¹, Yunting Liu¹, Qihong Liu¹, Dongfang Liu¹, Gangyi Yang¹ and Ke Li^{1*}

¹ Department of Endocrinology, the Second Affiliated Hospital, Chongqing Medical University, Chongqing, China,

² Department of Endocrinology and Neurology, Jiulongpo People's Hospital, Chongqing, China

OPEN ACCESS

Edited by:

Ningning Hou,
Affiliated Hospital of Weifang Medical
University, China

Reviewed by:

Mengwei Zang,
The University of Texas Health Science
Center at San Antonio, United States
Min Sun,
Nanjing Medical University, China

*Correspondence:

Ke Li
like@hospital.cqmu.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Renal Endocrinology,
a section of the journal
Frontiers in Endocrinology

Received: 16 February 2022

Accepted: 14 April 2022

Published: 19 May 2022

Citation:

Han H, Chen Y, Yang H, Cheng W,
Zhang S, Liu Y, Liu Q, Liu D, Yang G
and Li K (2022) Identification and
Verification of Diagnostic Biomarkers
for Glomerular Injury in Diabetic
Nephropathy Based on Machine
Learning Algorithms.
Front. Endocrinol. 13:876960.
doi: 10.3389/fendo.2022.876960

Diabetic nephropathy (DN) is regarded as the leading cause of end-stage renal disease worldwide and lacks novel therapeutic targets. To screen and verify special biomarkers for glomerular injury in patients with DN, fifteen datasets were retrieved from the Gene Expression Omnibus (GEO) database, correspondingly divided into training and testing cohorts and then merged. Using the limma package, 140 differentially expressed genes (DEGs) were screened out between 81 glomerular DN samples and 41 normal ones from the training cohort. With the help of the ConsensusClusterPlus and WGCNA packages, the 81 glomerular DN samples were distinctly divided into two subclusters, and two highly associated modules were identified. By using machine learning algorithms (LASSO, RF, and SVM-RFE) and the Venn diagram, two overlapping genes (PRKAR2B and TGFBI) were finally determined as potential biomarkers, which were further validated in external testing datasets and the HFD/STZ-induced mouse models. Based on the biomarkers, the diagnostic model was developed with reliable predictive ability for diabetic glomerular injury. Enrichment analyses indicated the apparent abnormal immune status in patients with DN, and the two biomarkers played an important role in the immune microenvironment. The identified biomarkers demonstrated a meaningful correlation between the immune cells' infiltration and renal function. In conclusion, two robust genes were identified as diagnostic biomarkers and may serve as potential targets for therapeutics of DN, which were closely associated with multiple immune cells.

Keywords: diabetic nephropathy, glomerular injury, biomarker, diagnostic model, machine learning algorithm

INTRODUCTION

Diabetic nephropathy (DN) is a serious cause of end-stage renal disease, resulting in heavy economic and medical burdens. Tubulointerstitial lesions, glomerular basement membrane thickening, mesangial matrix accumulation, and nodular glomerulosclerosis are the basic pathological features of DN (1). The current treatment strategy is either to strengthen glucose

control or to reduce glomerular intracapsular pressure to slow the progression of renal injury (2, 3). In fact, because of the individual heterogeneity of DN, not all patients can benefit from these drugs. Genome-wide expression profiles can be easily obtained from public databases, analyzed, and visualized on the R platform, thanks to the advancement and widespread application of bioinformatics analysis and high-throughput sequencing technology (4, 5). The changes in gene expression profiles involved in the initiation and progression of DN have been identified by high-throughput microarray technology (6).

According to the flow chart shown in **Supplementary Figure S1**, gene expression profiles of DN patients and normal samples were obtained and analyzed to identify the differentially expressed genes (DEGs). Highly associated modules were identified to determine the critical biomarkers, and a diagnostic model was developed based on the biomarkers. Moreover, enrichment analysis was performed to explore the potential mechanisms of the identified biomarkers in DN. It particularly illustrates the relationship between the biomarkers and immune cell infiltration.

MATERIALS AND METHODS

Data Collection and Preprocessing

A total of fifteen human microarray datasets, namely GSE96804, GSE47183-GPL11670, GSE47183-GPL14663, GSE99339-GPL19109, GSE99339-GPL19184, GSE104948-GPL22945, GSE104948-GPL24120, GSE30122, GSE1009, GSE30528, GSE30529, GSE47184-GPL11670, GSE47184-GPL14663, GSE104954-GPL22945, and GSE104954-GPL24120, were downloaded from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>). More details of the collected datasets are presented in **Table 1**. After eliminating the batch effects by the Surrogate Variable Analysis (SVA) algorithm (7), seven glomerular DN (GDN) datasets (GSE96804, GSE47183-GPL11670, GSE47183-GPL14663, GSE99339-GPL19109, GSE99339-GPL19184, GSE104948-GPL22945, GSE104948-GPL24120), three GDN datasets (GSE30122, GSE1009, GSE30528), and five tubulointerstitial DN

(TDN) datasets (GSE30529, GSE47184-GPL11670, GSE47184-GPL14663, GSE104954-GPL22945, GSE104954-GPL24120) were merged, normalized, and utilized as the GDN training cohort, GDN testing cohort, and TDN testing cohort, respectively. The distribution patterns between DN and normal samples were visualized by principal component analysis (PCA).

Identification of DEGs

DEGs between GDN and normal subjects in the GDN training cohort were detected by using the limma R package (8) with $|\log_2$ fold change (FC)| > 1 and adjusted $p < 0.05$ as the cutoff threshold. Meanwhile, Gene Ontology (GO) enrichment analysis of DEGs was conducted using the clusterProfiler package. Gene Set Enrichment Analysis (GSEA) was also performed to investigate the significant differences in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways between GDN and normal samples, with the Molecular Signature Database (MSigDB)-derived gene sets “c2.cp.kegg.v7.4.symbols.gmt” selected as a reference. Enriched pathways with $p < 0.05$ and false discovery rate (FDR) < 0.25 were considered statistically significant.

Consensus Cluster Analysis

The ConsensusClusterPlus algorithm (9) was used to perform clustering analysis to identify potential subclusters of the GDN samples from the GDN training cohort. The maximum cumulative distribution function (CDF) index was selected as the optimal k -value. Meanwhile, principal component analysis (PCA) was employed to verify this classification based on gene expression patterns among different subgroups.

Weighted Gene Coexpression Network Analysis

The Weighted Gene Coexpression Network Analysis (WGCNA) method (10) was applied to build potential modules related to different subclusters of the 81 GDN samples. After filtering abnormal samples and calculating the Pearson correlation coefficient, the correlation adjacency matrix was constructed. Highly associated modules were selected for subsequent analysis. Functional enrichments of the genes within given modules were

TABLE 1 | The essential information of included microarray datasets in this study.

GEO series	Normal	DN	Tissue	Data type
GSE96804	20	41	Glomerulus	Training
GSE47183-GPL11670	0	7	Glomerulus	Training
GSE47183-GPL14663	0	7	Glomerulus	Training
GSE99339-GPL19109	0	7	Glomerulus	Training
GSE99339-GPL19184	0	7	Glomerulus	Training
GSE104948-GPL22945	18	7	Glomerulus	Training
GSE104948-GPL24120	3	5	Glomerulus	Training
GSE30122	13	9	Glomerulus	Testing
GSE1009	3	3	Glomerulus	Testing
GSE30528	13	9	Glomerulus	Testing
GSE30529	12	10	Tubulointerstitium	Testing
GSE47184-GPL11670	0	7	Tubulointerstitium	Testing
GSE47184-GPL14663	4	11	Tubulointerstitium	Testing
GSE104954-GPL22945	18	7	Tubulointerstitium	Testing
GSE104954-GPL24120	3	10	Tubulointerstitium	Testing

performed to interpret the diverse biological effects based on the KEGG, GO, and Disease Ontology (DO) analyses using the ClusterProfiler, DOSE, and ggplot2 packages.

Diagnostic Gene Screening and Diagnostic Model Construction

The Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression (11), Support Vector Machine-Recursive Feature Elimination (SVM-RFE) (12), and Random Forest (RF) (13) algorithms were employed independently to screen the diagnostic genes from the selected modules. Ultimately, genes that overlapped among the three machine learning algorithms were regarded as diagnostic biomarkers. A receiver operating characteristic (ROC) curve was generated, and the area under the ROC curve (AUC) value was calculated to estimate the predictive utility of the identified biomarkers using the pROC package. The differential expression and predictive reliability of the biomarkers were further confirmed in the external testing cohorts. A diagnostic model was constructed using logistic regression analysis and visualized as a nomogram (14) to predict the glomerular injury in DN patients. The Concordance index (C-index), calibration curve, and decision curve analysis (DCA) were employed to visualize its discrimination performances. Besides, using the training datasets (Table 1), the expressions of the identified biomarkers were also explored in other chronic kidney diseases (CKD), including hypertensive nephropathy (HN) and systemic lupus erythematosus nephropathy (SLEN). Furthermore, based on the median expression level of each gene, 81 GDN samples from the GDN training dataset were divided into two groups (high- and low-expression group), and then Gene Set Variation Analysis (GSVA) was employed to clarify the enriched KEGG pathways with MSigDB gene sets “c2.cp.kegg.v7.4.symbols.gmt” used as a reference.

Verification and Clinical Correlation Analysis of the Identified Biomarkers

The expression patterns of identified biomarkers were reconfirmed by the Nephroseq v5 online database (<http://v5.nephroseq.org>) (15). A correlation analysis between the biomarkers and renal function was also carried out.

Evaluation of Immune Cell Infiltration

Based on the single-sample Gene-Set Enrichment Analysis (ssGSEA) method and the 29 gene sets of immune-related responses (16), the ssGSEA scores were quantified and designed to represent the activity and infiltrating fractions of immune cells and pathways in the GDN training cohort and the TDN testing cohort. The result of ssGSEA was shown as a heatmap. Furthermore, the cell-type identification by estimating relative subsets of RNA transcripts (CIBERSORT) algorithm (17) was performed to calculate the relative proportion of the infiltrating immune cells in each sample from the GDN training cohort and the TDN testing cohort. The abundances of infiltrating immune cells in DN patients and normal subjects were compared and visualized using the vioplot package. The differences in immune characteristics between the samples with low and high expression of the identified biomarkers were clarified. In the GDN training cohort, using the corplot

package, the correlations between the enrichment levels of infiltrating immune cells and the expressions of the diagnostic genes were also investigated.

Animal Experiments

A total of 15 male C57BL/6 mice (8 weeks old; ~25 g) were purchased from the Chongqing Medical University Animal Experiment Center (Chongqing, China). Mice were randomly divided into normal groups ($n = 5$) and high-glucose-induced renal injury models ($n = 10$) and given access to a normal chow diet (NCD) or a high-fat diet (HFD) for 4 weeks. A mouse model of hyperglycemia was induced by an intraperitoneal injection of streptozotocin (STZ; Sigma-Aldrich, USA). The random blood glucose levels ≥ 16.7 mmol/L 72 h after the injection were considered a successful establishment (18). At the end of 8 weeks, five NCD mice and six HFD/STZ-induced mice were fasted overnight, blood and 24-h urine samples were collected, and then mice were sacrificed. The kidney was harvested for subsequent study. All animal experiments were carried out following the Guide for the Care and Use of Laboratory Animals, and the procedures were approved by the Research Ethical Committee of Chongqing Medical University.

Blood glucose levels were measured using the Roche Dynamic Blood Glucose Monitoring System (Roche, Mannheim, Germany) by blood sampling from the tail vein. Urine albumin, blood urea nitrogen (BUN), and serum creatinine (Scr) were detected using an automatic biochemical analyzer (Hitachi, Tokyo, Japan). The obtained renal tissues were fixed, embedded, and cut into slices. Subsequently, hematoxylin and eosin (H&E), Masson, Periodic Acid-Silver (PAS), Oil Red O staining, and immunofluorescence (IF) staining for the selected biomarkers were performed. The stained slices were visualized and pictured with a light or fluorescence microscopy (Olympus, Tokyo, Japan). According to the manufacturer's instructions, the RT-qPCR was performed. The $2^{-\Delta\Delta Ct}$ method was used to quantify protein kinase cAMP-dependent regulatory type II beta (PRKAR2B) and transforming growth factor-beta-induced (TGFB1) expression with GAPDH as an internal control. The primer sequence is shown in Supplementary Table S1. A Western blot analysis was carried out. Primary antibodies against PRKAR2B (Santa Cruz, CA, USA) and antibodies against TGFB1 (Abcam, Cambridge, UK) were used, respectively.

Statistical Analysis

All statistical analysis was performed using the R software (version 3.6.3) or GraphPad Prism 8.0 (GraphPad Software, CA, USA). A Wilcoxon test was performed to compare immune cell infiltration and the identified biomarker expressions between normal subjects and DN patients. The logistic regression algorithm was used to develop the predictive model. A ROC curve was used to judge the diagnostic accuracy of selected biomarkers. Correlation analysis was realized by Pearson's analysis. Moreover, an unpaired *t*-test was used to analyze the RT-qPCR, Western blot data, biochemical detection data, the differential expression levels of the two biomarkers from the Nephroseq v5 online database, and the differential expression levels of the two biomarkers in other CKD.

If not specially indicated, $p < 0.05$ was defined as statistical significance.

RESULTS

Identification of DEGs and Enrichment Analysis

There was a clearly pronounced discrimination between GDN and normal samples (Figure 1A). A total of 140 DEGs were identified including 75 upregulated and 65 downregulated genes, displayed in the Volcano plot and heatmap (Figures 1B, C).

These DEGs were mainly involved in the biological processes associated with the extracellular structure organization and tumor necrosis factor production ($p < 0.05$, Figure 1D). The results of GSEA illustrated that metabolism-related pathways were enriched in the normal samples, while the immune-related signaling pathways were enriched in the GDN subjects (Figure 1E).

Unsupervised Cluster Construction and Key Module Identification

With the batch effects stripped, the consensus clustering was performed based on the gene expression profiles of the merged 81

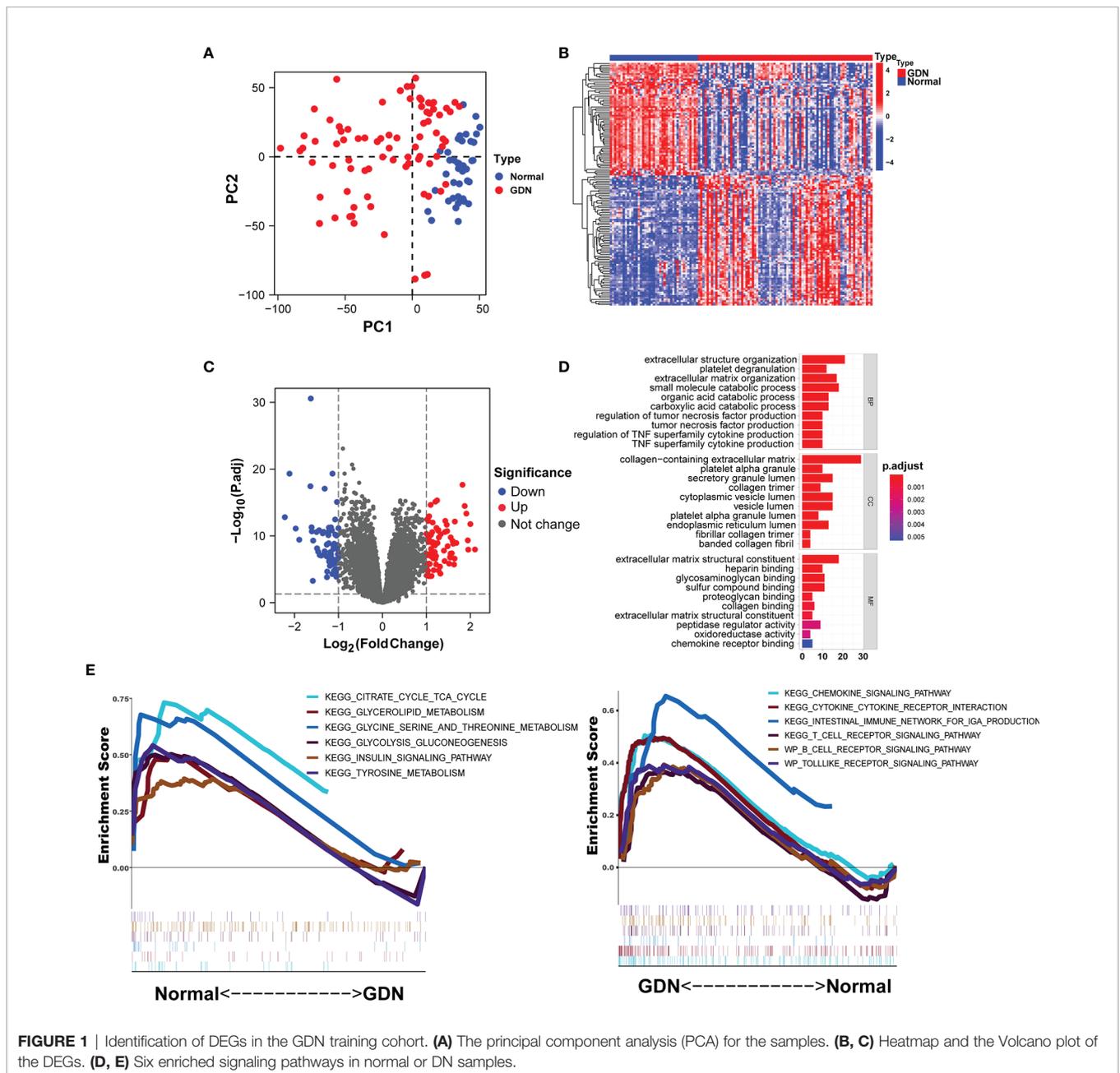
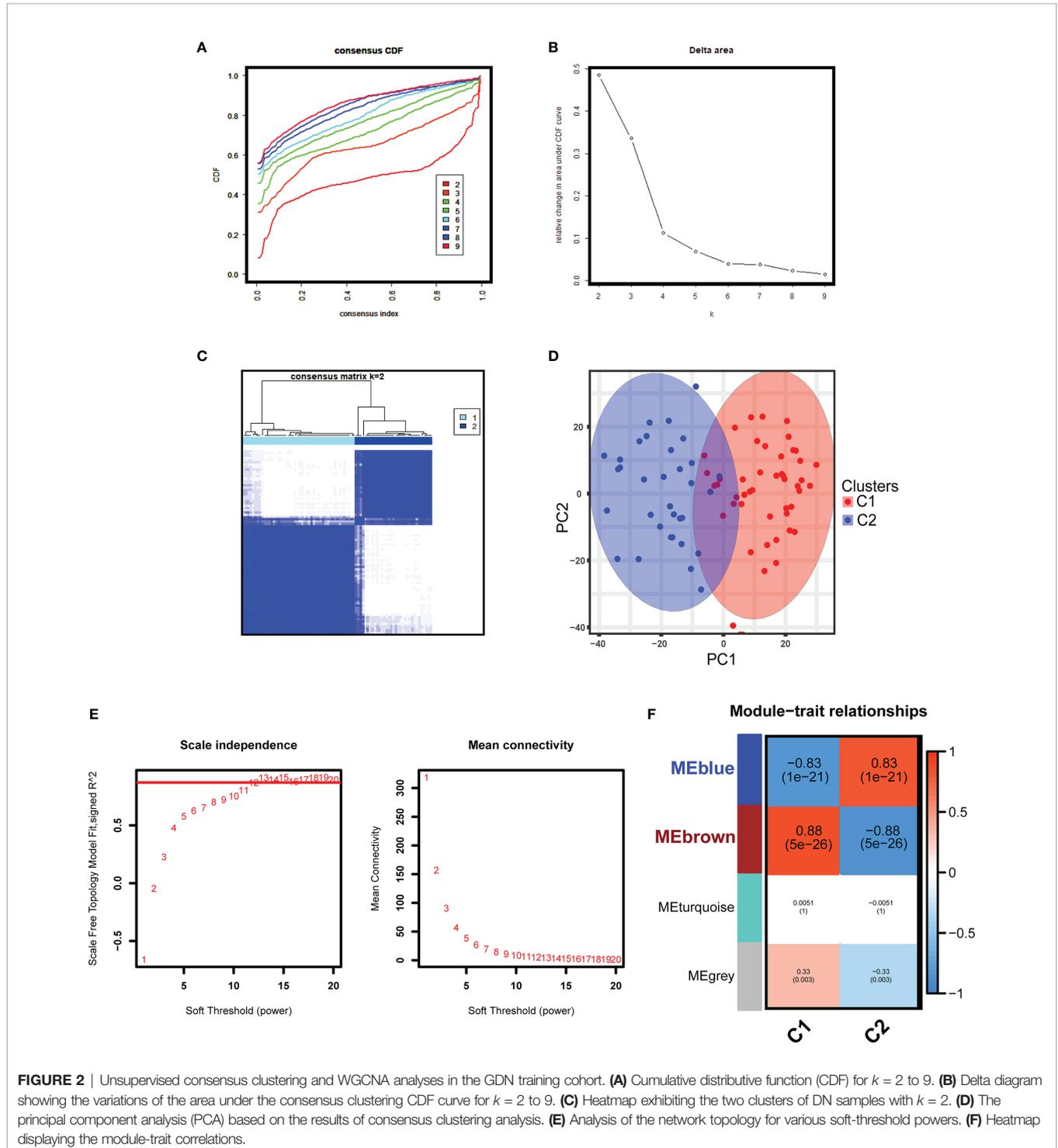


FIGURE 1 | Identification of DEGs in the GDN training cohort. **(A)** The principal component analysis (PCA) for the samples. **(B, C)** Heatmap and the Volcano plot of the DEGs. **(D, E)** Six enriched signaling pathways in normal or DN samples.

GDN samples in the GDN training cohort, and when $k = 2$, the classification was highly reliable and stable (Figures 2A–C). PCA confirmed that there was a distinct difference between the two subclusters (Figure 2D). GDN samples were divided into cluster 1 (C1, $N = 48$) and cluster 2 (C2, $N = 33$). With the soft-threshold power of $\beta = 12$ (scale-free $R^2 = 0.906$) set and the corresponding

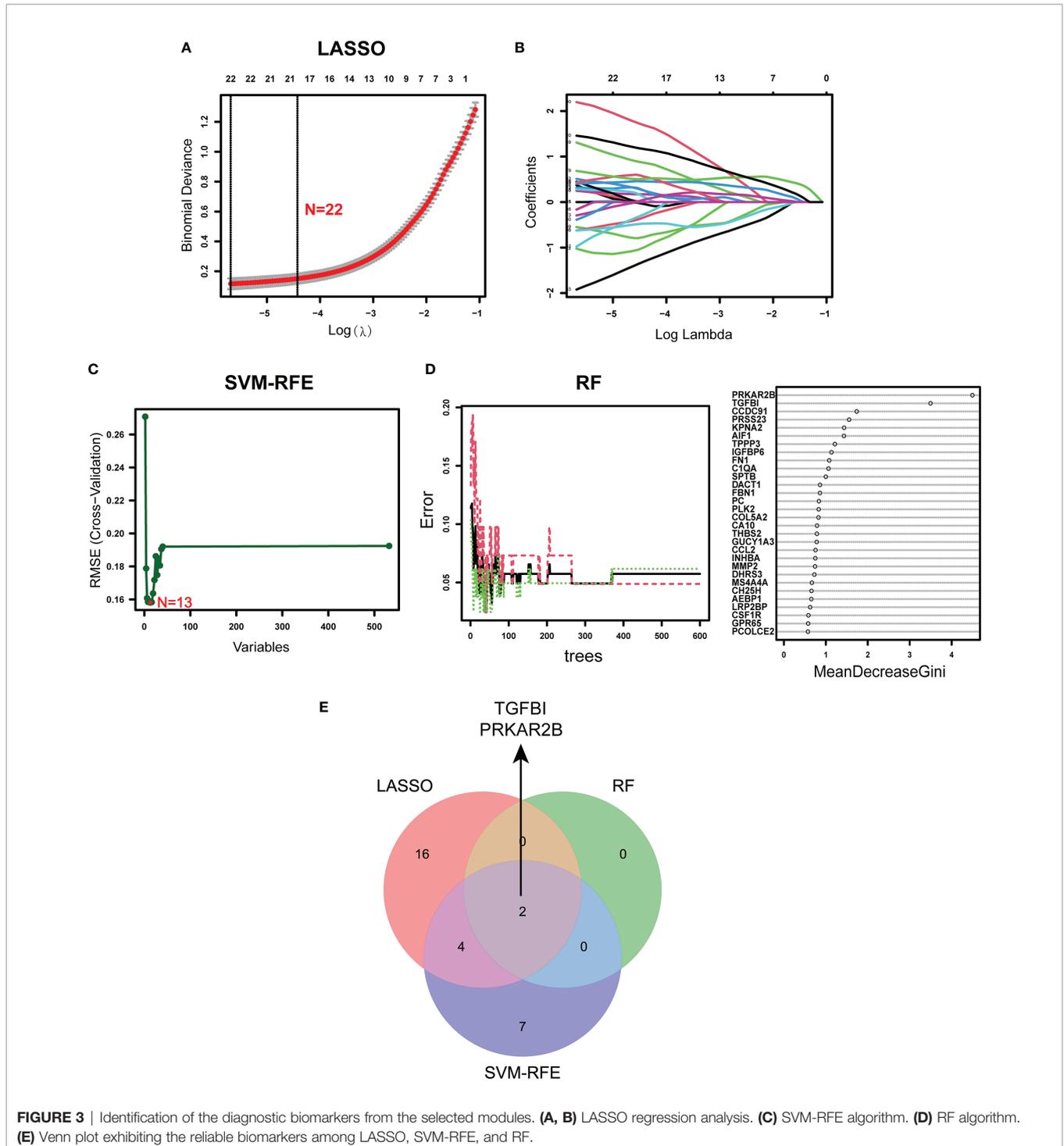
Pearson’s correlation coefficient calculated (Figure 2E), four modules were identified (Figure 2F). Brown and blue modules had the highest correlation with the subclusters, and therefore were selected as the associated modules for further analysis. The genes from the two selected modules were mainly responsible for extracellular structure organization and cytokine chemotaxis



reactions (**Supplementary Figure S2A**). KEGG analysis indicated that they were significantly enriched in complement and coagulation cascades, PI3K-Akt signaling pathway and cytokine-cytokine receptor interaction (**Supplementary Figure S2B**). DO analysis revealed that the genes were mostly involved in urinary system disease, urinary system cancer, and lung disease (**Supplementary Figure S2C**).

Diagnostic Biomarker Identification and Verification

Using the LASSO regression algorithm, 22 genes from the selected modules were identified as potential diagnostic biomarkers (**Figures 3A, B**). By SVM-RFE algorithm, 13 genes were extracted from these modules as candidate biomarkers (**Figure 3C**). Two diagnostic genes were identified by the RF



algorithm (Figure 3D). Two genes (PRKAR2B and TGFBI) were then overlapped *via* a Venn diagram, and served as robust diagnostic biomarkers (Figure 3E). Compared with normal control, decreased PRKAR2B expression ($p < 0.001$) and increased TGFBI expression ($p < 0.001$) were observed in the

glomerular samples from the GDN training cohort (Figure 4A). The results were validated in the GDN testing cohort, and the consistent gene expression patterns were obtained (Figure 4B). Interestingly, the expression of TGFBI was still significantly upregulated in tubulointerstitial samples from the TDN testing

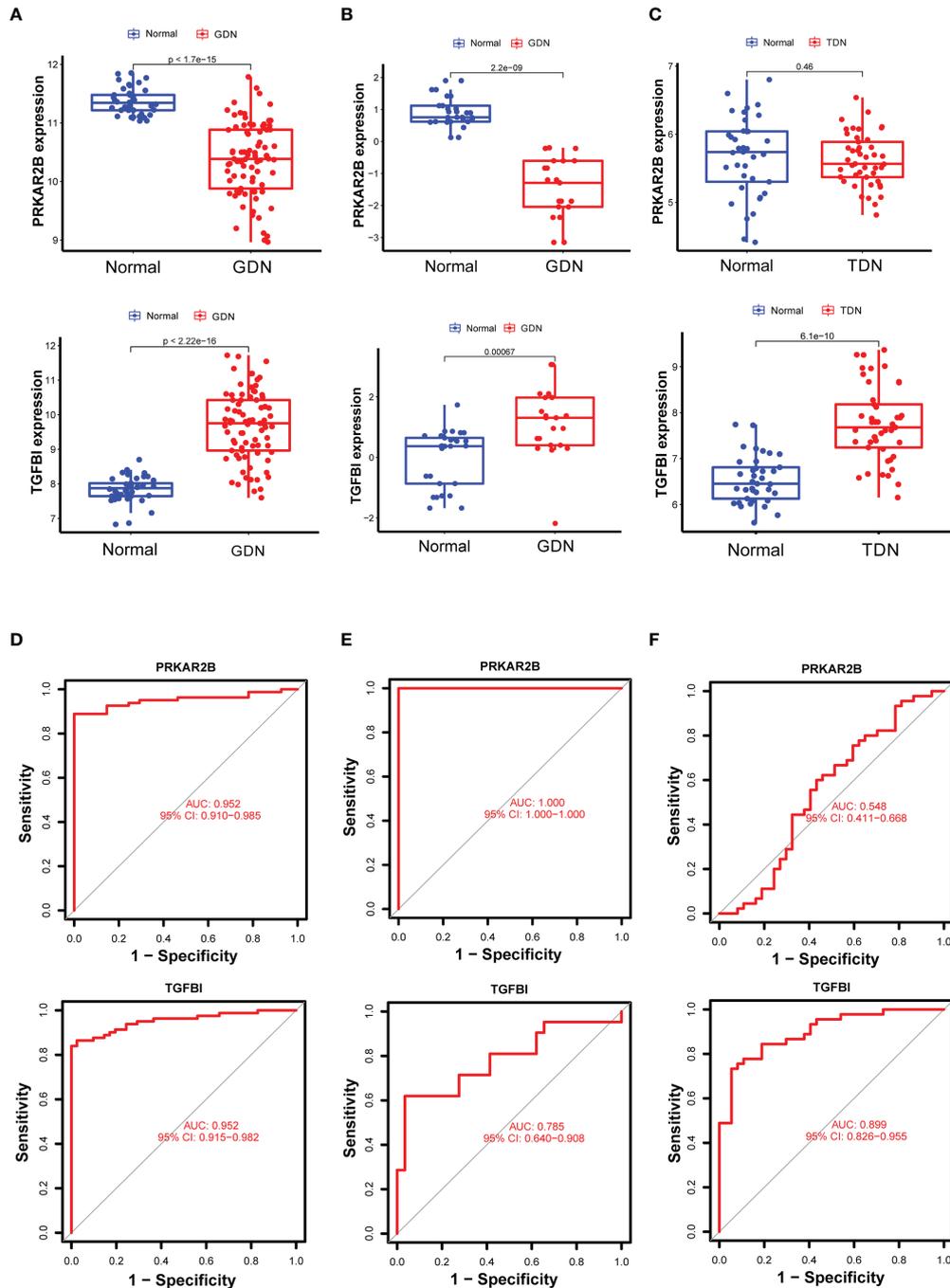


FIGURE 4 | Verification of the identified biomarkers. (A–C) Box plots for the differential expression analysis in the GDN training cohort, GDN testing cohort, and TDN testing cohort, respectively. (D–F) ROC curves for evaluating the diagnostic ability in the GDN training cohort, GDN testing cohort, and TDN testing cohort, respectively. $p < 0.05$ was considered statistically significant.

cohort ($p < 0.001$), while the expression of PRKAR2B had no significant change (Figure 4C). To estimate the predictive utility, the ROC curve was performed and found that the PRKAR2B and TGFBI illustrated a remarkably distinguishing efficiency with AUC values of 0.952 (95% CI: 0.910–0.985) and 0.952 (95% CI: 0.915–0.982) in the GDN training cohort, respectively (Figure 4D). Consistently, in the GDN testing cohort, the AUC value of PRKAR2B was 1.000 (95% CI: 1.000–1.000) and that of TGFBI was 0.785 (95% CI: 0.640–0.908) (Figure 4E). Unlike the low AUC value of PRKAR2B (0.548, 95% CI: 0.411–0.668), TGFBI still maintained a high AUC value of 0.899 (95% CI: 0.826–0.955) in the TDN testing cohort (Figure 4F). Furthermore, the similar expression patterns were also observed in HN and SLEN (Supplementary Figure S3).

Establishment of Nomogram

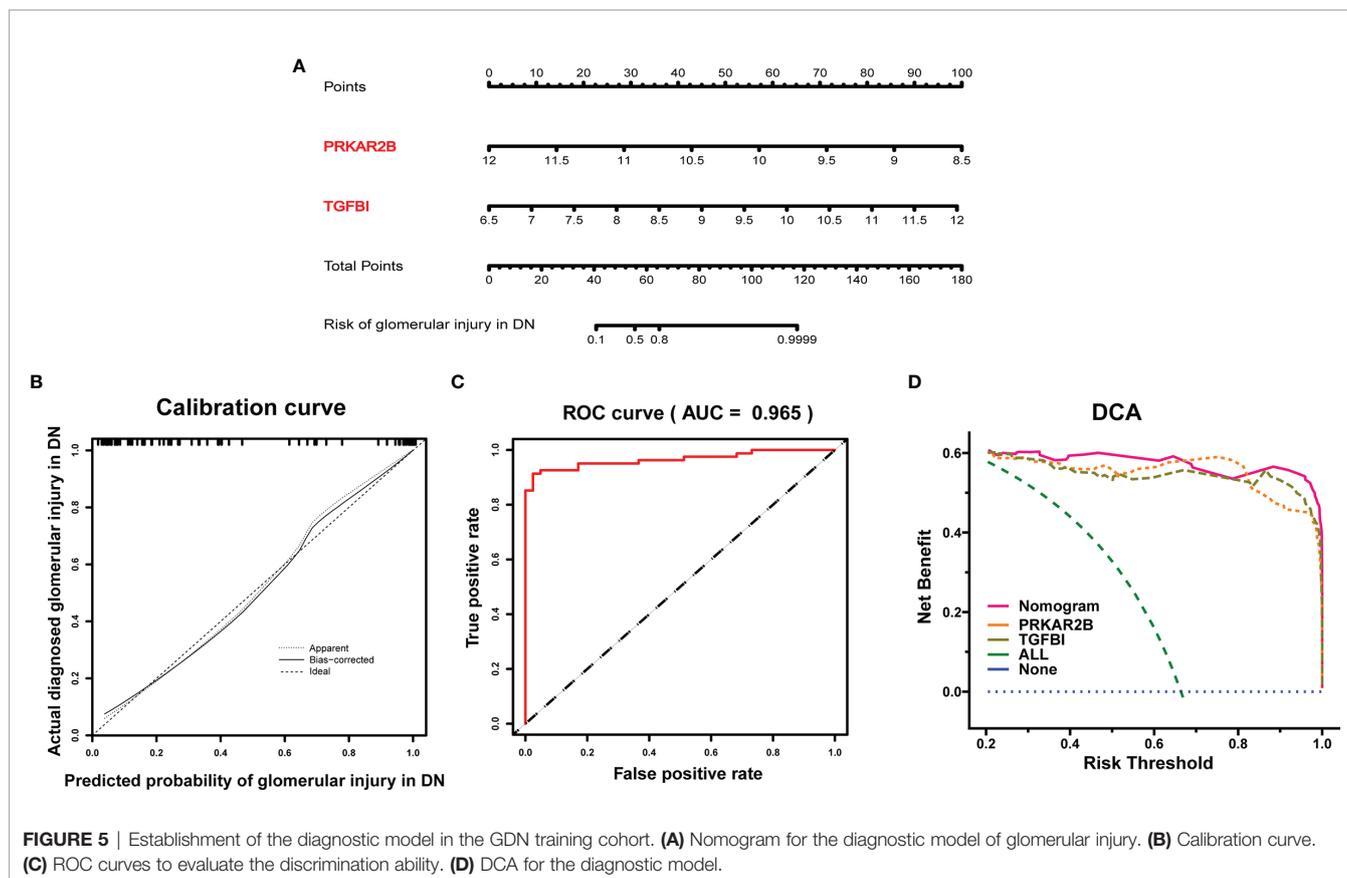
Based on the expressions of PRKAR2B and TGFBI from the GDN training cohort, a diagnostic model was constructed by logistic regression and visualized as a nomogram (Figure 5A). The C-index of the diagnostic model was 0.976 with an appropriate calibration plot. Also, the model showed a high AUC value (0.965), confirming the excellent prediction performance (Figures 5B, C). Additionally, DCA curves indicated the combined nomogram model showed the highest efficacy in predicting glomerular damage in DN patients compared with other single biomarker models (Figure 5D).

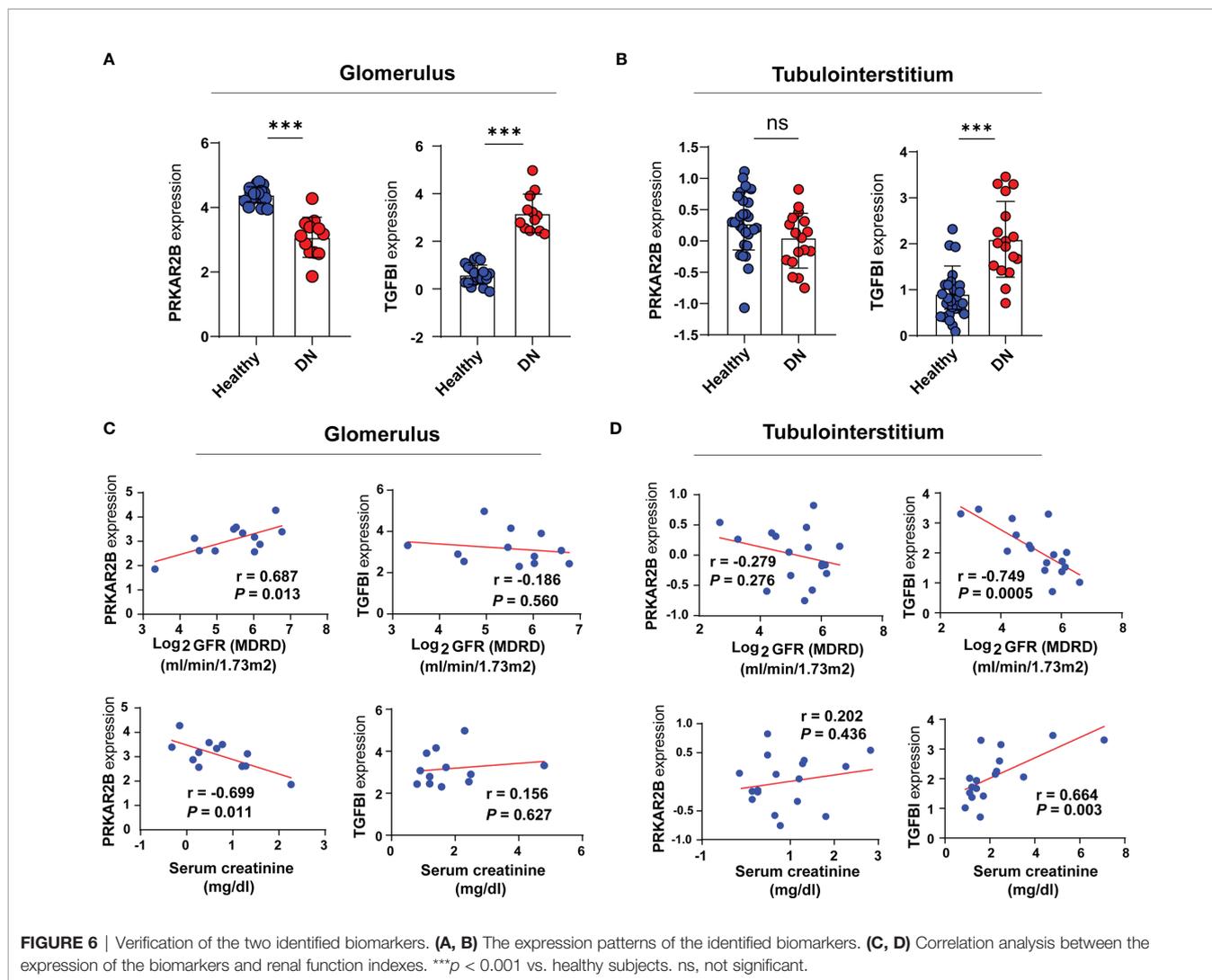
Expression Patterns and Clinical Correlation of the Biomarkers

Based on the Nephroseq v5 online tool, the expression patterns of both PRKAR2B and TGFBI in the glomerular and tubulointerstitial tissues of DN patients were further confirmed (Figures 6A, B). When compared with normal subjects, PRKAR2B expression was downregulated in DN glomerular tissue but not in DN tubulointerstitial tissue. The TGFBI expression was upregulated in both glomerular and tubulointerstitial tissue of DN patients. Correlation analysis revealed that PRKAR2B expression in DN glomerular tissue was positively correlated with glomerular filtration rate (GFR) ($r = 0.687$, $p = 0.013$) and negatively correlated with Scr ($r = -0.699$, $p = 0.011$) (Figure 6C). The TGFBI expression in DN tubulointerstitial tissue was found to be negatively correlated with GFR ($r = -0.749$, $p = 0.0005$) and positively correlated with Scr ($r = 0.664$, $p = 0.003$) (Figure 6D). Curiously, the expression of TGFBI in DN glomerular tissue was not associated with GFR and Scr. It suggested that the biomarkers were related to renal function in patients with DN, whereas their roles may be different.

Correlation Between the Two Biomarkers and Immune Cell Infiltration

The immune infiltration landscape in DN was obviously changed (Supplementary Figure S4). According to the GSVA results, the gene sets in the GDN samples with high PRKAR2B expression





were markedly associated with multiple activated metabolism-related pathways and immune suppression biological functions, as well as the GDN samples with low expression of TGFBI (Figure 7A). Thus, given the roles of PRKAR2B and TGFBI in immune regulations, their effects on the immune cells' infiltration and biological processes were also explored. A proliferation of neutrophils, regulatory T cells (Tregs), macrophages, and plasmacytoid dendritic cells (pDCs) were observed. In addition, the activities of check-point, tumor-infiltrating lymphocytes (TIL), chemokine C-C-Motif receptor (CCR), T-cell coinhibition, and type II interferon (IFN) response were markedly enhanced in the GDN subjects with low PRKAR2B expression or subjects with high TGFBI expression (Figure 7B). Correlation analysis revealed that the infiltration of naive B cells was most positively correlated with PRKAR2B and was most negatively correlated with TGFBI. However, the infiltration of gamma-delta T cells was most negatively correlated with PRKAR2B and was most positively correlated with TGFBI ($p < 0.001$). More details were exhibited in Figure 7C.

Validation in Animal Models

According to the treatment schedule (Figure 8A), four mice in the HFD+STZ group did not meet the established protocols and were excluded. The levels of blood glucose, Scr, BUN, and 24 h urinary protein were significantly elevated in the HFD/STZ-induced mice compared with the NCD mice ($p < 0.01$, Figure 8B). As shown in Figure 8C, glomerular hypertrophy, proliferation of glomerular mesangial cells, dilation of the mesangial matrix, and irregular thickening of the glomerular and tubular basement membrane were observed in the renal tissue of HFD/STZ-induced mouse model. Masson staining revealed the formations of renal blue-stained extracellular collagen, mostly in the glomerular tissue. Oil Red O staining showed the number of lipid droplets increased, and the lipid accumulation in the glomerulus was more obvious than that in the tubulointerstitium. Thus, the HFD combined with high-glucose-induced renal injury model was considered successfully established. The downregulated PRKAR2B expression in glomerular tissue (Figure 8D) and the upregulated TGFBI

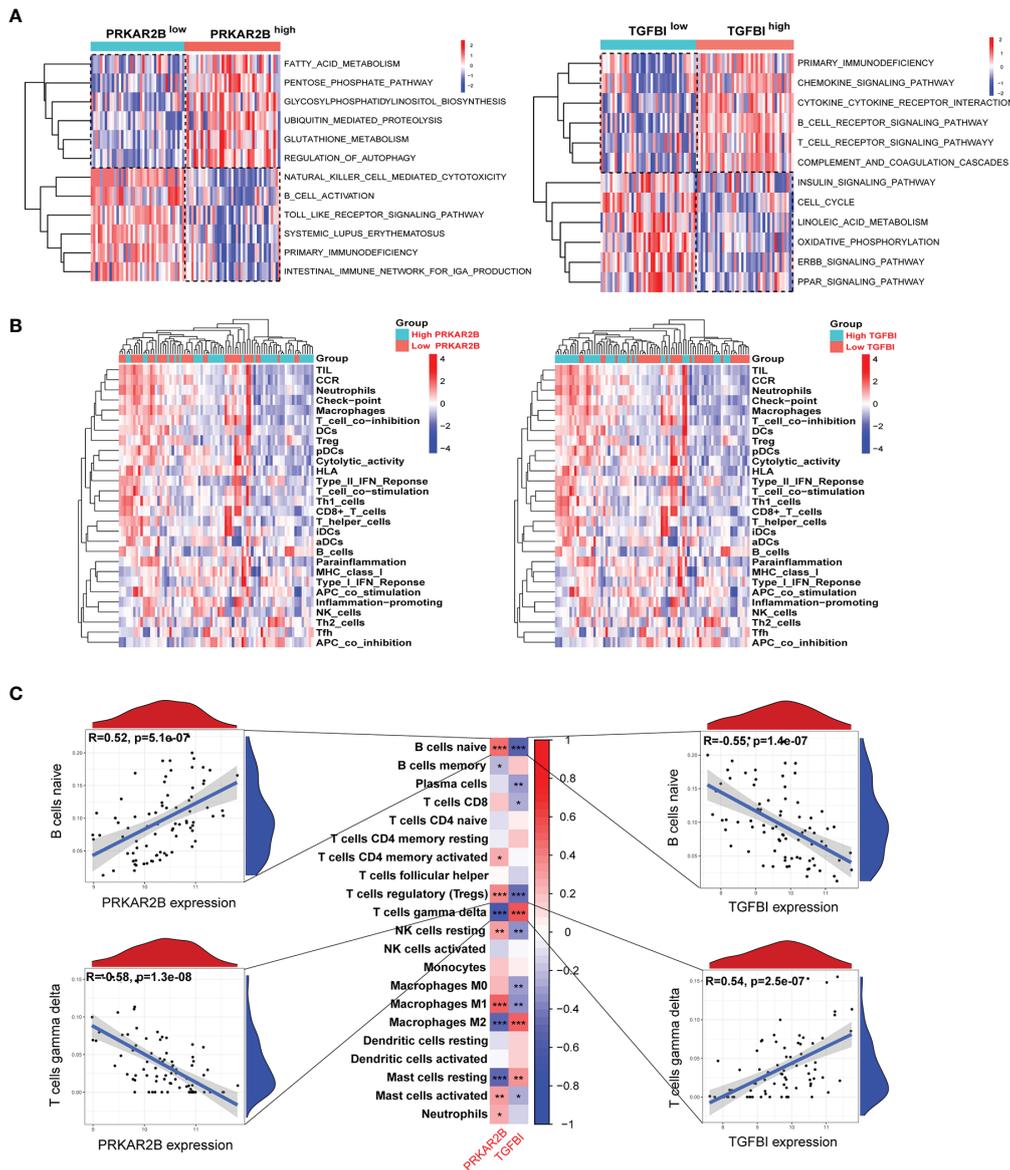


FIGURE 7 | The association between the biomarkers and immune infiltration in the GDN samples. **(A)** Heatmap of metabolism and immune-related gene sets by GSVA. **(B)** Heatmap of the immune landscape by ssGSEA. **(C)** Heatmap of the correlations between the biomarkers and infiltrating immune cells. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

expression in both glomerular and tubulointerstitial tissues of the mouse model were observed (Figure 8E, F). Moreover, the reduced PRKAR2B expression and increased TGFBI expression were also confirmed in the renal tissues of the mouse model by RT-qPCR and Western blot ($p < 0.01$, Figure 8G, H).

DISCUSSION

Diabetic nephropathy results from the interactions of multiple genes. However, its potential mechanisms remain unclear. Recently, a large number of studies have focused on the screening

of related biomarkers. Wang et al. analyzed five DN-associated gene datasets and identified fibronectin 1 (FN1) and complement component 3 (C3) as the immune infiltration-related biomarkers for DN (19). Wang et al. revealed the different pathological abnormalities between glomerulus and kidney tubules in DN and indicated that the changes of key regulated genes in methylation status might contribute to the pathogenesis of DN (20). However, although many efforts have been made to explore novel targets for DN, the present knowledge seems to be insufficient. Potential biomarkers with high specificity and sensitivity are still urgently required.

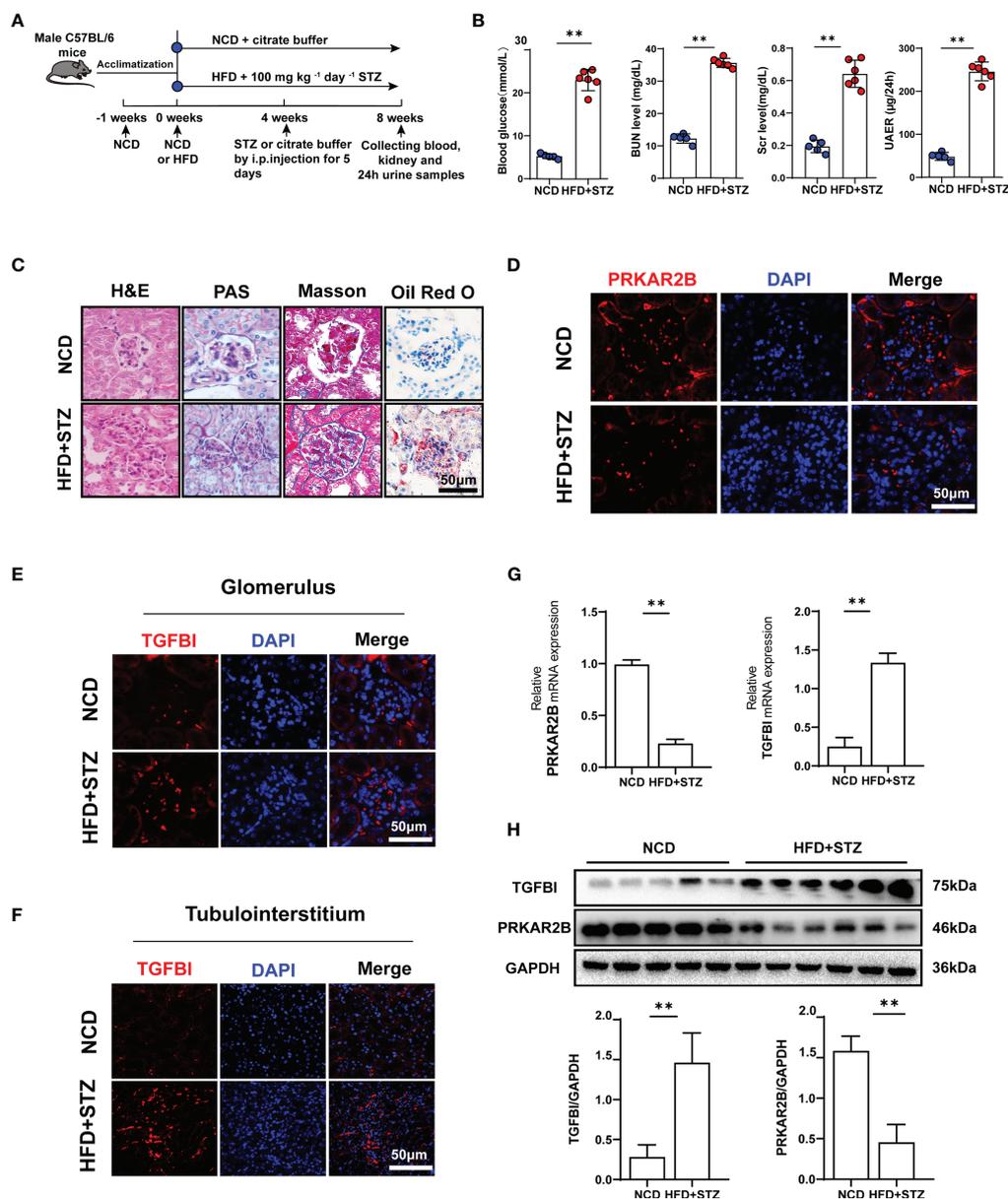


FIGURE 8 | Verification in animal experiments. **(A)** The treatment protocol for mouse model. **(B)** Blood glucose, Scr, BUN, and 24 h proteinuria levels. **(C)** Renal pathological sections stained with H&E, PAS, Masson, and Oil Red O. **(D)** Immunofluorescence staining of PRKAR2B in renal glomerular tissue. **(E, F)** Immunofluorescence staining of TGFBI in glomerular and tubulointerstitial tissues. **(G)** The mRNA expression of PRKAR2B and TGFBI in kidney tissue (RT-qPCR). **(H)** The protein levels of PRKAR2B and TGFBI in kidney tissue (Western blot). ** $p < 0.01$ vs. NCD mice. NCD, normal chow diet; HFD, high-fat diet; STZ, streptozotocin.

PRKAR2B is a cAMP-dependent protein kinase (PKA) (21) regulatory subunit that is abundantly expressed in various tumor tissues (22). However, there are few studies on the role of PRKAR2B in the progression of DN. Our study identified that PRKAR2B, with an excellent diagnostic value (AUC >0.95), was downregulated in the glomerulus but there was no significant change in the tubulointerstitium. TGFBI is a secretory protein induced by TGF- β in various cells and can be detected in serum and urine (23, 24). It was demonstrated that TGFBI was involved

in the fibrotic processes of chronic cyclosporine-induced nephropathy by affecting the synthesis and degradation of the extracellular matrix (25). In the present study, the expression of TGFBI was upregulated in both glomerular and tubulointerstitial tissues, and it was proved to have a reliable diagnostic ability for DN. It was reported that the expression of TGFBI was prominently increased in the kidneys of diabetic patients, whereas the concentration of TGFBI in urine was also raised (26). Elevated urinary TGFBI concentration has been shown to

predict the prognosis of DN (27). This evidence enhanced the accessibility and feasibility of the clinical applications of TGFBI as a diagnostic marker. However, it is unclear why there was no significant correlation between TGFBI expression in glomerular tissue and renal function indexes (such as GFR and Scr) in DN patients. Most notably, a novel diagnostic model combining the two biomarkers was developed with a high AUC value and favorable calibration, which exhibited excellent accuracy and reliability for estimating the glomerular damage in DN patients. Compared with any other single biomarker, the above model showed the highest efficacy for glomerular injury prediction in the GDN training cohort.

In this study, we found that the downregulated PRKAR2B expression in glomerular tissue may indicate the deterioration of kidney function in patients with DN, and so did the upregulated TGFBI in tubulointerstitial tissue. However, similar expression patterns of PRKAR2B and TGFBI were also found in patients with HN or SLEN, which suggested the differential expressions of PRKAR2B and TGFBI were not specific for DN but related to the renal injury.

It had been reported that extracellular matrix organization and extracellular matrix structural constituent lead to the accelerated deposition of extracellular matrix and renal fibrosis in DN (28). In this study, DEGs were demonstrated to be involved in this process in DN glomerular tissue. Multiple metabolism-related pathways were mainly enriched in normal samples, while immune inflammation pathways were mostly concentrated in GDN samples. It confirmed the notion that metabolic disorders and abnormal immune inflammation responses play a critical role in DN (29). Meanwhile, both PRKAR2B and TGFBI were disclosed to be involved in immune-related pathways and cell functions in the glomerular injury of DN. Moreover, both of them were associated with various immune cells such as naïve B cells, gamma delta T cells, Tregs, resting NK cells, resting mast cells, and macrophages. Previous studies reported that the deposition of macrophages, an important feature of DN, could be discovered in the kidney tissue of DN patients, indicating a decline in renal function (30). Mast cells were reported to participate in renal interstitial fibrosis, and the density of mast cells was related to serum creatinine levels in DN (31). It was reported that increased Tregs contributed to the improvement of DN and promoted the transplant tolerance to DN-induced renal allografts (32, 33). However, the roles of naïve B cells and gamma delta T cells in the pathological processes of DN have not been reported. Overall, the infiltrating immune cells are involved in the development and progression of DN. Improving abnormal immune status by targeting PRKAR2B and TGFBI may be a promising approach for the treatment of DN.

Some limitations need to be considered. First, different pathological stages of DN may affect the results of the study. Second, because of the potential heterogeneity from different annotation platforms and clinical covariates of samples, the batch effects cannot be completely eliminated among datasets. Third, the sample size may not be large enough. Finally, the present study was based on public data, so the biological functions of the two biomarkers need to be verified by further experiments.

In this study, using WGCNA, LASSO, SVM-RFE, and RF algorithms, PRKAR2B and TGFBI were identified as the potential biomarkers of DN. A diagnostic model combining PRKAR2B and TGFBI was established to evaluate the risk of diabetic glomerular injury with high sensitivity and accuracy. The potential association with infiltrating immune cells was also demonstrated, providing a fresh perspective on their roles in DN. Therefore, the findings may shed light on the management and treatment of patients with DN.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by the Research Ethical Committee of Chongqing Medical University.

AUTHOR CONTRIBUTIONS

HH, YC, and KL: conceptualization and methodology. HH, YC, HY, SZ, and WC: software and data curation. YL and QL: validation. DL and GY: reviewed and edited the manuscript. KL is the guarantor of this work and, as such, has full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This study is supported by the National Natural Science Foundation of China (81671381), Natural Science Foundation of Chongqing (cstc2021jcyj-msxmX0169), Science and Health Joint Medical Research Project of Chongqing (2022GDRC018), High-Level Medical Reserved Personnel Train Project of Chongqing, and Kuanren Talents Program of the Second Affiliated Hospital of Chongqing Medical University.

ACKNOWLEDGMENTS

The authors thank the contributors to the GEO database for sharing data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fendo.2022.876960/full#supplementary-material>

REFERENCES

- Tung CW, Hsu YC, Shih YH, Chang PJ, Lin CL. Glomerular Mesangial Cell and Podocyte Injuries in Diabetic Nephropathy. *Nephrol (Carlton)* (2018) 23 Suppl 4:32–7. doi: 10.1111/nep.13451
- Wang K, Hu J, Luo T, Wang Y, Yang S, Qing H, et al. Effects of Angiotensin-Converting Enzyme Inhibitors and Angiotensin II Receptor Blockers on All-Cause Mortality and Renal Outcomes in Patients With Diabetes and Albuminuria: A Systematic Review and Meta-Analysis. *Kidney Blood Press Res* (2018) 43:768–79. doi: 10.1159/000489913
- Palmer SC, Tendal B, Mustafa RA, Vandvik PO, Li S, Hao Q, et al. Sodium-Glucose Cotransporter Protein-2 (SGLT-2) Inhibitors and Glucagon-Like Peptide-1 (GLP-1) Receptor Agonists for Type 2 Diabetes: Systematic Review and Network Meta-Analysis of Randomised Controlled Trials. *BMJ* (2021) 372:m4573. doi: 10.1136/bmj.m4573
- Pareek CS, Smoczynski R, Tretyn A. Sequencing Technologies and Genome Sequencing. *J Appl Genet* (2011) 52:413–35. doi: 10.1007/s13353-011-0057-x
- Barrett T, Edgar R. Gene Expression Omnibus: Microarray Data Storage, Submission, Retrieval, and Analysis. *Methods Enzymol* (2006) 411:352–69. doi: 10.1016/S0076-6879(06)11019-8
- Li G, Zhang J, Liu D, Wei Q, Wang H, Lv Y, et al. Identification of Hub Genes and Potential ceRNA Networks of Diabetic Nephropathy by Weighted Gene Co-Expression Network Analysis. *Front Genet* (2021) 12:767654. doi: 10.3389/fgene.2021.767654
- Parker HS, Leek JT, Favorov AV, Conside M, Xia X, Chavan S, et al. Preserving Biological Heterogeneity With a Permuted Surrogate Variable Analysis for Genomics Batch Correction. *Bioinformatics* (2014) 30:2757–63. doi: 10.1093/bioinformatics/btu375
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res* (2015) 43:e47. doi: 10.1093/nar/gkv007
- Wilkerson MD, Hayes DN. ConsensusClusterPlus: A Class Discovery Tool With Confidence Assessments and Item Tracking. *Bioinformatics* (2010) 26:1572–3. doi: 10.1093/bioinformatics/btq170
- Langfelder P, Horvath S. WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559
- Zhao E, Xie H, Zhang Y. Predicting Diagnostic Gene Biomarkers Associated With Immune Infiltration in Patients With Acute Myocardial Infarction. *Front Cardiovasc Med* (2020) 7:586871. doi: 10.3389/fcvm.2020.586871
- Lin X, Yang F, Zhou L, Yin P, Kong H, Xing W, et al. A Support Vector Machine-Recursive Feature Elimination Feature Selection Method Based on Artificial Contrast Variables and Mutual Information. *J Chromatogr B Analyt Technol BioMed Life Sci* (2012) 910:149–55. doi: 10.1016/j.jchromb.2012.05.020
- Guo L, Wang Z, Du Y, Mao J, Zhang J, Yu Z, et al. Random-Forest Algorithm Based Biomarkers in Predicting Prognosis in the Patients With Hepatocellular Carcinoma. *Cancer Cell Int* (2020) 20:251. doi: 10.1186/s12935-020-01274-z
- Deng M, Yin Y, Zhang Q, Zhou X, Hou G. Identification of Inflammation-Related Biomarker Lp-PLA2 for Patients With COPD by Comprehensive Analysis. *Front Immunol* (2021) 12:670971. doi: 10.3389/fimmu.2021.670971
- Lay AC, Hale LJ, Stowell-Connolly H, Pope RJP, Nair V, Ju W, et al. IGF1BP-1 Expression Is Reduced in Human Type 2 Diabetic Glomeruli and Modulates Beta1-Integrin/FAK Signalling in Human Podocytes. *Diabetologia* (2021) 64:1690–702. doi: 10.1007/s00125-021-05427-1
- Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* (2013) 39:782–95. doi: 10.1016/j.immuni.2013.10.003
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust Enumeration of Cell Subsets From Tissue Expression Profiles. *Nat Methods* (2015) 12:453–7. doi: 10.1038/nmeth.3337
- Li Y, Hou JG, Liu Z, Gong XJ, Hu JN, Wang YP, et al. Alleviative Effects of 20 (R)-Rg3 on HFD/STZ-Induced Diabetic Nephropathy via MAPK/NF-kappaB Signaling Pathways in C57BL/6 Mice. *J Ethnopharmacol* (2021) 267:113500. doi: 10.1016/j.jep.2020.113500
- Wang Y, Zhao M, Zhang Y. Identification of Fibronectin 1 (FN1) and Complement Component 3 (C3) as Immune Infiltration-Related Biomarkers for Diabetic Nephropathy Using Integrated Bioinformatic Analysis. *Bioengineered* (2021) 12:5386–401. doi: 10.1080/21655979.2021.1960766
- Wang YZ, Xu WW, Zhu DY, Zhang N, Wang YL, Ding M, et al. Specific Expression Network Analysis of Diabetic Nephropathy Kidney Tissue Revealed Key Methylated Sites. *J Cell Physiol* (2018) 233:7139–47. doi: 10.1002/jcp.26638
- Adachi K, Hamer JE. Divergent cAMP Signaling Pathways Regulate Growth and Pathogenesis in the Rice Blast Fungus *Magnaporthe Grisea*. *Plant Cell* (1998) 10:1361–74. doi: 10.1105/tpc.10.8.1361
- Sha J, Xue W, Dong B, Pan J, Wu X, Li D, et al. PRKAR2B Plays an Oncogenic Role in the Castration-Resistant Prostate Cancer. *Oncotarget* (2017) 8:6114–29. doi: 10.18632/oncotarget.14044
- Ween MP, Oehler MK, Ricciardelli C. Transforming Growth Factor-Beta-Induced Protein (TGFB1)/(betaig-H3): A Matrix Protein With Dual Functions in Ovarian Cancer. *Int J Mol Sci* (2012) 13:10461–77. doi: 10.3390/ijms130810461
- Yang L, Cui R, Li Y, Liang K, Ni M, Gu Y. Hypoxia-Induced TGFB1 as a Serum Biomarker for Laboratory Diagnosis and Prognosis in Patients With Pancreatic Ductal Adenocarcinoma. *Lab Med* (2020) 51:352–61. doi: 10.1093/labmed/lmz063
- Li C, Lim SW, Choi BS, Lee SH, Cha JH, Kim IS, et al. Inhibitory Effect of Pravastatin on Transforming Growth Factor Beta1-Inducible Gene H3 Expression in a Rat Model of Chronic Cyclosporine Nephropathy. *Am J Nephrol* (2005) 25:611–20. doi: 10.1159/000089905
- Cha DR, Kim IS, Kang YS, Han SY, Han KH, Shin C, et al. Urinary Concentration of Transforming Growth Factor-Beta-Inducible Gene-H3 (Beta Ig-H3) in Patients With Type 2 Diabetes Mellitus. *Diabetes Med* (2005) 22:14–20. doi: 10.1111/j.1464-5491.2004.01295.x
- Ha SW, Kim HJ, Bae JS, Jeong GH, Chung SC, Kim JG, et al. Elevation of Urinary Betaig-H3, Transforming Growth Factor-Beta-Induced Protein in Patients With Type 2 Diabetes and Nephropathy. *Diabetes Res Clin Pract* (2004) 65:167–73. doi: 10.1016/j.diabres.2004.01.007
- Hills CE, Siamantouras E, Smith SW, Cockwell P, Liu KK, Squires PE. TGFBeta Modulates Cell-To-Cell Communication in Early Epithelial-to-Mesenchymal Transition. *Diabetologia* (2012) 55:812–24. doi: 10.1007/s00125-011-2409-9
- Hou Y, Wang Q, Han B, Chen Y, Qiao X, Wang L. CD36 Promotes NLRP3 Inflammasome Activation via the mtROS Pathway in Renal Tubular Epithelial Cells of Diabetic Kidneys. *Cell Death Dis* (2021) 12:523. doi: 10.1038/s41419-021-03813-6
- Nguyen D, Ping F, Mu W, Hill P, Atkins RC, Chadban SJ. Macrophage Accumulation in Human Progressive Diabetic Nephropathy. *Nephrol (Carlton)* (2006) 11:226–31. doi: 10.1111/j.1440-1797.2006.00576.x
- Okoń K, Stachura J. Increased Mast Cell Density in Renal Interstitium is Correlated With Relative Interstitial Volume, Serum Creatinine and Urea Especially in Diabetic Nephropathy But Also in Primary Glomerulonephritis. *Pol J Pathol* (2007) 58:193–7.
- Eller K, Kirsch A, Wolf AM, Sopper S, Tagwerker A, Stanzl U, et al. Potential Role of Regulatory T Cells in Reversing Obesity-Linked Insulin Resistance and Diabetic Nephropathy. *Diabetes* (2011) 60:2954–62. doi: 10.2337/db11-0358
- Dousdampanis P, Trigka K, Mouzaki A. Tregs and Kidney: From Diabetic Nephropathy to Renal Transplantation. *World J Transplant* (2016) 6:556–63. doi: 10.5500/wjt.v6.i3.556

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Han, Chen, Yang, Cheng, Zhang, Liu, Liu, Liu, Yang and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.