



OPEN ACCESS

EDITED BY

Yuanzeng Min,
University of Science and Technology of
China, China

REVIEWED BY

Rong Zhou,
Tongji University, China
Qi Feng Liu,
First People's Hospital of Kunshan, China

*CORRESPONDENCE

Masha Huang
✉ martha0126@163.com
Shan Mou
✉ shan_mou@126.com

[†]These authors have contributed equally to
this work

SPECIALTY SECTION

This article was submitted to
Inflammation,
a section of the journal
Frontiers in Immunology

RECEIVED 22 December 2022

ACCEPTED 06 February 2023

PUBLISHED 17 February 2023

CITATION

Xia J, Hou Y, Cai A, Xu Y, Yang W, Huang M
and Mou S (2023) An integrated co-
expression network analysis reveals novel
genetic biomarkers for immune cell
infiltration in chronic kidney disease.
Front. Immunol. 14:1129524.
doi: 10.3389/fimmu.2023.1129524

COPYRIGHT

© 2023 Xia, Hou, Cai, Xu, Yang, Huang and
Mou. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

An integrated co-expression network analysis reveals novel genetic biomarkers for immune cell infiltration in chronic kidney disease

Jia Xia^{1†}, Yutong Hou^{2†}, Anxiang Cai¹, Yingjie Xu², Wen Yang²,
Masha Huang^{2*} and Shan Mou^{1*}

¹Department of Nephrology, Molecular Cell Lab for Kidney Disease, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, ²Department of Biochemistry and Molecular Cell Biology, Shanghai Key Laboratory for Tumor Microenvironment and Inflammation, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Background: Chronic kidney disease (CKD) is characterized by persistent damage to kidney function or structure. Progression to end-stage leads to adverse effects on multiple systems. However, owing to its complex etiology and long-term cause, the molecular basis of CKD is not completely known.

Methods: To dissect the potential important molecules during the progression, based on CKD databases from Gene Expression Omnibus, we used weighted gene co-expression network analysis (WGCNA) to identify the key genes in kidney tissues and peripheral blood mononuclear cells (PBMC). Correlation analysis of these genes with clinical relevance was evaluated based on Nephroseq. Combined with a validation cohort and receiver operating characteristic curve (ROC), we found the candidate biomarkers. The immune cell infiltration of these biomarkers was evaluated. The expression of these biomarkers was further detected in folic acid-induced nephropathy (FAN) murine model and immunohistochemical staining.

Results: In total, eight genes (*CDCP1*, *CORO1C*, *DACH1*, *GSTA4*, *MAFB*, *TCF21*, *TGFB3*, and *TGIF1*) in kidney tissue and six genes (*DDX17*, *KLF11*, *MAN1C1*, *POLR2K*, *ST14*, and *TRIM66*) in PBMC were screened from co-expression network. Correlation analysis of these genes with serum creatinine levels and estimated glomerular filtration rate from Nephroseq showed a well clinical relevance. Validation cohort and ROC identified *TCF21*, *DACH1* in kidney tissue and *DDX17* in PBMC as biomarkers for the progression of CKD. Immune cell infiltration analysis revealed that *DACH1* and *TCF21* were correlated with eosinophil, activated CD8 T cell, activated CD4 T cell, while the *DDX17* was correlated with neutrophil, type-2 T helper cell, type-1 T helper cell, mast cell, etc. FAN murine model and immunohistochemical staining confirmed that these three molecules can be used as genetic biomarkers to distinguish CKD patients

from healthy people. Moreover, the increase of TCF21 in kidney tubules might play important role in the CKD progression.

Discussion: We identified three promising genetic biomarkers which could play important roles in the progression of CKD.

KEYWORDS

biomarkers, CKD, WGCNA, TCF21, DACH1, PBMC, DDX17

Introduction

Chronic kidney disease (CKD) is a global public health issue with a prevalence of 13.4% (1). It is clinically defined as renal structure abnormalities or dysfunction (estimated glomerular filtration rate, eGFR < 60 ml/min/1.73m²) that has persisted for more than 3 months (2). CKD has a poor prognosis and can easily progress to end-stage renal disease (ESRD). Renal biopsy is an essential tool for diagnosing the pathology of CKD. Owing to the complex etiology and long-term cause, the molecular basis of CKD is not completely known, and it is difficult to predict patient responses to treatment (3, 4). With the development of omics technologies in the past decade, transcriptional bioinformatics analyses of chronic diseases have improved our understanding of molecular processes involved in CKD and have identified some novel biomarkers (5–7). In addition, integrated analyses using appropriate methods have identified some candidate genes based on the transcriptome data from such single-gene studies (8, 9). Kidney often suffers pathogenic immune responses against autoantigens in kidney or renal complication of systemic autoimmune diseases, which drive the renal disease, such as lupus nephropathy, membranous nephropathy and glomerulonephritis (10). The immune cell infiltration in kidney worsens the renal function and aggravates renal fibrosis (11). On the other hand, uremic retention solutes in blood have toxic effect on immune cells and contribute to systemic inflammatory and immune dysfunction (12). It is valuable to give new insight into the immune cell infiltration and find novel genetic biomarker for elucidating the molecular mechanism of CKD.

However, due to the invasion, frequent renal biopsy greatly increases the risk of complications. In end-stage CKD, many uremic toxins remain in the blood. Peripheral blood mononuclear cells (PBMC) respond to such environmental stimuli and undergo functional changes (13). Therefore, as a non-invasive method, blood tests can help identify key genes or pathways involved in CKD progression (14).

For bioinformatics analysis, the algorithm and dataset sample size have a major impact on analysis results. To our knowledge, no network analysis using multiple methods has been performed on CKD kidney tissue and PBMC samples to find key genes related with immune cell infiltration. Weighted gene co-expression network analysis (WGCNA) is an algorithm that can analyze gene

expression profiles to filter out disease-related modules and find hub genes. It is the most commonly used algorithm when searching for genes associated with clinical characteristics (15). In this study, we aimed to use the WGCNA algorithm in combination with differentially expressed genes (DEGs) and least absolute shrinkage and selection operator (LASSO) analysis, integrating three Gene Expression Omnibus (GEO) databases comprising PBMC and kidney tissue samples, to obtain reliable key genes related with immune cell infiltration evaluation for distinguishing the CKD patients from healthy people. In addition, we used online kidney disease databases (www.nephroseq.org), immunohistochemical (IHC) staining detection and murine CKD model for verification.

Results

Flowchart of bioinformatics analysis

The steps of our bioinformatics analysis are shown in the flowchart (Figure 1). Briefly, the CKD biomarker identification study was divided into the following main steps: (1) Peripheral blood mononuclear cell and renal tissue (TISSUE) data extraction and identification of DEGs; (2) Construction of co-expression network to identify hub genes; (3) Integrated analysis to extract key genes from DEGs and hub genes; (4) Expression validation to test credibility using validation group and external online cohort; (5) Multiple evaluation of candidate genes by online clinical database and murine experimental CKD model.

Identification of DEGs in PBMC and renal tissue from CKD datasets

First, we extracted transcriptome data from GEO datasets of CKD patients. In the clinic, blood tests and renal biopsies reveal various clinical features including abnormal kidney function, morphological changes and immunological dysfunction. In this study, GEO datasets of both PBMC and TISSUE samples were included and analyzed to explore the novel gene biomarkers in tissue and PBMC. The PBMC controls were from kidney-disease-free patients or healthy individuals and the TISSUE controls were from adjacent normal tissues of patients treated with tumor

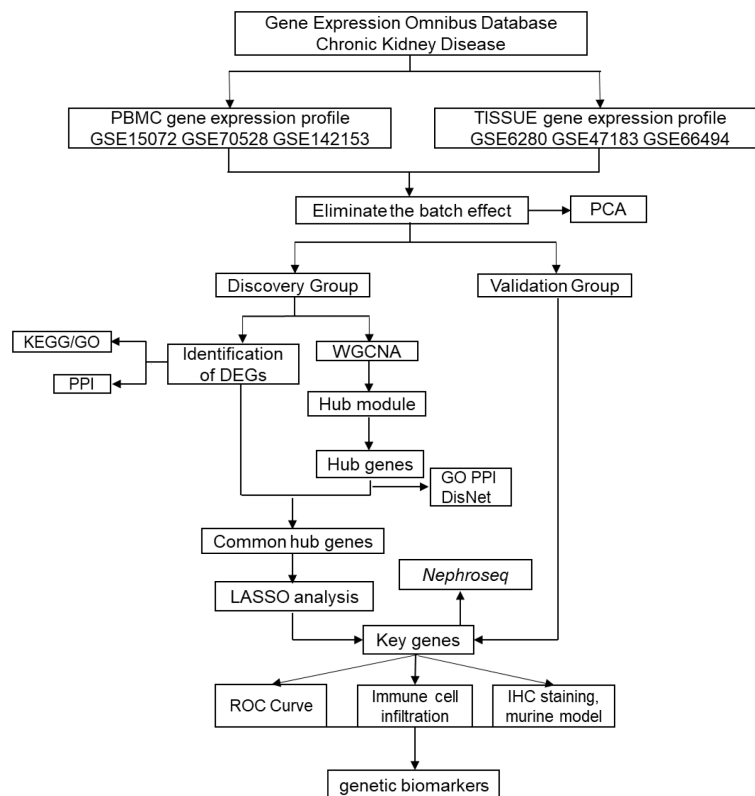


FIGURE 1

Flowchart to identify chronic kidney disease (CKD) biomarkers, including data extraction, processing and analysis.

nephrectomy. To remove the batch effect, the R package ComBat was used. Principal component analysis (PCA) showed the normalized GEO samples (GSMs) from different GEO datasets (Figure 2A). GSMs were randomly divided into discovery and validation cohort (approximately a 4:1 ratio, Table S1). The DEGs had to meet the selection criteria (adjusted $P < 0.05$). Preliminarily, we found 30 DEGs from PBMC and 142 DEGs from TISSUE samples (Table S2 and S3). Figure 2B shows the fold-change threshold ($|\log_{10}(FC)| > 0.3$ for PBMC and $|\log_{10}(FC)| > 0.52$ for TISSUE), and Figure 2C showed the ranking of DEGs. Overall, we identified many DEGs in PBMC and kidney tissues from CKD patients, as compared with levels in controls.

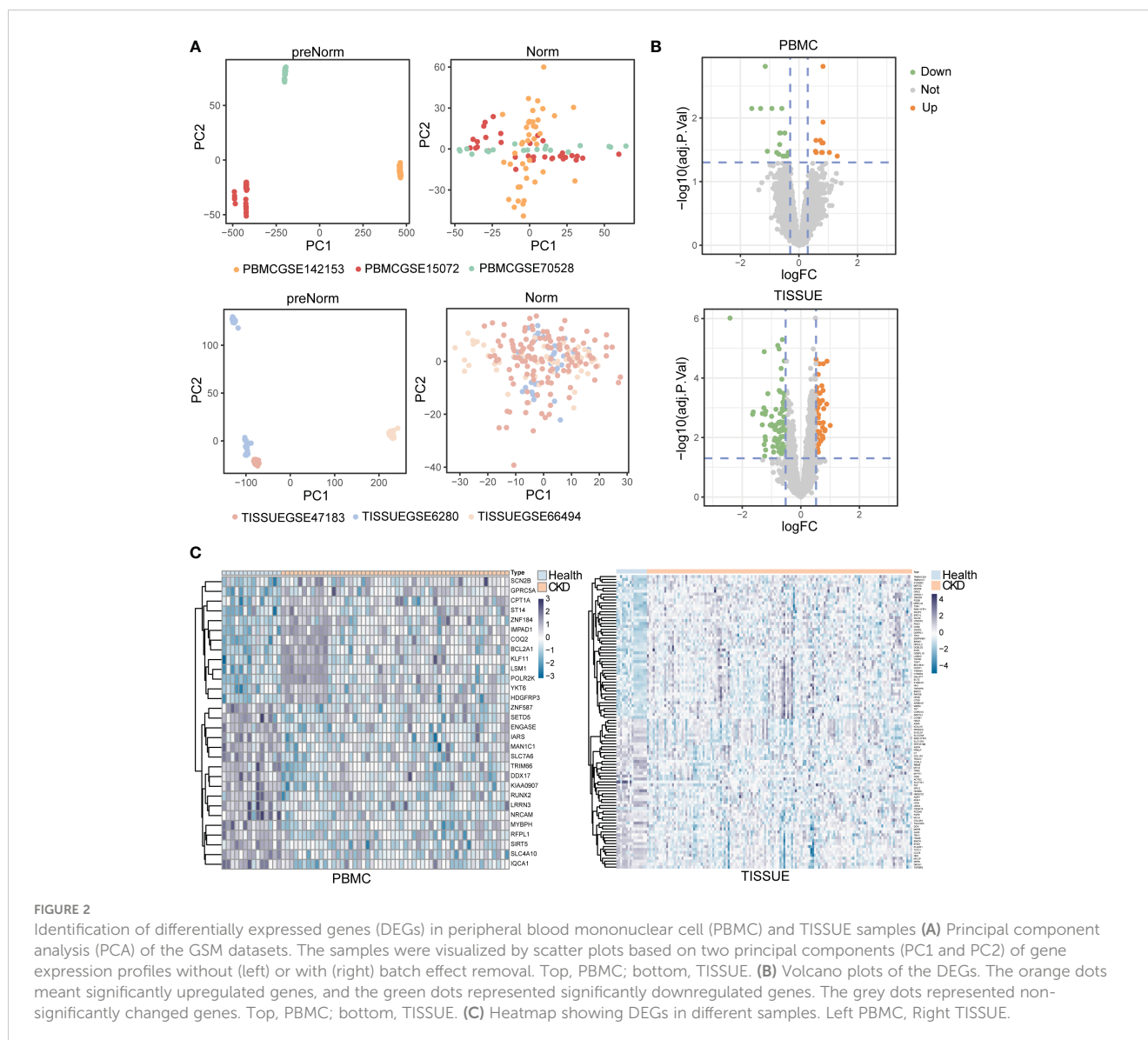
Kyoto encyclopedia of genes and genomics and gene ontology analysis of DEGs

To interpret the general biological properties of DEGs, we used STRING to create the protein–protein-interaction (PPI) network. We found a close interaction among DEGs from PMBC (7/30) and TISSUE samples (22/142) (Figures 3A, B). Furthermore, we analyzed the DEGs *via* Kyoto Encyclopedia of Genes and Genomics (KEGG) pathway and Gene Ontology (GO) biological process analysis. The top KEGG terms associated with PBMC and TISSUE samples are shown in Figures 3C, D. PBMC DEGs were enriched in transcriptional misregulation in cancer, ubiquinone and

other terpenoid-quinone biosynthesis and other glycan degradation, indicating the possible effect of the CKD status on energy metabolism and protein glycosylation. TISSUE DEGs were mainly involved in cell–ECM interactions, such as focal adhesion and ECM-receptor interaction. In addition, several signaling pathways were enriched, including TGF-beta signal pathway, AGE-RAGE signal pathway in diabetic complications, PI3K-Akt signal pathway and Hippo signaling pathway, which have been reported to play a role in CKD progression (16, 17). GO analysis results (Figures 3E, F) were similar to KEGG results. It can be seen that PBMC DEGs were involved in ketogenesis (cellular ketone metabolic process, ketone biosynthetic process), glycosylation processes (protein deglycosylation, hydrolase activity, hydrolyzing O-glycosyl compounds, etc.). In addition, TISSUE DEGs were enriched in ECM interactions and formation terms, such as collagen-containing extracellular matrix and complex of collagen trimers.

WGCNA highlights and functional analysis of CKD-associated gene co-expression modules

DEGs included only the most significantly regulated genes and other regulated genes may be missing from the transcripts. Here, we use WGCNA to robustly construct multiple gene co-expression modules. Hierarchical clustering was used to define branches of the



cluster dendrogram in multiple randomly color-coded modules (Figure 4A left, PBMC; Figure 4B, left, TISSUE). The heatmap of correlations between module eigengenes and clinical traits (CKD or not) is shown in Figure 4A (right, PBMC) and Figure 4B (right, TISSUE). The darkolivegreen1 module (Corresponding Correlation, $CC = -0.31$, $P = 0.01$) showed the highest negative correlation with CKD trait in PBMC (Figure 4A, right). The lightgreen module ($CC = -0.25$, $P = 0.002$) showed a high negative correlation with CKD trait in TISSUE samples (Figure 4B, right). The hub genes of each sample type were filtered from these modules that met the selection module member (MM) and gene significance (GS) criteria described in the methods (Table S4 and S5). Hub genes of the PBMC darkolivegreen1 module included *DDX17*, *KLF11*, *MAN1C1*, *POLR2K*, *ST14*, *TRIM66* and 121 other genes. Hub genes of the TISSUE lightgreen module include *CDCP1*, *CLIC5*, *CORO1C*, *DACH1*, *DENND2D*, *DPP6*, *GSTA4*, *MAFB*, *MAPK10*, *MYLIP*,

NEBL, *NES*, *PDLIM2*, *PFKP*, *PLA2R1*, *TCF21*, *TGFBR3* and *GIFI1*. The GO analysis of hub genes was performed on Metascape (<http://metascape.org/>). The top three GO terms of TISSUE hub genes were mesenchymal cell differentiation, actin cytoskeleton and DNA-binding transcription repressor activity, RNA polymerase II-specific (Figure 5A). DisNET analysis revealed that hub genes were closely related with glomerular disease (focal glomerulosclerosis and membranous glomerulonephritis) and renal carcinoma (Figure 5B). As for the hub genes of PBMC, the top three GO terms were leukocyte activation, positive regulation of cytokine production and positive regulation of leukocyte cell-cell adhesion (Figure 5C). Among them, The Molecular Complex Detection algorithm (MCODE) analysis enriched ribosome biogenesis and lymphocyte activation (Figure 5D). Above, the hub genes of TISSUE are closely correlated with DNA transcriptional dysregulation in kidney disease, while the hub genes of PBMC may actively affect the lymphocyte activation.

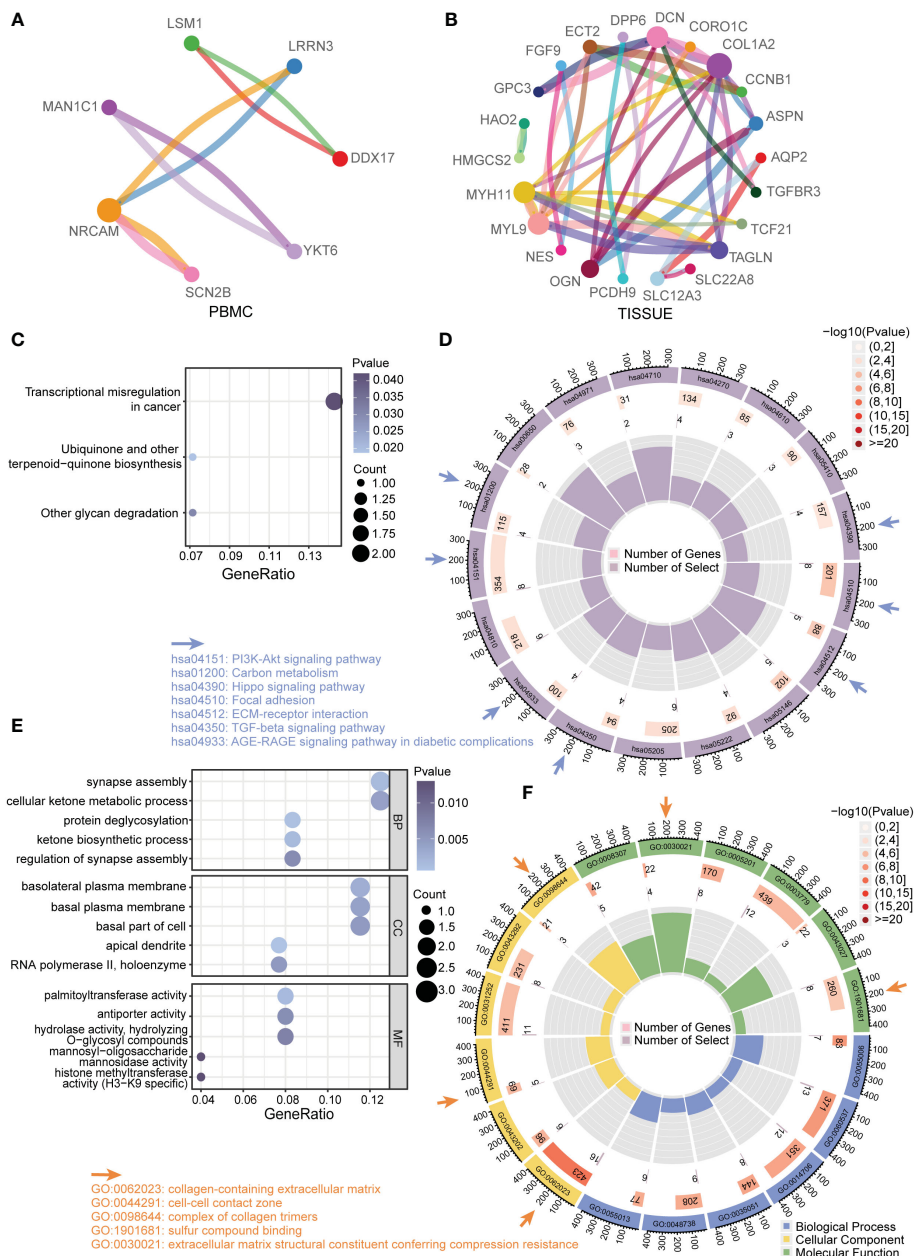


FIGURE 3

Functional enrichment analysis of DEGs (A) Protein-protein interaction (PPI) network of total DEGs from PBMC. Different colors of dots in the circle plot represented different proteins. The connectivity degree was represented by dot size. The edge width was proportional to combined score. (B) PPI network of total DEGs from TISSUE samples. (C) Enriched KEGG pathways among PBMC DEGs. The gene ratio was represented on the horizontal axis. The vertical axis indicated the KEGG signaling pathway terms, and the purple-to-blue gradually changing color indicated the change of significance from low to high. (D) Circular enrichment of KEGG pathways among TISSUE DEGs (hsa04510: Focal adhesion; hsa04512: ECM-receptor interaction; hsa05146: Amoebiasis; hsa05222: Small cell lung cancer; hsa05205: Proteoglycans in cancer; hsa04350: TGF-beta signaling pathway; hsa04933: AGE-RAGE signaling pathway in diabetic complications; hsa04810: Regulation of actin cytoskeleton; hsa04151: PI3K-Akt signaling pathway; hsa01200: Carbon metabolism; hsa00650: Butanoate metabolism; hsa04971: Gastric acid secretion; hsa04710: Circadian rhythm; hsa04270: Vascular smooth muscle contraction; hsa04610: Complement and coagulation cascades; hsa05410: Hypertrophic cardiomyopathy; hsa04390: Hippo signaling pathway). (E) Enriched GO terms among PBMC DEGs. (F) Circular enrichment of GO terms among TISSUE DEGs (GO:0055006: cardiac cell development; GO:0060537: muscle tissue development; GO:0014706: striated muscle tissue development; GO:0035051: cardiocyte differentiation; GO:0048738: cardiac muscle tissue development; GO:0055013: cardiac muscle cell development; GO:0062023: collagen-containing extracellular matrix; GO:0043202: lysosomal lumen; GO:0044291: cell-cell contact zone; GO:0031252: cell leading edge; GO:0043292: contractile fiber; GO:0098644: complex of collagen trimers; GO:0008307: structural constituent of muscle; GO:0030021: extracellular matrix structural constituent conferring compression resistance; GO:0005201: extracellular matrix structural constituent; GO:0003779: actin binding; GO:0043027: cysteine-type endopeptidase inhibitor activity involved in apoptotic process; GO:1901681: sulfur compound binding).

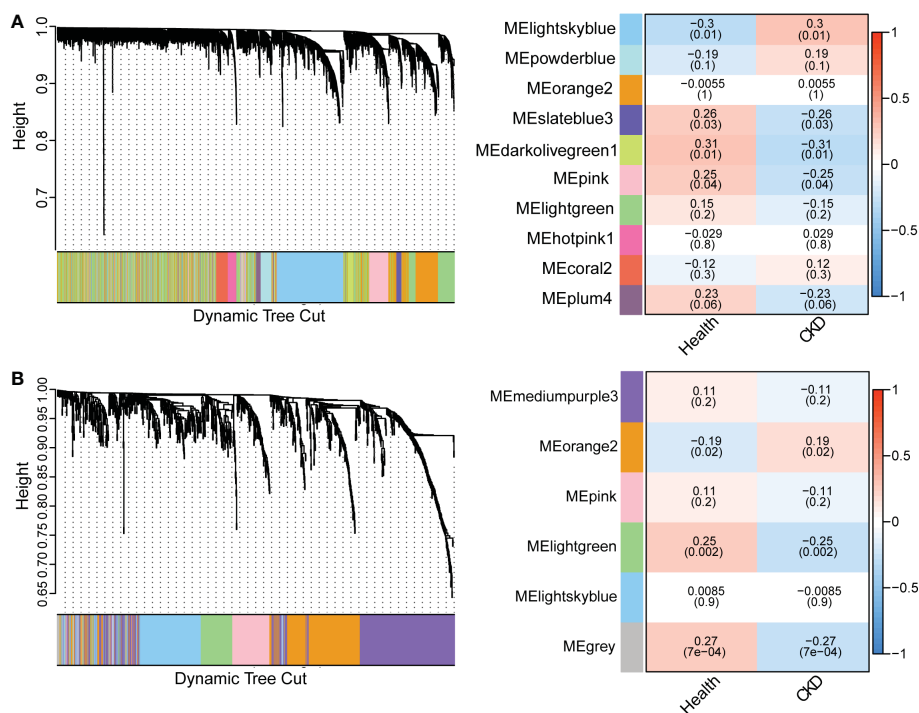


FIGURE 4

Weighted gene co-expression network analysis (WGCNA) revealing gene co-expression networks in samples from CKD patients (A) WGCNA analysis of PBMC samples. The left dendrogram represented the clusters of differentially expressed genes based on different metrics. Each branch represented one gene, and each color below branches represented one co-expression module. The right heatmap showed the correlation between gene modules and CKD. The correlation coefficient in each cube represented the correlation between gene modules and traits, which decreased from red to blue. (B) WGCNA analysis of TISSUE samples.

Identification and validation of common hub genes

Functional analysis based on hub genes did not exactly match the analysis based on DEGs. Therefore, to identify the key genes involved in CKD progression, we tried to define the common hub genes belonging to both DEGs and hub genes from WGCNA module (Figure 6A). 7 genes in PBMC and 14 genes in TISSUE were screened out. Using the LASSO regression algorithm, the common hub genes were reduced to six in PBMC (*DDX17*, *KLF11*, *MAN1C1*, *POLR2K*, *ST14* and *TRIM66*) and eight in TISSUE (*CDCP1*, *CORO1C*, *DACH1*, *GSTA4*, *MAFB*, *TCF21*, *TGFBR3* and *TGIF1*), which were identified as key genes (Figure 6B). Interestingly, *DDX17* and *MAN1C1* from PBMC hub genes and *CORO1C*, *TCF21* and *TGFBR3* from TISSUE hub genes were among the top-ranked genes in the PPI network. To examine whether the expression of these genes influenced CKD progression, we obtained clinical parameters (eGFR and serum creatinine, SCr) of these 8 tissue genes from www.nephroseq.org. Woroniecka Diabetes Glom database was used to evaluate the eGFR status, while the Ju CKD Glom was used to evaluate the SCr level. Consistently, the low-expressed genes, *GSTA4*, *MAFB*, *TGFBR3*, *DACH1* and *TCF21* expression was positively related with eGFR and negatively related with SCr. The high-expressed genes, *CDCP1*, *CORO1C* and *TGIF1* expression was negatively related with eGFR

and positively related with SCr (Figures 7A–H). The expression and significance of these hub genes were highlighted in Figure 7I.

To determine the association between these genes and CKD status, we examined these 14 key genes in the validation cohort. Among them, the expression of *DDX17* in PBMC, *DACH1* and *TCF21* in TISSUE samples showed similar characteristics with those in the discovery cohort (Figures 8A–C). Consistently, *DACH1* and *TCF21* mRNA level were also decreased in the Woroniecka diabetic nephropathy cohort (Figure 8D, *DACH1*, $P < 0.0001$; Figure 8E, *TCF21*, $P < 0.0001$). Moreover, as shown in Figures 8F–H, receiver operating characteristic curve (ROC) was used to investigate whether these three key genes could discriminate between healthy and CKD samples. The classification accuracy (area under the ROC curve, AUC) of these three key genes (*DDX17*, *DACH1*, *TCF21*) was 0.828, 0.825 and 0.981 in the discovery cohort and 0.885, 0.838 and 0.949 in the validation cohort, respectively, showing strong ability to discriminate between CKD and healthy individuals. By reason of the foregoing, we screened out *DDX17*, *DACH1* and *TCF21* as genetic biomarkers. The correlation between parameters of renal function and the expression of genetic biomarkers suggested they may play a renoprotective role. Briefly, the renal *DACH1* and *TCF21* expression was positively correlated with eGFR in CKD patients (*DACH1*, $P = 7.15e-5$, $R^2 = 0.554$; *TCF21*, $P = 5.34e-5$, $R^2 = 0.566$), whereas it was negatively correlated with SCr levels (*DACH1*, $P = 0.0022$, $R^2 = 0.366$; *TCF21*, $P = 5.04e-7$, $R^2 = 0.707$).

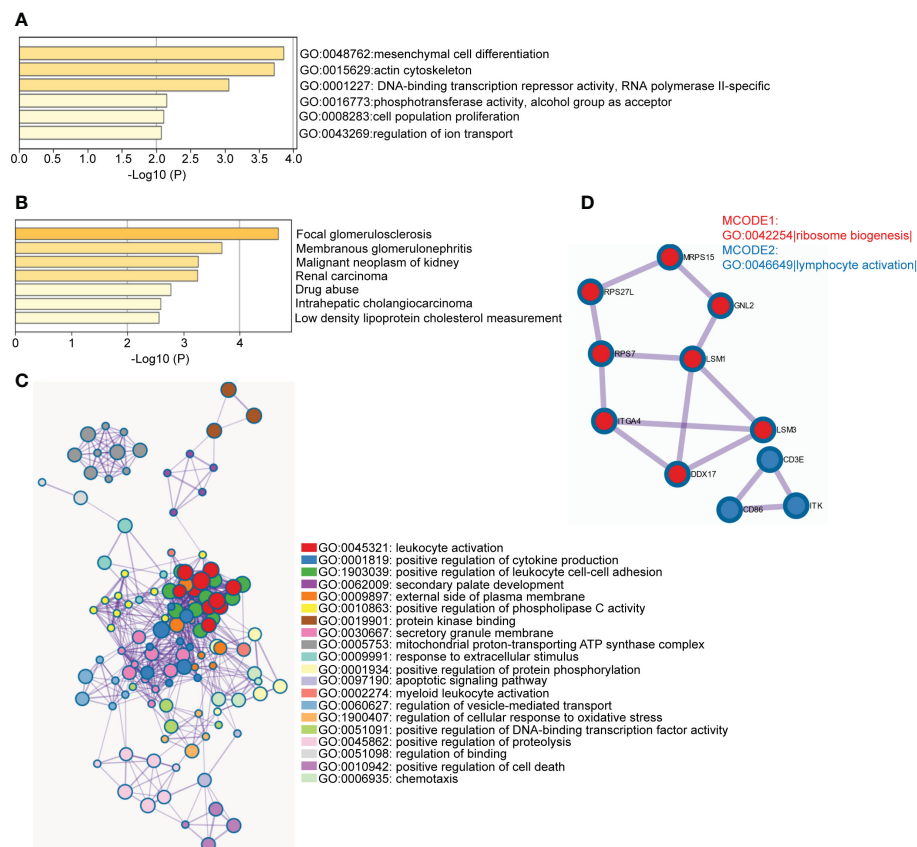


FIGURE 5

Functional enrichment analysis of hub genes in disease-related module (A) Enriched GO terms among TISSUE hub genes. The horizontal axis represented P-value of GO terms in log10 calculated on Metascape by default parameter. (B) Enriched DisGeNET terms among TISSUE hub genes. The horizontal axis represented P-value of GO terms in log10 calculated on Metascape. (C) Network of representative GO terms among PBMC hub genes. The clusters were calculated and visualized with Cytoscape using Metascape online platform by default parameter. The color of the node represented its cluster identity. One GO term from each cluster was selected to be shown as label. (D) Top MCODE terms of PBMC hub genes. All PPI among PBMC hub genes formed a network. The Molecular Complex Detection algorithm (MCODE) was used to identify the connected network components. The network was analyzed by GO enrichment to extract "biological meanings". One GO term was labelled to represent the MCODE (GO: 0042254: Ribosome biogenesis; GO: 0046649: lymphocyte activation).

Immune cell infiltration analysis of genetic biomarkers

The TISSUE genetic biomarkers *DACH1* and *TCF21* are tumor suppressor genes. They are low expressed in the Kidney Renal Clear Cell Carcinoma (KIRC) (Figures S1A, B). The KIRC patients with high *DACH1* and *TCF21* expression has a better prognosis (Figures S1C, D). On the other hands, PBMC hub genes are closely related with lymphocyte activation (leukocyte activation, positive regulation of cytokine production, positive regulation of leukocyte cell-cell adhesion, etc.) (Figure S1E). The chronic kidney diseases and kidney malignant tumors have been demonstrated to be linked to a more severe immune cell infiltration. We evaluate the immune cell infiltration of these genetic biomarkers. Immune cell infiltration analysis revealed that *DDX17* in PBMC was found to be correlated with neutrophil, type 1 helper cell, type 2 helper cell and mast cell (Figure 8I). The TISSUE *DACH1* and *TCF21* were both found to be correlated with eosinophil, activated CD8 T cell and activated CD4 T cell (Figure 8).

CKD murine model and immunohistochemical validation of genetic biomarker expression

To mimic renal dysfunction and tubulointerstitial fibrosis status during CKD, we developed a folic acid (FA)-induced CKD murine model. The results showed that SCr levels were increased significantly after a 3-day FA treatment (Figure 9A). At day 28, *dach1* mRNA levels in the kidneys of mice treated with FA were significantly decreased (Figure 9B, $P < 0.001$) and negatively correlated with SCr levels (Figure 9C, $P = 0.0264$, $R^2 = 0.4800$). Conversely, *tcf21* mRNA levels were significantly increased (Figure 9D, $P < 0.0001$) and positively correlated with SCr levels (Figure 9E, $P = 0.0128$, $R^2 = 0.5595$). In isolated PBMC, *ddx17* mRNA levels were significantly downregulated in the FA-treated kidney (Figure 9F, $P < 0.01$). These results showed that in the FA-induced CKD mouse model, *dach1* levels in tissue and *ddx17* levels in PBMC were in agreement with the changes in clinical samples. Interestingly, *tcf21* was significantly increased in the FA-induced

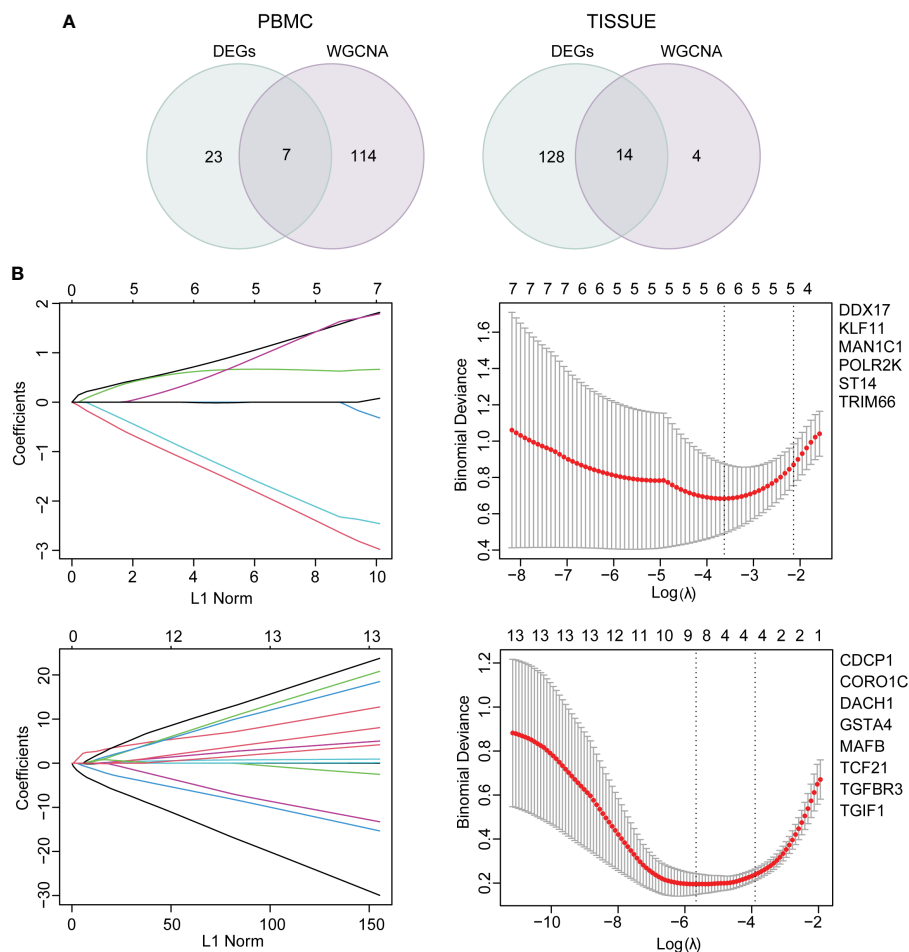


FIGURE 6

Common hub gene selection and least absolute shrinkage and selection operator (LASSO) analysis (A) The common hub genes shared between DEGs and hub genes were visualized in a Venn diagram. Left, PBMC; right, TISSUE. (B) The number of factors was determined by LASSO analysis. The procedure of LASSO Cox model fitting was shown in left panel. One curve represented a gene. The coefficient of each gene against the LC-norm was plotted with the lambda change. L1-norm represented the total absolute value of non-zero coefficients. A coefficient profile generated against the log (lambda) sequence was shown in the right panel. Continuous upright lines were the partial likelihood deviance \pm SE; The optimal values and gene symbols were depicted, based on the minimum criteria (lambda.min, left vertical dotted line) and 1-SE criteria (lambda.1se, right vertical dotted line). Top, PBMC; bottom, TISSUE. SE, standard error.

CKD mouse model, opposite to the transcriptome change in clinical samples. Further, IHC staining showed the protein expression of DACH1 and TCF21 in kidney of CKD patients and normal control. Both DACH1 and TCF21 were mainly expressed in the nucleus of glomerulus and renal tubule cells. DACH1 expression decreased in glomerulus and renal tubule of membranous nephropathy (MN) (Figures 9G, H, $P < 0.05$), while TCF21 expression increased in MN, especially in the renal tubule cells (Figures 9I, J, $P < 0.05$).

Discussion

CKD is a type of kidney disease in which kidney function deteriorates and/or the structure is abnormal. Owing to its heterogeneity in etiology, the mechanism underlying its occurrence and progression is still not well understood. It is known that the TGF- β /Smad pathway is usually activated in CKD (18). However, damage to kidney function causes uremic retention

solutes to remain in the body, resulting in adverse effects such as inflammation, immune dysfunction and oxidative stress (19–21), but the biological processes involved are not yet fully understood. In this study, we analyzed the gene expression profile of kidney tissues and PBMC from CKD patients, using various bioinformatics methods to find novel common genes or signaling pathways.

In the first-round screening, based on pathway enrichment analysis, signal molecules such as TGF- β , AGE-RAGE, PI3K-AKT and HIPPO play important roles in CKD progression and cross-talk with other molecules (16, 17, 22, 23). Based on GO analysis, the major functional terms are mainly enriched in cell adhesion and cell-extracellular matrix interactions, among others, which are involved in the critical pathological processes associated with CKD (24).

It is currently believed that WGCNA is better at identifying internal functional connections among regulated genes than DEG analysis. Wang et al. discovered that high expression levels of *CEBPZ*, *IFI16*, *LYAR*, *BRIX1*, *BMS1* and *DDX18* in the kidneys are

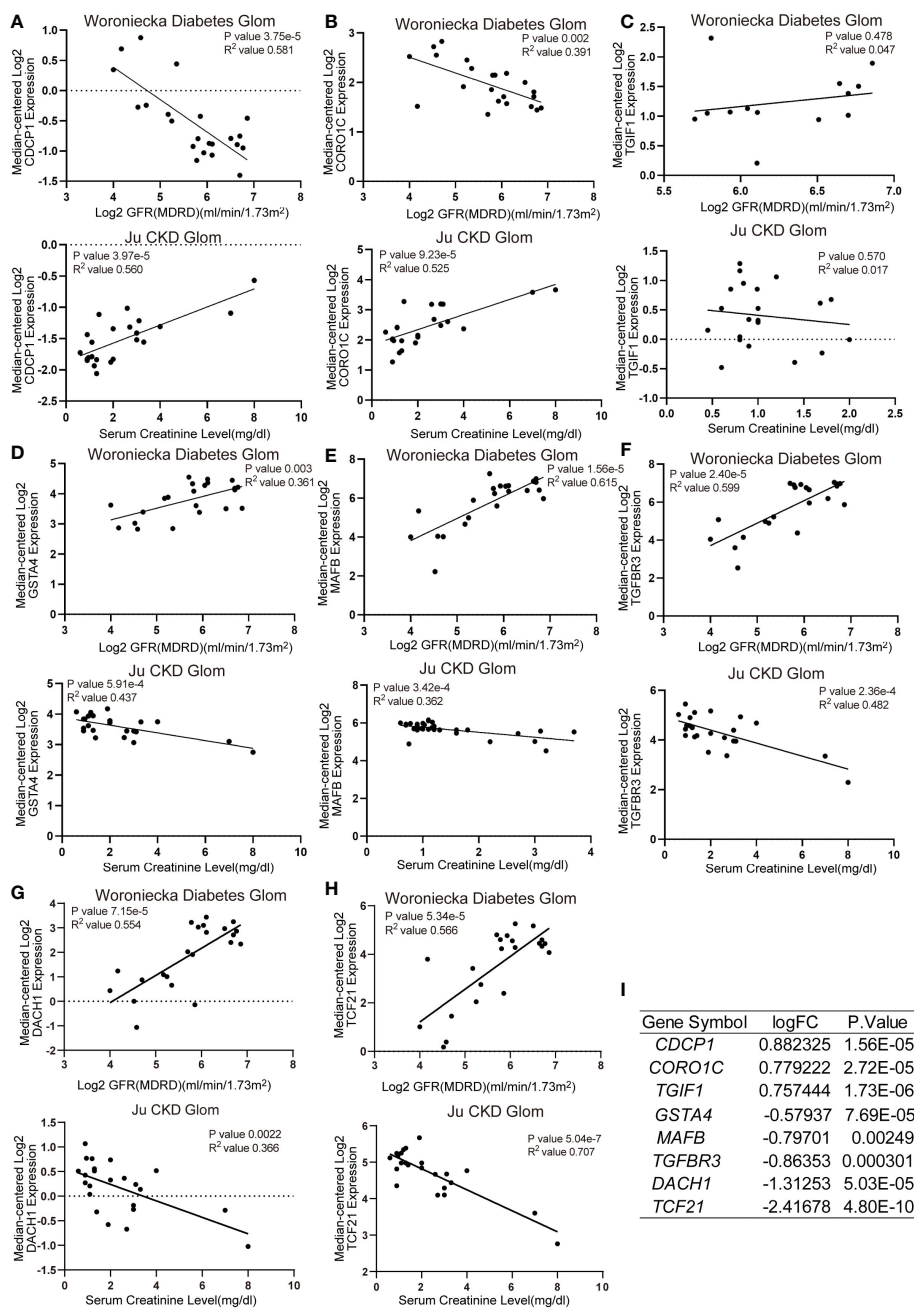


FIGURE 7 Correlation analysis of TISSUE common hub genes and CKD clinical parameters (A–H) The correlation of TISSUE genetic biomarker mRNA levels with estimated glomerular filtration rate (eGFR) in the Woroniecka Diabetes Glom Cohort (Top) or serum creatinine (SCr) level in the Ju CKD Glom Cohort (Bottom). (A) *CDCP1*; (B) *CORO1C*; (C) *TGIF1*; (D) *GSTA4*; (E) *MAFB*; (F) *TGFBR3*; (G) *DACH1*; (H) *TCF21*. Data were extracted from www.nephroseq.org. (I) The gene expression level of common hub genes in the discovery cohort.

potential key markers of CKD occurrence and progression (8). However, few studies combining WGCNA and clinical parameters using both kidney tissues and PBMC have been performed. In our study, 6 key genes (*KLF11*, *MAN1C1*, *POLR2K*, *ST14*, *TRIM66*, *DDX17*) in PBMC samples and 8 key genes (*CDCP1*, *CORO1C*, *GSTA4*, *MAFB*, *TGFBR3*, *TGIF1*, *TCF21* and *DACH1*) in TISSUE samples were identified by WGCNA. Several of these genes in tissue have been reported to contribute to the pathology and molecular changes in CKD. *MAFB* is a transcription factor that mediates renal

tubule development and macrophage maturation (25). *TGIF1* can bind to the MH1 domain of SMAD to inhibit TGF-β pathway activation (26). Most of them have a well correlation with SCr and eGFR in the CKD database. Among the PBMC common hub genes, several genes have been reported to be associated with kidney disease or hematologic disorder. *KLF11* is a Krüppel-type zinc finger protein whose deficiency enhances chemokine generation and fibrosis in murine unilateral ureteral obstruction (27) and highly induced by TGF-β (28). It has been reported that *KLF11* inhibits the activity of

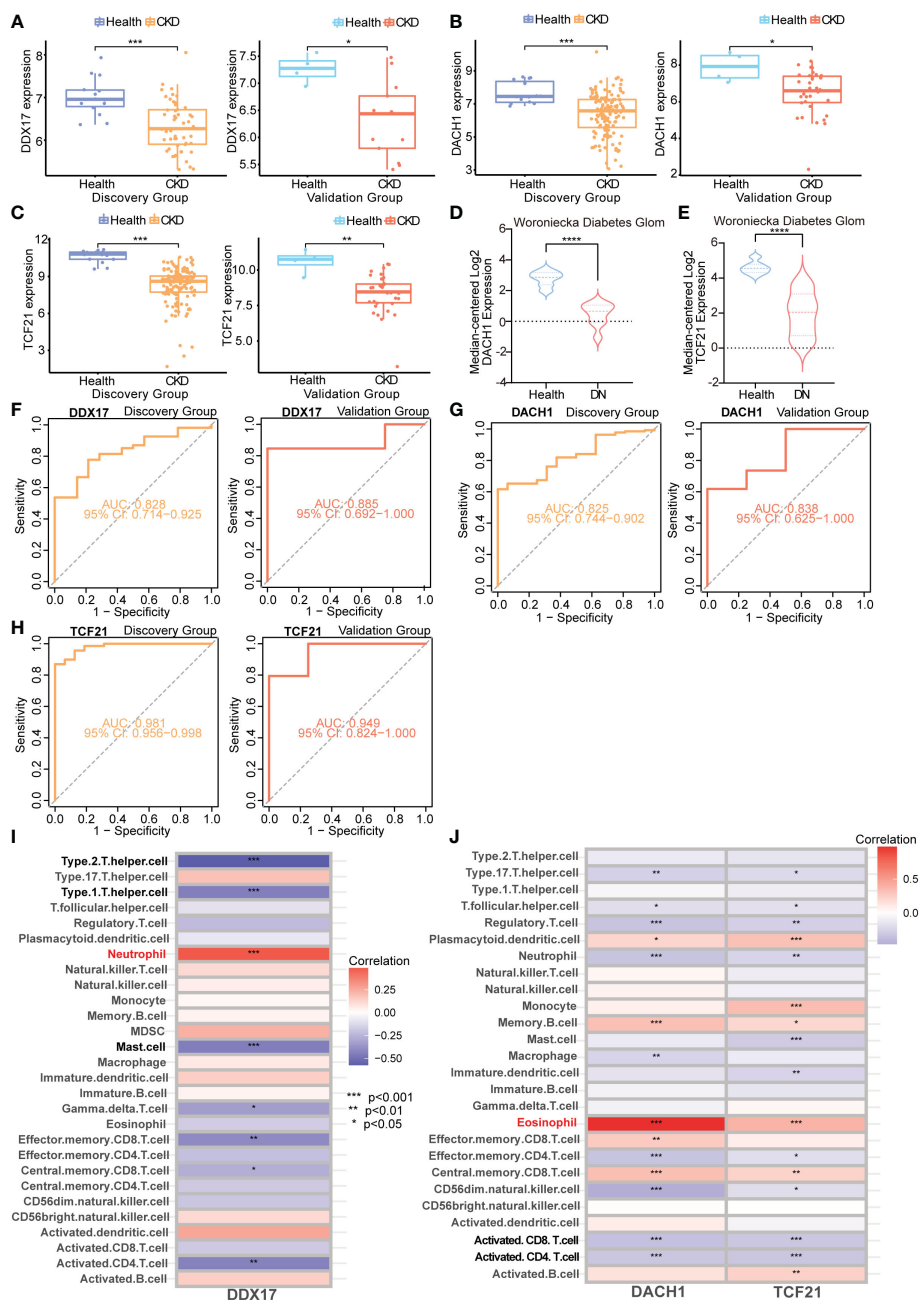


FIGURE 8
 Validation and immune infiltration evaluation of CKD genetic biomarkers from different samples (A–C) Box plots representing expression level of CKD genetic biomarkers in the discovery and validation cohorts. (D) mRNA level of *DACH1* in the Woroniecka Diabetes Glom Cohort. (E) mRNA level of *TCF21* in the Woroniecka Diabetes Glom Cohort. (F–H) Receiver operating characteristic (ROC) curve for the discovery and validation cohorts. (A, F) *DDX17* (PBMC); (B, G) *DACH1* (TISSUE); (C, H) *TCF21* (TISSUE). (I) Correlation heatmap demonstrating the relationship between *DDX17* (PBMC) and immune cells infiltration. (J) Correlation heatmap demonstrating the relationship between *DACH1* and *TCF21* (TISSUE) and immune cells infiltration. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

SMAD7 and enhances TGF- β pathway activation (29), through which KLF11 may influence the lymphocyte function. Besides, the highly upregulation of *ST14* promotes cancer cell invasion via imbalanced matriptase pericellular proteolysis (30, 31). The upregulation of *ST14* in PBMC may promote the inflammatory activation of endothelial cells in blood vessel, which is a common uremia-related complication. Thus, these molecules should be studied intensively. Finally, by examining the validation cohort

and ROC curve, we identified *TCF21* and *DACH1* in TISSUE samples and *DDX17* in PBMC as potential biomarkers for CKD. The correlation of *DACH1* and *TCF21* with clinical parameters (eGFR and SCr) in CKD patients suggested that these molecules are potential renoprotective biomarkers in kidney tissues.

Dachshund family transcription factor 1 (*DACH1*) has been previously described as a tumor suppressor that can inhibit breast cancer invasion and metastasis (32). In the past several years, many

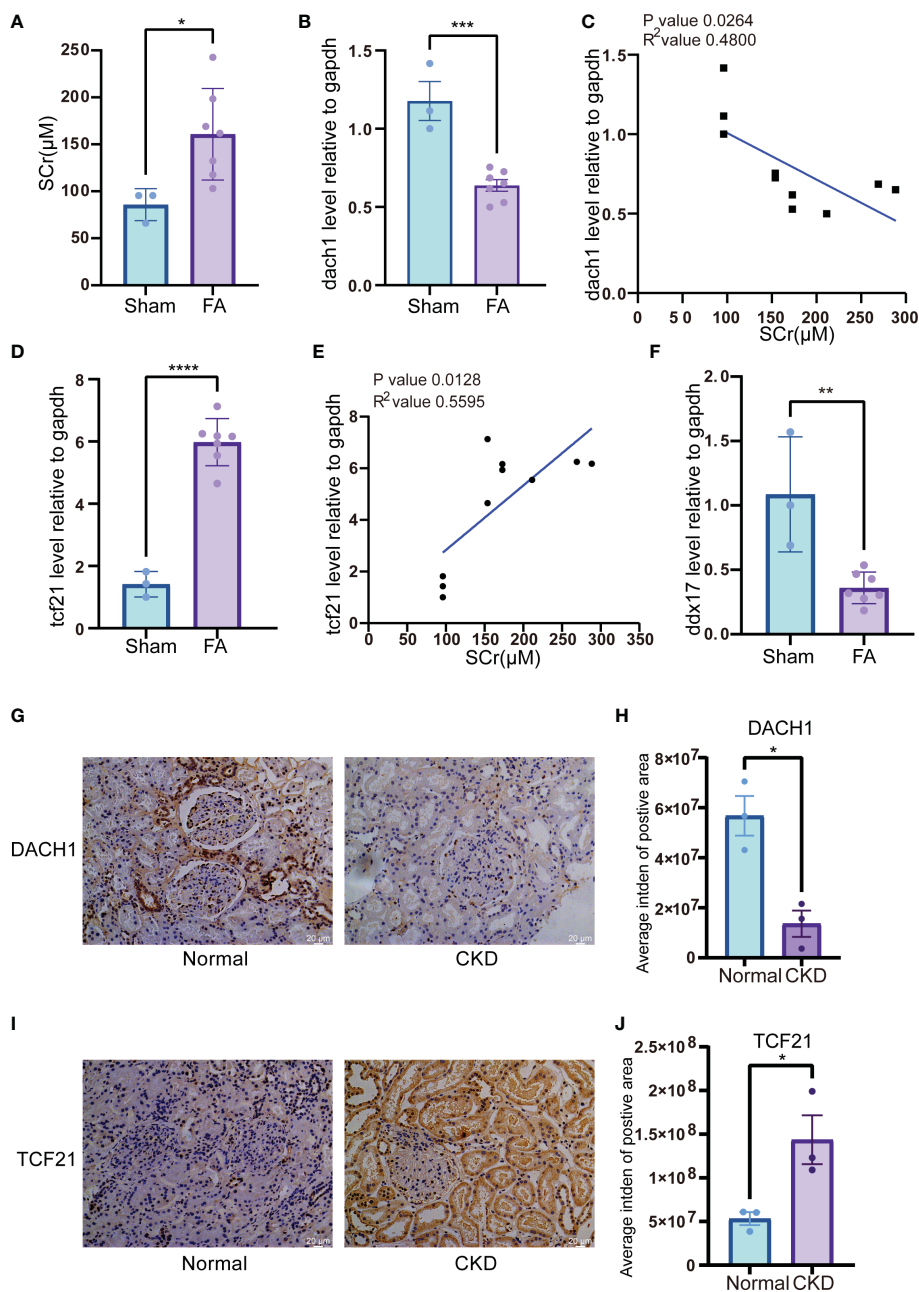


FIGURE 9

Expression of genetic biomarkers in folic acid (FA)-induced CKD murine model and membranous nephropathy (MN) patient's biopsy (A) SCr level in FA-induced CKD murine model. Blood serum was harvested 3 days after FA injection. (B) mRNA level of *dach1* in kidneys from FA-induced CKD murine model. (C) Correlation between *dach1* mRNA level in kidney tissue and SCr from FA-induced CKD murine model. (D) mRNA level of *tcf21* in kidney tissues from FA-induced CKD murine model. (E) Correlation between *tcf21* mRNA level in kidney tissues and SCr from FA-induced CKD murine model. (F) mRNA level of *ddx17* in PBMC from FA-induced CKD murine model. (G) IHC staining of DACH1 in normal kidney tissues and CKD tissues. The representative pictures of CKD were from the membranous nephropathy (MN) patient's biopsy. (H) Quantitative results of DACH1 expression in panel (G). The expression level was calculated by average integrated density (Intden) of positive area. (I) IHC staining of TCF21 in normal kidney tissues and CKD tissues. (J) Quantitative results of TCF21 expression in (I). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

studies have indicated that DACH1 is a renal-protective molecule. GWAS analysis showed that loss of DACH1 function was a susceptibility factor for renal fibrosis (33), and DACH1 can protect against podocyte damage in diabetic nephropathy model mice (34). Moreover, tubule-specific *DACH1*-knockout mice were more susceptible to renal damage and fibrosis in a FA-induced nephropathy model (33). In our study, this gene was identified

through transcriptome bioinformatics analysis and confirmed based on clinical parameters, FA-induced nephropathy model and the IHC staining in kidney tissues of CKD patient. This agreement with reported studies showed that our bioinformatics analysis was reliable.

Transcription factor 21 (TCF21) is a transcription factor that plays an important role in the differentiation of mesenchymal cells

and the development and maturation of gonads, muscle, kidney and other organs (35). *TCF21*-knockout mice developed kidney dysplasia at the embryonic stage and die after birth (36). Mice with podocyte-specific *TCF21*-knockout spontaneously developed proteinuria and exhibit FSGS (focal segmental glomerulosclerosis) lesions (37). Our bioinformatic analysis showed that *TCF21* mRNA levels were decreased in CKD samples and positively correlated with eGFR. However, in our FA-induced nephropathy murine model, *TCF21* mRNA levels were elevated compared to control group. The *TCF21* staining signal was mainly in the nucleus in the healthy kidney, while in our mild case of membranous nephropathy, the signal increased and appeared in the cytoplasm and brush border of renal tubules. Such inconsistency may be caused by the following reasons. First, the protective role of *TCF21* was mainly reported in the podocytes, not in the kidney tubules. This finding is supported by analyzing the tissue datasets of CKD cohort mainly from glomerular transcriptome (GSE47183 with 100 glomerular transcriptome samples and GSE66494 with 41 whole kidney transcriptome samples). Combining the two datasets may bias the display of key genes in the glomeruli during CKD progression. Recent study identified *TCF21* as a deactivation factor of fibrogenic HSCs in liver fibrosis (38). It might be a nephroprotective in tubulointerstitial fibrosis. In our mouse model, high dose of folic acid mainly destroys the tubules and leads to the consequent tubulointerstitial fibrosis. Therefore, in the acute injury stage and early CKD stage of our mouse model, the upregulated *TCF21* in tubule could play a protective role against kidney injury. In fact, *TCF21* protein levels were also elevated in the early stage of diabetic nephropathy in model mice (39). Second, *TCF21* was normally expressed in nuclei of podocytes and highly accumulate in both nuclei and cytoplasm of the injured podocytes in glomerular diseases, even was detected in urine (40). The severe injury of podocytes in CKD might lead to the loss of this cell population which lead to the reduction of total *TCF21* mRNA in kidney tissue. It is in line with our bioinformatics analysis and previous reports.

Among key genes from PBMC, we identified *DDX17*, which an ATP-dependent RNA helicase that is a coactivator of DNA-regulated transcription factors and is involved in mRNA transcription, splicing and maturation (41). *DDX17* is an immune-related gene defined in immunology database and analysis portal (ImmPort). In our study, it was expressed at low levels in the PBMC of CKD patients from the GEO database and in the PBMC of FA-induced CKD mouse model. In PBMC, *DDX17* plays an important role in innate immunity against virus invasion (42–44). *DDX17* is an essential mediator of sterile NLR4 inflammasome activation (45). Given the fact that the uremia-associated immune deficiency is a well-known complication of CKD and it increase the risk of virus-infection and virus-associated cancers (46), the low *DDX17* level in PBMC might associated with CKD progression.

As for the interrelationship of these three biomarkers. *DDX5/* *DDX17* complex can co-activate or co-repress transcription factor transcription. In kidney tissue, the expression of *DDX17* was also decreased in the Woroniecka Diabetes Glom database, which was in line with the *DACH1* and *TCF21* expression. There is no report on the relationship between these three genes in the occurrence and development of CKD. Mechanically, we speculated that the low

expression of *DDX17* might further down-regulate the activity of *TCF21* and *DACH1*, worsen the glomerular sclerosis and renal interstitial fibrosis. Given its critical role in innate immunity against virus invasion, the low *DDX17* level in PBMC may aggravate immune deficiency in ESRD (42–45), and may result in chronic inflammation and increased oxidative stress to exacerbate kidney injury and loss of renal function (47).

During the course of CKD, in addition to the abnormal activation of some signaling pathways such as TGF- β /Smad and PI3K-Akt, some central molecules are also lacking. The low activity of some key molecules that regulate kidney differentiation and development might lead to dysfunction. *DACH1* and *TCF21* are both important transcription factors for kidney development and maturation. Some studies have demonstrated the association between their expression and CKD and illuminated their glomerular-specific roles in kidney disease (33, 34, 36, 37, 48), but the exact mechanisms underlying their dysregulation in CKD are still elusive. In our study, GO analysis of DEGs and hub genes highlighted DNA transcriptional activity in CKD kidney tissues. Besides, the evaluation of immune cell infiltration showed a positive correlation of *DACH1*, *TCF21* expression with eosinophil, activated CD8 T cell and activated CD4 T cell. This analysis suggested that low expression of *DACH1* and *TCF21* may lead to pathological dysregulation through aberrant immune cell infiltration. Folic acid induced experimental nephropathy models undergo progression from acute kidney injury (AKI) to CKD with initial damage to proximal tubules, significant alterations to these two transcription factors suggest their overall impact on glomeruli and tubules.

Our study also had some limitations. For example, the included database did not distinguish between the different pathological types of CKD to find pathology-specific biomarkers. In addition, we mainly analyzed the mRNA levels of genes, not the protein level. In addition, many renal proteins can be detected noninvasively in urine. In a subsequent study, we hope to detect the protein levels of key genes under investigation.

Materials and methods

Clinical samples

In this study, we adopt three samples from healthy adjacent kidney tissues of individuals who were performed tumor nephrectomy, whereas other three samples as CKD group whose tissues were taken from CKD patients' biopsy. The pathologist confirmed these biopsy samples as membranous nephropathy (MN). The patients in this study consented under the ethics committee review of Renji Hospital affiliated to Shanghai Jiao Tong University, School of Medicine.

Folic acid-induced nephropathy murine model

C57bl/6J mice were obtained from Shanghai Jiao Tong University, School of Medicine. All mice were maintained with a

12-hour light-dark cycle and given food and water ad libitum. Procedures were performed in accordance with guidelines approved by the ethical committee of animal experiments, Shanghai Jiao Tong University, School of Medicine. Folic acid was purchased from Sangon Biotech and dissolved in 300 mM NaHCO₃. 12-week-old mice were intraperitoneally injected with FA (250 mg/kg) or NaHCO₃. Mice were euthanized and sacrificed 28 days after FA injection. Whole blood was collected in White's buffer (pH 6.4). Ficoll-Paque (GE healthcare) was used to separate the PBMC according to the manufacturer's instructions. Kidneys were collected and stored at -80°C for further study.

Microarray data collection from GEO database

We downloaded 6 gene expression datasets (GSE142153, GSE15072, GSE70528, GSE47183, GSE6280, GSE66494) from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). According to labeling information in the platform, probes were converted into their corresponding gene symbols. The GSE142153 dataset contained 7 CKD PBMC samples and 10 control PBMC samples. The GSE15072 dataset contained 26 CKD PBMC samples and 8 control PBMC samples. The GSE70528 dataset contained 11 CKD PBMC samples. The GSE47183 dataset contained 122 CKD kidney samples. The GSE6280 dataset contained 12 control kidney samples. The GSE66494 dataset contained 53 CKD kidney samples and 8 control kidney samples. In our study, the discovery cohorts from the GSE142153, GSE15072 and GSE70528 datasets were used to build co-expression networks and identify the predominant genes in the CKD PBMC samples. The discovery cohorts from the GSE47183, GSE6280 and GSE66494 datasets were used to build co-expression networks and identify the main genes associated with CKD kidney samples. The assignment of each sample to discovery or validation cohorts was shown in [Table S1](#). The procurement and application for all data were in accordance with the guidelines and principles of the GEO databases.

Data preprocessing

For both sample types (PBMC and TISSUE), each dataset was combined into a data matrix. ComBat from the R package sva was used to account for batch effects ([49](#), [50](#)). Whether the batch effect was removed was evaluated by PCA. The prcomp function in R was used to perform PCA. Data visualization was performed using the R packages ggplot2 (<https://ggplot2.tidyverse.org>) and RColorBrewer (<https://cran.r-project.org/web/packages/RColorBrewer/index.html>) unless otherwise noted.

DEGs identification

The R packages limma and edgeR were used to identify the DEGs between CKD and normal samples, respectively ([51](#), [52](#)). Genes with

an adjusted $P < 0.05$ and $|\log_{10}(FC)| > 0.3$ were selected as DEGs in PBMC. Genes with an adjusted $P < 0.05$ and $|\log_{10}(FC)| > 0.52$ were selected as DEGs in TISSUE. The R package pheatmap was used to generate heatmaps ([53](#)).

Protein–protein-interaction network construction

STRING database (<http://string-db.org>) was used to build the PPI network, where a combined score > 0.4 was considered statistically significant ([54](#)). The CytoHubba Cytoscape plugin was used to calculate the nodes using the connectivity degree method ([55](#)). We then used the netVisual_circle function in the R package CellChat to visualize the PPI network ([56](#)).

Functional enrichment analysis

We used the R packages ClusterProfiler and org.Hs.eg.db for Kyoto Encyclopedia of Genes and Genomics (KEGG) and Gene Ontology (GO) analysis ([57](#)). KEGG pathways or GO function terms with $P < 0.05$ were considered statistically significant. The R package circlize was used to create circos plots ([58](#)).

Weighted gene co-expression network analysis

The R package WGCNA was used to construct the co-expression network based on discovery cohorts ([15](#)). To merge modules that might be similar, 0.25 was defined as the cut-off height threshold. The phenotypes (CKD) were inputted into the co-expression network and the parameters modulus characteristic gene (ME), MM and GS were calculated. ME represented the important part in the PCA of each gene module and MM represented the connection between modules and genes. Correlation coefficients ≥ 0.50 and P -values < 0.05 were considered indicative of key modules for PBMC. Correlation coefficients ≥ 0.60 and P -values < 0.05 were considered indicative of key modules for kidney tissue. In the modular-trait correlation analysis, genes with high hub modularity were considered as hub genes. Hub genes of PBMC met the absolute values of $GS > 0.20$ and $MM > 0.50$. Hub genes of kidney tissue met the absolute values of $GS > 0.20$ and $MM > 0.60$.

Common hub gene selection and LASSO analysis

Common hub genes were defined as the overlap between DEGs and WGCNA hub genes. Venn diagrams were prepared using the R package venn. LASSO was a regression analysis algorithm that performs both gene selection and classification ([59](#)). To select hub genes which were credibly associated with CKD, a logistic LASSO

regression model was constructed based on common hub genes by R package glmnet. 10-fold cross-validation was performed for tuning parameter selection, and the partial likelihood deviance met the minimum criteria.

Evaluation of immune cell infiltration

A set of genes that mark each infiltrating immune cell type was obtained (60). The correlation between gene expressions and immune cells infiltration was calculated using Pearson correlation analysis.

The Cancer Genome Atlas verification of genetic biomarkers in TISSUE

GEPIA (<http://gepia.cancer-pku.cn/index.html>) is a customizable functionalities website for interactive analysis and visualization based on The Cancer Genome Atlas database (61). To further verify the two biomarkers of CKD kidney tissue, the GEPIA web server was used to plot gene expression level box plots between kidney renal clear cell carcinoma (KRIC) and normal tissues in the TCGA database. The patient data were grouped according to the transcripts per million (TPM) value. Log₂ (TPM + 1) was used for log-scale, and four-way analysis of variance (ANOVA) was applied. Overall survival analyses of biomarkers of kidney tissue were also performed using GEPIA.

Validation of CKD biomarkers

The R package ggpubr was used to generate the gene expression box plot for hub biomarkers. The R package pROC was used to plot the ROC curves (62). The AUC values were calculated to evaluate the sensitivity and specificity of model (63).

Kidney function

Serum creatinine (SCr) levels were evaluated through colorimetric assays based on Jaffe's reaction using deproteinized serum samples (Nanjing Jiancheng). Absorbance was measured at OD510 nm (BioTek) and analyzed accordingly.

Immunohistochemistry staining

5 μm slides cut from 4% paraformaldehyde fixed and paraffin embedded kidney tissues as were obtained from pathology department of Renji Hospital. All sections were de-paraffinized followed by heat-induced antigen retrieval on a heating block in Tris-EDTA buffer, PH = 9.0 for 15 min. Primary Rabbit DACH1 antibody (Proteintech) and Rabbit TCF21 antibody (Sigma Aldrich) was used in a final solution of 1:200 overnight at 4°C. Secondary

antibody was applied for 30 min at 37°C, and the color was developed using a diaminobenzidine peroxidase substrate kit (Dako REAL™ EnVision™, DAKO). Sections were then counterstained with hematoxylin, dehydrated and mounted. The expression of DACH1 and TCF21 were imaged by Leica Microscope X20.

Real-time PCR

Mouse kidney tissues were homogenized in Trizol reagent (TianGen). Total RNA was extracted and reverse transcribed into cDNA (HiScriptIII RT SuperMix, Vazyme). Real-time PCR was performed on LightCycler480 apparatus (Roche) using SYBR Green Mix (Yeasen). Mouse *gapdh* was used as internal control gene. The relative gene expression was analyzed using $2^{-\Delta\Delta CT}$ method. Primers were listed: *tcf21*-F, cgctcacttaaggcagatcc; *tcf21*-R, gtcaccacttccttcagtgca; *dach1*-F, cctgggaaccctgttactc; *dach1*-R, agatccaccattttgcactcatt; *ddx17*-F, gatcgggatcgtgacagggga; *ddx17*-R, agtcagcttctgacttctggat; *gapdh*-F, tggccttccgtgttctctac; *gapdh*-R, gaggctgttgaagtcga.

Statistical analyses

Data were shown as the Mean ± SEM. A two-tailed independent student's test was conducted to assess statistical significance. The significance was denoted as follows: **** $p < 0.0001$; *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; N.S., not significant.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Renji Hospital, School of Medicine, Shanghai Jiao Tong University. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by ethical committee of animal experiments, Shanghai Jiao Tong University, School of Medicine.

Author contributions

Conceptualization, SM and MH. Methodology, MH. Software, YH. Validation, YH and JX. Formal analysis, JX, YH and YX. Resources, AC. Writing—original draft preparation, JX. Writing—

review and editing, WY, YX and SM. Visualization, JX. Supervision, SM and MH. Project administration, SM and MH. Funding acquisition, SM. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by National Natural Science Foundation of China (81970574, 81770668), Shanghai Municipal Health Commission (ZXYXZ-201904) and Innovative research team of high-level local universities in Shanghai.

Acknowledgments

We thank all the members of Yang and Xu Labs for their insightful discussions. We thank Editage (<https://www.editage.com/>) for English-language editing support.

References

- Romagnani P, Remuzzi G, Glasscock R, Levin A, Jager KJ, Tonelli M, et al. Chronic kidney disease. *Nat Rev Dis Primers* (2017) 3:17088. doi: 10.1038/nrdp.2017.88
- Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. *Lancet (Lond Engl)* (2017) 389(10075):1238–52. doi: 10.1016/S0140-6736(16)32064-5
- Inrig JK, Califf RM, Tasneem A, Vegunta RK, Molina C, Stanifer JW, et al. The landscape of clinical trials in nephrology: a systematic review of clinicaltrials.gov. *Am J Kidney Dis* (2014) 63(5):771–80. doi: 10.1053/j.ajkd.2013.10.043
- O'Seaghdha CM, Fox CS. Genome-wide association studies of chronic kidney disease: what have we learned? *Nat Rev Nephrol* (2011) 8(2):89–99. doi: 10.1038/nrneph.2011.189
- Ou SM, Tsai MT, Chen HY, Li FA, Tseng WC, Lee KH, et al. Identification of galectin-3 as potential biomarkers for renal fibrosis by RNA-sequencing and clinicopathologic findings of kidney biopsy. *Front Med* (2021) 8:748225. doi: 10.3389/fmed.2021.748225
- Schena FP, Nistor I, Curci C. Transcriptomics in kidney biopsy is an untapped resource for precision therapy in nephrology: A systematic review. *Nephrol Dialysis Transplant* (2018) 33(7):1094–102. doi: 10.1093/ndt/gfx211
- Ju W, Nair V, Smith S, Zhu L, Shedden K, Song PXX, et al. Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. *Sci Trans Med* (2015) 7(316):316ra193. doi: 10.1126/scitranslmed.aac7071
- Wang J, Yin Y, Lu Q, Zhao YR, Hu YJ, Hu YZ, et al. Identification of important modules and hub gene in chronic kidney disease based on WGCNA. *J Immunol Res* (2022) 2022:4615292. doi: 10.1155/2022/4615292
- Zhou G, Zhang X, Wang W, Zhang W, Wang H, Xin G. Both peripheral blood and urinary miR-195-5p, miR-192-3p, miR-328-5p and their target genes PPM1A, RAB1A and BRSK1 may be potential biomarkers for membranous nephropathy. *Med Sci Monitor* (2019) 25:1903–16. doi: 10.12659/MSM.913057
- Kant S, Kronbichler A, Sharma P, Geetha D. Advances in understanding of pathogenesis and treatment of immune-mediated kidney disease: A review. *Am J Kidney Dis* (2022) 79(4):582–600. doi: 10.1053/j.ajkd.2021.07.019
- Chung KW, Dhillon P, Huang S, Sheng X, Shrestha R, Qiu C, et al. Mitochondrial damage and activation of the STING pathway lead to renal inflammation and fibrosis. *Cell Metab* (2019) 30(4):784–99.e5. doi: 10.1016/j.cmet.2019.08.003
- Cohen G. Immune dysfunction in uremia 2020. *Toxins* (2020) 12(7):439. doi: 10.3390/toxins12070439
- Yang J, Fang P, Yu D, Zhang L, Zhang D, Jiang X, et al. Chronic kidney disease induces inflammatory CD40+ monocyte differentiation via homocysteine elevation and DNA hypomethylation. *Circ Res* (2016) 119(11):1226–41. doi: 10.1161/CIRCRESAHA.116.308750
- Bernardor J, Alioli C, Meaux MN, Peyruchaud O, Machuca-Gayet I, Bacchetta J. Peripheral blood mononuclear cells (PBMCs) to dissect the underlying mechanisms of

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2023.1129524/full#supplementary-material>

- bone disease in chronic kidney disease and rare renal diseases. *Curr Osteoporos Rep* (2021) 19(6):553–62. doi: 10.1007/s11914-021-00707-6
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559
- Yuan Q, Tang B, Zhang C. Signaling pathways of chronic kidney diseases, implications for therapeutics. *Signal Transduct Targeted* (2022) 7(1):182. doi: 10.1038/s41392-022-01036-5
- Xu C, Wang L, Zhang Y, Li W, Li J, Wang Y, et al. Tubule-specific Mst1/2 deficiency induces CKD via YAP and non-YAP mechanisms. *J Am Soc Nephrol* (2020) 31(5):946–61. doi: 10.1681/ASN.2019101052
- Wu W, Wang X, Yu X, Lan HY. Smad3 signatures in renal inflammation and fibrosis. *Int J Biol Sci* (2022) 18(7):2795–806. doi: 10.7150/ijbs.71595
- Espi M, Koppe L, Fouque D, Thaunat O. Chronic kidney disease-associated immune dysfunctions: Impact of protein-bound uremic retention solutes on immune cells. *Toxins* (2020) 12(5):300. doi: 10.3390/toxins12050300
- Pletinck A, Glorieux G, Schepers E, Cohen G, Gondouin B, Van Landschoot M, et al. Protein-bound uremic toxins stimulate crosstalk between leukocytes and vessel wall. *J Am Soc Nephrol* (2013) 24(12):1981–94. doi: 10.1681/ASN.2012030281
- Vanholder R, Pletinck A, Schepers E, Glorieux G. Biochemical and clinical impact of organic uremic retention solutes: A comprehensive update. *Toxins* (2018) 10(1):33. doi: 10.3390/toxins10010033
- Stinghen AE, Massy ZA, Vlassara H, Striker GE, Boullier A. Uremic toxicity of advanced glycation end products in CKD. *J Am Soc Nephrol* (2016) 27(2):354–70. doi: 10.1681/ASN.2014101047
- Meng XM, Nikolic-Paterson DJ, Lan HY. TGF- β : the master regulator of fibrosis. *Nat Rev Nephrol* (2016) 12(6):325–38. doi: 10.1038/nrneph.2016.48
- Remuzzi G, Bertani T. Pathophysiology of progressive nephropathies. *New Engl J Med* (1998) 339(20):1448–56. doi: 10.1056/NEJM19981123392007
- Moriguchi T, Hamada M, Morito N, Terunuma T, Hasegawa K, Zhang C, et al. MafB is essential for renal development and F4/80 expression in macrophages. *Mol Cell Biol* (2006) 26(15):5715–27. doi: 10.1128/MCB.00001-06
- Guca E, Suñol D, Ruiz L, Konkol A, Cordero J, Torner C, et al. TGIF1 homeodomain interacts with smad MH1 domain and represses TGF- β signaling. *Nucleic Acids Res* (2018) 46(17):9220–35. doi: 10.1093/nar/gky680
- De Lorenzo SB, Vrieze AM, Johnson RA, Lien KR, Nath KA, Garovic VD, et al. KLF11 deficiency enhances chemokine generation and fibrosis in murine unilateral ureteral obstruction. *PLoS One* (2022) 17(4):e0266454. doi: 10.1371/journal.pone.0266454
- Spittau B, Wang Z, Boinska D, Kriegelstein K. Functional domains of the TGF- β -inducible transcription factor Tieg3 and detection of two putative nuclear localization signals within the zinc finger DNA-binding domain. *J Cell Biochem* (2007) 101(3):712–22. doi: 10.1002/jcb.21228

29. Gohla G, Krieglstein K, Spittau B. Tieg3/Klf1 induces apoptosis in OLI-neu cells and enhances the TGF-beta signaling pathway by transcriptional repression of Smad7. *J Cell Biochem* (2008) 104(3):850–61. doi: 10.1002/jcb.21669
30. Gao L, Liu M, Dong N, Jiang Y, Lin CY, Huang M, et al. Matriptase is highly upregulated in chronic lymphocytic leukemia and promotes cancer cell invasion. *Leukemia* (2013) 27(5):1191–4. doi: 10.1038/leu.2012.289
31. Chou FP, Chen YW, Zhao XF, Xu-Monette ZY, Young KH, Gartenhaus RB, et al. Imbalanced matriptase pericellular proteolysis contributes to the pathogenesis of malignant b-cell lymphomas. *Am J Pathol* (2013) 183(4):1306–17. doi: 10.1016/j.ajpath.2013.06.024
32. Wu K, Katiyar S, Li A, Liu M, Ju X, Popov VM, et al. Dachshund inhibits oncogene-induced breast cancer cellular migration and invasion through suppression of interleukin-8. *Proc Natl Acad Sci USA* (2008) 105(19):6924–9. doi: 10.1073/pnas.0802085105
33. Doka T, Huang S, Qiu C, Liu H, Guan Y, Hu H, et al. Transcriptome-wide association analysis identifies DACH1 as a kidney disease risk gene that contributes to fibrosis. *J Clin Invest* (2021) 131(10). doi: 10.1172/JCI141801
34. Cao A, Li J, Asadi M, Basgen JM, Zhu B, Yi Z, et al. DACH1 protects podocytes from experimental diabetic injury and modulates PTIP-H3K4Me3 activity. *J Clin Invest* (2021) 131(10). doi: 10.1172/JCI141279
35. Ao X, Ding W, Zhang Y, Ding D, Liu Y. TCF21: a critical transcription factor in health and cancer. *J Mol Med (Berlin Germany)* (2020) 98(8):1055–68. doi: 10.1007/s00109-020-01934-7
36. Ide S, Finer G, Maezawa Y, Onay T, Souma T, Scott R, et al. Transcription factor 21 is required for branching morphogenesis and regulates the gdnf-axis in kidney development. *J Am Soc Nephrol* (2018) 29(12):2795–808. doi: 10.1681/ASN.2017121278
37. Maezawa Y, Onay T, Scott RP, Keir LS, Dimke H, Li C, et al. Loss of the podocyte-expressed transcription factor Tcf21/Pod1 results in podocyte differentiation defects and FSGS. *J Am Soc Nephrol* (2014) 25(11):2459–70. doi: 10.1681/ASN.2013121307
38. Nakano Y, Kamiya A, Sumiyoshi H, Tsuruya K, Kagawa T, Inagaki Y. A deactivation factor of fibrogenic hepatic stellate cells induces regression of liver fibrosis in mice. *Hepatology* (2020) 71(4):1437–52. doi: 10.1002/hep.30965
39. Makino H, Miyamoto Y, Sawai K, Mori K, Mukoyama M, Nakao K, et al. Altered gene expression related to glomerulogenesis and podocyte structure in early diabetic nephropathy of db/db mice and its restoration by pioglitazone. *Diabetes* (2006) 55(10):2747–56. doi: 10.2337/db05-1683
40. Usui J, Yaguchi M, Yamazaki S, Takahashi-Kobayashi M, Kawamura T, Kaneko S, et al. Transcription factor 21 expression in injured podocytes of glomerular diseases. *Sci Rep* (2020) 10(1):11516. doi: 10.1038/s41598-020-68422-3
41. Lamm GM, Nicol SM, Fuller-Pace FV, Lamond AI. p72: a human nuclear DEAD box protein highly related to p68. *Nucleic Acids Res* (1996) 24(19):3739–47. doi: 10.1093/nar/24.19.3739
42. Moy RH, Cole BS, Yasunaga A, Gold B, Shankarling G, Varble A, et al. Stem-loop recognition by DDX17 facilitates miRNA processing and antiviral defense. *Cell* (2014) 158(4):764–77. doi: 10.1016/j.cell.2014.06.023
43. Zhang X, An T, Zhang X, Shen T, Li H, Dou L, et al. DDX17 protects hepatocytes against oleic acid/palmitic acid-induced lipid accumulation. *Biochem Biophys Res Commun* (2022) 612:169–75. doi: 10.1016/j.bbrc.2022.04.129
44. Bonaventure B, Goujon C. DEXH/D-box helicases at the frontline of intrinsic and innate immunity against viral infections. *J Gen Virol* (2022) 103(8). doi: 10.1099/jgv.0.001766
45. Wang SB, Narendran S, Hirahara S, Varshney A, Pereira F, Apicella I, et al. DDX17 is an essential mediator of sterile NLR4 inflammasome activation by retrotransposon RNAs. *Sci Immunol* (2021) 6(66):eabi4493. doi: 10.1126/sciimmunol.abi4493
46. Betjes MG. Immune cell dysfunction and inflammation in end-stage renal disease. *Nat Rev Nephrol* (2013) 9(5):255–65. doi: 10.1038/nrneph.2013.44
47. Vaziri ND. Oxidative stress in uremia: nature, mechanisms, and potential consequences. *Semin Nephrol* (2004) 24(5):469–73. doi: 10.1016/j.semnephrol.2004.06.026
48. Takahashi-Kobayashi M, Usui J, Yaguchi M, Yamazaki S, Kawamura T, Kaneko S, et al. Immunohistological score of transcription factor 21 had a positive correlation with its urinary excretion and proteinuria in immunoglobulin a nephropathy. *Histol Histopathol* (2021) 36(10):1093–8. doi: 10.14670/HH-18-367
49. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* (2012) 28(6):882–3. doi: 10.1093/bioinformatics/bts034
50. Chen C, Grennan K, Badner J, Zhang D, Gershon E, Jin L, et al. Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. *PLoS One* (2011) 6(2):e17238. doi: 10.1371/journal.pone.0017238
51. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
52. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (2010) 26(1):139–40. doi: 10.1093/bioinformatics/btp1616
53. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* (2016) 32(18):2847–9. doi: 10.1093/bioinformatics/btw313
54. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* (2021) 49(D1):D605–D12. doi: 10.1093/nar/gkaa1074
55. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* (2014) 8 Suppl 4:S11. doi: 10.1186/1752-0509-8-S4-S11
56. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* (2021) 12(1):1088. doi: 10.1038/s41467-021-21246-9
57. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118
58. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. *Bioinformatics* (2014) 30(19):2811–2. doi: 10.1093/bioinformatics/btu393
59. Tibshirani R. The lasso method for variable selection in the cox model. *Stat Med* (1997) 16(4):385–95. doi: 10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3
60. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* (2017) 18(1):248–62. doi: 10.1016/j.celrep.2016.12.019
61. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* (2017) 45(W1):W98–W102. doi: 10.1093/nar/gkx247
62. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf* (2011) 12:77. doi: 10.1186/1471-2105-12-77
63. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Med Res Methodol* (2017) 17(1):53. doi: 10.1186/s12874-017-0332-6