scientific reports

Check for updates

OPEN

Integrating bioinformatics and machine learning to identify glomerular injury genes and predict drug targets in diabetic nephropathy

Li Zhang[1,2,6], ZhenPeng Sun[4,6], Yao Yuan[5] & Jie Sheng[2,3]✉

Diabetes mellitus (DM) is a chronic metabolic disorder that poses significant challenges to public health. Among its various complications, diabetic nephropathy (DN) emerges as a critical microvascular complication associated with high mortality rates. Despite the development of diverse therapeutic strategies targeting metabolic improvement, hemodynamic regulation, and fibrosis mitigation, the precise mechanisms responsible for glomerular injury in DN are not yet fully elucidated. To explore these mechanisms, public DN datasets (GSE30528, GSE104948, and GSE96804) were obtained from the GEO database. We merged the GSE30528 and GSE104948 datasets to identify differentially expressed genes (DEGs) between DN and control groups using R software. Weighted gene co-expression network analysis (WGCNA) was subsequently employed to discern genes associated with DN in key modules. We utilized Venny software to pinpoint co-expressed genes shared between DEGs and key module genes. These co-expressed genes underwent gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) enrichment analyses. Through LASSO, SVM, and RF methods, we isolated five significant genes: FN1, C1orf21, CD36, CD48, and SRPX2. These genes were further validated using a logistic model and 10-fold cross-validation. The external dataset GSE96804 served to validate the identified biomarkers, while receiver operating characteristic (ROC) curve analysis assessed their diagnostic efficacy for DN. Additionally, GSE104948 facilitated comparison of biomarker expression levels between DN and five other kidney diseases, highlighting their specificity for DN. These biomarkers also enabled the identification and validation of two molecular subtypes characterized by distinct immune profiles. The Nephroseq v5 database corroborated the correlation between biomarkers and clinical data. Furthermore, the GSigDB database was employed to predict protein-drug interactions, with molecular docking confirming the therapeutic potential of these drug targets. Finally, a diabetic mouse model (BKS-db) was constructed, and RT-qPCR experiments validated the reliability of the identified biomarkers. The study identified five biomarkers with robust diagnostic predictive power for DN. Subtype classification based on these biomarkers revealed distinct enrichment pathways and immune cell infiltration profiles, underscoring the close relationship between these genes and immune functions in DN. Drug prediction and molecular docking analyses demonstrated excellent binding affinities of candidate drugs to target proteins. Differential expression analysis between DN and five other kidney diseases indicated that all biomarkers, except C1orf21, were highly expressed in DN. Notably, as the mouse model lacks the C1orf21 gene, RT-qPCR confirmed the upregulated expression of FN1, CD36, CD48, and SRPX2. This study successfully identified five biomarkers with potential diagnostic and therapeutic value for DN. These biomarkers not only offer insights into the regulatory mechanisms underlying glomerular injury but also provide a theoretical foundation for the development of diagnostic biomarkers and therapeutic targets related to DN-associated glomerular injury.

Keywords Diabetic nephropathy, Biomarkers, Bioinformatics, GEO, Glomerular injury, Drug prediction

[1]Department of Epidemiology and Statistics, College of Public Health, Zhengzhou University, Zhengzhou 450001, Henan, China. [2]School of Basic Medical Sciences, Chongqing Medical University, Chongqing 400016, China. [3]The

Joint International Research Laboratory of Reproduction and Development, Ministry of Education, Chongqing 400016, China. [4]Department of Urology, Xi'an Daxing Hospital, Xian, Shaanxi 710016, China. [5]Department of Pharmacology, College of Pharmacy, Army Medical University, Chongqing 400016, China. [6]Li Zhang and ZhenPeng Sun contributed equally to this work and shared first authorship. ✉email: shengjie@cqmu.edu.cn

Diabetes mellitus (DM) is a chronic metabolic disorder with profound public health implications, characterized primarily by persistent hyperglycemia. Its global prevalence continues to rise, posing a significant burden on healthcare systems. According to the World Health Organization (WHO), the number of individuals with type 2 diabetes (T2DM) is projected to reach approximately 629 million by 2045. Diabetes mellitus (DM) is associated with a spectrum of complications affecting multiple organ systems, including the cardiovascular, nervous, retinal, and renal systems, thereby contributing to increased morbidity and mortality rates[1]. Among these, diabetic nephropathy (DN) represents a major microvascular complication and the leading cause of end-stage renal disease (ESRD)[2].

The progression of DN is marked by several clinical hallmarks, including hypertension, increased urinary protein excretion, and a gradual decline in estimated glomerular filtration rate (eGFR). These pathological changes are influenced by genetic predisposition, hemodynamic disturbances, and metabolic dysregulation[3]. Histologically, DN is characterized by thickening of the glomerular and tubular basement membranes (BM), leading to progressive extracellular matrix (ECM) accumulation and tubulointerstitial fibrosis, ultimately resulting in irreversible kidney damage[4]. Impaired renal microcirculation further exacerbates glomerular capillary narrowing, culminating in glomerular obsolescence, declining filtration function, and, in advanced stages, azotemia and renal failure.

Microalbuminuria (MA) remains the primary clinical biomarker for DN diagnosis; however, its reliability in detecting early kidney injury and predicting disease progression remains debated. Individual variability and disease stage further complicate its diagnostic accuracy. Thus, there is an urgent need for more sensitive and specific biomarkers to enable early diagnosis and refined disease classification. Despite advancements in therapeutic strategies targeting metabolic, hemodynamic, and fibrotic pathways, the progression of renal dysfunction and the associated mortality rates in DN remain significant challenges, underscoring the necessity for deeper mechanistic insights and the development of targeted interventions.

Recent advances in high-throughput sequencing have facilitated the identification of molecular mechanisms and biomarkers associated with DN. Urinary and serum/plasma biomarkers have been shown to correlate with renal structural and functional alterations in DN patients[5]. To elucidate the molecular basis of glomerular injury in DN, this study integrates multi-omics data and bioinformatics analyses to systematically characterize DN-associated biomarkers and their functional relevance.

We analyzed multiple microarray datasets, including GSE30528, GSE104948, and GSE96804. Differentially expressed genes (DEGs) were identified, and weighted gene co-expression network analysis (WGCNA) was performed to delineate DN-associated gene modules in the GSE30528 and GSE104948 datasets. Three machine learning algorithms were employed to refine key biomarker selection. The diagnostic potential of these biomarkers was assessed via receiver operating characteristic (ROC) curve analysis using the external validation dataset GSE96804. Additionally, expression patterns of five biomarkers were compared between DN and other kidney diseases in the GSE104948 dataset. To explore molecular heterogeneity, consensus clustering was applied to identify DN subtypes based on patient gene expression profiles.

Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA) were utilized to elucidate the biological functions and pathway involvement of candidate genes. Protein-drug interaction and molecular docking analyses were conducted to assess the binding affinities of potential therapeutic compounds, providing insights into the regulatory mechanisms underlying glomerular injury in DN. Finally, key biomarkers were validated in a diabetic BKS-db mouse model using reverse transcription quantitative PCR (RT-qPCR), reinforcing the reliability of our bioinformatics predictions.

By integrating multi-omics analyses with experimental validation, this study delineates key molecular mechanisms and potential biomarkers implicated in glomerular injury in DN, offering theoretical insights and candidate targets for precision diagnostics and personalized therapeutic strategies. The overall workflow of this study is illustrated in Fig. 1.

Materials and methods

Data acquisition and integration

To identify key biomarkers associated with diabetic nephropathy (DN), the publicly available microarray datasets GSE30528 and GSE104948 were retrieved from the NCBI Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov) using the keywords "Diabetic Nephropathy" and "Glomerulus." RNA sequencing data were integrated with the corresponding clinical information for each sample. The GSE30528 dataset, generated using the GPL571 platform, comprised gene expression profiles from glomerular tissues of 9 DN patients and 13 healthy controls. The GSE104948 dataset, based on the GPL22945 platform, included transcriptomic data from glomerular tissues of 7 DN patients and 18 normal controls. To enhance statistical power, these datasets were merged into a single cohort, and batch effects were mitigated using the sva R package[6]. The efficiency of batch correction was assessed via principal component analysis (PCA).

For external validation, the GSE96804 dataset (GPL17586 platform), which contains gene expression data from glomerular tissues of 41 DN patients and 20 healthy controls, was utilized to confirm the diagnostic relevance of the identified biomarkers. Additionally, the GSE104948 dataset (incorporating GPL22945 and GPL24120 platforms) was employed to evaluate the differential expression of candidate biomarkers across DN and five other kidney diseases—IgA nephropathy, membranous nephropathy, hypertensive nephropathy, lupus nephritis, and focal segmental glomerulosclerosis—thereby ensuring biomarker specificity for DN.

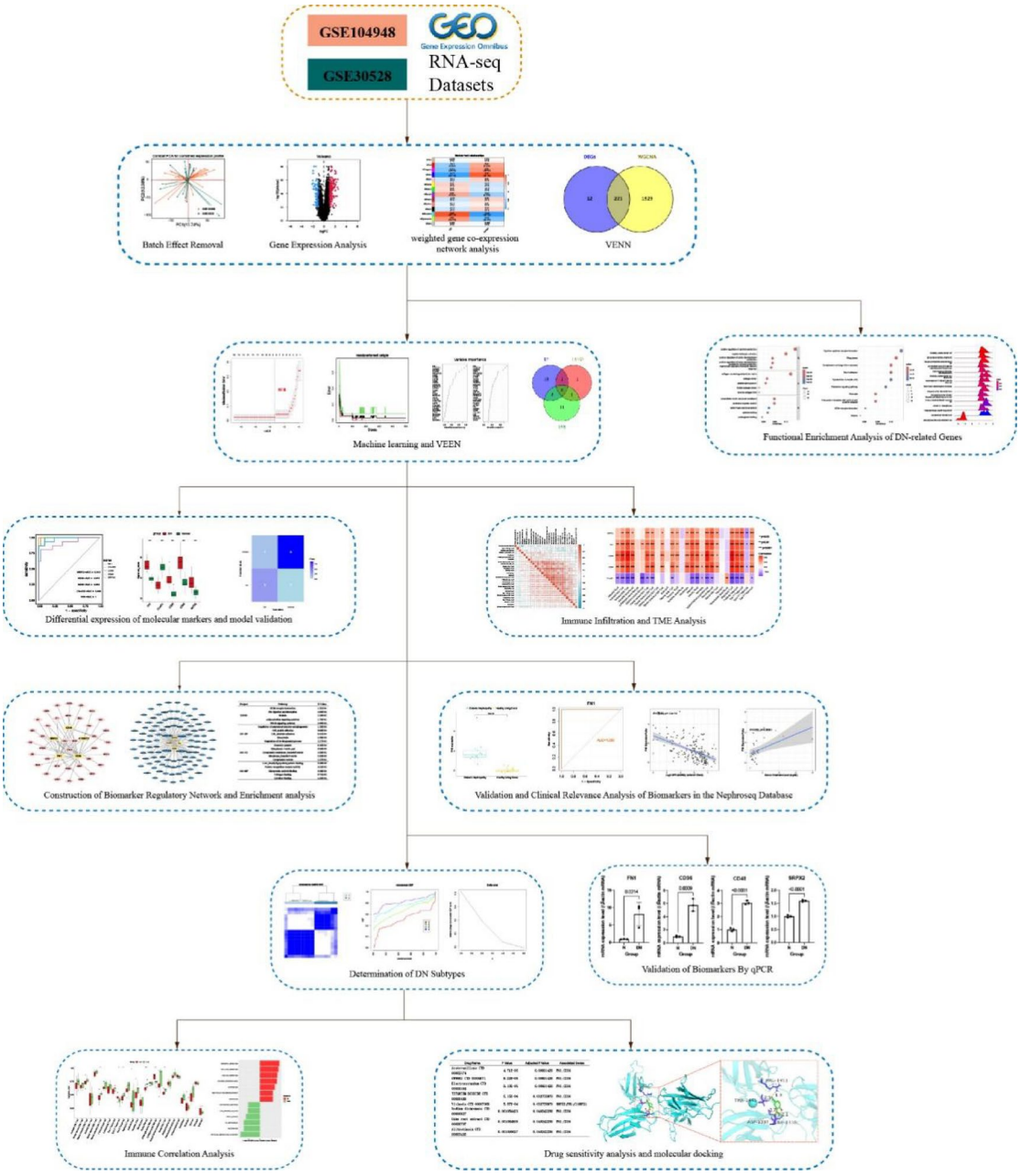

Fig. 1. A flowchart for the study to identify the hug genes and drug predictions.

Identification of differentially expressed genes (DEGs)

To identify key molecular alterations in DN, differentially expressed genes (DEGs) were determined using the limma R package, with statistical thresholds set at an adjusted $p < 0.01$ and $|\log_2\text{FoldChange}| \geq 1$. The identified DEGs were visualized through volcano plots and heatmaps, generated using the pheatmap and ggplot2 R packages[7].

Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) was performed to construct a gene co-expression network and to explore the relationships between gene modules and DN-related clinical phenotypes[8]. Key parameters for WGCNA were set as follows: minimum module size = 30, merging threshold = 0.25, and module membership > 0.8. Genes within highly correlated modules that overlapped with DEGs were designated as DN-related genes and were subjected to further analyses.

Functional enrichment analysis of DN-related genes

To elucidate the biological functions of DN-related genes, functional enrichment analyses, including Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG)[9–11], and Gene Set Enrichment Analysis

(GSEA), were conducted using the ClusterProfiler and ggplot2 R packages[12]. GO analysis encompassed biological processes, cellular components, and molecular functions, while KEGG pathway analysis provided insights into cellular and systemic regulatory mechanisms. Significance thresholds were set at $p < 0.05$ and false discovery rate (FDR) < 0.05. For GSEA, the c2.cp.kegg.v7.4.symbols gene set was utilized, with enrichment criteria defined as |normalized enrichment score (NES)| > 1, $p < 0.05$, and q < 0.25[13].

Machine learning analysis

To refine key diagnostic biomarkers for DN, advanced machine learning algorithms were applied to the identified DEGs. The GSE30528 and GSE104948 datasets were randomly partitioned into training and validation sets in a 7:3 ratio. Three distinct machine learning methods were employed: Least Absolute Shrinkage and Selection Operator (LASSO) regression was used to identify optimal feature genes by selecting the λ value with the lowest classification error[14]. Support Vector Machine-Recursive Feature Elimination (SVM-RFE)[15] iteratively removed less informative features to enhance classification performance. The random forest (RF) algorithms[16] analysis ranked feature importance and identified the most predictive biomarkers. Biomarkers consistently identified across all three algorithms were considered robust diagnostic candidates, with the results visualized using Venny (version 2.1.0, https://bioinfogp.cnb.csic.es/tools/venny/index.html). The diagnostic performance of these biomarkers was further validated through receiver operating characteristic (ROC) curve analysis using the roc R package, ensuring high sensitivity and specificity in distinguishing DN from healthy controls.

Validation of diagnostic biomarkers and model construction

To evaluate the predictive accuracy of the identified biomarkers and their association with diabetic nephropathy (DN), the GSE30528 and GSE104948 datasets were randomly partitioned into training and test sets in a 7:3 ratio. A logistic regression model was constructed using the training cohort and subsequently validated in the test cohort. The model's performance was assessed through receiver operating characteristic (ROC) curve analysis, with accuracy further evaluated using a confusion matrix. To ensure the robustness of the findings, a 10-fold cross-validation was performed. The GSE30528 and GSE104948 datasets were randomly divided into ten equal subsets, with each subset serving as the test cohort in turn, while the remaining subsets were used for model training. The validation cycle yielding the highest accuracy was designated as the test cohort, with the remaining data forming the training cohort. ROC analysis and confusion matrices were employed to validate the model's performance in both cohorts. For external validation, the GSE96804 dataset was used to compare gene expression differences between DN and normal controls, followed by ROC analysis to assess the model's diagnostic capability[17]. Additionally, the GSE104948 dataset was utilized to evaluate the expression differences of the identified biomarkers across DN and other kidney diseases, ensuring biomarker specificity.

Immune infiltration and functional analysis

Single-sample gene set enrichment analysis (ssGSEA)[18] was conducted using the GSVA R package to quantify the infiltration levels of 28 immune cell types in each sample. Three distinct scores—immune score, stromal score, and ESTIMATE score—were calculated using the ESTIMATE algorithm. Correlations between biomarker expression and immune infiltration levels were examined using the ggcorplot and boxplot R packages. Furthermore, immune cell infiltration patterns were compared with gene modules identified through WGCNA to elucidate immune-related functional associations.

Construction of MiRNA and transcription factor (TF) regulatory networks

To explore the regulatory mechanisms of the identified biomarkers, miRNA-biomarker interactions were predicted using the TarBase v9.0 database[19], while TF-gene interactions were retrieved from the JASPAR database[20]. The resulting regulatory networks were constructed using NetworkAnalyst (https://www.networkanalyst.ca/)[21] and visualized with Cytoscape v3.10.1.

Validation and clinical relevance analysis of biomarkers

The Nephroseq v5 online database (http://v5.nephrasq.org) was employed to investigate the clinical significance of the identified biomarkers, analyzing their correlations with key clinical parameters in glomerular tissue samples from DN patients.

Identification of DN molecular subtypes and GSVA analysis

Consensus clustering was performed using the ConsensusClusterPlus R package to classify DN samples into distinct molecular subtypes based on biomarker expression profiles[22]. The optimal number of clusters was determined through cumulative distribution function (CDF) curves. Differences in immune infiltration and pathway enrichment among the molecular subtypes were further examined using GSVA, with a significance threshold of $p < 0.05$.

Drug–gene interaction prediction and molecular Docking

Potential therapeutic agents targeting the identified biomarkers were predicted using the Drug Signatures Database (DSigDB, http://dsigdb.tanlab.org/DSigDBv1.0/)[23] via the Enrichr platform. Drug structures were obtained from the PubChem database (https://pubchem.ncbi.nlm.nih.gov), while protein structures of the candidate genes were retrieved from the Protein Data Bank (PDB, http://www.rcsb.org/). Molecular docking was performed using AutoDock V1.5.7[24] to evaluate drug–target interactions, and docking results were visualized using PyMOL.

Animal experiments

Six BKS-db mice were obtained from Beijing GemPharmatech Co., Ltd. and housed under standardized conditions (temperature: 20–26 °C, humidity: 40–70%, 12-h light/dark cycle). After 12 weeks of feeding, the animals were divided into two groups: an experimental group receiving intraperitoneal injections of streptozotocin (STZ) and a control group consisting of wild-type BKS-db mice injected with an equivalent volume of citrate buffer. Blood glucose levels were measured via tail vein sampling 5–7 days post-injection. Mice were considered diabetic if blood glucose levels in the experimental group were approximately threefold higher than in controls, and a statistically significant difference was observed ($p < 0.05$, t-test). Subsequent analyses were performed on diabetic mice.

All animal experiments were conducted in accordance with ethical guidelines and were approved by the Animal Management and Use Committee of Beijing Xiehe Jianhao Pharmaceutical Technology Development Co., Ltd. (approval number: TS24010-YX).

RNA extraction and real-time polymerase chain reaction (PCR)

Total RNA was extracted from kidney tissues using TRIzol reagent (Solarbio, Beijing, China). Complementary DNA (cDNA) synthesis was performed using a RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, Waltham, MA, USA). Quantification of four key genes was conducted using real-time polymerase chain reaction (PCR) with a SYBR Green reagent kit (Accurate Biology, Changsha, China) on a CFX PCR system (Bio-Rad, CA, USA). Primers were designed using Primer Premier 5.0 software (PREMIER Biosoft International, Palo Alto, CA, USA).

Statistical methods

All statistical analyses were performed using R software (version 4.1.2). The Wilcoxon rank-sum test was employed to compare biomarker expression levels between groups. Diagnostic performance was assessed using ROC curve analysis, with an area under the curve (AUC) approaching 1, indicating high sensitivity and specificity. Spearman correlation analysis was used to evaluate the associations between biomarkers and immune cell infiltration levels. A p-value < 0.05 was considered statistically significant.

Results

Integration and batch effect removal of DN datasets

The GSE30528 and GSE104948 datasets were retrieved from the GEO database and underwent rigorous batch correction and normalization. Supplementary Fig. 1 illustrates the data before and after normalization. The integrated dataset, comprising 16 DN samples and 31 control samples, was designated as the merged cohort. Principal Component Analysis (PCA) was conducted to visualize the inter-dataset variations and confirm the successful removal of batch effects (Fig. 2). To validate our findings, the external validation dataset GSE96804, consisting of 41 DN samples and 20 control samples, was also employed (Table 1).

Identification of differentially expressed genes (DEGs)

Differential expression analysis was conducted using the DESeq2 package in R, identifying a total of 79 upregulated genes and 154 downregulated genes as differentially expressed (DEGs). The thresholds for significance were set at $p < 0.01$ and $|\log FC| > 1$. The top 30 DEGs from each category were visualized in a heatmap (Fig. 3A), while the overall distribution of DEGs was represented in a volcano plot (Fig. 3B).

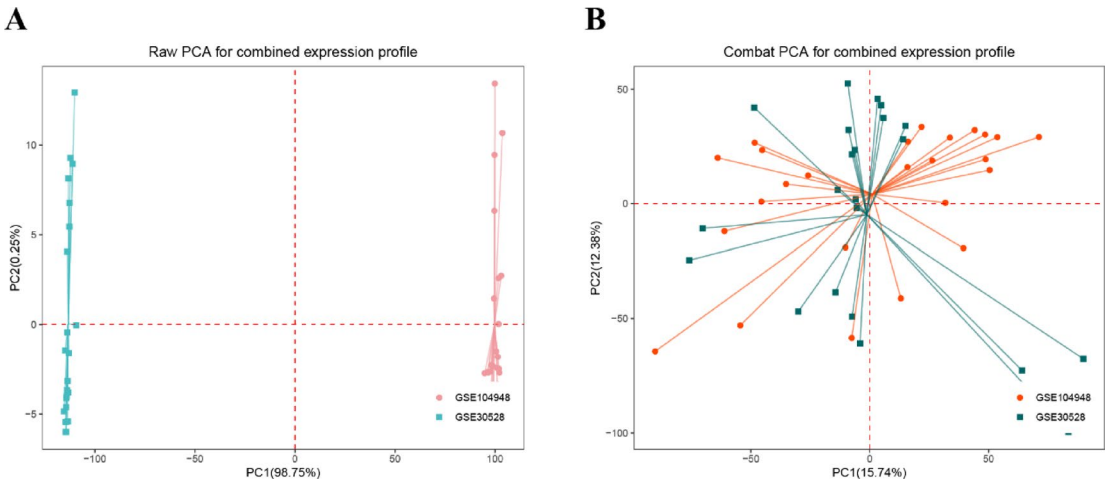

Fig. 2. Integration of datasets and removal of batch effects. (**A**) Principal component analysis (PCA) plot of datasets GSE30528 and GSE104948 before batch effect removal. (**B**) PCA plot of datasets GSE30528 and GSE104948 after batch effect removal.

Dataset	Sample size (Disease: Control)	Disease	Species	Sample type
GSE30528	9:13	DN	Human	Glomeruli
GSE104948 (GPL22945)	7:18	DN	Human	Glomeruli
GSE96804	41:20	DN	Human	Glomeruli
GSE104948(GPL22945,GPL24120)	12: 27: 21: 15: 32: 22	DN: IgA: MGN: HT: SLE: FSGS	Human	Glomeruli

Table 1. Summary of the datasets obtained.

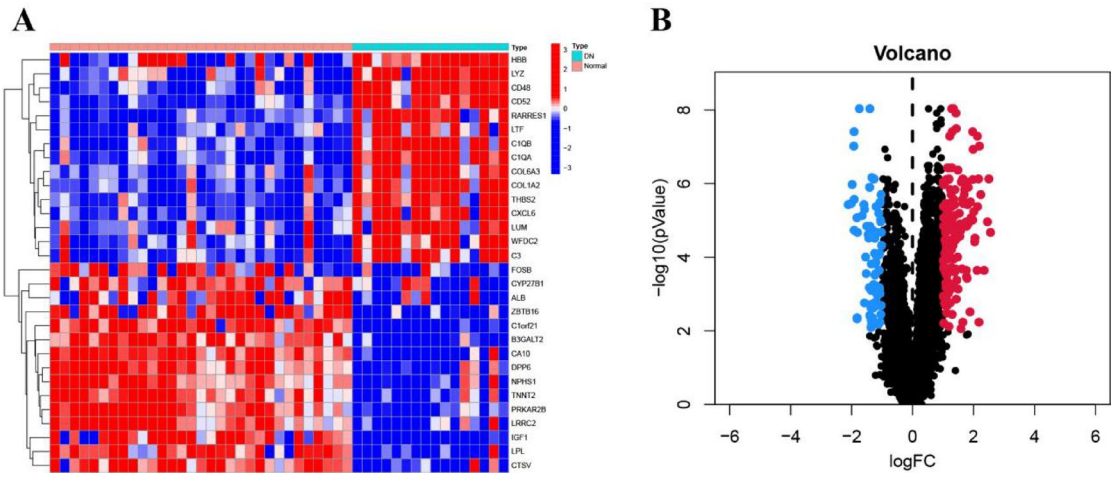

Fig. 3. Identification of differential expressed genes. (**A**) Heatmap of differentially expressed genes (DEGs) between DN and the control group. (**B**) Volcano plot of DEGs between DN and the control group.

WGCNA analysis

WGCNA was performed to explore the gene co-expression network and identify key modules associated with the DN phenotype. Using a dissimilarity measure (1-TOM), we clustered all expressed genes (Fig. 4A). The optimal soft-threshold power of 7 was determined, achieving a scale-free topology fit index greater than 0.85 for all networks, which was subsequently used for network construction (Fig. 4B). Gene modules were then identified, yielding a total of 14 modules (Fig. 4C). Among these, the turquoise and blue modules exhibited the strongest correlation with the DN phenotype (Fig. 4D,E). Genes from these two modules were intersected with DEGs to pinpoint DN-related genes, and the results were presented in a Venn diagram (Fig. 4F).

Functional enrichment analysis of DN-related genes

To investigate the biological functions underlying the DN-related genes, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. The GO analysis revealed that the DN-related genes are primarily involved in the regulation of cytokine and tumor necrosis factor signaling in biological processes (BP) (Fig. 5A). In cellular components (CC), these genes were predominantly enriched in pathways related to collagen fibers and collagen trimers, while in molecular function (MF), they were primarily associated with extracellular matrix activity and cytokine binding. KEGG pathway analysis further demonstrated that DN-related genes are significantly involved in cytokine-receptor interactions, phagosome formation, and chemokine signaling pathways (Fig. 5B). Gene Set Enrichment Analysis (GSEA) also revealed that these genes are strongly associated with renal aging, tumorigenesis, neutrophil responses, immune system activation, and transplant kidney rejection processes (Fig. 5C).

Identification of DN biomarkers through machine learning

To identify potential biomarkers for DN, we applied three distinct machine learning approaches. Using the Least Absolute Shrinkage and Selection Operator (LASSO) regression algorithm, 8 genes were selected as candidate biomarkers (Fig. 6A,B). Additionally, the Support Vector Machine-Recursive Feature Elimination (SVM-RFE) algorithm identified 23 candidate genes (Fig. 6C). The Random Forest (RF) algorithm highlighted the top 30 diagnostic genes (Fig. 6D,E).

To evaluate the performance of the three machine learning models, we employed receiver operating characteristic (ROC) curve analysis on the test cohort. The area under the curve (AUC) for each model was as follows: LASSO: 0.972 (Fig. 7A), SVM-RFE: 1.0 (Fig. 7B), and RF: 0.944 (Fig. 7C). These results highlight the strong predictive capacity of the models in identifying genes critical to the pathogenesis of diabetic nephropathy (DN). Ultimately, five genes—FN1, C1orf21, CD36, CD48, and SRPX2—overlapped across all three methods and were considered robust diagnostic biomarkers (Fig. 7D). To further validate the generalizability and robustness

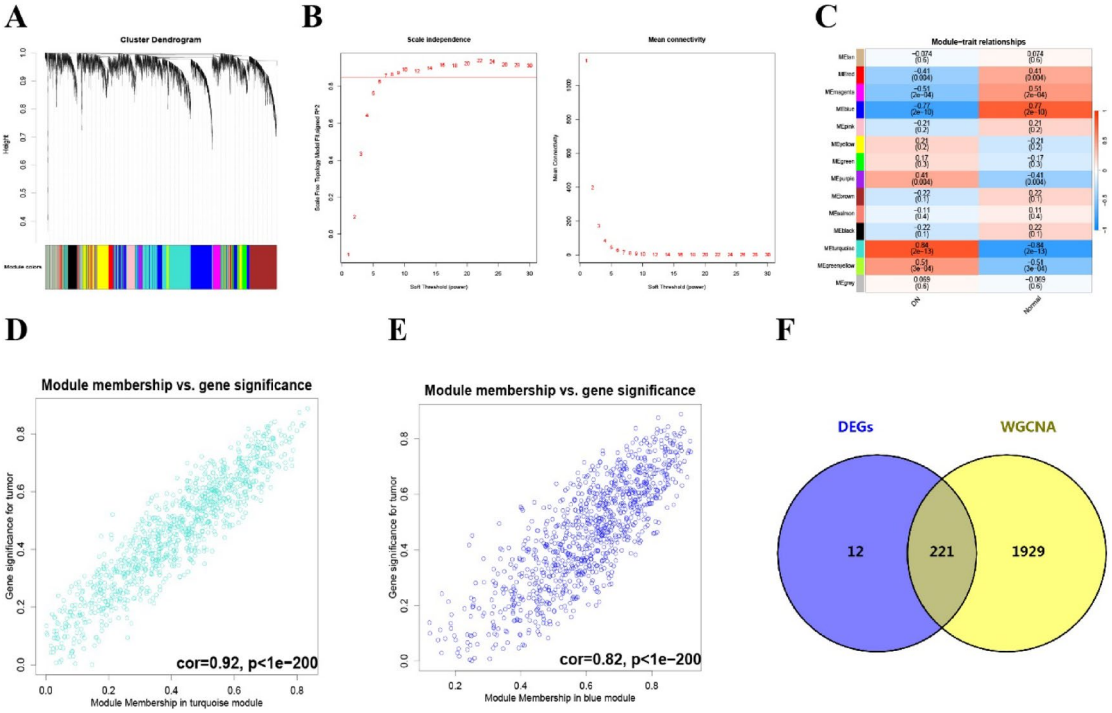

Fig. 4. Construction of the weighted gene co-expression networks. (**A**) Dendrogram of clustering all differentially expressed genes based on dissimilarity measure (1-TOM). (**B**) Network topology analysis for different soft-thresholding powers. (**C**) Correlation between different modules and clinical traits. (**D**) Significance of DN-related genes in the turquoise module. (**E**) Significance of DN-related genes in the blue module. (**F**) Venn diagram showing the overlap of DN-related genes between DEGs and WGCNA results.

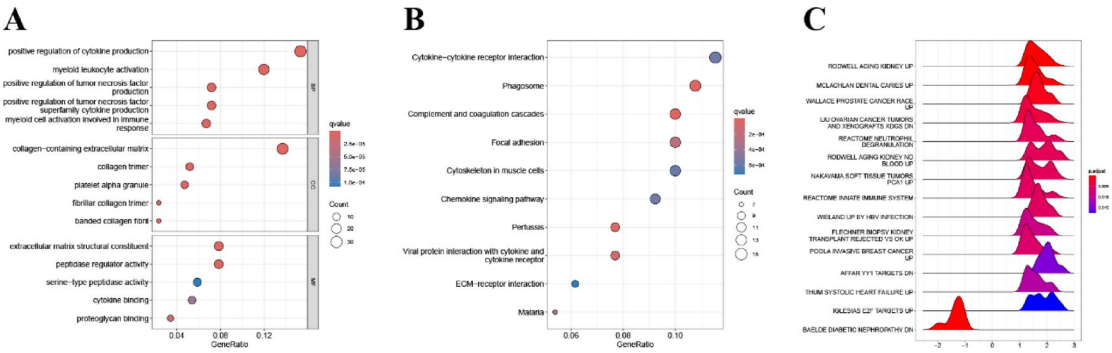

Fig. 5. Functional enrichment analysis of DN-related genes. (**A**) The top 5 Gene Ontology (GO) terms identified across three categories: molecular function (MF), cellular component (CC), and biological process (BP). (**B**) The top 10 Kyoto Encyclopedia of Genes and Genomes (KEGG) terms in enrichment analysis of DN-related genes. (**C**) The top 15 Gene set enrichment analysis (GSEA) terms of DN-related genes.

of our machine learning models, we employed an independent dataset (GSE96804) obtained from the Gene Expression Omnibus (GEO) database (Supplementary Fig. 2).

Diagnostic significance and specificity of key DN biomarkers

To assess the diagnostic utility of the five candidate biomarkers in diabetic nephropathy (DN), we conducted receiver operating characteristic (ROC) curve analysis. An AUC value greater than 0.7 was considered indicative of robust diagnostic potential. The results revealed that the area under the ROC curve (AUC) for all five biomarkers exceeded 0.9 in the merged cohort (Fig. 8A), indicating high diagnostic accuracy. Expression analyses revealed significant differences between DN and normal controls. Specifically, FN1, CD36, CD48, and SRPX2 exhibited elevated expression in the glomerular tissues of DN patients ($p < 0.001$), while C1orf21 expression was significantly reduced ($p < 0.001$) in the merged cohort (Fig. 8B).

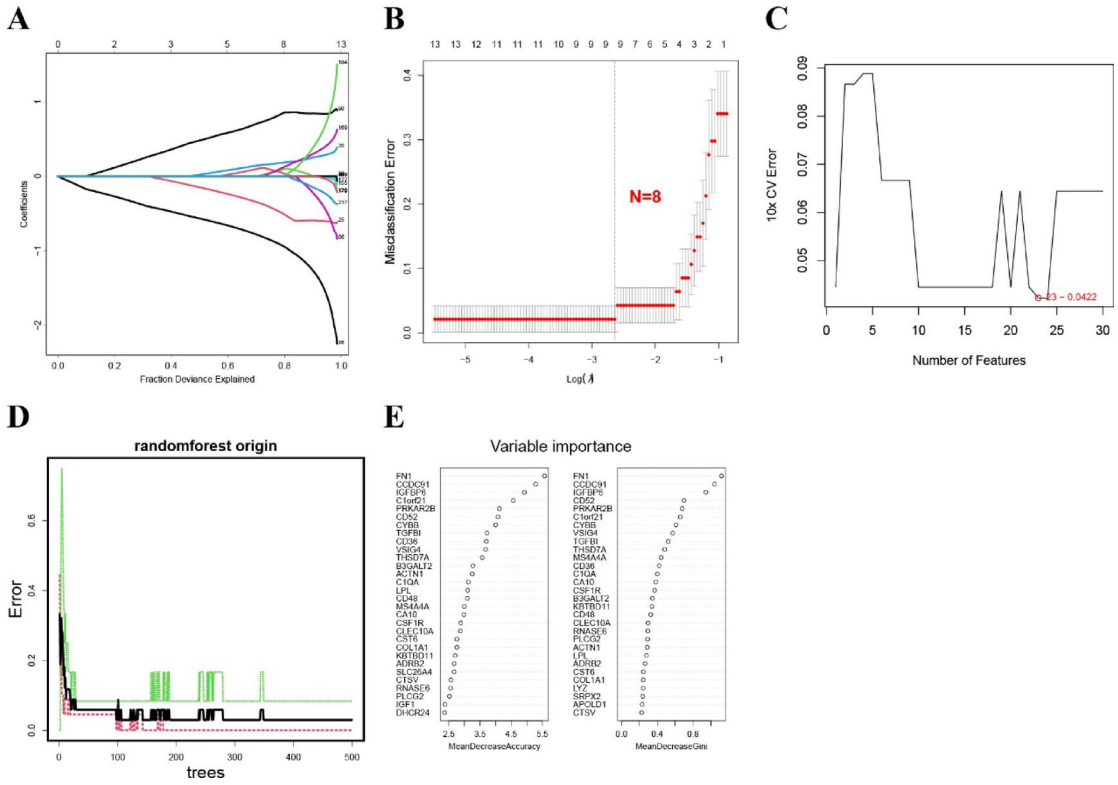

Fig. 6. Screening of DN biomarkers using machine learning methods. (**A**) LASSO regression analysis, resulting in eight feature genes. (**B**) Selection of the λ Parameter in 10-Fold Cross-Validation. (**C**) Support vector machine recursive feature elimination (SVM-RFE) algorithm, resulting in 23 genes. (**D**) Evaluation of Random Forest (RF) Model Performance. (**E**) Random forest (RF) algorithm, resulting in the top 30 feature genes.

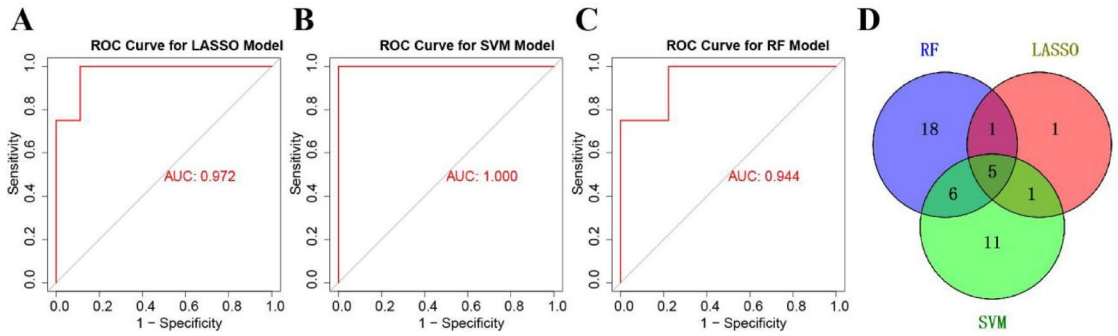

Fig. 7. Verification the effectiveness of three machine learning methods. (**A**) ROC curve of LASSO model in the test cohort. (**B**) ROC curve of SVM model in the test cohort. (**C**) ROC curve of RF model in the test cohort. (**D**) Venn diagram showing the overlap of DN biomarkers identified by the three machine learning methods.

These biomarkers demonstrated comparable diagnostic performance in the validation cohort (Fig. 8C). The differential expression of these genes was corroborated in an external validation cohort, showing consistent gene expression patterns (Fig. 8D).

To examine the specificity of these biomarkers, we compared their expression profiles in DN and five other kidney diseases (focal segmental glomerulosclerosis [FSGS], hypertensive nephropathy [HT], IgA nephropathy [IgA], membranous nephropathy [MGN], and lupus nephritis). The analysis revealed that FN1, CD36, CD48, and SRPX2 were markedly upregulated in DN tissues compared to other kidney disease types, whereas C1orf21 exhibited the opposite trend, highlighting its diagnostic specificity for DN (Fig. 8E).

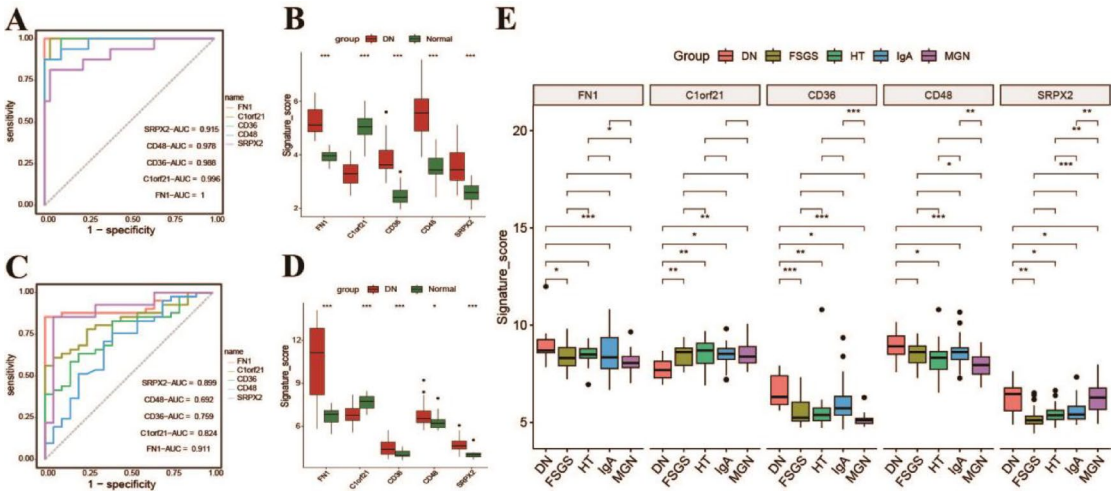

Fig. 8. ROC curve and expression level validation. (**A**, **B**) ROC curve and box plot of the expression levels of the five biomarkers between DN and the control group. (**C**, **D**) ROC curve and box plot of the expression levels of the five biomarkers between DN and the control group in an external dataset. (E) Boxplot of expression levels of five biomarkers between DN and other kidney diseases. *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$.

Model construction and validation

To further validate the five biomarkers in DN, we constructed a logistic regression model based on their expression profiles in the training cohort. ROC curve analysis of the model in the test cohort revealed an AUC of 0.767 (Fig. 9A). The model's performance was further validated via 10-fold cross-validation, where it exhibited strong predictive power, with an AUC of 0.875 in the test cohort and perfect accuracy (AUC = 1) in the training cohort (Fig. 9B,C). Additionally, confusion matrix analysis was performed in both cohorts (Fig. 9D–F), and accuracy, recall, precision, and F1-score metrics are summarized in Table 2. These results underscore the potential of the identified biomarkers in DN pathogenesis and diagnosis.

Immune infiltration and TME analysis

We next investigated immune cell infiltration patterns in DN, identifying a strong association between several immune cell types—regulatory T cells, memory B cells, macrophages, mast cells, myeloid-derived suppressor cells, and activated CD8 T cells—within DN samples (Fig. 10A). Notably, DN samples exhibited significantly higher levels of activated CD8 T cells, central memory CD8 T cells, and other immune cell populations compared to controls (Fig. 10B).

Subsequent analysis of the correlation between the five biomarkers and immune cell infiltration levels revealed that FN1, CD36, CD48, and SRPX2 were significantly positively correlated with the infiltration of most immune cells, while C1orf21 exhibited negative correlations (Fig. 10C). These findings suggest that these biomarkers may modulate immune cell dynamics in DN progression.

Furthermore, we assessed the association between these biomarkers and TME-related scores. DN samples showed significantly higher estimates in immune, stromal, and overall TME assessments compared to controls (Fig. 10D–F). Strong positive correlations were observed between FN1, CD36, CD48, and SRPX2 with all three TME scores, while C1orf21 showed significant negative correlations (Fig. 10G). WGCNA analysis revealed 14 distinct gene modules, with the turquoise and blue modules demonstrating strong correlations with immune cell infiltration levels (Supplementary Fig. 3).

Construction of biomarker regulatory network

To explore the transcriptional regulation of the five key biomarkers, we predicted their interactions with miRNAs and transcription factors (TFs). The most critical miRNAs identified were hsa-mir-145-5p, hsa-mir-661, hsa-let-7 g-5p, and hsa-mir-26b-5p, while the most significant TFs included FOXL1, FOXC1, and NFKB1 (Fig. 11A,B). Enrichment analysis of the TF regulatory network revealed associations with inflammatory pathways, Toll-like receptor signaling, and DNA binding (Table 3). In contrast, the miRNA regulatory network was predominantly linked to lipid metabolism, extracellular matrix remodeling, and cytokine binding (Table 4).

Validation and clinical relevance analysis of biomarkers in the Nephroseq database

To further validate the clinical relevance of the identified biomarkers, we explored their expression profiles in DN glomerular tissue samples using the Nephroseq database. Consistent with previous findings, FN1, CD36, CD48, and SRPX2 were significantly upregulated in DN, while C1orf21 expression was notably reduced (Fig. 12A). The ROC curve analysis yielded an AUC exceeding 0.9 for these five biomarkers, confirming their high diagnostic potential in the Nephroseq dataset (Fig. 12B).

Additionally, we observed that FN1, CD36, CD48, and SRPX2 were significantly negatively correlated with glomerular filtration rate (GFR), implying their potential involvement in glomerular damage in diabetic

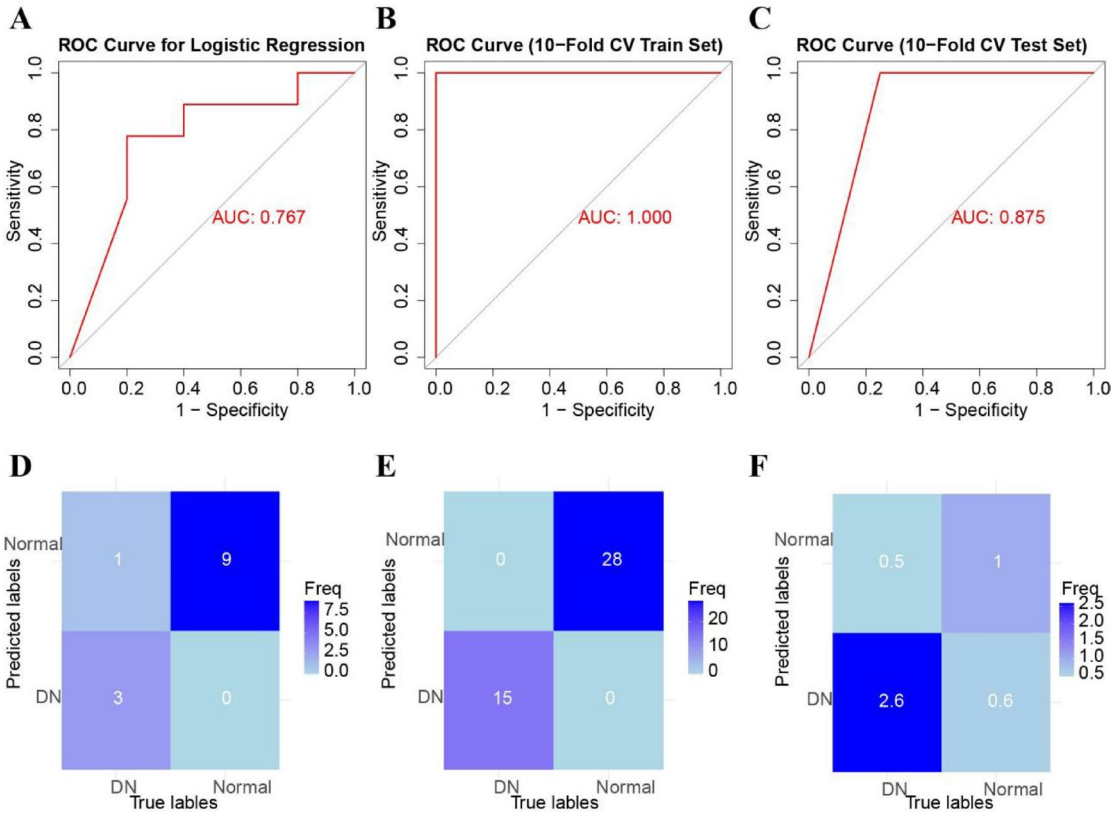

Fig. 9. The ROC curve and confusion matrix of the logistic model and the 10-fold cross validation according to the five key genes. (**A**) ROC curve of the logistic model in the test cohort for identifying DN from the normal tissues. (**B**) ROC curve of the 10-fold cross validation in the train cohort for identifying DN from the normal tissues. (**C**) ROC curve of the 10-fold cross validation in the test cohort for identifying DN from the normal tissues. (**D**) Confusion matrix of the logistic model in the test cohort. (**E**) Confusion matrix of the10-fold cross validation in the train cohort. (**F**) Confusion matrix of the10-fold cross validation in the test cohort.

	Logistic regression	k-fold cross-validation	
Index	Test cohort	Training cohort	Testing cohort
Accuracy	0.923	1	0.787
Precision	0.75	1	0.839
Recall	1	1	0.813
F1-score	0.857	1	0.825

Table 2. The confusion matrix index of logistic regression and k-fold cross-validation.

nephropathy (DN). Conversely, C1orf21 exhibited a significant positive correlation with GFR (Fig. 12C). FN1 and CD36 were also significantly positively correlated with serum creatinine levels (Fig. 12D), further suggesting their potential role in predicting poor prognosis in DN patients.

Determination of DN subtypes based on biomarkers

To elucidate the molecular subtypes associated with glomerular injury in diabetic nephropathy (DN), we classified DN samples into distinct molecular subtypes based on the expression profiles of five key biomarkers. Using an unsupervised clustering approach, we applied the ConsensusCluster R package to determine the optimal number of clusters. The consensus clustering matrix (Fig. 13A) and cumulative distribution function (CDF) plot (Fig. 13B) identified two clusters as the most stable configuration, with $k=2$ as the optimal number of clusters. The CDF curve demonstrated a continuous decline from $k=2$ to $k=9$, corroborating the previous results (Fig. 13C). As a result, DN samples were classified into two distinct molecular subtypes: CS1 and CS2.

Immune correlation analysis of DN subtypes

To better understand the molecular divergence between the two subtypes, we first assessed the differential expression of key biomarkers. Notably, expression levels of FN1, CD36, C1orf21, and SRPX2 were significantly

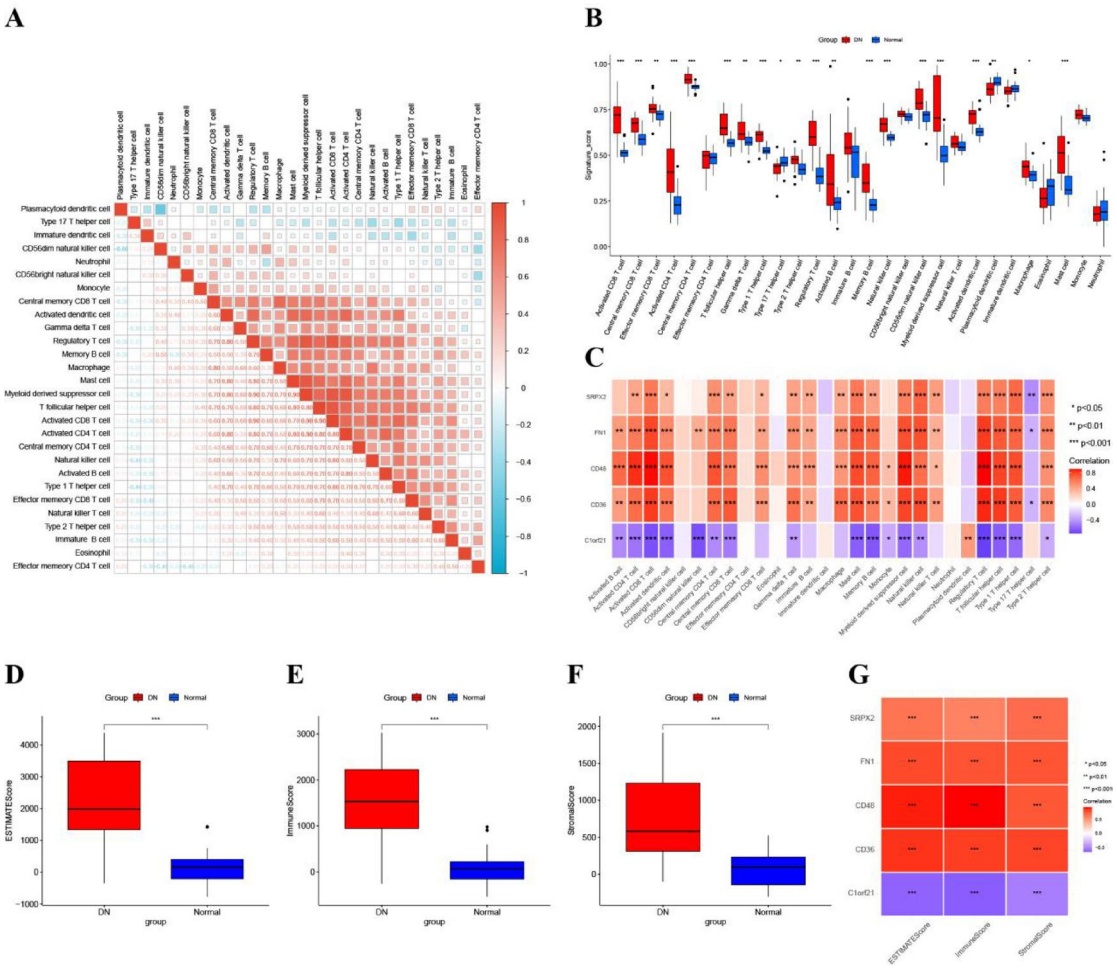

Fig. 10. Immune analysis. (**A**) Heatmap of correlations between various immune cells in the DN group. (**B**) Boxplot showing differences in immune cell infiltration levels between the DN and control groups. (**C**) Heatmap of correlations between the five DN biomarkers and immune cell infiltration levels in DN. (**D**) Differences in ESTIMATE scores between DN and the control group. (**E**) Differences in IMMUNE scores between DN and the control group. (**F**) Differences in STROMAL scores between DN and the control group. (**G**) Heatmap of correlations between the five biomarkers and the three immune-related ESTIMATE scores. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

downregulated in the CS1 subtype compared to the CS2 subtype (Fig. 14A). Immune infiltration analysis revealed significant alterations in immune cell populations within CS1. Specifically, the levels of effector memory CD4 + T cells, Th2 cells, natural killer (NK) cells, and natural killer T (NKT) cells were markedly reduced in the CS1 subtype, while Th17 cells and CD56-positive NK cells exhibited significant upregulation (Fig. 14B).

Subsequent GSVA functional enrichment analysis highlighted key molecular pathways that distinguished the two subtypes. In particular, fatty acid metabolism, bile acid metabolism, oxidative phosphorylation, and reactive oxygen species (ROS) pathways were significantly upregulated in CS1, suggesting that this subtype may be more closely associated with the pathophysiology of diabetic nephropathy (Fig. 14C).

Prediction of candidate drugs

We leveraged the Drug Signatures Database (DSigDB) to predict potential therapeutic agents for the identified biomarkers. Eight compounds were identified with an adjusted p-value < 0.05 (Table 5). Among these, acetovanillone, GW9662, electrocorundum, titanium dioxide, sodium dichromate, cube root extract, and alitretinoin were found to have significant associations with FN1 and CD36, while vitinoin showed interactions with multiple genes, suggesting its potential as a broad-spectrum therapeutic candidate.

Molecular docking

To investigate the binding affinity between the predicted therapeutic agents and their respective target proteins, we performed molecular docking using AutoDock V1.5.7 software. This analysis focused on evaluating the interactions between candidate drugs and the binding sites of protein targets encoded by the identified biomarkers. The docking results revealed that each drug interacted with its protein target via hydrogen bonds

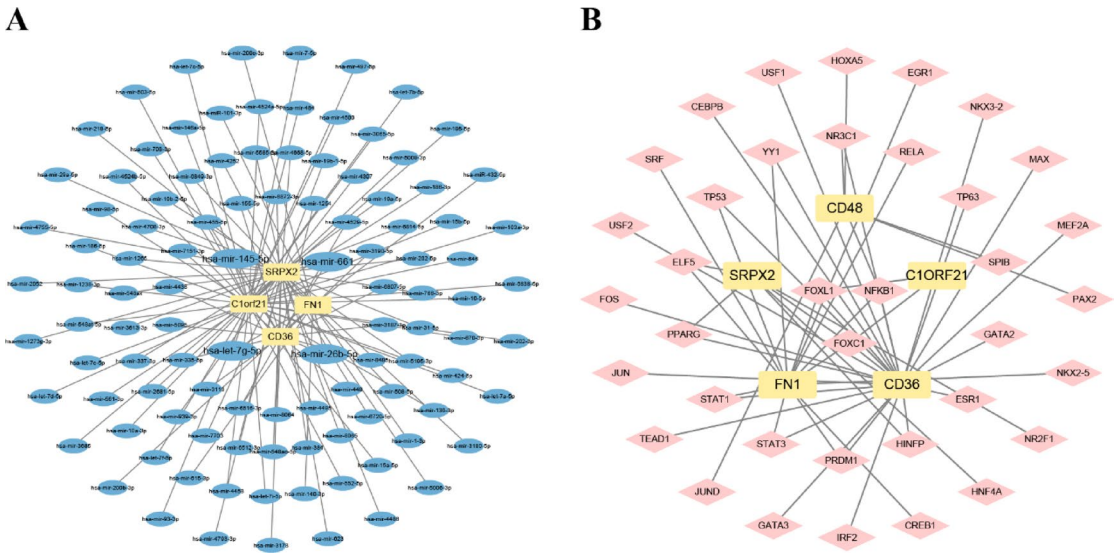

Fig. 11. Transcriptional regulatory network of DN biomarkers. (**A**) Regulatory network between DN biomarkers and microRNAs (miRNA). (**B**) Regulatory network between DN biomarkers and transcription factors (TF).

Project	Pathway	*P*.Value
KEGG	ECM-receptor interaction	1.11E-04
	Fat digestion and absorption	1.06E-02
	Malaria	1.26E-02
	Adipocytokine signaling pathway	1.78E-02
	PPAR signaling pathway	1.90E-02
GO: BP	Regulation of anatomical structure morphogenesis	1.18E-04
	Cell_matrix adhesion	3.66E-04
	Cell_substrate adhesion	8.41E-04
	Exocytosis	1.38E-03
	Regulation of developmental process	2.27E-03
GO: CC	Secretory granule	8.10E-04
	Cytoplasmic vesicle part	2.26E-03
	Cytoplasmic membrane_bounded vesicle	1.08E-02
	Membrane_bounded vesicle	1.25E-02
	Cytoplasmic vesicle	1.27E-02
GO: MF	Low_density lipoprotein particle binding	3.10E-03
	Pattern recognition receptor activity	3.52E-03
	Lipoprotein particle binding	5.38E-03
	Collagen binding	9.71E-03
	Cytokine binding	1.36E-02

Table 3. Enrichment analysis of the MiRNA regulatory network.

and strong electrostatic interactions. Notably, the binding energy between the CD36 gene and GW9662 was the lowest (-8.36 kcal/mol), indicating a particularly stable interaction (Fig. 15).

Validation of biomarkers in mice

To further validate the expression of the five biomarkers, we utilized 12-week-old BKS-db mice as an experimental model for diabetic nephropathy. Quantitative PCR (qPCR) was employed to measure the mRNA expression levels of FN1, CD36, CD48, and SRPX2. Consistent with the bioinformatics analysis, the results demonstrated significantly elevated expression levels of FN1, CD36, CD48, and SRPX2 in the DN mouse models compared to the normal controls (Fig. 16). However, due to the absence of the C1orf21 gene in mice, we were unable to confirm its expression in this model. To further corroborate the role of C1orf21, we performed additional validation using two external datasets, confirming its expression patterns in DN along with the other four biomarkers (Supplementary Fig. 4). All specific primers used in the study are listed in Table 6.

Project	Pathway	*P*.Value
KEGG	Prolactin signaling pathway	1.54E-09
	Osteoclast differentiation	3.85E-09
	AGE-RAGE signaling pathway in diabetic complications	1.94E-08
	Hepatitis B	2.61E-08
	Th17 cell differentiation	3.11E-08
GO: BP	MyD88: Mal cascade initiated on plasma membrane	6.23E-09
	Toll Like Receptor TLR1:TLR2 Cascade	6.23E-09
	Toll Like Receptor TLR6:TLR2 Cascade	6.23E-09
	Toll Like Receptor 2 (TLR2) Cascade	6.23E-09
	Activated TLR4 signalling	2.76E-08
GO: CC	Nucleoplasm	7.24E-17
	Nuclear lumen	9.98E-15
	Nuclear chromatin	1.04E-13
	Organelle lumen	4.45E-13
	Membrane_enclosed lumen	7.14E-13
GO: MF	Sequence_specific DNA binding	3.90E-34
	Transcription from RNA polymerase II promoter	1.76E-28
	DNA binding	8.33E-23
	Transcription factor binding	7.79E-19
	Positive regulation of transcription, DNA_dependent	6.25E-17

Table 4. Enrichment analysis of the TF regulatory network.

Discussion

Diabetic nephropathy (DN) is one of the most prevalent and severe complications of diabetes mellitus (DM), contributing significantly to increased morbidity and mortality among diabetic patients[25]. In China, an estimated 92.4 million adults—approximately 9.7% of the adult population—have diabetes, with 60.7% remaining undiagnosed or untreated[26]. Approximately one-third of diabetic individuals eventually develop DN after a prolonged latency period. However, whether to routinely screen for microalbuminuria or conduct predictive assessments for diabetic nephropathy, as part of an individualized treatment strategy, to allocate resources for intensified interventions remains a matter of ongoing debate[27]. While previous research on DN has predominantly focused on renal tubular interstitial damage and its underlying pathogenesis, our study offers novel insights into glomerular lesions and potential therapeutic strategies through advanced bioinformatics approaches. Despite considerable efforts to identify new therapeutic targets for DN, the available knowledge remains insufficient, underscoring the urgent need for highly specific and sensitive biomarkers, as well as targeted pharmacological interventions.

In this study, we obtained gene expression profiles from two DN-associated glomerular tissue datasets, comprising 16 DN samples and 31 normal controls. Differential expression analysis identified 233 differentially expressed genes (DEGs), which were further analyzed using weighted gene co-expression network analysis (WGCNA) to identify key non-conserved modules. The turquoise and blue modules exhibited the strongest correlation with the DN phenotype, and the intersecting genes from these modules were further examined. Gene Ontology (GO) enrichment analysis revealed that these genes were primarily associated with cytokine regulation, tumor necrosis factor signaling, and extracellular matrix-related processes. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis indicated that the genes were enriched in cytokine-receptor interactions, phagocytosis, and chemokine signaling pathways, while Gene Set Enrichment Analysis (GSEA) demonstrated their involvement in kidney aging, cancer-related pathways, and immune system regulation.

Through integrative machine learning approaches, including least absolute shrinkage and selection operator (LASSO) regression, support vector machine recursive feature elimination (SVM-RFE), and random forest (RF) analysis, we identified five key hub genes—FN1, CD36, CD48, SRPX2, and C1orf21—and validated their expression using independent external datasets. Receiver operating characteristic (ROC) curve analysis demonstrated that all five genes exhibited excellent diagnostic potential. Furthermore, in the GSE104948 dataset, we assessed the differential expression of these biomarkers across DN and five other kidney diseases. Notably, C1orf21 was significantly downregulated in DN compared to other renal diseases, whereas the remaining four genes exhibited high expression in DN. These findings highlight the specificity of these biomarkers in DN and their potential utility as diagnostic indicators. Finally, experimental validation via quantitative PCR (qPCR) confirmed the upregulation of FN1, CD36, CD48, and SRPX2 in DN mouse models, further corroborating our bioinformatics findings.

FN1 (fibronectin 1) is a major component of the extracellular matrix (ECM) and plays a crucial role in cell adhesion, migration, differentiation, host defense, and metastasis. FN1 promotes the secretion of M2 macrophage markers (IL-10) and reduces the production of the M1 marker TNF-α in a concentration-dependent manner[28]. FN1 is an important component of the ECM, and studies have shown that mutations in the FN1 gene may lead to thinning of the glomerular basement membrane and thin basement membrane nephropathy. The α5β1 integrin

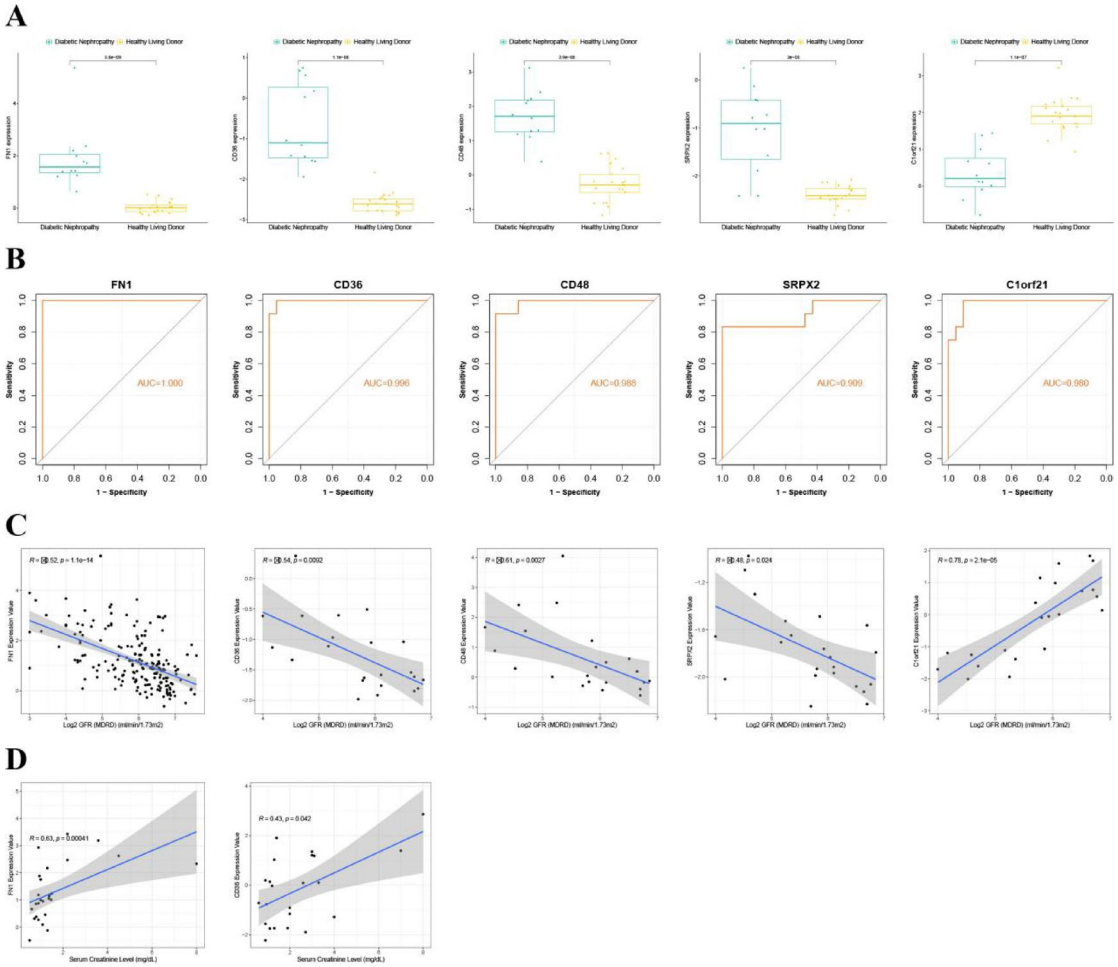

Fig. 12. Validation and clinical relevance analysis of DN biomarkers in the nephroseq database. (**A**) Validation of the expression characteristics of DN biomarkers. (**B**) Validation of the ROC curve characteristics of DN biomarkers. (**C**) Scatter plot showing the correlation between the expression levels of DN biomarkers and glomerular filtration rate (GFR). (**D**) Scatter plot showing the correlation between the expression levels of DN biomarkers and serum creatinine levels.

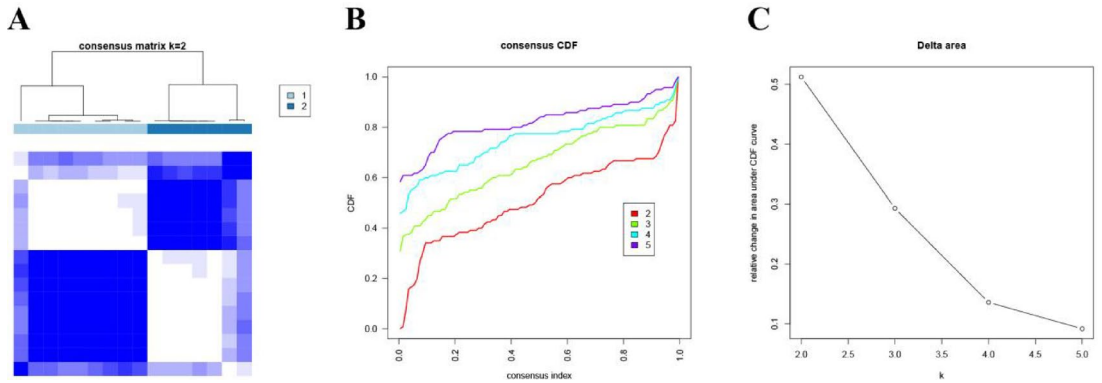

Fig. 13. Unsupervised Consensus Clustering in the DN Merged Cohort. (**A**) Heatmap of the two DN sample groups when k = 2. (**B**) Cumulative Distribution Function (CDF) from k = 2 to 9. (**C**) Delta plot showing the change in the area under the CDF curve from k = 2 to k = 9 in consensus clustering.

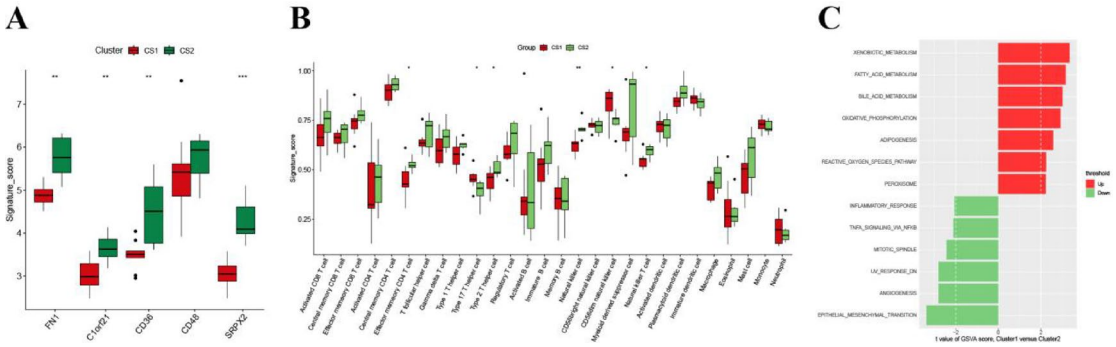

Fig. 14. Immune and differential analysis in DN subtypes. (**A**) Expression differences of biomarkers in the two obtained DN subtypes. (**B**) Boxplot showing the differences in immune cell infiltration levels between the two DN subtypes. (**C**) GSVA functional enrichment analysis between the two obtained DN subtypes. (**D**) *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Drug name	*P* Value	Adjusted *P* Value	Associated Genes
Acetovanillone CTD 00002374	4.71E-05	0.00891428	FN1; CD36
GW9662 CTD 00004071	8.22E-05	0.00891428	FN1; CD36
Electrocorundum CTD 00005364	9.10E-05	0.00891428	FN1; CD36
TITANIUM DIOXIDE CTD 00000489	5.15E-04	0.032755879	FN1; CD36
Vitinoin CTD 00007069	5.57E-04	0.032755879	SRPX2; FN1; C1ORF21
Sodium dichromate CTD 00000827	0.001054423	0.040242238	FN1; CD36
Cube root extract CTD 00006707	0.001084805	0.040242238	FN1; CD36
Alitretinoin CTD 00003402	0.001095027	0.040242238	FN1; CD36

Table 5. Predicted candidate drugs.

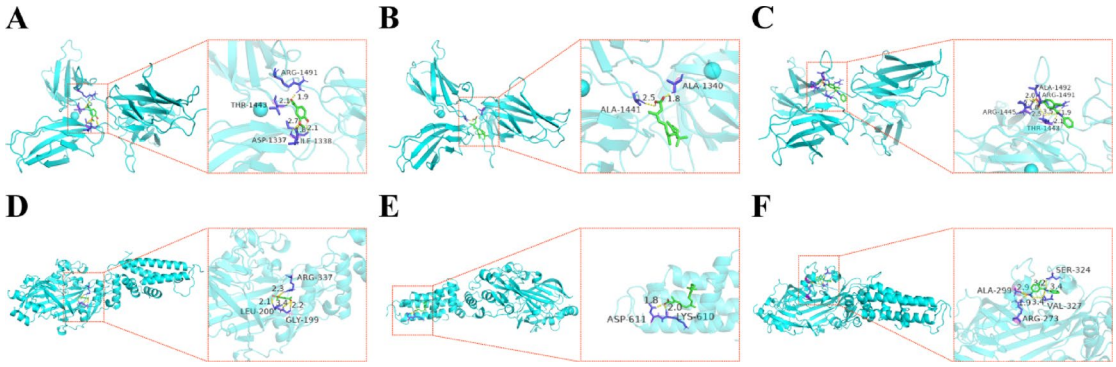

Fig. 15. Docking results of proteins and small molecules. (**A**). FN1 docking with Acetovanillone (affinity: −4.6); (**B**). FN1 docking with GW9662 (affinity: −7.57); (**C**). FN1 docking with Alitretinoin (affinity: −6.9); (**D**). CD36 docking with Acetovanillone (affinity: −5.14); (**E**). CD36 docking with GW9662 (affinity: −8.36); (**F**). CD36 docking with Alitretinoin (affinity: −7.8).

receptor binds to FN dimers, facilitating the assembly of FN into a fibrillar matrix. This interaction induces conformational changes in FN, promoting FN-FN interactions that lead to the formation of new fibrils[29]. Through continuous deposition, these fibrils mature into a stable, insoluble matrix, serving as a scaffold for the deposition of other extracellular matrix components[30]. FN1 deposition can cause glomerular lesions and sclerosis, manifesting as renal dysfunction and hematuria[31]. FN1 has been implicated in DN progression due to its role in fibrosis and glomerular dysfunction. It aligns with previously identified biomarkers like TGF-β1 and COAL41, which are also involved in extracellular matrix remodeling and fibrosis in DN[32]. Unlike TGF-β1, which is primarily a signaling molecule, FN1 directly contributes to structural changes in the kidney, making it a more specific marker for tissue damage. Previous studies primarily focused on the mechanistic role of FN1 in ECM accumulation[33], our study extends its significance by identifying FN1 as a predictive biomarker for early-stage DN, FN1 expression was upregulated in DN samples, which negatively correlated with the GFR

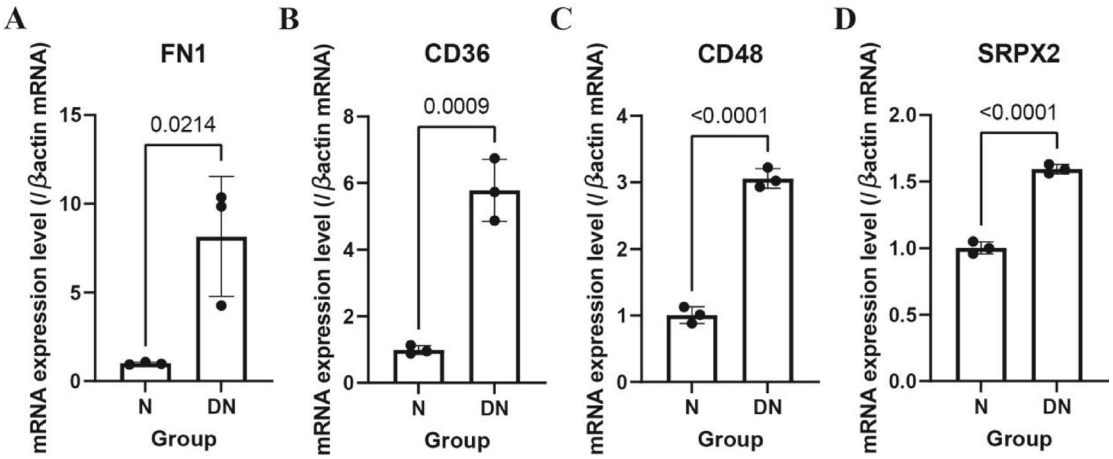

Fig. 16. Validation of the expression level of key genes using RT-qPCR. FN1, (**B**) CD36, (**C**) CD48 and (**D**) SRPX2. $p < 0.05$; unpaired Student's t-test.

Gene	Species	Primer sequence (5′ to 3′)
FN1	Mouse	Forward: ACATGGCTTTAGGCGGACAA
		Reverse: TTCGGCAGGTATGGTCTTGG
CD36	Mouse	Forward: ATGGGCTGTGATCGGAACTG
		Reverse: ACTGGCATGAGAATGCCTCC
CD48	Mouse	Forward: ACATGAGAGTGCTGCGTGAA
		Reverse: ACTAGCCAAGTTGCAGTCCA
SRPX2	Mouse	Forward: AGAGTCCCCCGATGGTGTTA
		Reverse: ACAGGAGAGCTCACAACGTG

Table 6. Primers used for real-time PCR.

and positively correlated with serum creatinine. Additionally, our analysis revealed novel regulatory networks involving FN1, suggesting that its role in DN may be more complex than previously understood and could serve as a key gene for glomerular tissue lesions in DN.

CD36 is a multifunctional receptor and a fatty acid transporter, widely expressed in various tissues and organs. It primarily mediates the uptake and translocation of long-chain fatty acids and is expressed in proximal and distal tubular epithelial cells, podocytes, and mesangial cells in the kidney. Podocytes are an important component of the glomerular filtration barrier, and one of the main pathological features of DN is podocyte damage leading to proteinuria. CD36-mediated lipid deposition and inflammation may exacerbate podocyte damage and contribute to the development of DN[34–36]. Elevated blood glucose levels can upregulate the expression of CD36, which subsequently binds to TGF-β1, thereby initiating the fibrotic process[37]. Besides, CD36 mediates lipid-induced toxicity, further aggravating fibrosis via the Wnt/β-catenin signaling pathway[38]. Additionally, CD36 serves as a signal hub in immune responses, mediating interactions between macrophages and endothelial cells[39,40]. Studies have shown that reduced CD36 expression may protect diabetic mice from oxidative stress and kidney damage[41]. CD36 is involved in lipid metabolism and inflammation, similar to biomarkers like MCP-1 and NF-κB, which are linked to inflammatory pathways in DN[42]. CD36 is unique in its role as a fatty acid transporter, contributing to lipid accumulation in renal cells, a process less emphasized in traditional DN biomarkers[43,44]. Studies primarily focused on the role of CD36 in lipid metabolism and oxidative stress, our study extends its significance by identifying CD36 as a predictive biomarker for early-stage DN and upregulation of CD36 was significantly associated with decreased GFR and increased serum creatinine. This distinction highlights the dual role of CD36—not only as a contributor to disease progression but also as a potential tool for early intervention.

CD48 encodes an immunoglobulin receptor involved in immune cell adhesion and activation. Experimental evidence has shown that CD48, a member of the signaling lymphocyte activation molecule (SLAM) family, interacts with CD244 to modulate immune responses. This interaction influences the cytotoxicity and effector functions of CD8 + T cells, playing a pivotal role in immune regulation. Similar to biomarkers like IL-6 and TNF-α, which are key players in the inflammatory response in DN. It has been reported CD48 is more specific to immune cell activation and may serve as a marker for immune-mediated kidney injury, a less explored aspect of DN pathology[45,46]. It has also been implicated in mediating acute rejection in transplanted kidneys[47]. CD48 may be implicated in humoral immunity and macrophage activation, with integrated polygenic regulation enhancing the inflammatory cascade, thereby accelerating the progression of DN[48,49]. Our findings support the role of CD48

in DN and further validate its potential as a therapeutic target for immune modulation in DN. Additionally, SRPX2 is mainly expressed in perivascular regions and promotes neovascularization[50]. SRPX2 also facilitates cell proliferation and migration.SRPX2 is involved in cell adhesion and angiogenesis, similar to biomarkers like VEGF and ANGPT2, which are associated with vascular dysfunction in DN[51]. Specifically, SRPX2 expression is upregulated in renal endothelial cells under hyperglycemic conditions, where it disrupts vascular integrity by promoting endothelial-to-mesenchymal transition (EndMT). SRPX2 is less studied in DN but may highlight its potential as a therapeutic target in DN.C1orf21 may play a role in cellular stress responses, similar to biomarkers like HSP70 or SOD2, which are involved in oxidative stress and inflammation in DN.C1orf21 is less studied compared to established oxidative stress markers, and its specific mechanistic role in DN remains unclear. But our study has shown that C1orf21 expression is downregulated in DN glomerular tissue[52] and introduces C1orf21 as a novel candidate for further investigation. This finding aligns with our results. However, the roles of CD48, SRPX2, and C1orf21 in DN, particularly in glomerular lesions, remain underexplored, and further investigation into their mechanisms could aid the development of new DN therapies.

Previous research has indicated that the pathogenesis of DN is complex and multifactorial, involving many pathways and mediators[53]. We found significant upregulation of immune, matrix, and estimated scores in DN compared to the control group, and there were also significant differences in immune cell infiltration levels between the two groups. Most immune cells, including CD8 + T cells, CD4 + T cells, killer cells, B cells, and macrophages, showed high infiltration in DN patients[54]. In type 2 diabetic patients' kidneys, significant increases in CD4+, CD8+, and CD20 + T cells have been reported, correlating with proteinuria levels[55]. CD4 + T cells produce IFN-γ and TNF-α locally in the kidneys, exacerbating the occurrence and progression of DN[56,57]. Moreover, CD4 + T cells can amplify local inflammatory responses by activating the JAK/STAT pathway or secreting cytokines such as TGF-β, IL-6, and STAT3[58,59]. Chronic inflammation in DN triggers the release of damage-associated molecular patterns (DAMPs)[60], which subsequently activate B cells and drive cytokine production. Effector B cells further modulate the onset and progression of local inflammation through the secretion of IL-6 and TNF[61,62]. Additionally, elevated glucose levels trigger the ROS-p38 MAPK pathway in macrophages, leading to the release of TNF-α and promoting podocyte apoptosis[63]. High glucose enhances the expression of glycolytic enzymes, activates NF-κB in bone marrow-derived macrophages, and regulates macrophage activation, thereby exacerbating the inflammatory response[64,65]. Infiltrating macrophages mediate kidney damage by releasing lysosomal enzymes, growth factors, and vascular endothelial growth factors, thereby promoting DN progression. Interestingly, our study showed strong positive correlations among various immune cells, especially between T cells, CD4 + T cells, CD8 + T cells, B cells, and macrophages. This aligns with previous reports, emphasizing the importance of immune-related cells in the pathogenesis of DN. Furthermore, by analyzing the correlations between the five biomarkers and immune cells, we found significant associations with CD4 + T cells, CD8 + T cells, neutrophils, dendritic cells, mast cells, memory B cells, and helper T cells, confirming the findings from earlier studies that these immune cells play an important role[66]. Identifying these immune cells' impact on renal damage in DN will help us identify reliable biomarkers and new therapeutic targets for DN patients.

Transcription factors (TFs) are major regulators of the expression of multiple target genes, generating feedback regulation in biological processes. Hypoxia-inducible factor (HIF-1) has been shown to mediate renal tubular interstitial fibrosis and tubular damage in type 1 diabetic mice[67,68]. Yin-Yang factor 1 (YY1) accelerates renal fibrosis in db/db mice by upregulating α-SMA expression and epithelial-to-mesenchymal transition (EMT)[69]. FOXC1 mediates normal development of the renal ureter, and mutations in FOXC1 result in complex kidney and ureteral phenotypes[70]. NFKB1 regulates membranous nephropathy and inflammatory progression in DN via the NF-κB pathway[71,72]. FOXL1 inhibits tumor cell proliferation in renal clear cell carcinoma[73]. miRNAs also play crucial roles in disease progression. Previous studies have shown that miR-1187 induces podocyte injury through autophagy, promoting the progression of DN[74]. miR-92a-1-5p mediates exosome transfer from proximal tubular cells, leading to mesangial cell damage in DN[75]. Our study found that hsa-miR-145-5p, hsa-miR-661, hsa-let-7 g-5p, and hsa-miR-26b-5p may also be involved in the occurrence and progression of DN.

Based on the gene expression of the five biomarkers identified above, we classified two different DN subtypes. The results showed that the five biomarkers were all expressed at low levels in the CS1 subtype. Further analysis of the relationship between biomarkers and immune infiltration revealed that most immune cells also exhibited low expression in the CS1 subtype. In contrast to CS1, the immune infiltration levels of effector memory CD4 T cells, type 2 T helper cells, and natural killer T cells were higher in the CS2 subtype, while the immune infiltration of Th17 cells and CD56dim natural killer cells was lower. GSVA enrichment analysis indicated that various metabolic pathways, such as fatty acid metabolism, exogenous substance metabolism, and oxidative phosphorylation, were significantly upregulated in CS1. The metabolic abnormalities, such as those in fatty acid metabolism, could lead to increased blood lipids, reduced insulin effectiveness, and an elevated risk of cardiovascular diseases, suggesting that the CS1 subtype may exacerbate the development of diabetic nephropathy.

In our study, we predicted the gene-drug interactions of five biomarkers and identified three high-value potential compounds for DN treatment using molecular docking. Acetovanillone, a naturally occurring inhibitor of NADPH oxidase[76], ameliorates cardiac and acute lung injury by suppressing oxidative stress and modulating inflammatory signaling pathways[77,78].Studies have demonstrated that Acetovanillone ameliorates DN by inhibiting NADPH oxidase activity, leading to reduced reactive oxygen species (ROS) production and subsequent oxidative stress[79]. This reduction in oxidative stress contributes to improved renal function in diabetic models. Additionally, Acetovanillone has been shown to suppress NLRP3 inflammasome activation, thereby attenuating renal inflammation and fibrosis associated with DN. These findings suggest that targeting oxidative stress and inflammatory pathways with apocynin may offer a promising therapeutic approach for managing DN. In parallel, GW9662—a selective PPARγ antagonist—has been shown to exert a protective effect

against drug-induced renal injury when administered in combination with losartan[80]. In contrast, another study demonstrated that GW9662 inhibited the anti-fibrotic effects of pioglitazone in renal interstitium, highlighting its potential impact on fibrotic processes[81]. Recent studies suggest that GW9662 has been shown to inhibit renal fibrosis in DN by modulating PPARγ-mediated signaling pathways[82]. Specifically, GW9662 reduces the expression of fibrotic markers such as collagen and fibronectin, and suppresses TGF-β1 signaling, thereby attenuating renal fibrosis. Moreover, alitretinoin, a vitamin A derivative currently approved for the treatment of chronic hand eczema[83], prevents fibrosis by mitigating NFκB activation and inhibiting its nuclear translocation in the glomerulus[84]. Studies have shown that Alitretinoin treatment significantly reduced the expression of inflammatory cytokines and oxidative stress markers in a DN model and Alitretinoin attenuated renal fibrosis by downregulating TGF-β1 signaling and reducing the accumulation of extracellular matrix components. While these findings underscore the potential of these agents to modulate key pathogenic pathways in DN and fibrosis[83]. Future studies should focus on in vitro and in vivo experiments to validate the biological relevance and therapeutic potential of these targets. For example, cell-based assays could be used to assess the effects of target inhibition on disease-related phenotypes, while animal models could be employed to evaluate the in vivo efficacy and safety of potential therapeutic compounds.

The biomarkers FN1, C1orf21, CD36, CD48, and SRPX2 identified in this study exhibit considerable potential for clinical translation in diabetic nephropathy (DN), offering new avenues for early diagnosis and targeted therapeutic strategies. Early diagnosis of DN remains a critical challenge, and the integration of multiple biomarkers could significantly improve diagnostic accuracy. For instance, previous studies developed a urine exosome-based multi-marker panel for early DN detection, which demonstrated superior sensitivity and specificity compared to traditional albuminuria testing[85]. Their approach leveraged the unique molecular signatures of exosomes, which reflect the pathophysiological state of renal cells, to identify early-stage DN with greater accuracy. Similarly, our study suggests that a panel combining FN1, CD36, and SRPX2 could provide a comprehensive assessment of DN risk. Furthermore, we use the logistic algorithms to analyze these biomarkers and develop predictive models for early DN detection. This approach could leverage the complementary information provided by multiple biomarkers, address the limitations of current diagnostic methods and pave the way for more effective and targeted interventions. Additionally, non-invasive methods, such as measuring CD48 levels in urine exosomes, patients might benefit from early interventions targeting immune activation and inflammation, potentially slowing disease progression[86]. Personalized treatment strategies based on biomarker profiles have the potential to revolutionize DN management. Similarly, recent research suggested that SRPX2, which is involved in angiogenesis and vascular dysfunction, could serve as a therapeutic target for anti-angiogenic treatments[87]. Furthermore, C1orf21's association with oxidative stress opens the possibility of using antioxidant therapies in patients with high C1orf21 expression[88]. To translate these biomarkers into clinical practice, large-scale, multi-center validation studies are essential. For example, studies conducted extensive validation of FN1 as a diagnostic marker, paving the way for its clinical adoption. Similarly, our findings need to be validated in diverse patient cohorts to ensure generalizability. Additionally, the development of cost-effective and high-throughput detection methods, such as ELISA or mass spectrometry-based assays, is critical for widespread implementation. Collaborations with diagnostic companies and regulatory agencies, such as the FDA or EMA, will be necessary to obtain approval for clinical use[89]. Finally, integrating these biomarkers into clinical guidelines and educating healthcare providers will facilitate their adoption in routine practice. In conclusion, the biomarkers FN1, C1orf21, CD36, CD48, and SRPX2 hold significant promise for improving the diagnosis and management of DN. By enabling early detection and facilitating personalized treatment strategies, these biomarkers could transform the clinical approach to DN, ultimately improving patient outcomes and reducing the burden on healthcare systems. Biomarker-driven approaches have the potential to revolutionize the management of complex diseases like DN. With further research and validation, these biomarkers could become integral components of DN care, paving the way for more precise and effective interventions.

In summary, we conducted a comprehensive and systematic bioinformatics analysis and identified five potential biomarkers for DN, which are closely related to clinical indicators. The clinical model based on these biomarkers demonstrated high performance in DN diagnosis. We also constructed two molecular subtypes of DN, which showed significant differences in immune cell infiltration levels and various metabolic pathways, epithelial-mesenchymal transition, and angiogenesis. Enrichment analysis revealed the functional characteristics of these genes, and drug prediction and molecular docking suggested the potential therapeutic value of these genes. However, our study has certain limitations. Our study relies on publicly available datasets, which inevitably introduce certain biases due to inherent heterogeneity in sample composition, data acquisition protocols, and experimental conditions. Although we have implemented rigorous preprocessing and batch correction methods to mitigate these effects, such variability may still limit the generalizability of our findings. Therefore, future studies should focus on performing functional validations in cellular or animal models to confirm our findings and elucidate the underlying biological mechanisms. Furthermore, it will be crucial to verify the robustness and clinical applicability of these biomarkers in larger, independent cohorts to establish their relevance in clinical practice.

Data availability

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.

Received: 18 December 2024; Accepted: 7 May 2025
Published online: 15 May 2025

References

1. Tanase, D. M. et al. Role of gut microbiota on onset and progression of microvascular complications of type 2 diabetes (T2DM). *Nutrients* **12**, 3719. https://doi.org/10.3390/nu12123719 (2020).
2. Zhang, L. et al. Trends in chronic kidney disease in China. *N. Engl. J. Med.* **375**, 905–906. https://doi.org/10.1056/NEJMc1602469 (2016).
3. Mafi, A. et al. Metabolic and genetic response to probiotics supplementation in patients with diabetic nephropathy: A randomized, double-blind, placebo-controlled trial. *Food Funct.* **9**, 4763–4770. https://doi.org/10.1039/c8fo00888d (2018).
4. Samsu, N. Diabetic nephropathy: Challenges in pathogenesis, diagnosis, and treatment. *Biomed. Res. Int.* 1497449. https://doi.org/10.1155/2021/1497449 (2021).
5. Colhoun, H. M. & Marcovecchio, M. L. Biomarkers of diabetic kidney disease. *Diabetologia* **61**, 996–1011. https://doi.org/10.1007/s00125-018-4567-5 (2018).
6. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The Sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883. https://doi.org/10.1093/bioinformatics/bts034 (2012).
7. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313 (2016).
8. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A. L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555. https://doi.org/10.1126/science.1073374 (2002).
9. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. https://doi.org/10.1093/nar/28.1.27 (2000).
10. Kanehisa, M. Toward Understanding the origin and evolution of cellular organisms. *Protein Sci. Publ. Protein Soc.* **28**, 1947–1951. https://doi.org/10.1002/pro.3715 (2019).
11. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–d592. https://doi.org/10.1093/nar/gkac963 (2023).
12. Joly, J. H., Lowry, W. E. & Graham, N. A. Differential gene set enrichment analysis: A statistical approach to quantify the relative enrichment of two gene sets. *Bioinformatics* **36**, 5247–5254. https://doi.org/10.1093/bioinformatics/btaa658 (2021).
13. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **16**, 284–287. https://doi.org/10.1089/omi.2011.0118 (2012).
14. Zhang, J. et al. Identification of TYR, TYRP1, DCT and LARP7 as related biomarkers and immune infiltration characteristics of vitiligo via comprehensive strategies. *Bioengineered* **12**, 2214–2227. https://doi.org/10.1080/21655979.2021.1933743 (2021).
15. Sundermann, B. et al. Support vector machine analysis of functional magnetic resonance imaging of interoception does not reliably predict individual outcomes of cognitive behavioral therapy in panic disorder with agoraphobia. *Front. Psychiatry* **8**, 99. https://doi.org/10.3389/fpsyt.2017.00099 (2017).
16. Ambale-Venkatesh, B. et al. Cardiovascular event prediction by machine learning: The multi-ethnic study of atherosclerosis. *Circul. Res.* **121**, 1092–1101. https://doi.org/10.1161/circresaha.117.311312 (2017).
17. Li, Z. et al. Identification of potential early diagnostic biomarkers of Sepsis. *J. Inflamm. Res.* **14**, 621–631. https://doi.org/10.2147/jir.S298604 (2021).
18. Bindea, G. et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782–795. https://doi.org/10.1016/j.immuni.2013.10.003 (2013).
19. Skoufos, G. et al. TarBase-v9.0 extends experimentally supported miRNA-gene interactions to cell-types and virally encoded MiRNAs. *Nucleic Acids Res.* **52**, D304–d310. https://doi.org/10.1093/nar/gkad1071 (2024).
20. Rauluseviciute, I. et al. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **52**, D174–D182. https://doi.org/10.1093/nar/gkad1059 (2024).
21. Zhou, G. et al. NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* **47**, W234–w241. https://doi.org/10.1093/nar/gkz240 (2019).
22. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinf. (Oxford England)* **26**, 1572–1573. https://doi.org/10.1093/bioinformatics/btq170 (2010).
23. Yoo, M. et al. DSigDB: drug signatures database for gene set analysis. *Bioinf. (Oxford England)* **31**, 3069–3071. https://doi.org/10.1093/bioinformatics/btv313 (2015).
24. Morris, G. M., Huey, R. & Olson, A. J. Using AutoDock for ligand-receptor docking. *Curr. Protoc. Bioinform.* **8**, 14. https://doi.org/10.1002/0471250953.bi0814s24 (2008),
25. Valencia, W. M. & Florez, H. How to prevent the microvascular complications of type 2 diabetes beyond glucose control. *BMJ (Clinical Res. ed.)* **356**, i6505. https://doi.org/10.1136/bmj.i6505 (2017).
26. Yang, W. et al. Prevalence of diabetes among men and women in China. *N. Engl. J. Med.* **362**, 1090–1101. https://doi.org/10.1056/NEJMoa0908292 (2010).
27. Susztak, K. & Böttinger, E. P. Diabetic nephropathy: A frontier for personalized medicine. *J. Am. Soc. Nephrol.* **17**, 361–367. https://doi.org/10.1681/asn.2005101109 (2006).
28. Yoshida, Y., Kang, K., Chen, G., Gilliam, A. C. & Cooper, K. D. Cellular fibronectin is induced in ultraviolet-exposed human skin and induces IL-10 production by monocytes/macrophages. *J. Invest. Dermatol.* **113**, 49–55. https://doi.org/10.1046/j.1523-1747.1999.00623.x (1999).
29. Singh, P., Carraher, C. & Schwarzbauer, J. E. Assembly of fibronectin extracellular matrix. *Annu. Rev. Cell Dev. Biol.* **26**, 397–419. https://doi.org/10.1146/annurev-cellbio-100109-104020 (2010).
30. Vega, M. E., Kastberger, B., Wehrle-Haller, B. & Schwarzbauer, J. E. Stimulation of fibronectin matrix assembly by lysine acetylation. *Cells* **9**. https://doi.org/10.3390/cells9030655 (2020).
31. Li, Z. et al. Screening of the key genes and signalling pathways for diabetic nephropathy using bioinformatics analysis. *Front. Endocrinol.* **13**, 864407. https://doi.org/10.3389/fendo.2022.864407 (2022).
32. Mok, H., Al-Jumaily, A. & Lu, J. Plasmacytoma variant translocation 1 (PVT1) gene as a potential novel target for the treatment of diabetic nephropathy. *Biomedicines* **10**. https://doi.org/10.3390/biomedicines10112711 (2022).
33. Qin, B. & Cao, X. LncRNA PVT1 regulates high glucose-induced viability, oxidative stress, fibrosis, and inflammation in diabetic nephropathy via miR-325-3p/Snail1 Axis. *Diabetes Metab. Syndr. Obes.* **14**, 1741–1750. https://doi.org/10.2147/dmso.S303151 (2021).
34. Okamura, D. M., López-Guisa, J. M., Koelsch, K., Collins, S. & Eddy, A. A. Atherogenic scavenger receptor modulation in the tubulointerstitium in response to chronic renal injury. *Am. J. Physiol. Renal. Physiol.* **293**, F575–585. https://doi.org/10.1152/ajprenal.00063.2007 (2007).
35. Hua, W. et al. CD36 mediated fatty acid-induced podocyte apoptosis via oxidative stress. *PloS One.* **10**, e0127507. https://doi.org/10.1371/journal.pone.0127507 (2015).
36. Ruan, X. Z., Varghese, Z., Powis, S. H. & Moorhead, J. F. Human mesangial cells express inducible macrophage scavenger receptor. *Kidney Int.* **56**, 440–451. https://doi.org/10.1046/j.1523-1755.1999.00587.x (1999).
37. Yang, Y. L. et al. CD36 is a novel and potential anti-fibrogenic target in albumin-induced renal proximal tubule fibrosis. *J. Cell. Biochem.* **101**, 735–744. https://doi.org/10.1002/jcb.21236 (2007).
38. Li, X. et al. Advanced oxidation protein products promote lipotoxicity and tubulointerstitial fibrosis via CD36/β-Catenin pathway in diabetic nephropathy. *Antioxid. Redox. Signal.* **31**, 521–538. https://doi.org/10.1089/ars.2018.7634 (2019).

39. Kumar, S., Gowda, N. M., Wu, X., Gowda, R. N. & Gowda, D. C. CD36 modulates proinflammatory cytokine responses to plasmodium falciparum glycosylphosphatidylinositols and merozoites by dendritic cells. *Parasite Immunol.* **34**, 372–382. https://doi.org/10.1111/j.1365-3024.2012.01367.x (2012).
40. Silverstein, R. L. & Febbraio, M. CD36, a scavenger receptor involved in immunity, metabolism, angiogenesis, and behavior. *Sci. Signal.* **2**, re3 (2009). https://doi.org/10.1126/scisignal.272re3
41. Xu, M. et al. Identification and validation of immune and oxidative stress-related diagnostic markers for diabetic nephropathy by WGCNA and machine learning. *Front. Immunol.* **14**, 1084531. https://doi.org/10.3389/fimmu.2023.1084531 (2023).
42. Niu, H. et al. CD36 deletion ameliorates diabetic kidney disease by restoring fatty acid oxidation and improving mitochondrial function.
43. Niu, H. et al. CD36 deletion ameliorates diabetic kidney disease by restoring fatty acid oxidation and improving mitochondrial function. *Ren. Fail.* **45**, 2292753. https://doi.org/10.1080/0886022x.2023.2292753 (2023).
44. An, Z. et al. Prognostic value of serum Interleukin-6, NF-κB plus MCP-1 assay in patients with diabetic nephropathy. *Dis. Mark.* **2022** 4428484. https://doi.org/10.1155/2022/4428484 (2022).
45. Xu, D., Jiang, C., Xiao, Y. & Ding, H. Identification and validation of disulfidptosis-related gene signatures and their subtype in diabetic nephropathy. *Front. Genet.* **14**, 1287613. https://doi.org/10.3389/fgene.2023.1287613 (2023).
46. Qin, M. & Qiu, Z. Changes in TNF-α, IL-6, IL-10 and VEGF in rats with ARDS and the effects of dexamethasone. *Exp. Ther. Med.* **17**, 383–387. https://doi.org/10.3892/etm.2018.6926 (2019).
47. Zhang, J. et al. CCR7 and CD48 as Predicted Targets in Acute Rejection Related to M1 Macrophage after Pediatric Kidney Transplantation. *J. Immunol. Res.* 6908968. https://doi.org/10.1155/2024/6908968 (2024).
48. Xu, B. et al. Investigation of the mechanism of complement system in diabetic nephropathy via bioinformatics analysis. *J. Diabetes Res.* **2021**, 5546199. https://doi.org/10.1155/2021/5546199 (2021).
49. Klessens, C. Q. F. et al. Macrophages in diabetic nephropathy in patients with type 2 diabetes. *Nephrol. Dial. Transplant.* **32**, 1322–1329 (2017). https://doi.org/10.1093/ndt/gfw260
50. Puchalski, R. B. et al. An anatomic transcriptional atlas of human glioblastoma. *Science* **360**, 660–663. https://doi.org/10.1126/science.aaf2666 (2018).
51. Tan, H. et al. Glabridin, a bioactive component of licorice, ameliorates diabetic nephropathy by regulating ferroptosis and the VEGF/Akt/ERK pathways. *Mol. Med.* **28**, 58. https://doi.org/10.1186/s10020-022-00481-w (2022).
52. Xiaojun, W., Feixue, N. & Yuxue, X. Screening and identification of core genes for diabetic nephropathy. *Acta Universitatis Medicinalis Anhui* **59**, 610–618. https://doi.org/10.19405/j.cnki.issn1000-1492.2024.04.009 (2024).
53. Pérez-Morales, R. E. et al. Inflammation in diabetic kidney disease. *Nephron* **143**, 12–16. https://doi.org/10.1159/000493278 (2019).
54. Zhou, H., Mu, L., Yang, Z. & Shi, Y. Identification of a novel immune landscape signature as effective diagnostic markers related to immune cell infiltration in diabetic nephropathy. *Front. Immunol.* **14**, 1113212. https://doi.org/10.3389/fimmu.2023.1113212 (2023).
55. Moon, J. Y. et al. Aberrant recruitment and activation of T cells in diabetic nephropathy. *Am. J. Nephrol.* **35**, 164–174. https://doi.org/10.1159/000334928 (2012).
56. Chen, M. M. et al. Shen-Qi-Jiang-Tang granule ameliorates diabetic nephropathy via modulating tumor necrosis factor signaling pathway. *J. Ethnopharmacol.* **303**, 116031. https://doi.org/10.1016/j.jep.2022.116031 (2023).
57. Kim, S. M. et al. Targeting T helper 17 by mycophenolate mofetil attenuates diabetic nephropathy progression. *Transl. Res. J. Lab. Clin. Med.* **166**, 375–383. https://doi.org/10.1016/j.trsl.2015.04.013 (2015).
58. Feigerlová, E. & Battaglia-Hsu, S. F. IL-6 signaling in diabetic nephropathy: From pathophysiology to therapeutic perspectives. *Cytokine Growth Factor Rev.* **37**, 57–65. https://doi.org/10.1016/j.cytogfr.2017.03.003 (2017).
59. Han, Q. et al. Higher density of CD4 + T cell infiltration predicts severe renal lesions and renal function decline in patients with diabetic nephropathy. *Front. Immunol.* **15**, 1474377. https://doi.org/10.3389/fimmu.2024.1474377 (2024).
60. Duran-Salgado, M. B. & Rubio-Guerra, A. F. Diabetic nephropathy and inflammation. *World J. Diabetes* **5**, 393–398. https://doi.org/10.4239/wjd.v5.i3.393 (2014).
61. Agrawal, S. & Gupta, S. TLR1/2, TLR7, and TLR9 signals directly activate human peripheral blood Naive and memory B cell subsets to produce cytokines, chemokines, and hematopoietic growth factors. *J. Clin. Immunol.* **31**, 89–98. https://doi.org/10.1007/s10875-010-9456-8 (2011).
62. Navarro-González, J. F., Mora-Fernández, C., Muros de Fuentes, M. & García-Pérez, J. Inflammatory molecules and pathways in the pathogenesis of diabetic nephropathy. *Nat. Rev. Nephrol.* **7**, 327–340. https://doi.org/10.1038/nrneph.2011.51 (2011).
63. Guo, Y. et al. Infiltrating macrophages in diabetic nephropathy promote podocytes apoptosis via TNF-α-ROS-p38MAPK pathway. *Oncotarget* **8**, 53276–53287. https://doi.org/10.18632/oncotarget.18394 (2017).
64. Xu, X.et al. High glucose induced-macrophage activation through TGF-β-activated kinase 1 signaling pathway. *Inflamm. Res.* **65**, 655–664 https://doi.org/10.1007/s00011-016-0948-8 (2016).
65. Gao, L., Zhong, X., Jin, J., Li, J. & Meng, X. M. Potential targeted therapy and diagnosis based on novel insight into growth factors, receptors, and downstream effectors in acute kidney injury and acute kidney injury-chronic kidney disease progression. *Signal. Transduct. Target. Ther.* **5**, 9. https://doi.org/10.1038/s41392-020-0106-1 (2020).
66. Sun, Y., Dai, W. & He, W. Identification of key immune-related genes and immune infiltration in diabetic nephropathy based on machine learning algorithms. *IET Syst. Biol.* **17**, 95–106. https://doi.org/10.1049/syb2.12061 (2023).
67. Nayak, B. K. et al. HIF-1 mediates renal fibrosis in OVE26 type 1 diabetic mice. *Diabetes* **65**, 1387–1397. https://doi.org/10.2337/db15-0519 (2016).
68. Jiang, N. et al. HIF-1α ameliorates tubular injury in diabetic nephropathy via HO-1-mediated control of mitochondrial dynamics. *Cell Prolif.* **53**, e12909. https://doi.org/10.1111/cpr.12909 (2020).
69. Yang, T. et al. YY1: A novel therapeutic target for diabetic nephropathy orchestrated renal fibrosis. *Metab. Clin. Exp.* **96**, 33–45. https://doi.org/10.1016/j.metabol.2019.04.013 (2019).
70. Motojima, M. et al. Characterization of kidney and skeleton phenotypes of mice double heterozygous for Foxc1 and Foxc2. *Cells Tissues Organs.* **201**, 380–389. https://doi.org/10.1159/000445027 (2016).
71. Xie, J. et al. The genetic architecture of membranous nephropathy and its potential to improve non-invasive diagnosis. *Nat. Commun.* **11**, 1600. https://doi.org/10.1038/s41467-020-15383-w (2020).
72. Schmid, H. et al. Modular activation of nuclear factor-kappaB transcriptional programs in human diabetic nephropathy. *Diabetes* **55**, 2993–3003. https://doi.org/10.2337/db06-0477 (2006).
73. Yang, F. Q. et al. Foxl1 inhibits tumor invasion and predicts outcome in human renal cancer. *Int. J. Clin. Exp. Pathol.* **7**, 110–122 (2014).
74. Chen, B. & He, Q. miR-1187 induces podocyte injury and diabetic nephropathy through autophagy. *Diabetes Vasc. Dis. Res.* **20**, 14791641231172139. https://doi.org/10.1177/14791641231172139 (2023).
75. Tsai, Y. C. et al. Proximal tubule-derived exosomes contribute to mesangial cell injury in diabetic nephropathy via miR-92a-1-5p transfer. *Cell. Commun. Signal.* **21**, 10. https://doi.org/10.1186/s12964-022-00997-y (2023).
76. Simons, J. M., Hart, B. A., Vai Ching, I., Van Dijk, T. R., Labadie, R. P. & H. & Metabolic activation of natural phenols into selective oxidative burst agonists by activated human neutrophils. *Free Radic. Biol. Med.* **8**, 251–258. https://doi.org/10.1016/0891-5849(90)90070-y (1990).

77. Abd El-Ghafar, O. A. M. et al. Acetovanillone prevents cyclophosphamide-induced acute lung injury by modulating PI3K/Akt/mTOR and Nrf2 signaling in rats. *Phytother. Res.* **35**, 4499–4510. https://doi.org/10.1002/ptr.7153 (2021).
78. Hassanein, E. H. M., Bakr, A. G., El-Shoura, E. A. M., Ahmed, L. K. & Ali, F. E. M. Acetovanillone augmented the cardioprotective effect of carvedilol against cadmium-induced heart injury via suppression of oxidative stress and inflammation signaling pathways. *Sci. Rep.* **13**, 5278. https://doi.org/10.1038/s41598-023-31231-5 (2023).
79. Yang, X. et al. Apocynin attenuates acute kidney injury and inflammation in rats with acute hypertriglyceridemic pancreatitis. *Dig. Dis. Sci.* **65**, 1735–1747. https://doi.org/10.1007/s10620-019-05892-0 (2020).
80. Şahin, S., Aydın, A., Göçmen, A. Y. & Kaymak, E. Evaluation of the protective effect of Losartan in acetaminophen-induced liver and kidney damage in mice. *Naunyn. Schmiedebergs Arch. Pharmacol.* **397**, 5067–5078. https://doi.org/10.1007/s00210-023-02937-0 (2024).
81. Wang, Z. et al. Pioglitazone downregulates Twist-1 expression in the kidney and protects renal function of Zucker diabetic fatty rats. *Biomed. Pharmacother.* **118**, 109346. https://doi.org/10.1016/j.biopha.2019.109346 (2019).
82. Seargent, J. M., Yates, E. A. & Gill, J. H. GW9662, a potent antagonist of PPARgamma, inhibits growth of breast tumour cells and promotes the anticancer effects of the PPARgamma agonist Rosiglitazone, independently of PPARgamma activation. *Br. J. Pharmacol.* **143**, 933–937. https://doi.org/10.1038/sj.bjp.0705973 (2004).
83. Napolitano, M. et al. Alitretinoin for the treatment of severe chronic eczema of the hands. *Expert Opin. Pharmacother.* **23**, 159–167. https://doi.org/10.1080/14656566.2021.1998457 (2022).
84. Guan, Y. Targeting peroxisome proliferator-activated receptors (PPARs) in kidney and urologic disease. *Minerva Urol. Nefrol.* **54**, 65–79 (2002).
85. Yu, W. et al. Exosome-based liquid biopsies in cancer: Opportunities and challenges. *Ann. Oncol.* **32**, 466–477. https://doi.org/10.1016/j.annonc.2021.01.074 (2021).
86. Old, R. W., Crea, F., Puszyk, W. & Hultén, M. A. Candidate epigenetic biomarkers for non-invasive prenatal diagnosis of down syndrome. *Reprod. Biomed. Online* **15**, 227–235. https://doi.org/10.1016/s1472-6483(10)60713-4 (2007).
87. Anwer, M. et al. Sushi repeat-containing protein X-linked 2: A novel phylogenetically conserved hypothalamo-pituitary protein. *J. Comp. Neurol.* **526**, 1806–1819. https://doi.org/10.1002/cne.24449 (2018).
88. Chen, H., Su, X., Li, Y., Dang, C. & Luo, Z. Identification of metabolic reprogramming-related genes as potential diagnostic biomarkers for diabetic nephropathy based on bioinformatics. *Diabetol. Metab. Syndr.* **16**, 287. https://doi.org/10.1186/s13098-024-01531-5 (2024).
89. Altman, J. et al. A candidate panel of eight urinary proteins shows potential of early diagnosis and risk assessment for diabetic kidney disease in type 1 diabetes. *J. Proteom.* **300**, 105167. https://doi.org/10.1016/j.jprot.2024.105167 (2024).

Author contributions
Z.L. conceived and designed the study and analyzed the data for GEO datasets. ZP.S. wrote the manuscript. Y.Y. and J.S. modified the manuscript. All the authors read and approved the final manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval
All animal experiments followed ARRIVE guidelines and the NIH Guide for the Care and Use of Laboratory Animals (NIH Publication No. 8023, revised 1978).

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-01628-5.

Correspondence and requests for materials should be addressed to J.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.