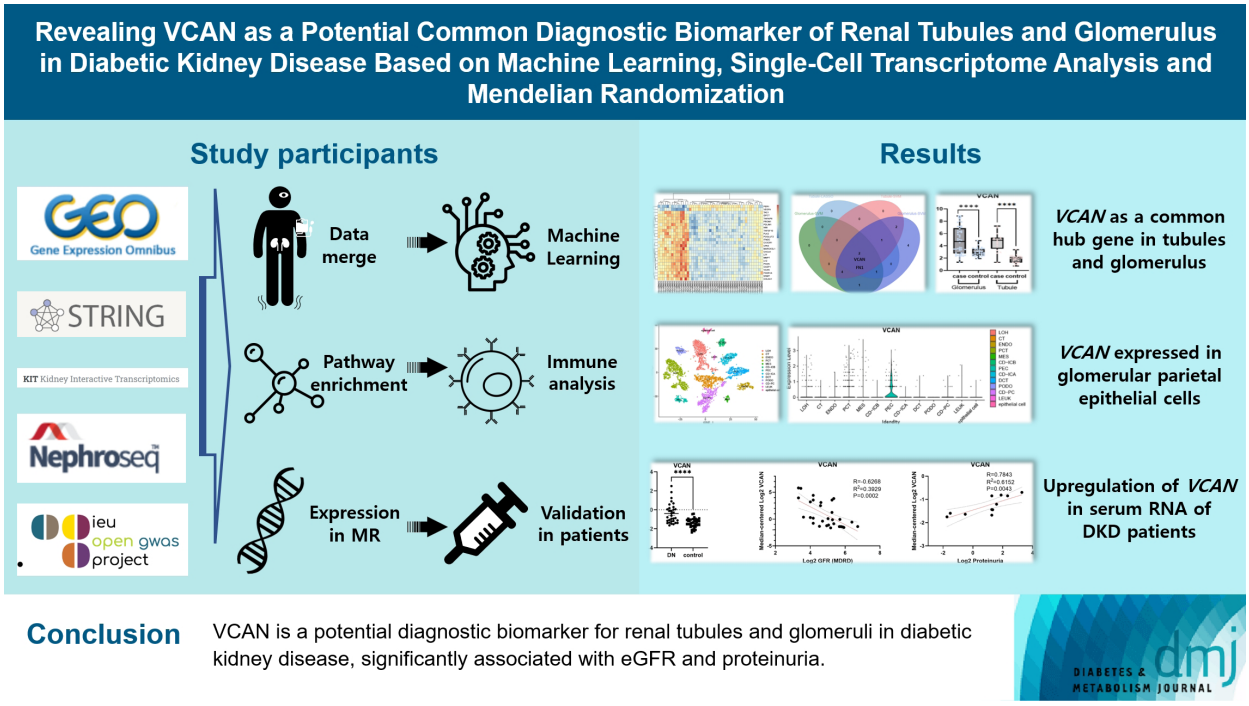
dmj
DIABETES & METABOLISM JOURNAL

Revealing VCAN as a Potential Common Diagnostic Biomarker of Renal Tubules and Glomerulus in Diabetic Kidney Disease Based on Machine Learning, Single-Cell Transcriptome Analysis and Mendelian Randomization

Li Jiang[1,*], Jie Jian[2,*], Xulin Sai[3], Xiai Wu[1]

Diabetes Metab J 2025;49:407-420 | *https://doi.org/10.4093/dmj.2024.0233*



Highlights

- **Machine learning algorithms identified VCAN as common in glomerular and tubular DKDs.**
- **VCAN was highly expressed in glomerular and proximal tubular epithelial cells.**
- **Mendelian randomization analysis showed serum VCAN protein as a risk factor for DKD.**
- **VCAN showed strong diagnostic potential for eGFR and proteinuria in DKD.**

How to cite this article:
Jiang L, Jian J, Sai X, Wu X. Revealing VCAN as a Potential Common Diagnostic Biomarker of Renal Tubules and Glomerulus in Diabetic Kidney Disease Based on Machine Learning, Single-Cell Transcriptome Analysis and Mendelian Randomization. Diabetes Metab J 2025;49:407-420. https://doi.org/10.4093/dmj.2024.0233

Original Article
Basic and Translational Research

Diabetes Metab J 2025;49:407-420
https://doi.org/10.4093/dmj.2024.0233
pISSN 2233-6079 · eISSN 2233-6087

dmj
DIABETES & METABOLISM JOURNAL

Revealing VCAN as a Potential Common Diagnostic Biomarker of Renal Tubules and Glomerulus in Diabetic Kidney Disease Based on Machine Learning, Single-Cell Transcriptome Analysis and Mendelian Randomization

Li Jiang[1,*], Jie Jian[2,*], Xulin Sai[3], Xiai Wu[1]

[1]Diabetes Department of Integrated Chinese and Western Medicine, China National Center for Integrated Traditional Chinese and Western Medicine, China-Japan Friendship Hospital, Beijing,
[2]Mental Health Center of Dongcheng District, Beijing,
[3]Dongzhimen Hospital Affiliated to Beijing University of Chinese Medicine, Beijing, China

Background: Diabetic kidney disease (DKD) is recognized as a significant complication of diabetes mellitus and categorized into glomerular DKDs and tubular DKDs, each governed by distinct pathological mechanisms and biomarkers.
Methods: Through the identification of common features observed in glomerular and tubular lesions in DKD, numerous differentially expressed gene were identified by the machine learning, single-cell transcriptome and mendelian randomization.
Results: The diagnostic markers versican (VCAN) was identified, offering supplementary options for clinical diagnosis. VCAN significantly highly expressed in glomerular parietal epithelial cell and proximal convoluted tubular cell. It was mainly involved in the up-regulation of immune genes and infiltration of immune cells like mast cell. Mendelian randomization analysis confirmed that serum VCAN protein levels were a risky factor for DKD, while there was no reverse association. It exhibited the good diagnostic potential for estimated glomerular filtration rate and proteinuria in DKD.
Conclusion: VCAN showed the prospects into DKD pathology and clinical indicator.

Keywords: Diabetic nephropathies; Kidney glomerulus; Kidney tubules; Machine learning; Mendelian randomization analysis; Single-cell gene expression analysis; VCAN protein, human

INTRODUCTION

Diabetic kidney disease (DKD), a complication of diabetes mellitus, poses a significant public health challenge globally. With the prevalence of diabetes on the rise, particularly type 2 diabetes mellitus which comprises the majority of cases, the burden of DKD has also surged [1]. According to the International Diabetes Federation (IDF) Diabetes Atlas (10th), 537 million people worldwide have diabetes, and this number is projected to reach 643 million by 2030, and 783 million by 2045 [2]. Up to 40% of people living with diabetes develop CKD. Among individuals with diabetes, DKD stands as one of the leading causes of end-stage renal disease, necessitating dialysis, or kidney transplantation for survival [3]. Early prediction and assessment of the onset and progression of DKD are crucial for effective prevention and treatment strategies.

According to the anatomical differences of different parts of the human kidney, DKD is subdivided into glomerular diabetic kidney diseases (GDKDs) and tubular diabetic kidney diseases (TDKDs) [4]. GDKDs primarily involve structural and

Corresponding authors: Xiai Wu https://orcid.org/0009-0006-3662-1978
Diabetes Department of Integrated Chinese and Western Medicine, China National Center for Integrated Traditional Chinese and Western Medicine, China-Japan Friendship Hospital, 2 Yinghuayuan E St, Chaoyang 100013, Beijing, China
E-mail: 20180931313@bucm.edu.cn

*Li Jiang and Jie Jian contributed equally to this study as first authors.

Received: May 5, 2024; Accepted: Sep. 7, 2024

functional abnormalities in the glomerulus, the filtering units of the kidney. In this subtype, there is a reduction in glomerular filtration capacity, often accompanied by thickening of the glomerular basement membrane and immune deposition, leading to proteinuria and eventual decline in kidney function [5,6]. In contrast, TDKDs predominantly affect the renal tubules, which are responsible for re-absorption and secretion of various substances. In TDKDs, dysfunction of tubular transporters and channels occurs, disrupting the normal handling of electrolytes, water, and other solutes by the tubular epithelial cells [7]. This dysfunction can result in abnormalities such as electrolyte imbalances, urinary concentrating defects, and impaired acid-base balance [8].

The typical progression of DKD is primarily based on the glomerulopathy, known as the glomerular hypothesis [9,10]. In recent years, with the emergence of the tubular hypothesis of nephron filtration [11,12], increasing evidence suggests that changes in tubular function may also play a significant role in the development of DKD. Based on different hypotheses regarding glomerular and tubular mechanisms, it is advisable to identify prognostic and diagnostic biomarkers to monitor the occurrence and progression of DKD, such as α1-microglobulin, kidney injury molecule 1 (proximal tubule markers), nephrin, and vascular endothelial growth factor (podocyte markers) [13,14]. However, there is currently limited research on shared biomarkers between tubules and glomerulus.

We analyzed the gene expression profiles of TDKD (GSE30529, GSE104954) and GDKD (GSE96804, GSE30528) obtained from the National Center for Biotechnology Information-Gene Expression Omnibus (NCBI-GEO) database. Based on the identification of differentially expressed genes (DEGs), we obtained a set of genes co-expressed in both tubules and glomerulus. Using comprehensive bioinformatics methods, we further investigated the molecular factors driving the pathogenesis of DKD. The co-expressed gene was also utilized for single-sample gene set enrichment analysis (ssGSEA) to explore the relationship between immune infiltration and DKD pathology. Finally, we performed the two-sample Mendelian randomization (MR) and clinical data correlation analysis to certify the clinical significance of hub genes.

METHODS

Data collection and merge

Gene expression datasets encompassing GSE96804, GSE30528, GSE30529, and GSE104954, wherein GSE96804 and GSE30528 delineated glomerular disease, while GSE30529 and GSE104954 characterized renal tubular lesions. Details regarding their platforms and sample annotation were provided in Supplementary Table 1. The R software packages "limma" and "sva" were employed to merge the data queues corresponding to glomerulus and renal tubules, respectively [15]. The batch effect was mitigated using the ComBat function [16]. Genes exhibiting log2 fold change >1 and $P<0.05$ were designated as DEGs [17].

Weighted gene co-expression network analysis

Weighted correlation network analysis (WGCNA) was utilized for the identification of modules comprising closely correlated genes and their interactions with phenotypic traits [18,19]. Following module delineation, their correlation with disease grouping (e.g., DKD and control) was evaluated via correlation analysis, thereby facilitating the identification of gene modules that may manifest differential expression in the context of disease [20]. WGCNA was separately conducted on renal tubule and glomerular datasets by screening the top 25% genes of variables. After removing the outlier sample, the most significantly expressing genes were integrated into the DEG datasets.

Functional enrichment and protein-protein interaction network construction

The Gene Ontology (GO) annotation of genes sourced from the R software package "org.Hs.eg.db" was employed, encompassing categories pertaining to biological processes, cellular components, and molecular functions [21]. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was predicated upon the "enrichKEGG" function. Statistical significance was determined by considering a $P<0.05$ and a false discovery rate <0.2 [22]. Interaction networks among DEGs-encoded proteins namely protein-protein interaction (PPI) were elucidated utilizing the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database [23]. A minimum required interaction score of 0.9 was stipulated, ensuring the absence of disconnected nodes within the network.

The filter of diagnostic biomarker based on machine learning

The PPI from the STRING database were imported into Cytoscape. With the aid of the "cytohubba" plug-in software, six kinds of algorithms including density of maximum neighbor-

hood component, maximal clique centrality, degree, bottleneck, edge percolated component, and closeness were obtained to forecast key genes [24]. Then, the Least Absolute Shrinkage and Selection Operator (LASSO) regression and support vector machine-recursive feature elimination (SVM-RFE) were implemented to discern genes significantly associated with DKD [25]. The receiver operating characteristic (ROC) analysis for these hub genes was executed. The area under the curve (AUC) values exceeding 0.65 in both the glomerular and tubular datasets were deemed to exhibit commendable diagnostic efficacy [26]. The chosen hub-gene would undergo validation in an external dataset GSE99325 (tubular tissue) and GSE30122 (kidney tissue).

Immune infiltration analysis

The R packages GSEABase, gene set variation analysis (GSVA), tibble, and tidyverse were loaded to facilitate the analysis. A total of 28 immune cell gene expression matrix tables were imported and linked with the glomerular and renal tubule datasets [27], enabling the assessment of immune cell infiltration in specific tissues. The comparisons were made between immune infiltrations in glomerulus and tubules, elucidating common immune patterns across both tissue types. The correlation between hub genes and immune cells were also analyzed in conjunction with ggcorrplot package.

Expression level of VCAN in single-cell transcriptome data

The expression level of versican (VCAN) in single-cell transcriptome data would be conducted in two different platform. Firstly, the gene would be imported in the Kidney Integrative Transcriptomics (K.I.T.) platform (http://humphreyslab.com/SingleCell/) based on the original data from human diabetic kidney [28]. Secondly, three signature DKD samples and three control samples from the GSE195460 dataset would be aggregated and analyzed. The single-cell transcriptome data were pre-processed using the "Seurat" package [29]. The data standardization and principal component analysis (PCA) clustering were carried out before the cell annotation. The cell trajectory analysis was conducted using the BiocGenerics and monocle packages to explore differentiation status and dynamic gene expression.

Two-sample Mendelian randomization

The phenotype-related single nucleotide polymorphisms (SNPs), as instrumental variables (IVs) derived from the genome-wide association study catalog, formed the basis of MR analysis. The data of VCAN originated from the whole-genome sequencing analysis of the metabolic proteome (ebi-a-GCST90010358); thus, existed as plasma protein exposure rather than mRNA. The data of DKD rooted in a cross-population atlas of genetic associations for 220 human phenotypes (ebi-a-GCST90018832). The flow chart of analysis was shown in Supplementary Fig. 1. The fundamental three assumptions could be accomplished by several methods as illustrated in Supplementary Fig. 2 [30]. The inverse variance weighting (IVW) served as the primary method for estimating causal effects, with a significance threshold of $P<1.03\times10^{-4}$ (after Bonferroni correction) denoting causality and $P<0.05$ indicating potential risk factors [31].

Clinical correlation

Correlations between the expression levels of hub genes and clinical indicators among patients diagnosed with DKD were investigated using Nephroseq (http://v5.nephroseq.org). Clinical indicators encompassed glomerular filtration rate (GFR) estimated by Modification of Diet in Renal Disease (MDRD) equation, proteinuria levels, blood glucose concentrations, and serum creatinine levels. The data of hub-gene originated from the affymetrix expression arrays from the diabetic kidney samples and expressed as serum mRNA. Clinical data from diverse origins were integrated following necessary corrections.

RESULTS

Screening of common DEGs in the glomerulus and tubules

The merging procedure of glomerulus or tubules dataset were illustrated in Supplementary Fig. 3. The differential expression was shown in Fig. 1. The volcano plot of glomerulus datasets showed 66 up-regulated genes and 148 down-regulated genes (Fig. 1A), while there were 221 up-regulated genes and 72 down-regulated genes in tubule dataset (Fig. 1B). A total of 38 up-regulated and 26 down-regulated genes were obtained (Fig. 1C).

Recognize of key gene dataset by WGCNA and functional enrichment

The WGCNA from glomerulus merge data distinguished four distinct clusters denoted as module (ME) blue (613 genes), ME brown (385 genes), ME turquoise (684 genes), and ME grey (1,235 genes) (Fig. 2A-C). The ME brown cluster exhibited the most robust negative association with the DKD group (F=

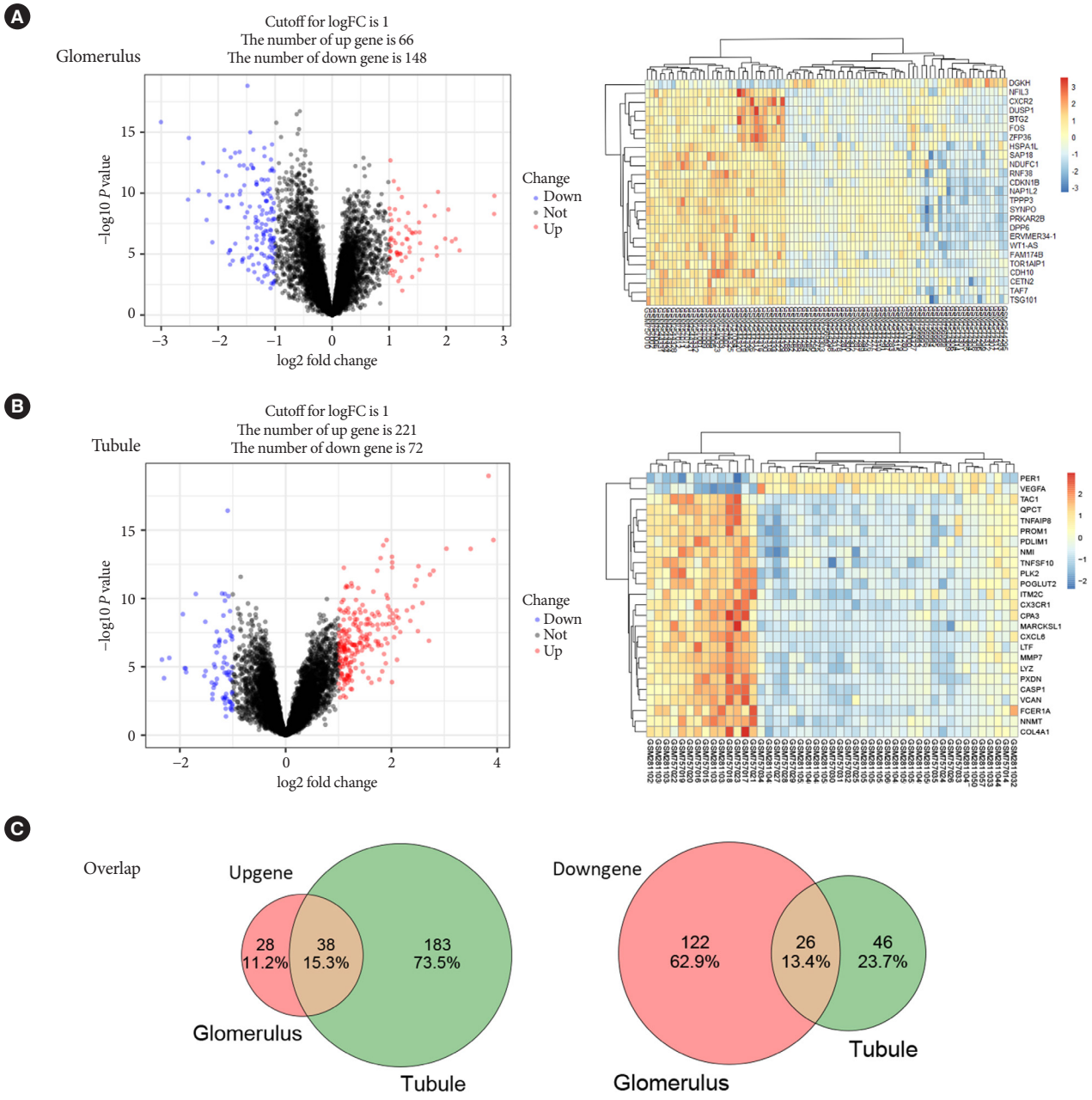


Fig. 1. Identification of differentially expressed genes in the two merge datasets. (A) Glomerulus datasets. (B) Tubule datasets. (C) The Venn plot of intersected gene between glomerulus and tubule. FC, fold change.

−0.66, *P*=2e-11) (Fig. 2D). In the tubule merge data, five distinct clusters were identified. They were namely ME turquoise (506 genes), ME yellow (342 genes), ME blue (447 genes), ME brown (344 genes), and ME grey (1,337 genes)(Fig. 2E-G).The ME grey cluster exhibited the strongest positive association with the DKD group (F=0.83, *P*=2e-11) (Fig. 2H). An intersection was computed between the dataset from glomerulus and tubule and merged with the preliminary DEGs originated from differential expression analysis, resulting in the formation of a new DEGs dataset. The GO and KEGG enrichment analysis were depicted in Supplementary Fig. 4, which revealed that the primary pathological pathways encompassed extracellular matrix (ECM)-receptor interaction, focal adhesion, phosphoinositide 3-kinase (PI3K)-Akt signaling pathway and oth-

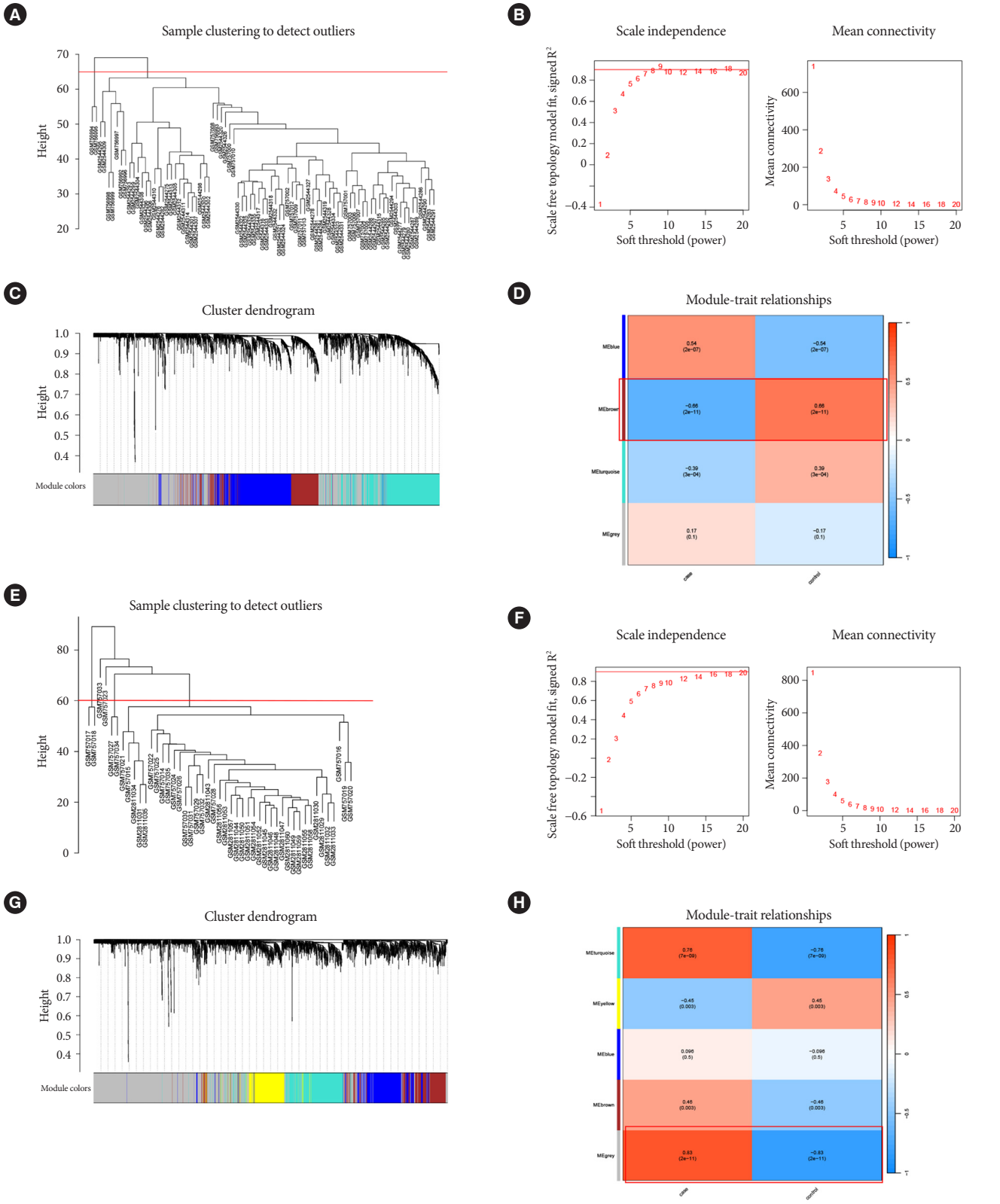


Fig. 2. Identification of key expressed genes by weighted correlation network analysis. (A, B, C, D) The glomerulus merge data. (E, F, G, H) The tubule merge data. ME, module.

ers. The top five representative enriched terms in GO and KEGG analysis were shown in Supplementary Table 2.

Common immune infiltration pattern of glomerulus and tubules

The immune infiltration analysis revealed that DEGs were primarily enriched in central memory CD4 T-cells and plasmacytoid dendritic cells in the glomerulus dataset (Fig. 3A). When compared to the control group, DKD patients exhibited significant activation in activated CD4 T-cells ($P<0.001$), regulatory T-cells ($P<0.001$), memory B-cells ($P<0.001$), and mast cells ($P<0.001$) (Fig. 3B). In the tubule dataset, DEGs were predominantly concentrated in immature dendritic cells, CD56dim natural killer cells, central memory CD4 T-cells, and plasmacytoid dendritic cells (Fig. 3C). In DKD patients compared to the control group, natural killer cells ($P<0.001$), effector memory CD8 T-cells ($P<0.001$), activated dendritic cells ($P<0.001$), and mast cells ($P<0.001$) showed significant activation (Fig. 3D). Thus, the mast cells participated in the common immune infiltration mode in both glomerulus and renal tubules.

Screening of hub genes by machine learning and network analysis

The PPI network diagram was shown in Supplementary Fig. 5. After comprehensive calculation of Cytoscape, 15 core genes, namely fibronectin 1 (FN1), collagen type I alpha 2 chain (COL1A2), decorin (DCN), collagen type III alpha 1 chain (COL3A1), lumican (LUM), CD44, VCAN, collagen type VI alpha 3 chain (COL6A3), integrin subunit alpha V (ITGAV), galectin 3 (LGALS3), cartilage oligomeric matrix protein (COMP), nephrosis 1 (NPHS1), nephrosis 2 (NPHS2), synaptopodin (SYNPO), podocalyxin like (PODXL) were subjected to LAS-

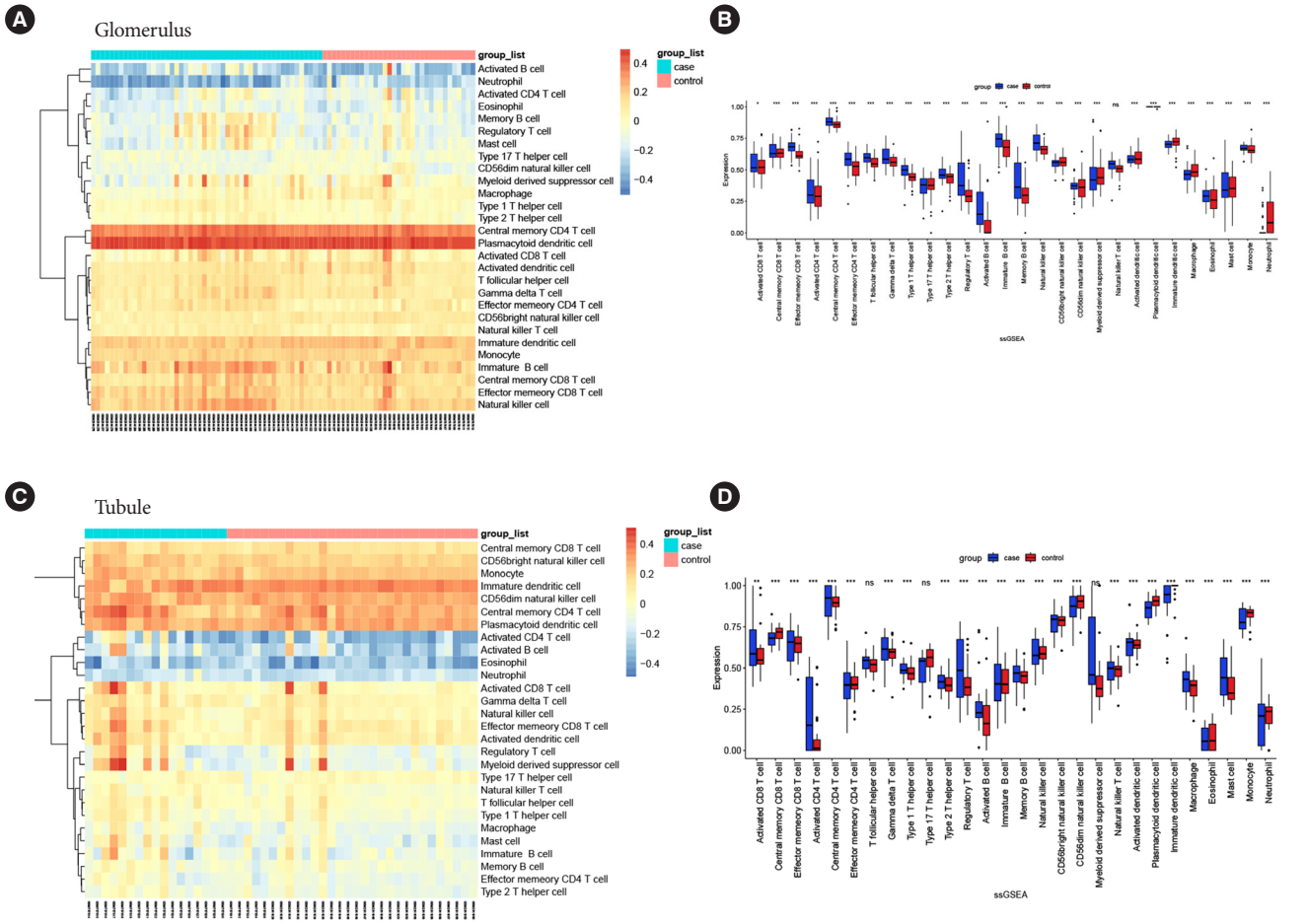


Fig. 3. The immune infiltration pattern of glomerulus and tubules. (A, B) The immune infiltration of glomerular tissue. (C, D) The immune infiltration of tubular tissue. ssGSEA, single-sample gene set enrichment analysis.

SO regression and SVM-RFE analysis, as shown in Supplementary Fig. 6. By intersecting the hub genes screened in both LASSO and SVM algorithm, the hub-gene co-expressed in glomerulus and tubule in patients with DKD could be determined (Supplementary Fig. 7). They were namely VCAN and FN1.

Validation and the immune infiltration pattern of the hub-gene

The ROC analysis revealed that in the glomerulus dataset, VCAN and FN1 both exhibited high diagnostic accuracy, with the AUC values exceeding 0.7 (Supplementary Fig. 8A). In the tubule dataset, VCAN and FN1 also demonstrated high diagnostic accuracy, with all AUC values surpassing 0.9. They showed significant up-regulation in DKD patients (Supplementary Fig. 8B). In the test dataset 1 (GSE99325, tubule) and dataset 2 (GSE30122, glomerulus+tubule). VCAN both showed significant up-regulation in DKD patients (Supplementary Fig. 8C and D). The associations between VCAN and immune cells were shown in Supplementary Fig. 9. The immune cells infiltrating both glomerulus and tubules mediated by VCAN primarily comprisedw the mast cells, regulatory T-cells and activated CD4 T-cells, with neutrophils being the principal suppressive cells.

Expression level of VCAN in single-cell transcriptome data

The quality control results of the GSE195460 dataset were shown in Supplementary Fig. 10, with low inter-sample heterogeneity and no significant batch effect. The elbow plot showed that the curve reached a plateau when principal component (PC) arrived at about 15–20 (Fig. 4A). Within the diabetic nephropathy (DN) group, the merged data could be effectively reduced to 16 clusters (Fig. 4B) and into 13 types of cells (Fig. 4C), they were namely mesenchymal cell (MES), glomerular parietal epithelial cell (PEC), proximal convoluted tubular cell (PCT), loop of Henle cell (LOH), distal convoluted tubular cell (DCT), convoluted tubular cell (CT), collecting duct-principal cell (CD-PC), collecting duct-intercalated cell type A (CD-ICA), collecting duct-intercalated cell type B (CD-ICB), endothelial cell (ENDO), and leukocyte (LEUK) and epithelial cells. Fig. 4D showed the correlation between cell markers and clusters. In the DN group, VCAN expression was significantly enriched in PEC and PCT cells (Fig. 4E and F). while in the control group, it was not significant (Fig. 4G and H).

The K.I.T. database showed that cell types were annotated as follows: MES, PEC, PCT, LOH, DCT, CT, CD-PC, CD-ICA, CD-ICB, ENDO, and LEUK (Fig. 4I). In the control group, VCAN expression was detected in CD-ICB and PEC cells (Fig. 4J). While in the DN group, VCAN exhibited significantly heightened expression levels primarily within PEC and MES cell populations (Fig. 4K-M).

The pseudotime analysis of PEC and PCT cells were shown in Supplementary Fig. 11. The expression of VCAN peaked in intermediate PEC cells and was lowest in immature PEC cells, indicating an enrichment pattern associated with cell differentiation. The PCT cells experienced a trajectory through two pivotal branching points. VCAN expressed highest in intermediate PCT cells and lowest in mature PCT cells. The immune score was shown in Supplementary Fig. 12. ENDO and LEUK were the principal cell types involved in the immune response, while PEC and PCT were moderately involved in the immune response. Most of the cells participated in the immune up-regulation reaction, and VCAN also mainly mediated the infiltration of immune cells rather than inhibition.

Two-sample bidirectional Mendelian randomization

There were 60 SNPs and 98 SNPs included in terms of VCAN-DKD and DKD-VCAN. The mean F-statistic was 19.14 and 18.71, respectively, indicating the robustness of IVs used in MR analysis. The MR calculations of VCAN-DKD and its reverse correlation were shown in the Supplementary Table 3. In the VCAN-DKD association analysis, the results based on IVW showed that odds ratio 1.088 (95% confidence interval, 0.151 to 1.017; $P=0.014 < 0.05$), and the beta direction of the five analysis methods was consistent, suggesting that VCAN-DKD had potential causal association. Serum VCAN protein was a risky factor for DKD. In the DKD-VCAN analysis, the beta direction of the five calculation methods was inconsistent, suggesting that there was no reverse causal association between them.

The scatter plot, funnel plot and leave-one-out plot were shown in Fig. 5. Sensitivity analysis was shown in Supplementary Table 4. There was positive causal association from VCAN to DKD (Fig. 5A). The funnel plot showed bilateral symmetry of SNPS, indicating no significant bias (Fig. 5B). The leave-one-out plot showed that all SNPS were located on one side of the equivalent line, indicating no significant heterogeneity in MR analysis (Fig. 5C). There was no reverse causality between VCAN and DKD (Fig. 5D-F).

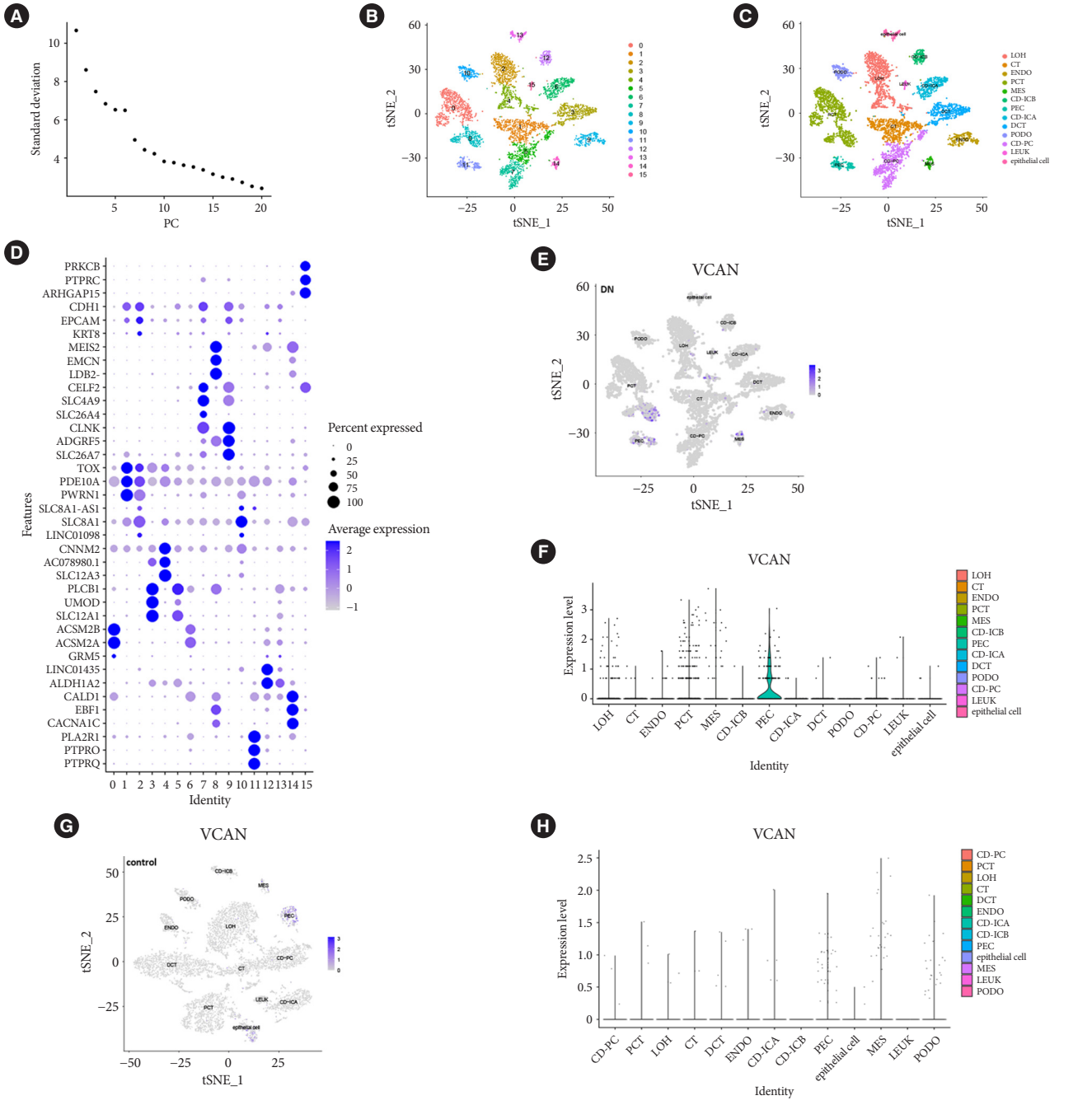


Fig. 4. Expression level of versican (VCAN) in single cell transcriptome data. (A) The elbow plot before annotation. (B) The primary cluster plot. (C) The cluster plot with manual annotation: mesenchymal cell (MES), glomerular parietal epithelial cell (PEC), proximal convoluted tubular cell (PCT), loop of Henle cell (LOH), distal convoluted tubular cell (DCT), convoluted tubular cell (CT), collecting duct-principal cell (CD-PC), collecting duct-intercalated cell type A (CD-ICA), collecting duct-intercalated cell type B (CD-ICB), endothelia cell (ENDO), leukocyte (LEUK). (D) The correlation between cell markers and clusters. (E) The distribution of VCAN in kidney cells of diabetic nephropathy (DN) group. (F) The expression of VCAN in DN group. (G) The distribution of VCAN in kidney cells of control group. (H) The expression of VCAN in control group. (I) The cluster plot with default annotation in Kidney Integrative Transcriptomics (K.I.T.) database. (J) The distribution of VCAN in the control group. (K) The distribution of VCAN in DN group. (L) The violin plot of VCAN expression. (M) The dot plot of VCAN expression. PODO, podocyte.

(*Continued to the next page*)

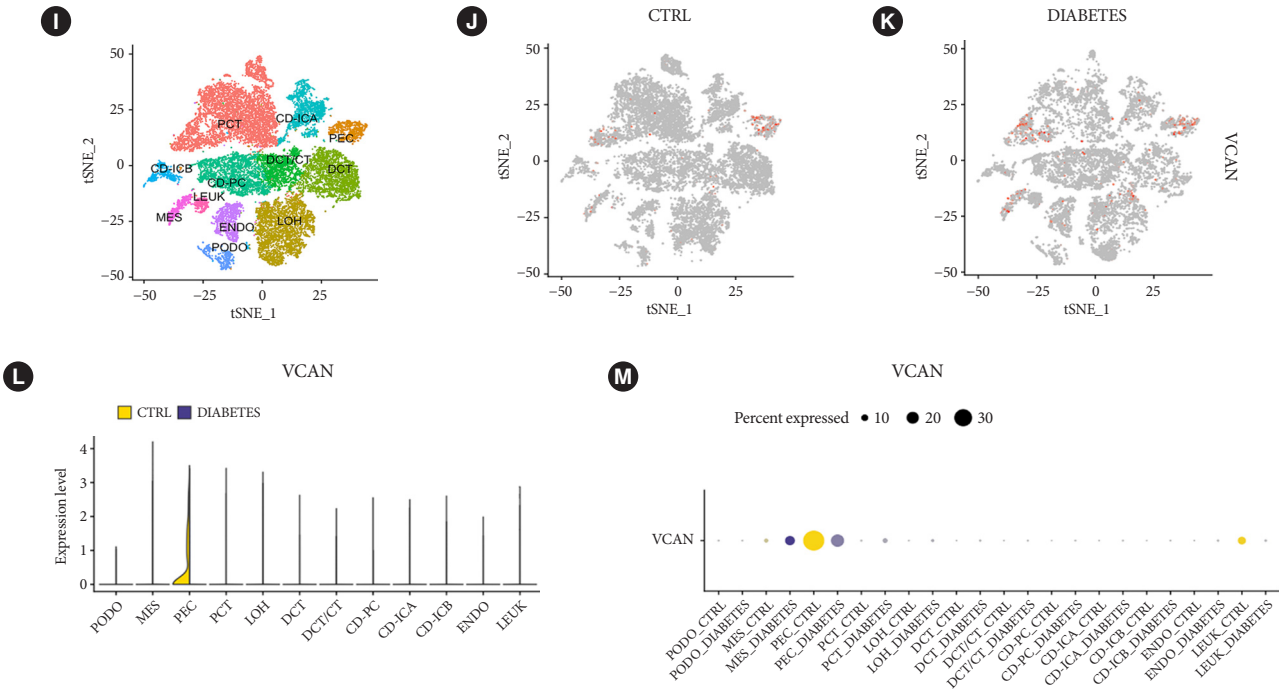


Fig. 4. Continued.

Expression level of hub genes in clinical DKD patients

It was observed that the expression levels of VCAN in serum RNA of DKD patients were notably up-regulated, exhibiting significant negative correlations with GFR (Fig. 6A). Furthermore, VCAN demonstrated a positive correlation with urinary protein levels, while FN1 showed no significant correlation (Fig. 6B).

DISCUSSION

This study was aimed to identify the shared biomarkers between tubules and glomerulus in DKD patients. The differential analysis yielded 66 up-regulated genes and 148 down-regulated genes from the glomerular dataset, and 221 up-regulated genes and 72 down-regulated genes from the renal tubule dataset. The intersection of these datasets resulted in the identification of the initial set of DEG. The WGCNA led to the second part of DEGs, contributing to the following functional annotation.

The GO analysis of the study showed that the common gene targets were most enriched in the ECM, extracellular structure and other related pathways. The significance of ECM in the pathogenesis of DKD has already been elucidated by recent research in pathology and molecular biology [32]. This dysregulated ECM turnover leads to glomerular and tubular fibrosis, impairing renal function and promoting the development of DKD-associated complications such as albuminuria and glomerulosclerosis. Additionally, the PI3K-Akt signaling pathway played a significant role in the progression and evolution of DKD by exerting deleterious effects on both the glomerulus and tubules. In the glomerulus, increased activation of Akt leads to glomerular hypertrophy and hyperfiltration, resulting in increased intraglomerular pressure and subsequent damage to the glomerular filtration barrier [33]. In the renal tubules, enhanced Akt signaling promoted tubular epithelial cell hypertrophy and dedifferentiation, leading to tubular epithelial cell dysfunction and impaired re-absorption of filtered proteins and electrolytes [34].

Leveraging various machine learning techniques, we ultimately pinpointed the VCAN as the most significant diagnostic biomarker. VCAN, also known as chondroitin sulfate proteoglycan 2 (CSPG2), is a large ECM proteoglycan involved in the pathological processes of DKD [35]. It consists of multiple domains, including a central hyaluronic acid-binding domain and chondroitin sulfate attachment sites, thus promoting ECM remodeling and inflammation reaction [36]. It can also modulate the activity of matrix metalloproteinases and tissue inhibitors of metalloproteinases, thereby regulating ECM turnover

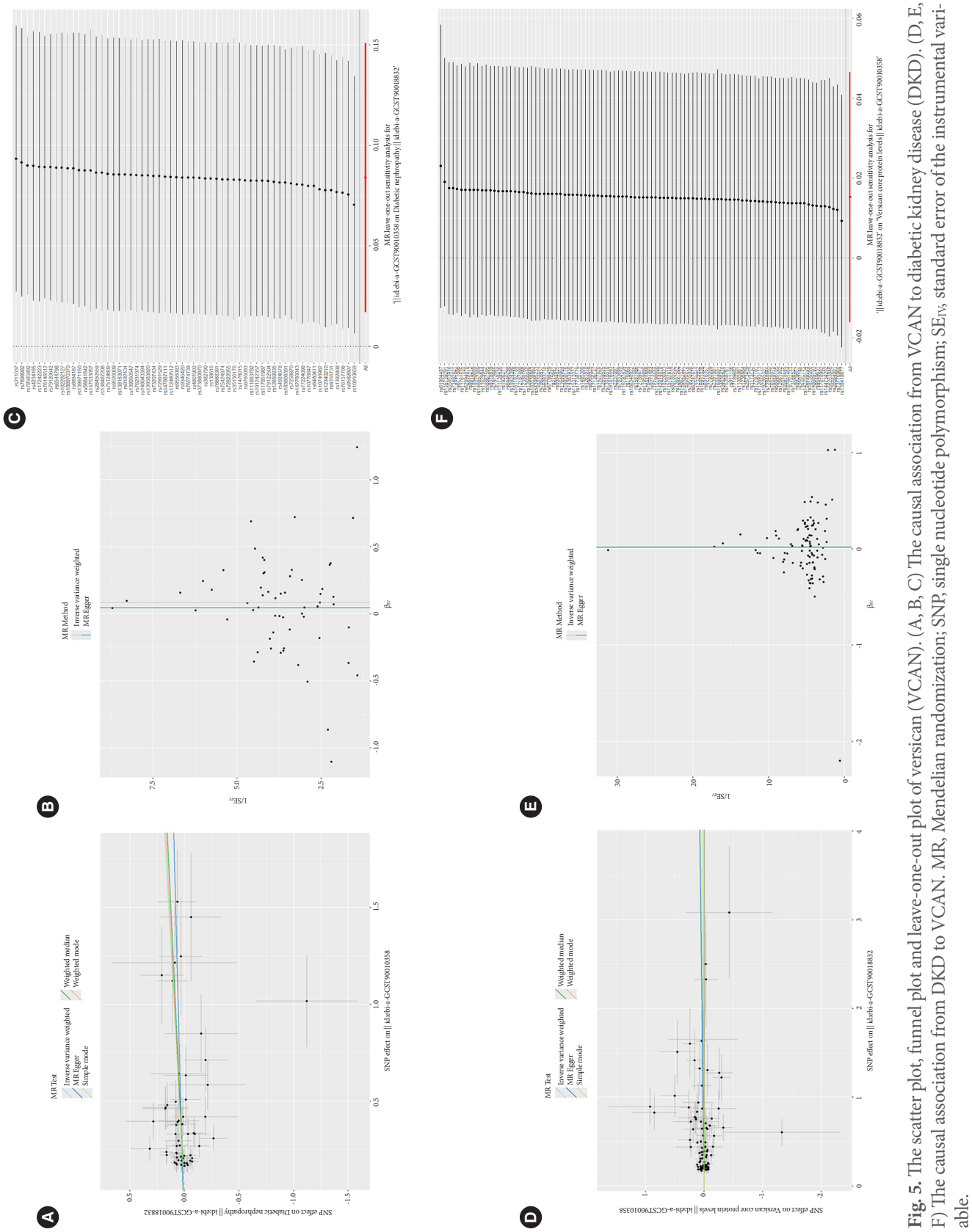


Fig. 5. The scatter plot, funnel plot and leave-one-out plot of versican (VCAN). (A, B, C) The causal association from VCAN to diabetic kidney disease (DKD). (D, E, F) The causal association from DKD to VCAN. MR, Mendelian randomization; SNP, single nucleotide polymorphism; SE_{IV}, standard error of the instrumental variable.

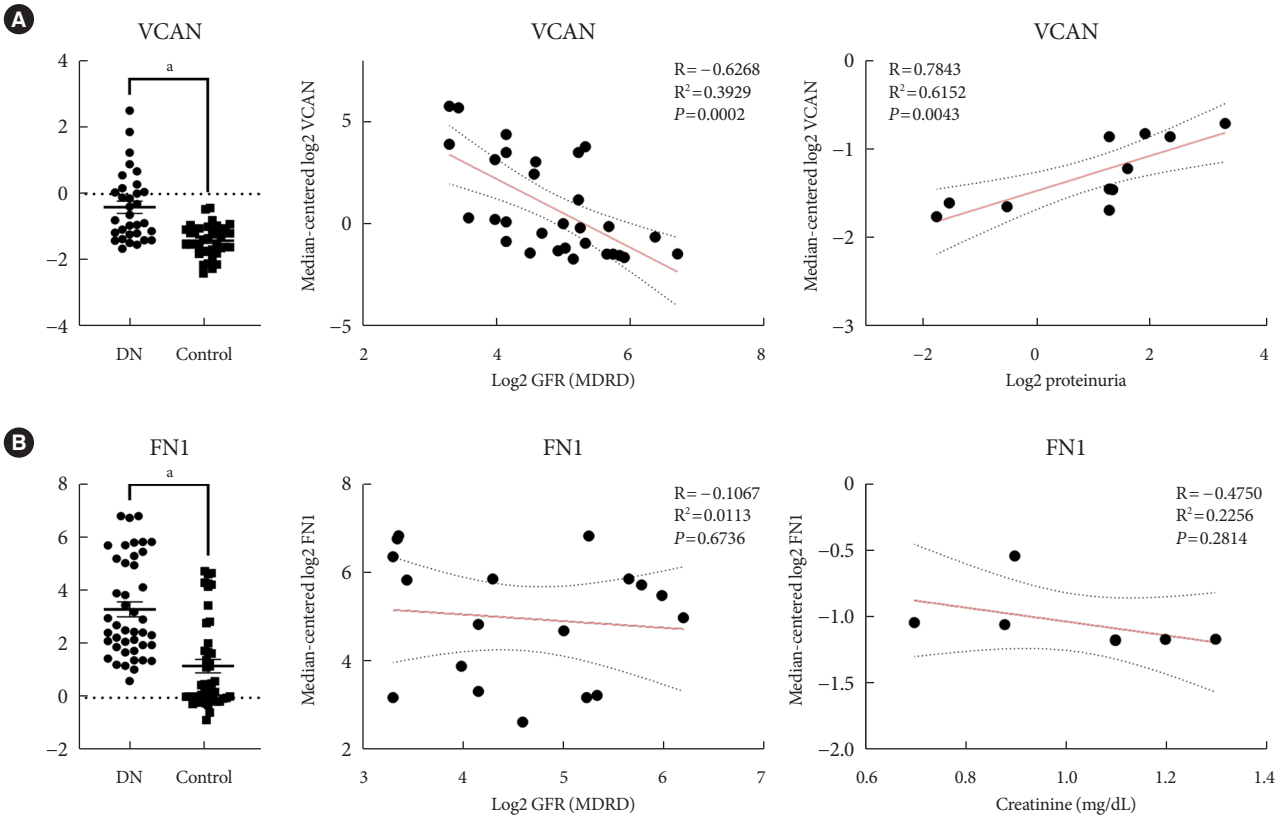


Fig. 6. Clinical correlation of hub-gene. (A) Serum versican (VCAN) mRNA levels in diabetic kidney disease (DKD) patients. (B) Serum fibronectin 1 (FN1) mRNA level in DKD patients. DN, diabetic nephropathy; GFR, glomerular filtration rate; MDRD, Modification of Diet in Renal Disease. [a]$P<0.0001$ (unpaired *t*-test).

and fibrosis. In renal tubules, VCAN contributes to ECM deposition and fibroblast activation, resulting in tubular fibrosis and dysfunction. In glomerulus, VCAN promotes ECM accumulation, thickening of the glomerular basement membrane, leading to impaired filtration function and structural damage [37,38]. It has been considered as the essential biomarker of ECM in DKD [39].

The result of immune infiltration showed that the VCAN might cause immune damage to the glomerulus and renal tubules by mediating mast cell recruitment. Mast cell-derived cytokines stimulate resident renal cells, such as mesangial cells, endothelial cells, and tubular epithelial cells, to produce additional inflammatory mediators, perpetuating renal inflammation and tissue damage. It also contributes to renal fibrosis by releasing profibrotic factors, such as transforming growth factor-β, fibroblast growth factor, and platelet-derived growth factor [40].

The single-cell transcriptome results further showed that VCAN was significantly enriched in glomerular PEC and PCT cells, which were the typical cells of glomerulus and renal tubules participated in the DKD pathology. The expression of VCAN peaked in intermediate PEC cells and intermediate PCT cells, which has been found as the key cell subsets or stages in the progression of kidney disease [41,42]. The immune score showed that VCAN mainly mediated the infiltration of immune cells rather than inhibition, which was consistent with the result of ssGSEA, further suggesting the role of VCAN as a biomolecule in renal immune injury. Two-sample bidirectional MR showed the serum VCAN protein was a potential risk factor for DKD. Clinical association analyses further reinforce this impression. It was observed that the expression levels of VCAN in serum RNA of DKD patients were notably up-regulated, exhibiting significant negative correlations with GFR and positive correlation with urinary protein levels.

Previous studies have confirmed the correlation between VCAN and DKD progression. Xu et al. [43], Wang et al. [44],

and Gao et al. [39] all identified VCAN as a key hub-gene by bioinformatics, which was up-regulated in human diabetic kidney tissues. However, they applied fewer datasets and directly integrated the transcriptome expression profiling by array from tubular and glomerular tissues, which lacked the quantitative and locative analysis of genes at the single-cell level. In this study, the transcriptome data of tubular and glomerular tissues were merged separately, and the co-expressed genes of glomerular and tubular tissues were obtained by intersecting them through multiple machine learning algorithms. Through the single-cell transcriptomics and MR, we deepened the role of VCAN in glomerular PEC and PCT cells in DKD patients. We also found that VCAN acted as a risk exposure factor for DKD rather than DKD triggering the changes of VCAN. As a clinical biomarker, it showed a significant positive correlation with proteinuria, which was also consistent with previous observation [45].

In summary, through the identification of common features observed in glomerular and tubular lesions in DKD, numerous distinct genes were identified, emphasizing roles in ECM formation and inflammatory pathways. Additionally, the diagnostic markers VCAN was identified, offering supplementary options for clinical diagnosis. VCAN significantly highly expressed in glomerular PEC and PCT cell. It was mainly involved in the up-regulation of immune genes and infiltration of immune cells like mast cell. MR analysis confirmed that serum VCAN protein levels were a risk factor for DKD, while there was no reverse association. It exhibited the good diagnostic potential for estimated GFR and proteinuria in DKD, thereby offering novel insights into DKD pathology and potentially diagnostic marker for DKD.

SUPPLEMENTARY MATERIALS

Supplementary materials related to this article can be found online at https://doi.org/10.4093/dmj.2024.0233.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

AUTHOR CONTRIBUTIONS

Conception or design: L.J., J.J.
Acquisition, analysis, or interpretation of data: all authors.
Drafting the work or revising: L.J., X.S., X.W.
Final approval of the manuscript: all authors.

ORCID

Li Jiang *https://orcid.org/0000-0003-3827-2178*
Jie Jian *https://orcid.org/0009-0002-1960-6515*
Xiai Wu *https://orcid.org/0009-0006-3662-1978*

FUNDING

This work was supported by National High Level Hospital Clinical Research Funding (2022-NHLHCRF-LX-02-0101) and National Natural Science Foundation of China (No.82004357).

ACKNOWLEDGMENTS

All authors would like to thank the public and shared dataset from the Gene Expression Omnibus (GEO) database and express gratitude for the funding.

REFERENCES

1. Alicic RZ, Rooney MT, Tuttle KR. Diabetic kidney disease: challenges, progress, and possibilities. Clin J Am Soc Nephrol 2017;12:2032-45.
2. Magliano DJ, Boyko EJ. IDF Diabetes Atlas. 10th ed. Brussels: International Diabetes Federation; 2021.
3. Gupta S, Dominguez M, Golestaneh L. Diabetic kidney disease: an update. Med Clin North Am 2023;107:689-705.
4. Chen SJ, Lv LL, Liu BC, Tang RN. Crosstalk between tubular epithelial cells and glomerular endothelial cells in diabetic kidney disease. Cell Prolif 2020;53:e12763.
5. Oliva-Damaso N, Mora-Gutierrez JM, Bomback AS. Glomerular diseases in diabetic patients: implications for diagnosis and management. J Clin Med 2021;10:1855.
6. Dong R, Xu Y. Glomerular cell cross talk in diabetic kidney diseases. J Diabetes 2022;14:514-23.
7. Wang Y, Jin M, Cheng CK, Li Q. Tubular injury in diabetic kidney disease: molecular mechanisms and potential therapeutic perspectives. Front Endocrinol (Lausanne) 2023;14:1238927.
8. Zeni L, Norden AGW, Cancarini G, Unwin RJ. A more tubulocentric view of diabetic kidney disease. J Nephrol 2017;30:701-17.

9. Oberg CM, Lindstrom M, Grubb A, Christensson A. Potential relationship between eGFRcystatin C/eGFRcreatinine-ratio and glomerular basement membrane thickness in diabetic kidney disease. Physiol Rep 2021;9:e14939.
10. Lassen E, Daehn IS. Molecular mechanisms in early diabetic kidney disease: glomerular endothelial cell dysfunction. Int J Mol Sci 2020;21:9456.
11. Vallon V, Thomson SC. The tubular hypothesis of nephron filtration and diabetic kidney disease. Nat Rev Nephrol 2020;16: 317-36.
12. Panchapakesan U, Pollock C. The primary cilia in diabetic kidney disease: a tubulocentric view? Int J Biochem Cell Biol 2020; 122:105718.
13. Jung CY, Yoo TH. Pathophysiologic mechanisms and potential biomarkers in diabetic kidney disease. Diabetes Metab J 2022; 46:181-97.
14. Colhoun HM, Marcovecchio ML. Biomarkers of diabetic kidney disease. Diabetologia 2018;61:996-1011.
15. Zeng H, Fu Y, Shen L, Quan S. Integrated analysis of multiple microarrays based on raw data identified novel gene signatures in recurrent implantation failure. Front Endocrinol (Lausanne) 2022;13:785462.
16. Bai J, Pu X, Zhang Y, Dai E. Renal tubular gen e biomarkers identification based on immune infiltrates in focal segmental glomerulosclerosis. Ren Fail 2022;44:966-86.
17. Wang Z, Song M, Li Y, Chen S, Ma H. Differential color development and response to light deprivation of fig (Ficus carica L.) syconia peel and female flower tissues: transcriptome elucidation. BMC Plant Biol 2019;19:217.
18. Fan G, Jin Z, Wang K, Yang H, Wang J, Li Y, et al. Identification of four hub genes in venous thromboembolism via weighted gene coexpression network analysis. BMC Cardiovasc Disord 2021;21:577.
19. Zhang T, Wong G. Gene expression data analysis using Hellinger correlation in weighted gene co-expression networks (WGCNA). Comput Struct Biotechnol J 2022;20:3851-63.
20. Toubiana D, Puzis R, Sadka A, Blumwald E. A genetic algorithm to optimize weighted gene co-expression network analysis. J Comput Biol 2019;26:1349-66.
21. Cui CY, Liu X, Peng MH, Liu Q, Zhang Y. Identification of key candidate genes and biological pathways in neuropathic pain. Comput Biol Med 2022;150:106135.
22. Liang W, Sun F, Zhao Y, Shan L, Lou H. Identification of susceptibility modules and genes for cardiovascular disease in diabetic patients using WGCNA analysis. J Diabetes Res 2020; 2020:4178639.
23. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 2023; 51:D638-46.
24. Xu M, Zhou H, Hu P, Pan Y, Wang S, Liu L, et al. Identification and validation of immune and oxidative stress-related diagnostic markers for diabetic nephropathy by WGCNA and machine learning. Front Immunol 2023;14:1084531.
25. Huang J, Zhang J, Zhang F, Lu S, Guo S, Shi R, et al. Identification of a disulfidptosis-related genes signature for prognostic implication in lung adenocarcinoma. Comput Biol Med 2023; 165:107402.
26. Ayoub I, Wolf BJ, Geng L, Song H, Khatiwada A, Tsao BP, et al. Prediction models of treatment response in lupus nephritis. Kidney Int 2022;101:379-89.
27. Ota M, Nagafuchi Y, Hatano H, Ishigaki K, Terao C, Takeshima Y, et al. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. Cell 2021;184:3006-21.
28. Wilson PC, Wu H, Kirita Y, Uchimura K, Ledru N, Rennke HG, et al. The single-cell transcriptomic landscape of early human diabetic nephropathy. Proc Natl Acad Sci U S A 2019;116: 19619-25.
29. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 2019;20: 163-72.
30. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. Wellcome Open Res 2023;4:186.
31. Li P, Wang H, Guo L, Gou X, Chen G, Lin D, et al. Association between gut microbiota and preeclampsia-eclampsia: a two-sample Mendelian randomization study. BMC Med 2022;20: 443.
32. Chen Y, Zou H, Lu H, Xiang H, Chen S. Research progress of endothelial-mesenchymal transition in diabetic kidney disease. J Cell Mol Med 2022;26:3313-22.
33. Wang H, Gao L, Zhao C, Fang F, Liu J, Wang Z, et al. The role of PI3K/Akt signaling pathway in chronic kidney disease. Int Urol Nephrol 2024;56:2623-33.
34. Zhang Y, Jin D, Kang X, Zhou R, Sun Y, Lian F, et al. Signaling pathways involved in diabetic renal fibrosis. Front Cell Dev Biol 2021;9:696542.

35. Kawashima H, Li YF, Watanabe N, Hirose J, Hirose M, Miyasaka M. Identification and characterization of ligands for L-selectin in the kidney. I. Versican, a large chondroitin sulfate proteoglycan, is a ligand for L-selectin. Int Immunol 1999;11:393-405.
36. Rudnicki M, Perco P, Neuwirt H, Noppert SJ, Leierer J, Sunzenauer J, et al. Increased renal versican expression is associated with progression of chronic kidney disease. PLoS One 2012;7:e44891.
37. Wight TN, Kang I, Evanko SP, Harten IA, Chang MY, Pearce OMT, et al. Versican: a critical extracellular matrix regulator of immunity and inflammation. Front Immunol 2020;11:512.
38. Wu YJ, La Pierre DP, Wu J, Yee AJ, Yang BB. The interaction of versican with its binding partners. Cell Res 2005;15:483-94.
39. Gao Z, Aishwarya S, Li XM, Li XL, Sui LN. Identification of key candidate genes and chemical perturbagens in diabetic kidney disease using integrated bioinformatics analysis. Front Endocrinol (Lausanne) 2021;12:721202.
40. Zhang X, Chao P, Zhang L, Xu L, Cui X, Wang S, et al. Single-cell RNA and transcriptome sequencing profiles identify immune-associated key genes in the development of diabetic kidney disease. Front Immunol 2023;14:1030198.
41. Liu WB, Huang GR, Liu BL, Hu HK, Geng J, Rui HL, et al. Single cell landscape of parietal epithelial cells in healthy and diseased states. Kidney Int 2023;104:108-23.
42. Ina K, Kitamura H, Nagai K, Tatsukawa S, Fujikura Y. Ultrastructural and functional changes of the proximal tubular epithelial cells in the renal cortex from spontaneously diabetic KKAy mice. J Electron Microsc (Tokyo) 1999;48:443-8.
43. Xu Q, Li B, Wang Y, Wang C, Feng S, Xue L, et al. Identification of VCAN as hub gene for diabetic kidney disease immune injury using integrated bioinformatics analysis. Front Physiol 2021;12:651690.
44. Wang Y, Tan J, Xu C, Wu H, Zhang Y, Xiong Y, et al. Identification and construction of lncRNA-associated ceRNA network in diabetic kidney disease. Medicine (Baltimore) 2021;100:e26062.
45. Li S, Li N, Li L, Zhan J. Sex difference in the association between serum versican and albuminuria in patients with type 2 diabetes mellitus. Diabetes Metab Syndr Obes 2023;16:3631-9.

Supplementary Table 1. Characteristic of dataset included in the analysis

Dataset	Database	Platform	Sample	Tissue
GSE96804	GEO	GPL17586	41 Cases of DKD, 20 cases of control	Glomerulus
GSE30528	GEO	GPL571	9 Cases of DKD, 13 cases of control	Glomerulus
GSE30529	GEO	GPL571	10 Cases of DKD, 12 cases of control	Tubules
GSE104954	GEO	GPL24120	7 Cases of DKD, 18 cases of control	Tubules

GEO, Gene Expression Omnibus; DKD, diabetic kidney disease.

Supplementary Table 2. Top five representative enriched terms in GO and KEGG analysis

GO	Description	Count	%	*P* value	*q* value
BP	Extracellular matrix organization	13	8.13	3.33E-06	0.002157498
	Extracellular structure organization	13	8.13	3.45E-06	0.002157498
	External encapsulating structure organization	13	8.13	3.69E-06	0.002157498
	Negative regulation of locomotion	15	9.38	4.34E-06	0.002157498
	Negative regulation of cell migration	14	8.75	4.59E-06	0.002157498
CC	Collagen-containing extracellular matrix	22	13.10	1.03E-11	2.17E-09
	Basement membrane	7	4.17	8.02E-06	0.000843834
	Endoplasmic reticulum lumen	12	7.14	1.72E-05	0.001210251
	Complex of collagen trimers	4	2.40	2.83E-05	0.001445218
	Lamellipodium membrane	4	2.40	3.43E-05	0.001445218
MF	Extracellular matrix structural constituent	13	7.80	3.14E-09	1.02E-06
	Glycosaminoglycan binding	14	8.59	2.03E-08	3.31E-06
	Enzyme inhibitor activity	15	7.36	7.58E-07	8.24E-05
	Peptidase regulator activity	12	6.13	1.02E-06	8.36E-05
	Endopeptidase inhibitor activity	10	6.13	3.37E-06	0.000220041
KEGG	ECM-receptor interacti on	9	9.90	3.30E-07	5.27E-05
	Focal adhesion	10	11.1	4.68E-05	0.002495322
	AGE-RAGE signaling pathway in diabetic complications	7	7.70	8.04E-05	0.002764122
	PI3K-Akt signaling pathway	13	14.30	8.64E-05	0.002764122
	Human papillomavirus infection	11	12.10	0.00065257	0.016334073

GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; BP, biological process; CC, cellular component; MF, molecular function; ECM, extracellular matrix; AGE-RAGE, advanced glycation end-products-receptor for advanced glycation end-products; PI3K, phosphoinositide 3-kinase.

Supplementary Table 3. Two-sample bidirectional Mendelian randomization between VCAN and DKD

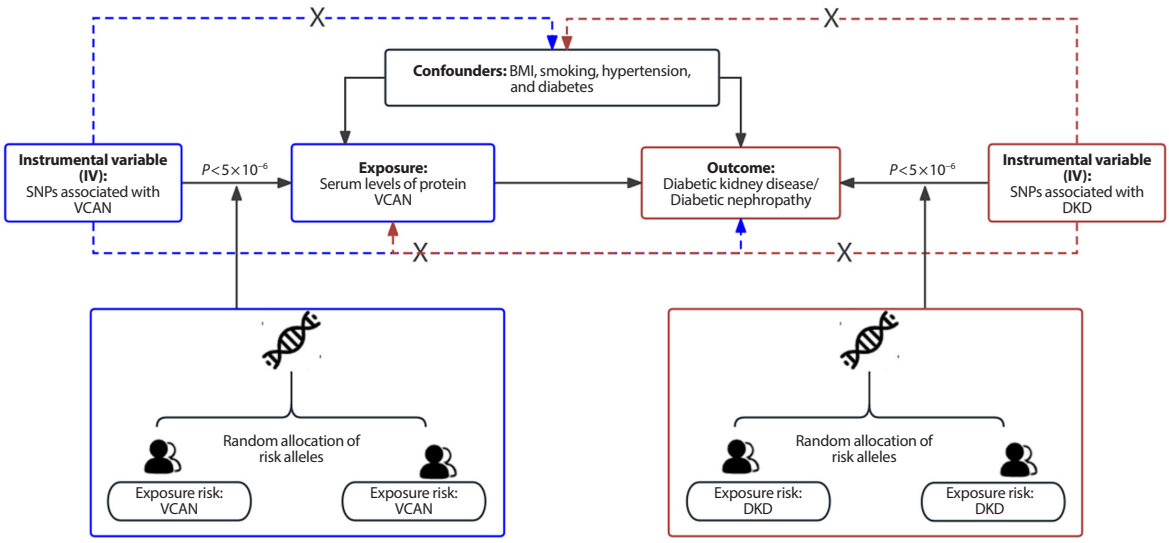
Exposure	Outcome	Method	SNPs	Beta	Standard error	*P* value	OR (95% CI)
VCAN	DKD	MR Egger	60	0.045	0.067	0.502	1.046 (0.176–0.918)
		Weighted median	60	0.084	0.049	0.085	1.087 (0.179–0.989)
		Inverse variance weighted	60	0.084	0.034	0.014	1.088 (0.151–1.017)
		Simple mode	60	0.096	0.095	0.313	1.101 (0.281–0.915)
		Weighted mode	60	0.090	0.070	0.200	1.095 (0.227–0.955)
DKD	VCAN	MR Egger	98	0.020	0.024	0.412	
		Weighted median	98	–0.001	0.026	0.960	
		Inverse variance weighted	98	0.015	0.016	0.338	
		Simple mode	98	0.008	0.054	0.884	
		Weighted mode	98	0.001	0.025	0.983	

VCAN, versican; DKD, diabetic kidney disease; SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval.

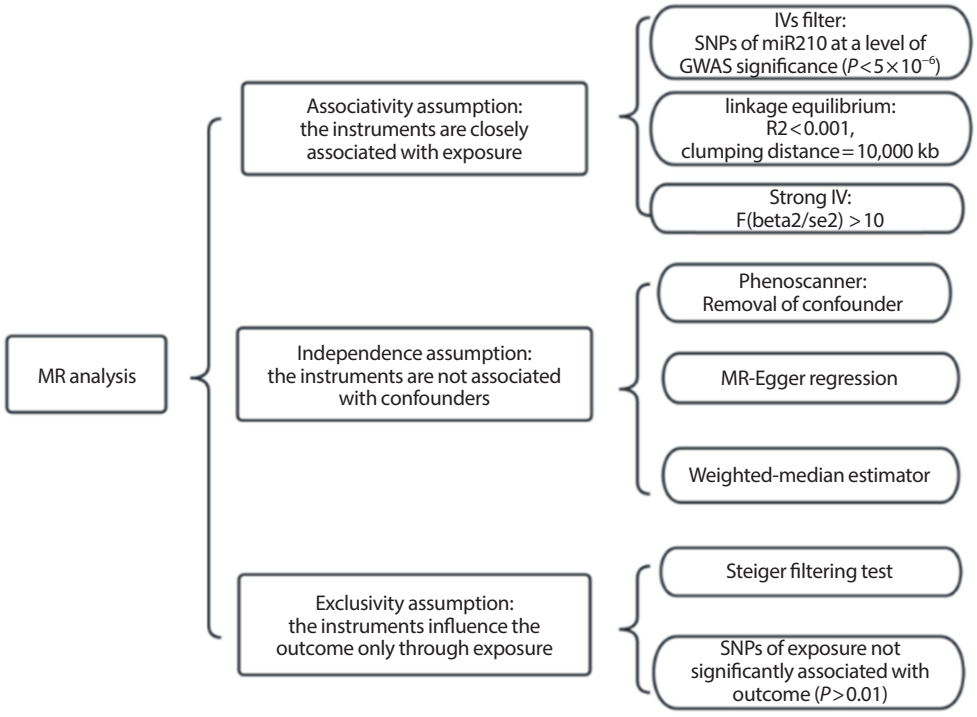
Supplementary Table 4. Sensitivity analysis

	MR analysis causal estimate	SD	T-stat	Main MR results *P* value	Global test	Global test *P* value
MR-presso						
VCAN-DKD	0.084	0.034	2.458	0.017	68.506	0.275
DKD-VCAN	0.015	0.016	0.957	0.341	112.609	0.143
Heierogeneity	Method	Q	Q_df	Q_*P* value		
VCAN-DKD	IVW	66.525	59.000	0.234		
DKD-VCAN	IVW	110.433	97.000	0.166		
Plciotropy	Egger_intercept	SE	*P* value			
VCAN-DKD	0.013	0.019	0.499			
DKD-VCAN	–0.003	0.010	0.807			

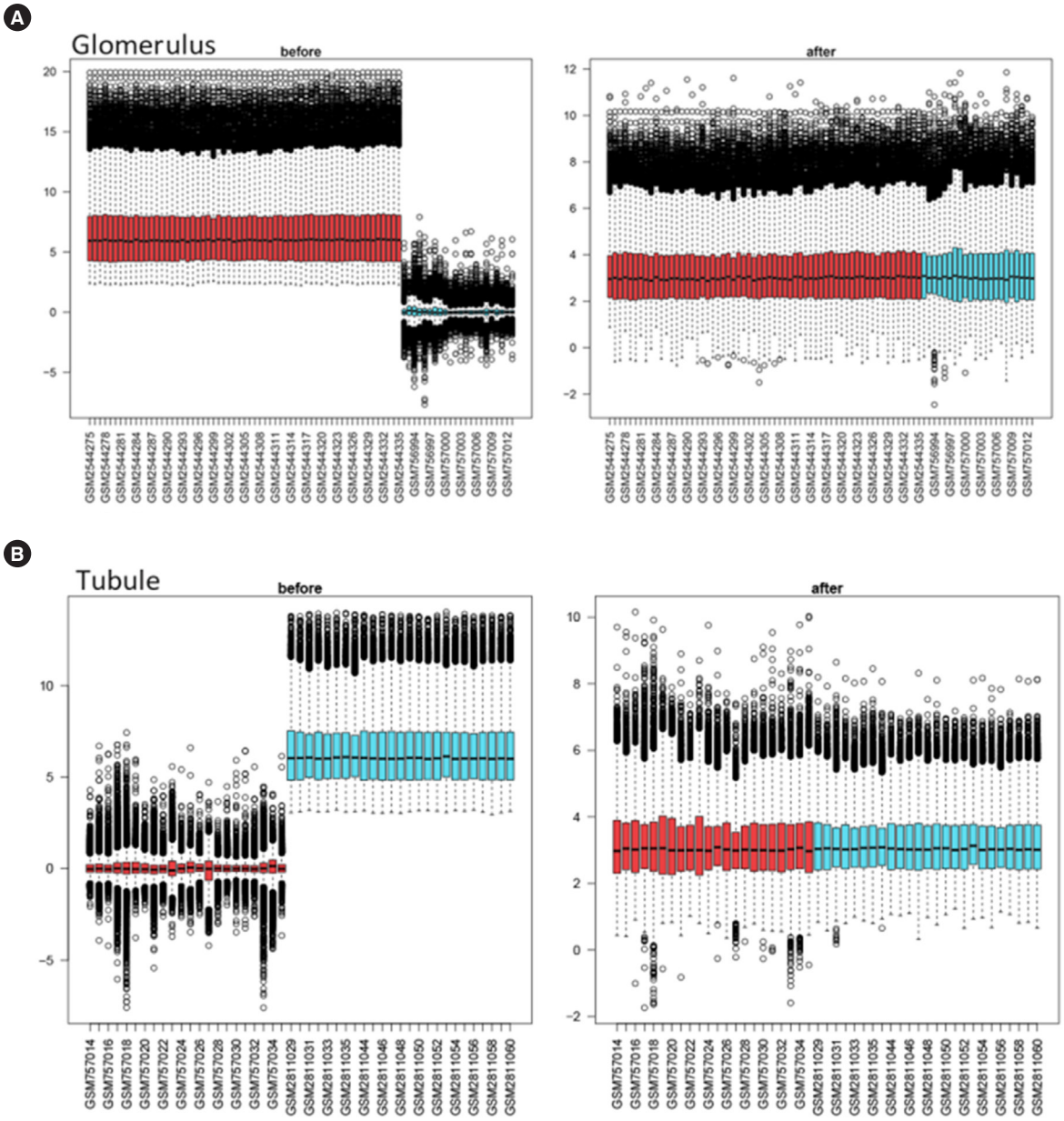
MR, Mendelian randomization; SD, standard deviation; VCAN, versican; DKD, diabetic kidney disease; Q, Cochran's q statistic; df, degrees of freedom; IVW, inverse variance weighting; SE, standard error.



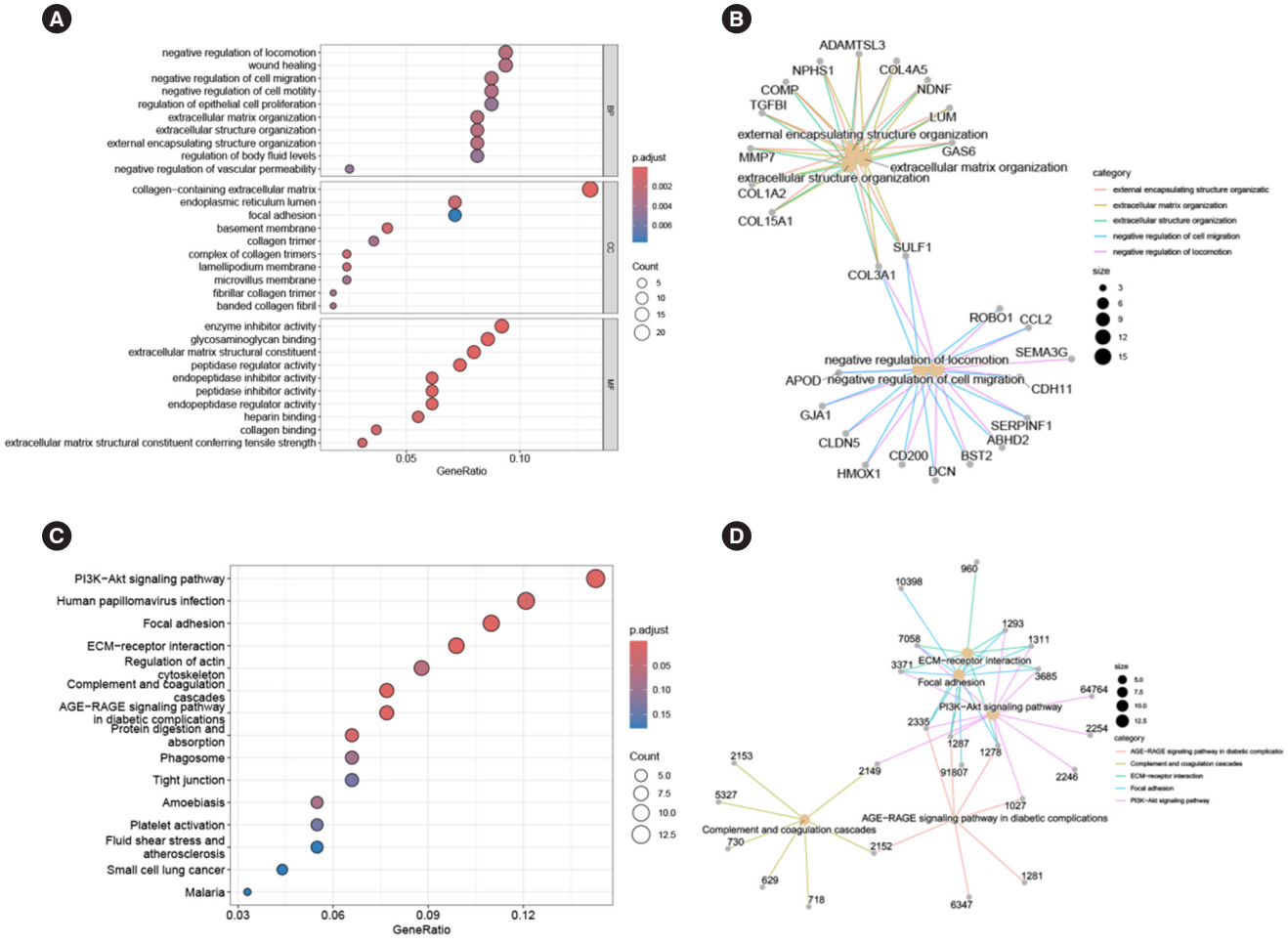
Supplementary Fig. 1. Flow chart of Mendelian randomization. BMI, body mass index; SNP, single nucleotide polymorphism; VCAN, versican; DKD, diabetic kidney disease.



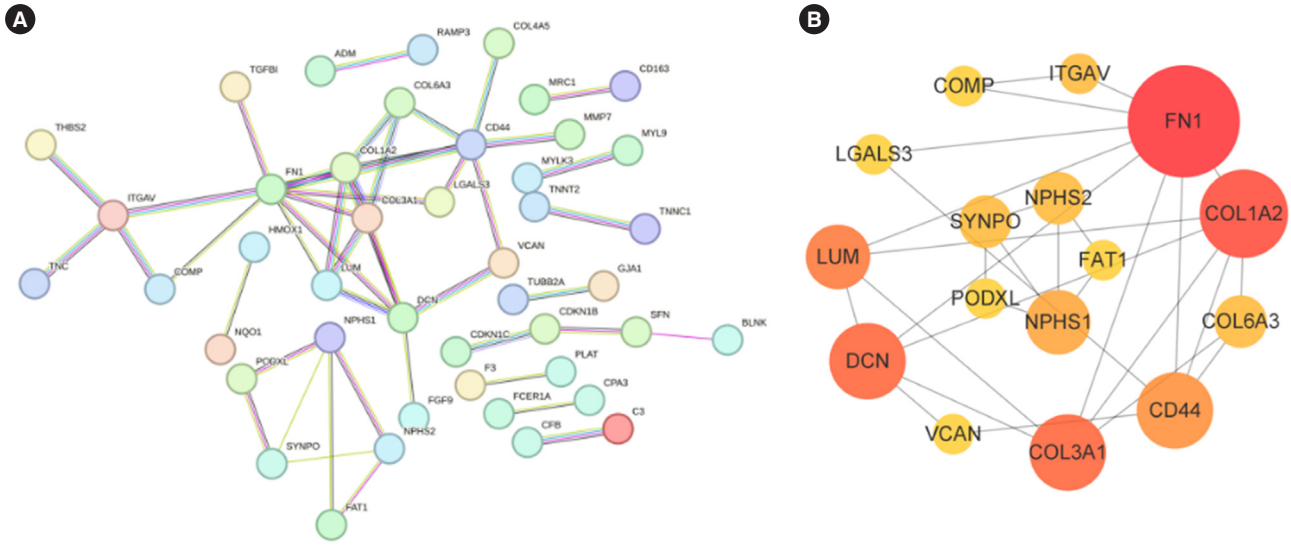
Supplementary Fig. 2. Hypotheses-test method. MR, Mendelian randomization; IV, instrumental variable; SNP, single nucleotide polymorphism; GWAS, genome-wide association study.



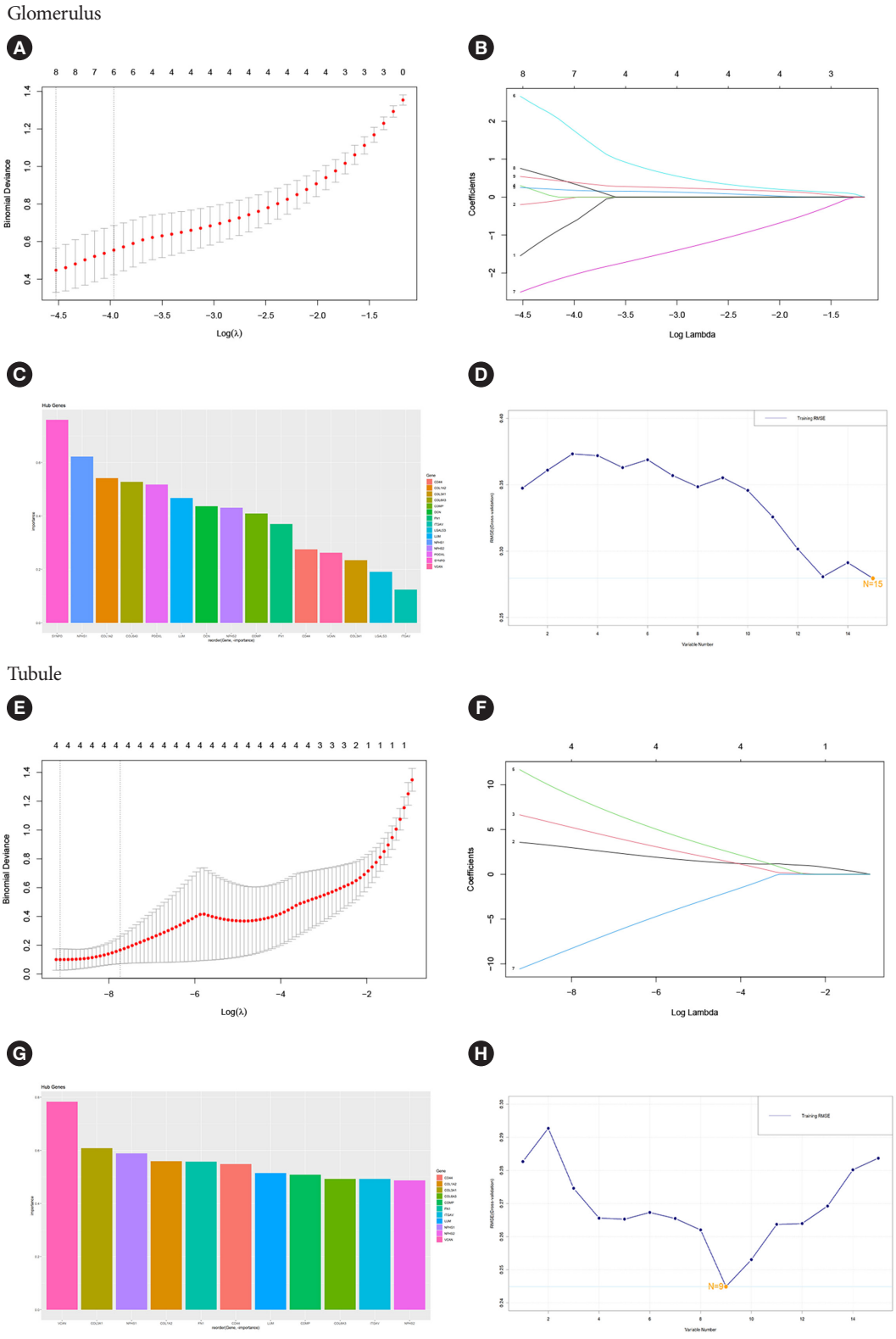
Supplementary Fig. 3. The merge data after removing the batch effect. (A) Glomerulus merge data consist of GSE96804 (red) and GSE30528 (blue) in total of 61 diabetic kidney disease (DKD) samples and 22 healthy samples. (B) Tubules merge data consist of GSE30529 (red) and GSE104954 (blue) in total of 22 DKD samples and 25 healthy samples.



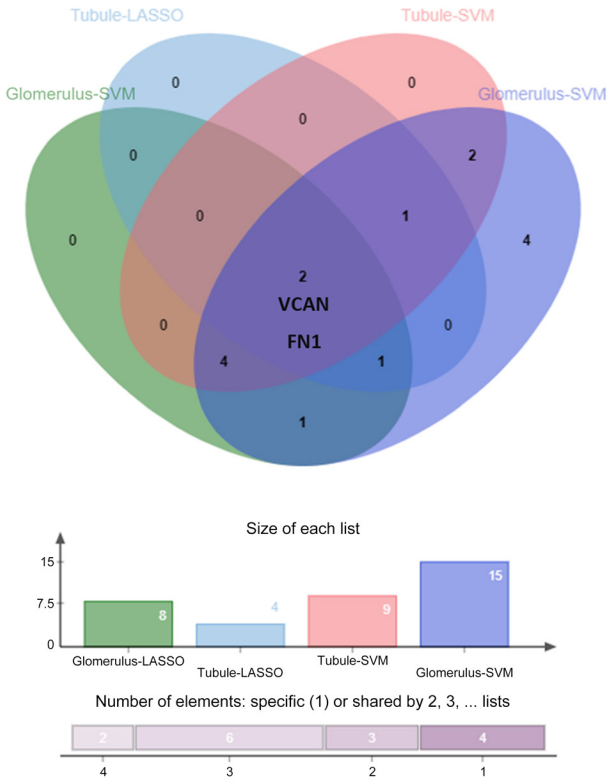
Supplementary Fig. 4. Functional enrichment by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. (A, B) GO analysis based on biological process (BP), cellular component (CC), and molecular function (MF). (C, D) KEGG analysis. PI3K, phosphoinositide 3-kinase; ECM, extracellular matrix; AGE-RAGE, advanced glycation end-products-receptor for advanced glycation end-products.



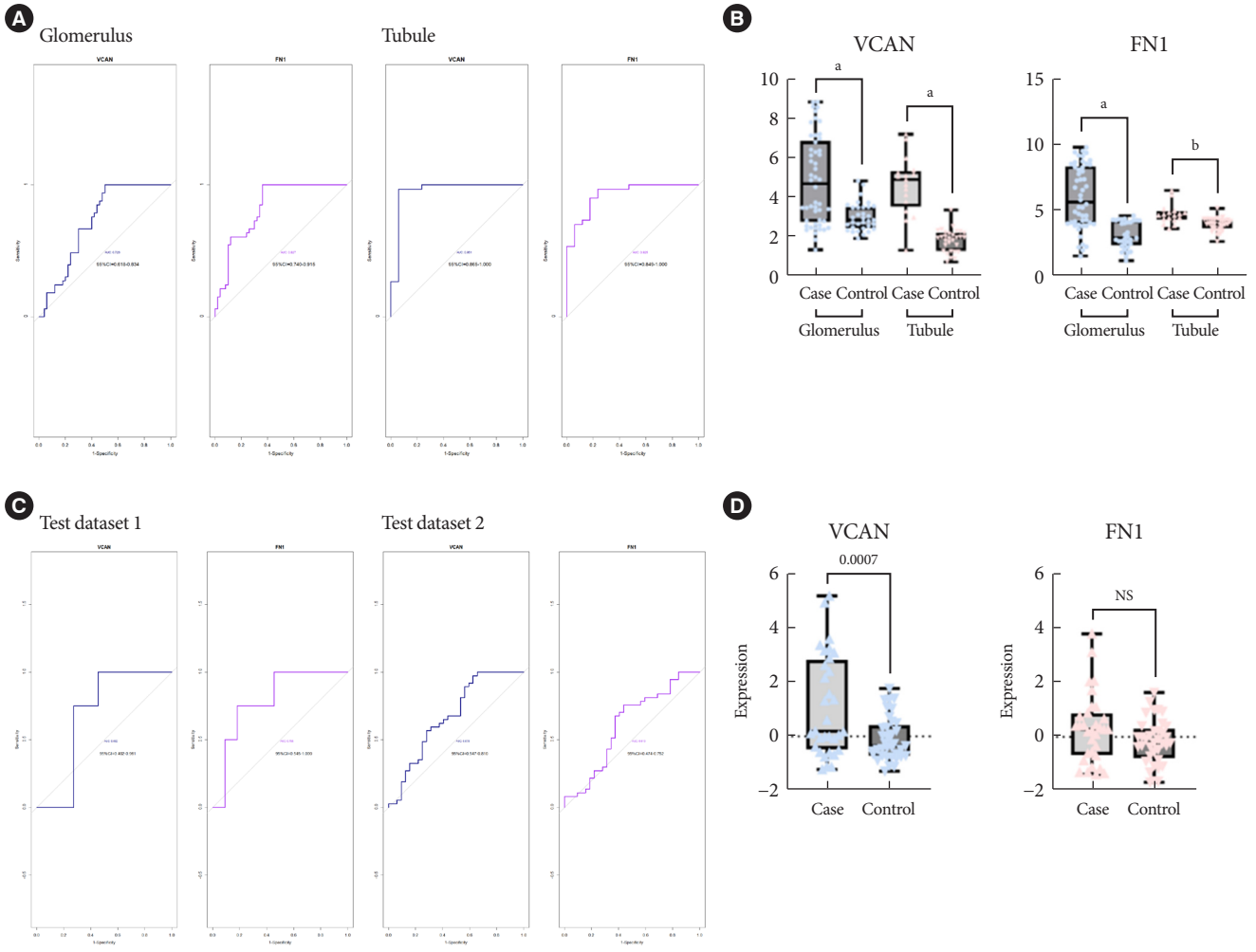
Supplementary Fig. 5. Protein-protein interaction (PPI) network and hub-gene network. (A) PPI network: The nodes with a degree value greater than 0.9 which were interrelated have been selected. (B) Hub-gene network: 15 hub-gene were determined with the aid of the "cytohubba" package in Cytoscape. The darker the color, the larger the node, and the greater the probability that the node occupies the first place in the six algorithms.



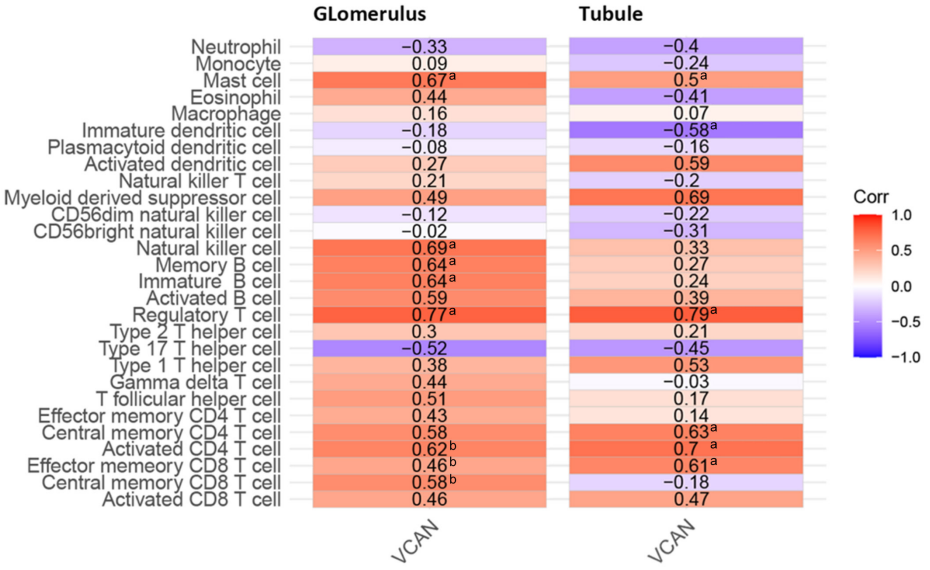
Supplementary Fig. 6. Least Absolute Shrinkage and Selection Operator (LASSO) regression and support vector machine (SVM). (A, B, E, F) LASSO regression. (C, D, G, H) The SVM-recursive feature elimination (RFE) analysis.



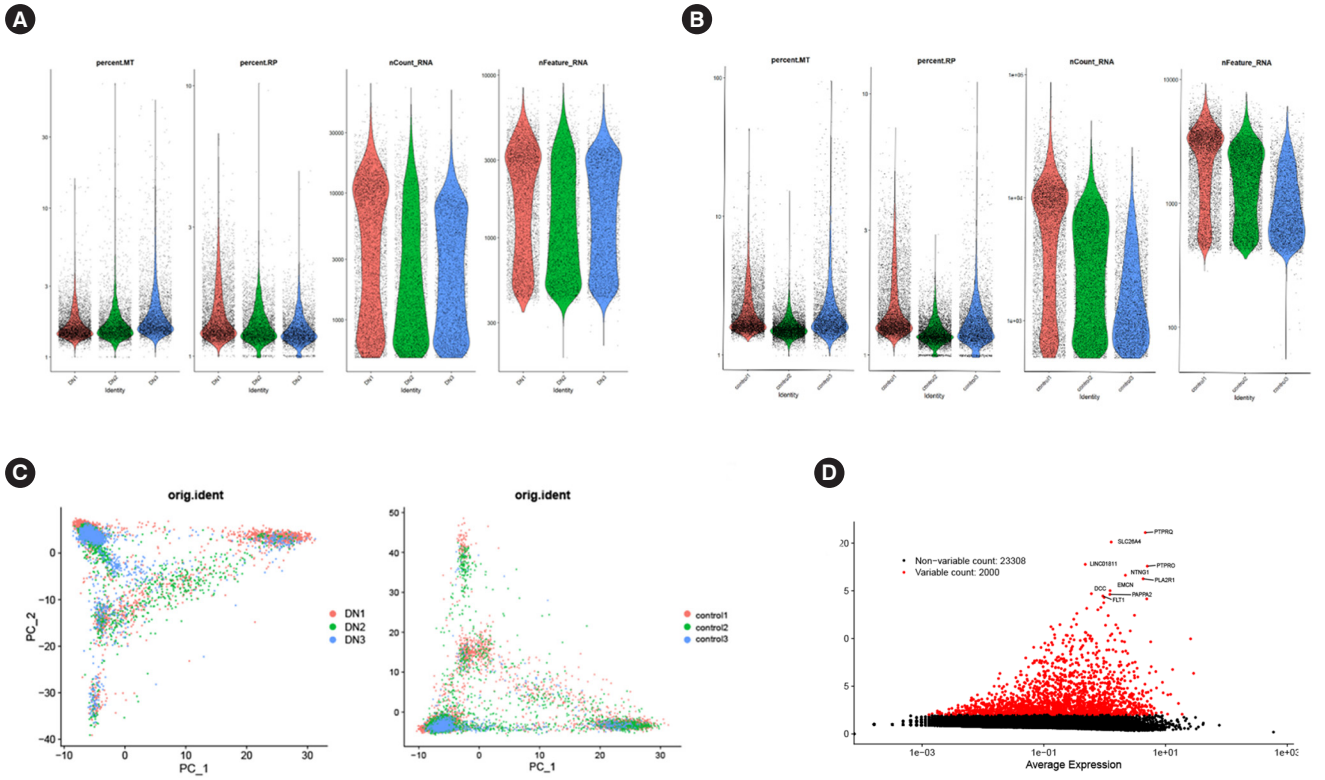
Supplementary Fig. 7. Intersection of the hub genes. Versican (VCAN) and fibronectin 1 (FN1) were the co-expressed genes in glomerulus and tubule in patients with diabetic kidney disease. LASSO, Least Absolute Shrinkage and Selection Operator; SVM, support vector machine.



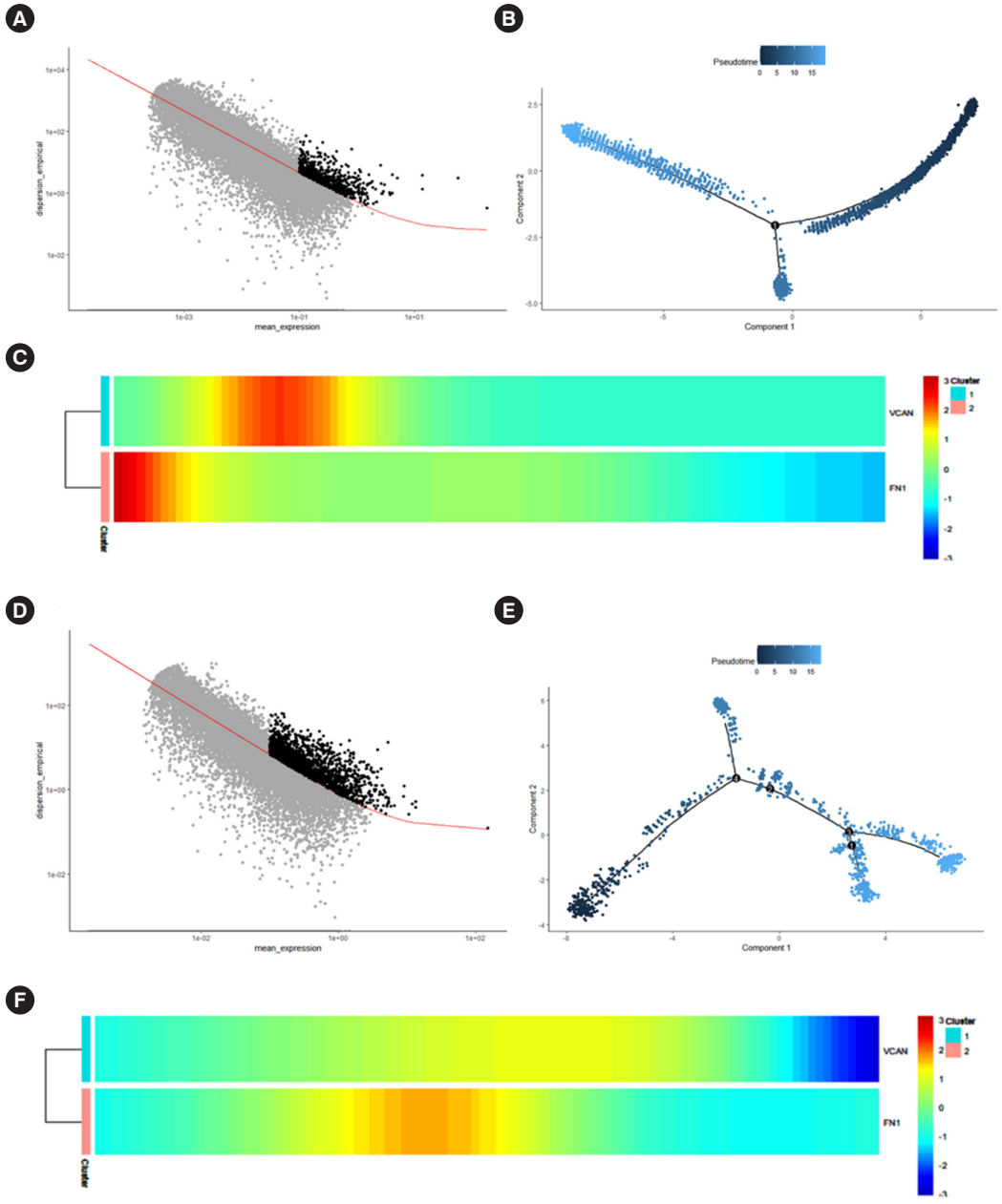
Supplementary Fig. 8. Receiver operating characteristic curve. (A, B) Internal dataset. (C, D) External test dataset. VCAN, versican; FN1, fibronectin 1; NS, not significant. [a]$P<0.0001$, [b]$P<0.001$, unpaired *t*-test.



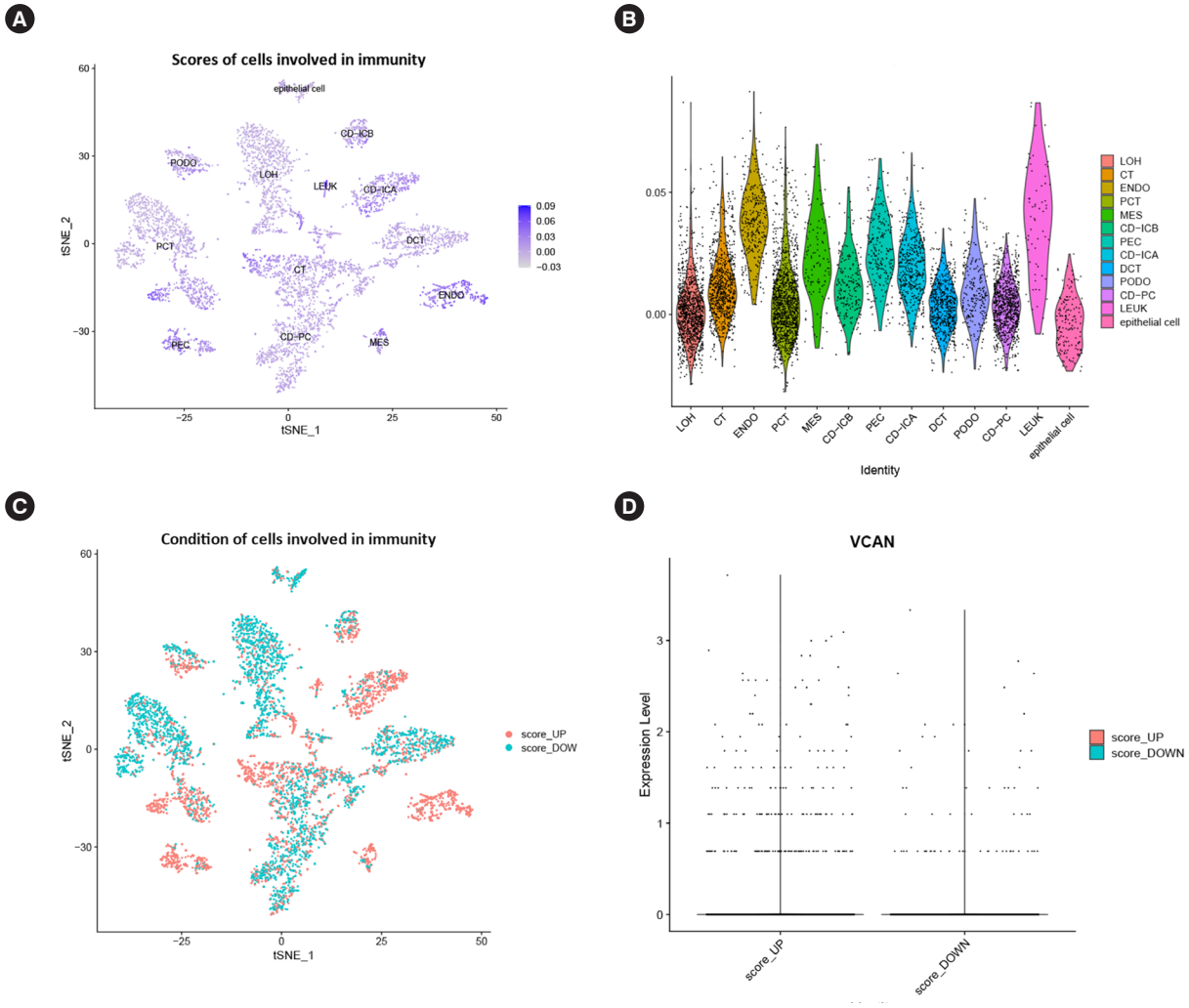
Supplementary Fig. 9. The immune infiltration pattern of versican (VCAN) in glomerulus and tubule. [a]$P<0.0001$, [b]$P<0.001$, unpaired *t*-test.



Supplementary Fig. 10. Quality control of single-cell transcriptome dataset. (A) After mitochondrial and ribosome genes were filtered in the three samples of diabetic nephropathy (DN), the DN feature performed well. (B) After mitochondrial and ribosome genes were filtered in the three samples of control, the control feature performed well. (C) Batch effect was eliminated after standardization in the DN and control groups. There were no significant differences between the samples. (D) The top 2,000 highly variable genes were shown. PC, principal component.



Supplementary Fig. 11. The pseudotime analysis. (A, B, C) Parietal epithelial cell (PEC) cells. (D, E, F) Proximal convoluted tubular (PCT) cells. VCAN, versican; FN1, fibronectin 1.



Supplementary Fig. 12. The immune score. (A) The immunity-participated degree of each cell cluster. The darker the color, the more involved in the immune reaction of diabetic kidney disease. (B) Endothelial cell (ENDO) and leukocyte (LEUK) were the principal cell types involved in the immune response, while parietal epithelial cell (PEC) and proximal convoluted tubular cell (PCT) were moderately involved in the immune response. (C) The differentiation of cell cluster involved in the immune directions. The red indicated that it was involved in the up-regulation of immune genes and the infiltration of immune cells. Blue represented involvement in the down-regulation of immune genes and suppression of immune cells. (D) Versican (VCAN) was mainly involved in the up-regulation of immune genes and the infiltration of immune cells. LOH, loop of Henle cell; CT, convoluted tubular cell; MES, mesenchymal cell; CD-ICB, collecting duct-intercalated cell type B; CD-ICA, collecting duct-intercalated cell type A; DCT, distal convoluted tubular cell; PODO, podocyte; CD-PC, collecting duct-principal cell; tSNE, t-distributed stochastic neighbor embedding.