Identification of key pathways and genes in different types of chronic kidney disease based on WGCNA

YUHE GUO[1*], JUNJIE MA[1*], LANYAN XIAO[2], JIALI FANG[1], GUANGHUI LI[1], LEI ZHANG[1], LU XU[1], XINGQIANG LAI[1], GUANGHUI PAN[1] and ZHENG CHEN[1]

[1]Department of Organ Transplantation, Second Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong 510260; [2]Department of Center Laboratory, The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong 510700, P.R. China

Received September 21, 2018; Accepted March 15, 2019

DOI: 10.3892/mmr.2019.10443

Abstract. Chronic kidney disease (CKD) is a highly heterogeneous nephrosis that occurs when the structure and function of the kidney is damaged. Gene expression studies have been widely used to elucidate various biological processes; however, the gene expression profile of CKD is currently unclear. The present study aimed to identify diagnostic biomarkers and therapeutic targets using renal biopsy sample data from patients with CKD. Gene expression data from 30 patients with CKD and 21 living donors were analyzed by weighted gene co-expression network analysis (WGCNA), in order to identify gene networks and profiles for CKD, as well as its specific characteristics, and to potentially uncover diagnostic biomarkers and therapeutic targets for patients with CKD. In addition, functional enrichment analysis was performed on co-expressed genes to determine modules of interest. Four co-expression modules were constructed from the WGCNA. The number of genes in the constructed modules ranged from 269 genes in the Turquoise module to 60 genes in the Yellow module. All four co-expression modules were correlated with CKD clinical traits ($P<0.05$). For example, the Turquoise module, which mostly contained genes that were upregulated in CKD, was positively correlated with CKD clinical traits, whereas the Blue, Brown and Yellow modules were negatively correlated with clinical traits. Functional enrichment analysis revealed that the Turquoise module was mainly enriched in genes associated with the 'defense response', 'mitotic cell cycle' and 'collagen catabolic process' Gene Ontology (GO) terms, implying that genes involved in cell cycle arrest and fibrogenesis were upregulated in CKD. Conversely, the Yellow module was mainly enriched in genes associated with 'glomerulus development' and 'kidney development' GO terms, indicating that genes associated with renal development and damage repair were downregulated in CKD. The hub genes in the modules were acetyl-CoA carboxylase α, cyclin-dependent kinase 1, Wilm's tumour 1, NPHS2 stomatin family member, podocin, JunB proto-oncogene, AP-1 transcription factor subunit, activating transcription factor 3, forkhead box O1 and v-abl Abelson murine leukemia viral oncogene homolog 1, which were confirmed to be significantly differentially expressed in CKD biopsies. Combining the eight hub genes enabled a high capacity for discrimination between patients with CKD and healthy subjects, with an area under the receiver operating characteristic curve of 1.00. In conclusion, this study provided a framework for co-expression modules of renal biopsy samples from patients with CKD and living donors, and identified several potential diagnostic biomarkers and therapeutic targets for CKD.

Correspondence to: Dr Zheng Chen or Dr Guanghui Pan, Department of Organ Transplantation, Second Affiliated Hospital of Guangzhou Medical University, 250 Changgang East Road, Guangzhou, Guangdong 510260, P.R. China
E-mail: docchenzheng@163.com
E-mail: panguanghui@126.com

*Contributed equally

Abbreviations: WGCNA, weighted gene co-expression network analysis; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; GO, Gene Ontology; STRING, Search Tool for the Retrieval of Interacting Genes; DEGs, differentially expressed genes; MM, module membership; GS, gene significance; DUSP1, dual specificity phosphatase 1; ACACA, acetyl-CoA carboxylase α; CDK1, cyclin-dependent kinase 1; CKS2, cyclin-dependent kinases regulatory subunit 2; CCNB1, cyclin B1; COL4A2, collagen type IV α2; SERPINE2, serpin family E member 2; CDKN1C, cyclin-dependent kinase inhibitor 1C; NPHS1, NPHS1 adhesion molecule, nephrin; ATF3, activating transcription factor 3; FOXO1, forkhead box O1; ABL1, v-abl Abelson murine leukemia viral oncogene homolog 1; WT1, Wilm's tumour 1; PLCE1, phospholipase C ε1

Key words: weighted gene co-expression network analysis, chronic kidney disease, co-expression module

Introduction

Chronic kidney disease (CKD) is one of the most common types of nephrosis worldwide, and the number of patients with CKD has increased rapidly in recent years (1,2). CKD is a highly heterogeneous disease in which the structure and function of the kidney is damaged (3-5). Traditionally, kidney failure is considered the eventual outcome of CKD, and

generally the symptoms are caused by a reduction in kidney function (6,7). When symptoms become severe, the consequent end-stage kidney failure can only be treated by transplantation and dialysis. Over the past three decades, clinical and experimental studies have extended our understanding of the causes of CKD (8-11). Most forms of CKD eventually progress to end-stage kidney disease; however, the mechanisms underlying the progression of CKD remain poorly understood. Gene expression studies have been successfully applied to elucidate various biological processes, including cancer (12-14), angiocardiopathy (15,16), asthma (17,18), and chronic obstructive pulmonary disease (COPD) (19,20); these studies are useful for the identification of early detection biomarkers and therapeutic targets.

Weighted gene co-expression network analysis (WGCNA) is a novel methodology used to study relationships between clinical traits and gene expression profiles (21,22). WGCNA converts gene expression data into co-expression networks (modules), groups co-expressed genes with common biological functions or associations, and provides co-expression networks that may be responsible for clinical traits of interest. This technique has been successfully used to identify potential biomarkers and therapeutic targets for numerous biological processes, including cancer, COPD and asthma (18,19,23).

The present study aimed to identify the genetic mechanisms underlying CKD using renal biopsy sample data from patients with CKD and living donors. Genome-wide expression data were obtained from 30 patients with CKD (13 with minimal change disease and 17 with membranous glomerulonephropathy) and 21 living donors. WGCNA was applied to associate co-expression networks with extensive clinical traits, including disease status and disease type. The biological functions were further analyzed using gene co-expression networks, and co-expression networks that were significantly related to disease status and disease type were highlighted. Functional enrichment analysis was used to study the modules of interest, and hub genes in each module were identified and displayed using Search Tool for the Retrieval of Interacting Genes (STRING), which provided useful information for determining the dominant genes in these modules. The present study provided co-expression modules for renal biopsy samples from patients with CKD and may be beneficial for obtaining a better understanding of the mechanisms underlying CKD.

Materials and methods

Expression analysis of microarray data from renal biopsy samples from patients with CKD and living donors. RNA sequencing data from renal biopsy samples from patients with CKD and the clinical traits of these patients were downloaded from National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) DataSets (www.ncbi.nlm.nih.gov/gds) under accession no. GSE104954 (24). The combined cohort contained a total of 51 tubulointerstitial samples from patients with CKD (n=30; 13 with minimal change disease and 17 with membranous glomerulonephropathy) and healthy living donors (n=21). The sequence data and clinical traits were supplied by the European Renal Biopsy cDNA Bank. Prior to performing the WGCNA calculation, microarray probes were annotated by mapping to gene symbols using R (version 3.3.4; www.r-project.org). Probes matching more than one gene symbol were eliminated from the cohort, and average expression levels were calculated for genes with multiple probes.

Identification of differentially expressed genes (DEGs). The raw data files used in this study contained CEL files. The analysis was conducted using R language (version 3.3.4; www.r-project.org) and Bioconductor (www.bioconductor.org). The microarray signal intensity was normalized using robust multi-array average, and DEGs were identified using the R software extension package 'DESeq' (www.bioconductor.org/packages/release/bioc/html/DESeq.html) (25).

Construction of co-expression networks (modules) in patients with CKD and living donors. In this study, to reduce the amount of unnecessary calculations, only genes that exhibited a ≥1.2-fold change were chosen to construct the co-expression modules. First, hierarchical clustering of samples was analyzed using the flashClust function (21,26). Then, the soft thresholding power β-value was screened during module construction by the pickSoftThreshold function of the WGCNA algorithm (21). A set of candidate powers (ranging between 1 and 30) was applied to test the average connectivity degrees of different modules and their independence. A suitable power value was selected if the degree of independence was >0.9. Once the power value was selected, the WGCNA algorithm, which is an R software extension package (cran.r-project.org/web/packages/WGCNA/index.html), was performed to construct co-expression networks (modules); the minimum module size was set to 30. Co-expression networks or modules were defined as branches of a hierarchical clustering tree, and each module was assigned a unique color label. The correlations between each module were analyzed and visualized using the heatmap tool package (https://cran.r-project.org/web/packages/pheatmap/index.html).

Relating co-expression modules to external clinical traits. Firstly, the module eigengene was defined as the first principal component of the expression matrix for a given module. The module eigengene can be considered an average gene expression level for all genes in each module. Subsequently, clinical information, including disease status and disease type, was converted into numerical values, after which, a regression analysis was performed between the module eigengene values and the clinical information. Module membership (MM) was defined as the association between a gene and a given module, and gene significance (GS) was defined as the correlation of genes with clinical traits. Genes with high GS for a clinical trait and MM were considered to be candidates for subsequent analysis. All analyses were conducted using the WGCNA package.

Functional enrichment analysis of co-expression modules. The established modules were sorted according to the number of genes they contained; subsequently, functional enrichment analysis was performed for each individual module. Gene Ontology (GO) (27,28) analysis was performed using the extension R package 'ClusterProfiler' (https://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html) with a

correction; P≤0.05 was set as the threshold (29). Subsequently, the modules of interested were visualized using STRING (string-db.org/cgi/input.pl), and only experimentally confirmed interactions with an interaction score >0.4 were selected as significant. For each module, genes with the maximum intra-modular connectivity were considered intra-modular hub genes.

Statistical analysis. Statistical analysis was performed using SPSS 22.0 (SPSS, Inc.) and R software 3.3.4. Graphs were generated using GraphPad Prism 6.0 (GraphPad Software, Inc.). Student's t-test and Mann-Whitney U test were used to evaluate significant differences between the patients with CKD and healthy subjects. Receiver operating characteristic (ROC) curves were generated to evaluate the diagnostic accuracy of each hub gene, and the area under the curve (AUC) was used to evaluate sensitivity and specificity. All P-values were two sided, and P<0.05 was considered to indicate a statistically significant difference.

Results

Identification of DEGs in tubulointerstitial samples from patients with CKD. A total of 51 tubulointerstitial samples, including 30 samples from patients with CKD and 21 healthy individuals, were analyzed. All relevant gene expression data and clinical information were analyzed using the R software and its extension packages. Using a log2-fold change ≥2 and P<0.05 as cutoff values, a total of 54 DEGs were identified in patients with CKD, of which 16 genes were upregulated and 38 genes were downregulated. When using the criteria of P<0.05 and 1.2-fold change, 489 genes were upregulated and 399 genes were downregulated. A volcano plot of the log2fold change vs. the P-value (-log10 P-value) for all probe sets is shown in Fig. 1, and hierarchical clustering of the 50 most significant DEGs, including the 25 top upregulated and downregulated genes, was visualized using a heatmap (Fig. 2). Red represents increased expression, whereas blue represents decreased expression. The most upregulated genes included lactotransferrin, hemoglobin subunit β, regenerating family member 1α, C-C motif chemokine ligand 20 and apolipoprotein C1, whereas Fos proto-oncogene, AP-1 transcription factor subunit, pyruvate dehydrogenase kinase 4, MAF bZIP transcription factor F, dual specificity phosphatase 1 (DUSP1) and FosB proto-oncogene, AP-1 transcription factor subunit were the most downregulated genes in the CKD samples.

Construction of co-expression modules using tubulointerstitial samples from patients with CKD. The expression values of 887 genes with a fold change of 1.2 (P<0.05) in 51 renal biopsy samples from patients with CKD and living donors were used to construct co-expression networks (modules) with WGCNA. Hierarchical clustering of the samples was analyzed using the flashClust function; the clustering results are shown in Fig. 3. All 51 samples clustered well and were mainly divided into two clusters; the first cluster was comprised of GSM2811026, GSM2811027, GSM2811028, GSM2811043, GSM2811045, GSM2811046, GSM2811047, GSM2811048, GSM2811049, GSM2811050, GSM2811051, GSM2811052, GSM2811053, GSM2811054, GSM2811055, GSM2811056, GSM2811057,

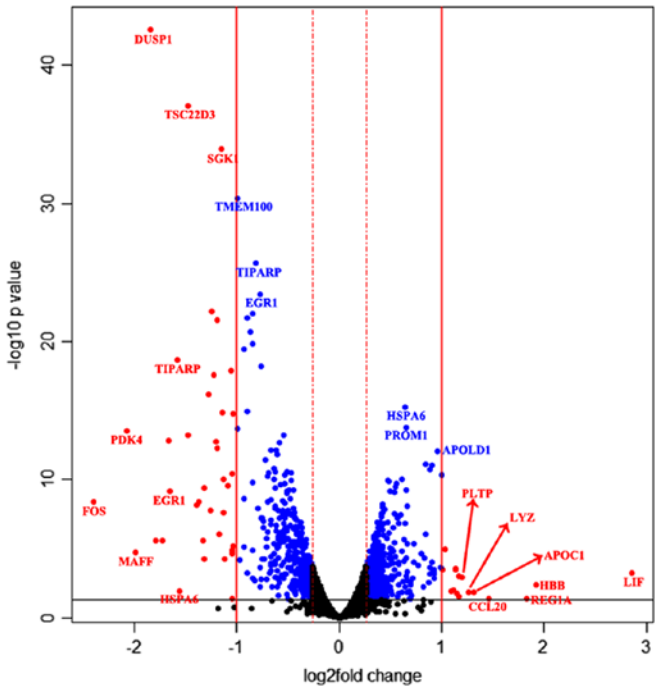

Figure 1. Volcano plot indicating the upregulated and downregulated genes in tubulointerstitial samples from patients with CKD. The horizontal axis represents the fold change between healthy living donors and patients with CKD. The vertical axis represents the P-values of the differences between healthy donors and patients with CKD, as determined using Student's t-test. The genes most relevant for CKD are highlighted in red (2-fold change) or blue (1.2-fold change). CKD, chronic kidney disease.

GSM2811058, GSM2811059 and GSM2811060, and contained most of the samples from the living donors, whereas the remaining samples yielded the second cluster. Subsequently, the power value, which mainly determined the independence and the average connectivity degrees of the co-expression networks, was screened using a set of candidate numbers (ranging between 1 and 30). The power value 12, which was the lowest power value for the scale with an independence degree of up to 0.9, was selected to construct a hierarchical clustering tree for the 887 genes (Fig. 4). Four co-expression modules were identified, with a range in size from 269 genes in the Turquoise module to 60 genes in the Yellow module. Furthermore, an extra module (Grey), which contained genes that did not belong to any of the other four modules, was also defined (Fig. 5 and Table I). Interactions between the four co-expression modules were also analyzed and are shown in Fig. 6.

Relationships between co-expression modules and clinical traits. The corresponding clinical trait information was downloaded from NCBI GEO DataSets, and unwanted information was removed prior to analysis. The correlations between the co-expression modules and the measured clinical traits were quantified based on the correlations between the module eigengenes and the clinical traits (Fig. 7 and Table I). Furthermore, the eigengene dendrogram and heatmap were used to demonstrate groups of correlated eigengenes (Fig. 8). The results demonstrated that the four co-expression modules were highly related to CKD clinical traits. Of the four modules, one module (Turquoise) was

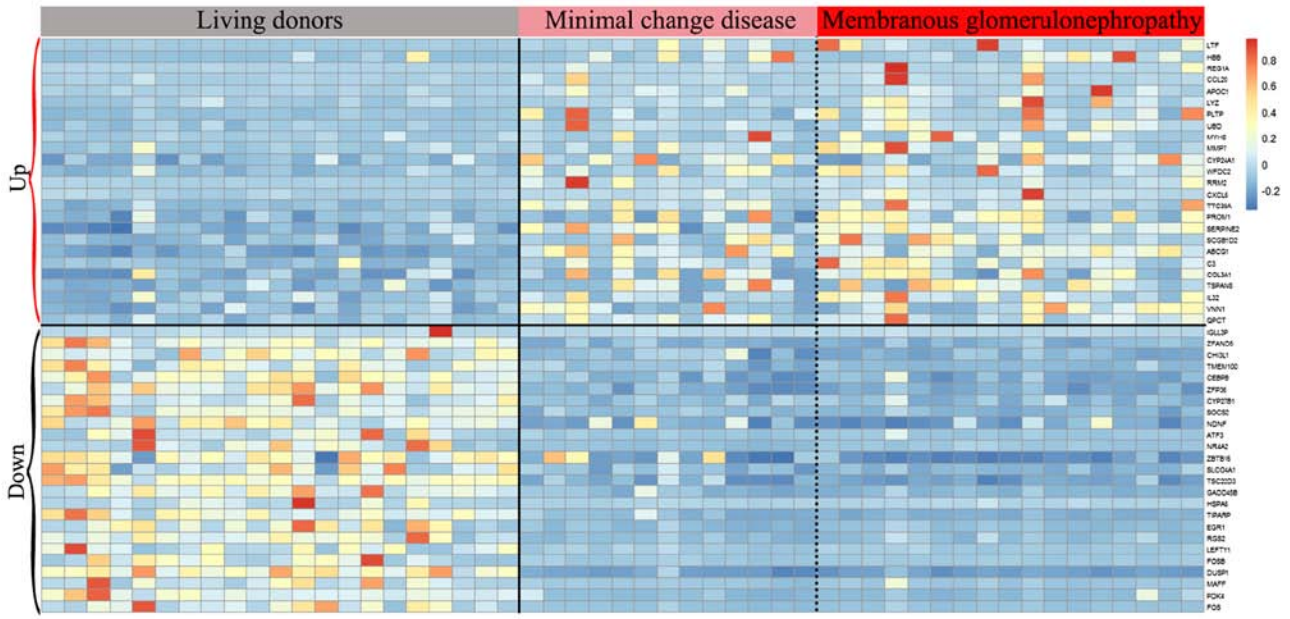

Figure 2. Heatmap of differentially expressed genes in tubulointerstitial samples from patients with CKD. The heatmap shows differentially expressed genes (P<0.05) between the healthy donors and patients with CKD from the GSE104954 dataset. The grey, pink and red bars above the heatmap indicate the groups of tubulointerstitial samples. Genes with higher expression in the CKD tubulointerstitial samples are shown in the upper part of the heatmap, and genes with lower expression in the CKD tubulointerstitial samples are shown in the lower part. Blue represents decreased relative expression, whereas red represents increased relative expression. CKD, chronic kidney disease.

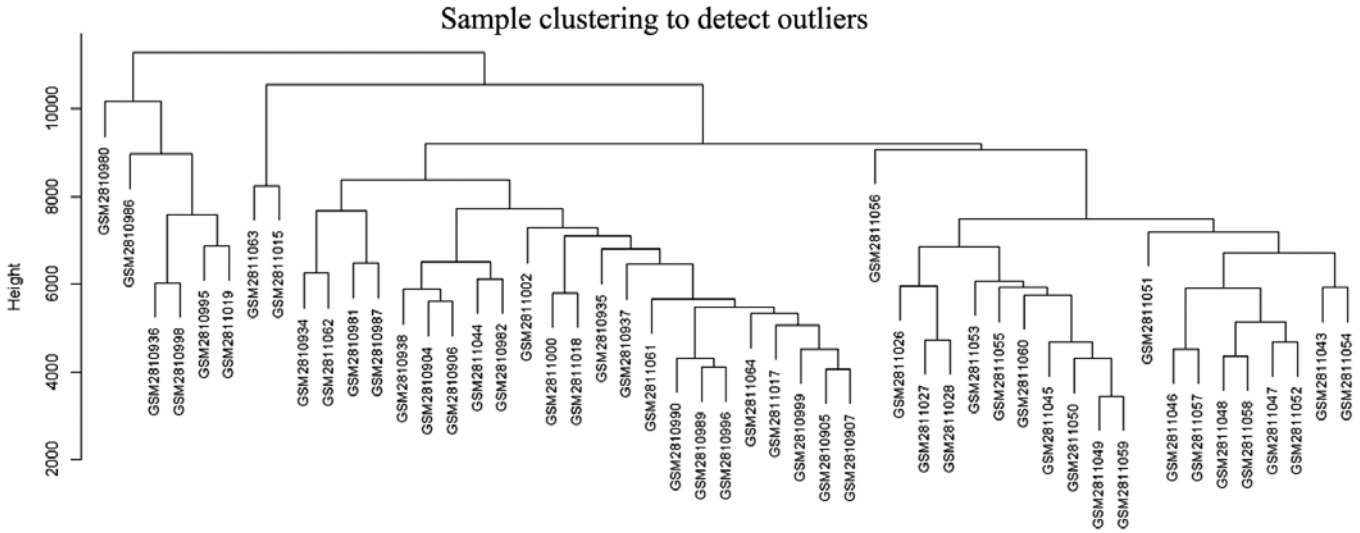

Figure 3. Sample clustering to detect outliers. All samples were well clustered.

positively correlated with the CKD disease_status (correlation, 0.62; P=1x10^{-6}) and CKD disease_type (correlation, 0.63, P=9x10^{-7}) (Table I); this module mostly contained genes that were overexpressed in samples from patients with CKD. Conversely, the Yellow (correlation, -0.72, P=3x10^{-9} and correlation, -0.66, P=1x10^{-7}), Blue (correlation, -0.87, P=6x10^{-17} and correlation, -0.8, P=1x10^{-12}) and Brown (correlation, -0.79, P=7x10^{-12} and correlation, -0.7, P=1x10^{-8}) modules were negatively correlated with CKD clinical traits (Fig. 7 and Table I), and the genes in these modules were predominantly downregulated in patients with CKD. Subsequently, the WGCNA algorithm was used to calculate GS vs. MM. The results demonstrated that the Turquoise, Yellow, Blue and Brown genes that were most significantly associated with the CKD clinical traits (GS) were also the most important elements of MM, as demonstrated by the genes included in the upper right region of the graphs shown in Fig. 9. Notably, genes [including acetyl-CoA carboxylase α (ACACA), collagen type IV α2 (COL4A2), TNF-α-induced protein 8, β-1,4-galactosyltransferase 5, serpin family E member 2 (SERPINE2), NPHS1 adhesion molecule, nephrin (NPHS1), cyclin-dependent kinase inhibitor 1C (CDKN1C), phospholipase C ε1 (PLCE1), DUSP1, ZFP36 ring finger protein, neural precursor cell expressed, developmentally down-regulated 9, and TSC22 domain family member 3] that were presented in the upper right region were highly

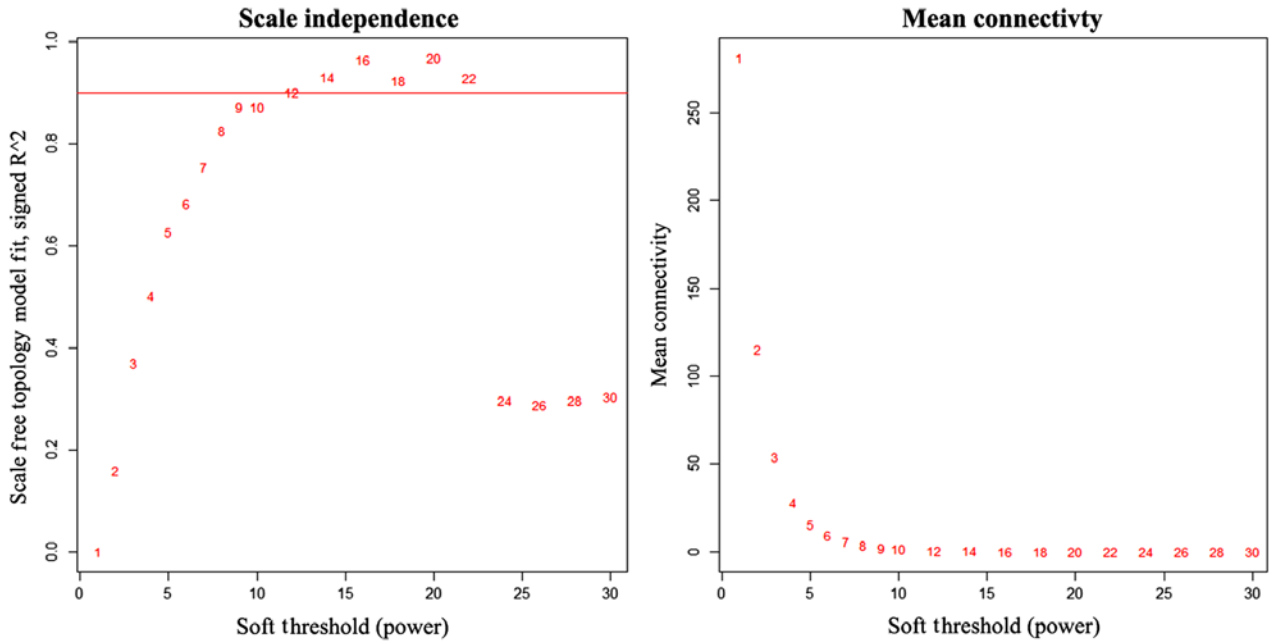

Figure 4. Analysis of network topology for a set of soft thresholding powers. The left graph displays the scale of the free fit index (y-axis) as a function of the soft thresholding power (x-axis). The right graph shows the mean connectivity (degree, y-axis) as a function of the soft thresholding power (x-axis).

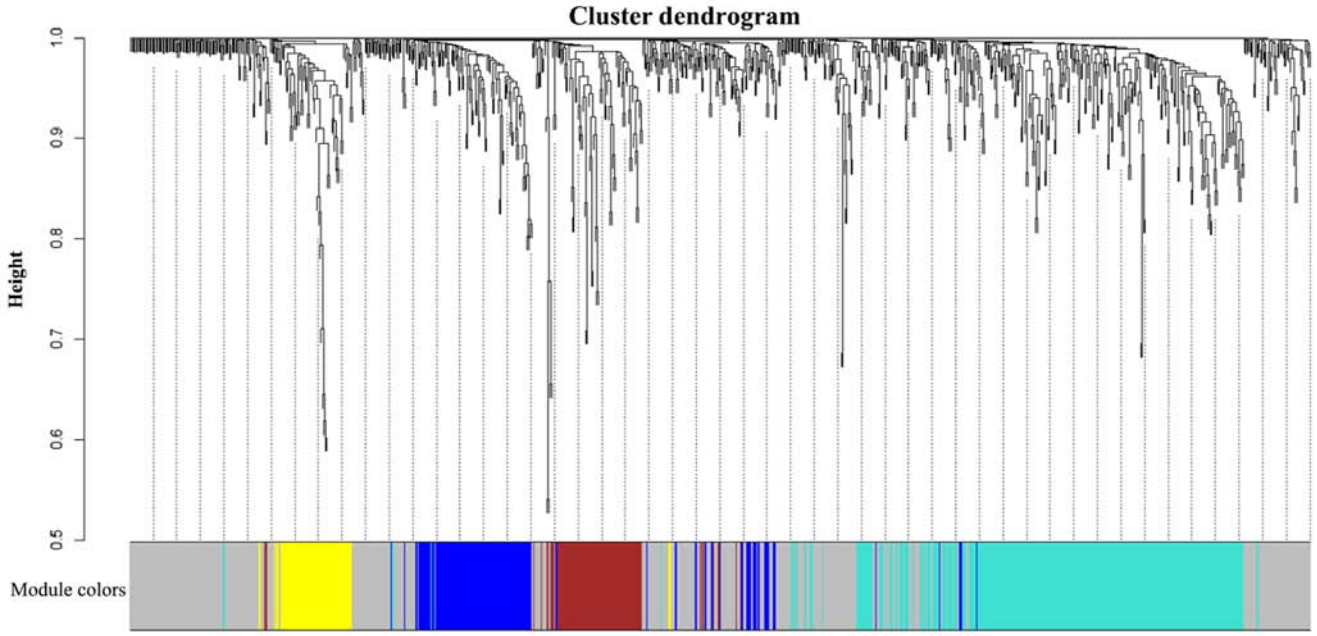

Figure 5. Clustering dendrograms of genes with dissimilarity based on topological overlap, together with the assigned module colors. As a result, four co-expression modules were constructed and shown with distinctive colors.

correlated to the trait of interest (CKD) and key components in the underlying biological function.

Functional enrichment analysis of the modules of interest. GO enrichment analysis was performed for the genes in the four significant co-expression modules (Fig. 10 and Table II). Genes in the Turquoise module were mainly enriched in 'GO:0006952 defense response', 'GO:0030574 collagen catabolic process' and 'GO:0000278 mitotic cell cycle', whereas genes in the Yellow module were mainly enriched in 'GO:0032835 glomerulus development' and 'GO:0007275 multicellular organismal development', genes in the Brown module were enriched in 'GO:0009605 response to external stimulus', and genes in the Blue module were enriched in 'GO:0009719 response to endogenous stimulus' and 'GO:0010941 regulation of cell death'.

Module visualization and hub genes. The four significant modules were further visualized using the STRING database. Only genes with a minimum interaction score of >0.4 were considered significant. The intramodular connectivity was quantified for each gene. Genes with high intramodular

Table I. Correlation of module eigengene with CKD clinic trait.

		Disease status		Disease type	
Module color	Genes	Correlation (r)	P-value	Correlation (r)	P-value
Turquoise	269	0.62	$1x10^{-6}$	0.63	$9x10^{-7}$
Yellow	60	-0.72	$3x10^{-9}$	-0.66	$1x10^{-7}$
Blue	112	-0.87	$6x10^{-17}$	-0.8	$1x10^{-12}$
Brown	79	-0.79	$7x10^{-12}$	-0.7	$1x10^{-8}$

Modules identified by weighted gene co-expression analysis are listed together with the number of genes they contained. Four co-expression modules were significantly correlated to CKD disease status and disease type. CKD, chronic kidney disease.

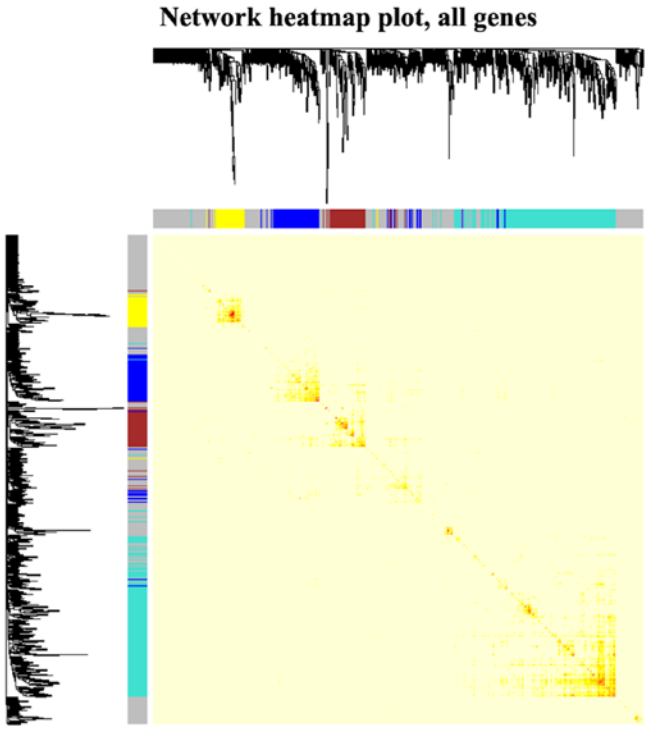

Figure 6. Visualization of the gene co-expression modules using a heatmap plot. This plot shows the topological overlap matrix among all differentially expressed genes in this analysis. Light colors represent lower overlap, and a darker red color indicates higher overlap. Blocks of darker colors along the diagonal indicate co-expression modules.

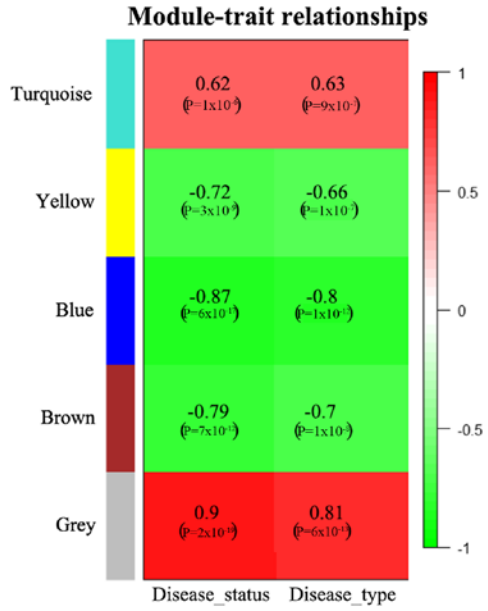

Figure 7. Module-trait relationships. Each row shows a module eigengene, and each column corresponds to a clinical trait. Each cell contains the corresponding correlation and P-value. The table is color-coded by correlation according to the color legend.

connectivity were considered intramodular hub genes. The hub genes ACACA, cyclin-dependent kinase 1 (CDK1), Wilm's tumor 1 (WT1), NPHS2 stomatin family member, podocin (NPHS2), JunB proto-oncogene, AP-1 transcription factor subunit (JUNB), activating transcription factor 3 (ATF3), forkhead box O1 (FOXO1), and v-abl Abelson murine leukemia viral oncogene homolog 1 (ABL1) in the four modules are shown in Fig. 11.

Verification of hub gene expression and ROC curve analysis. Analysis of hub gene expression was performed using renal biopsy sample data from patients with CKD and living donors. ACACA, CDK1 and ABL1 were expressed at significantly higher levels in patients with CKD, whereas the other five hub genes (NPHS2, JUNB, ATF3, WT and FOXO1) were significantly downregulated (Fig. 12). Subsequently, ROC curve analysis was used to evaluate the diagnostic prediction values of the hub genes for CKD. This analysis revealed that the AUC for ACACA was 0.86 (P<0.0001). At the optimal cut-off value of 0.55, the sensitivity and specificity were 80 and 71%, respectively. Similar results were obtained for CDK1, NPHS2, JUNB, ATF3, WT1, FOXO1 and ABL1 (Fig. 13 and Table III). When combined, these hub genes possessed a high ability to discriminate between patients with CKD and living donors, with an AUC of 1.00.

Discussion

The main objective of the study was to utilize a global approach for construction of a gene co-expression network that predicted clusters of candidate genes involved in CKD pathogenesis. In the present study, four co-expression modules were constructed using the WGCNA algorithm, which was used to study the relationship between CKD gene expression and clinical traits. One module (Turquoise) contained mostly upregulated genes and was significantly positively

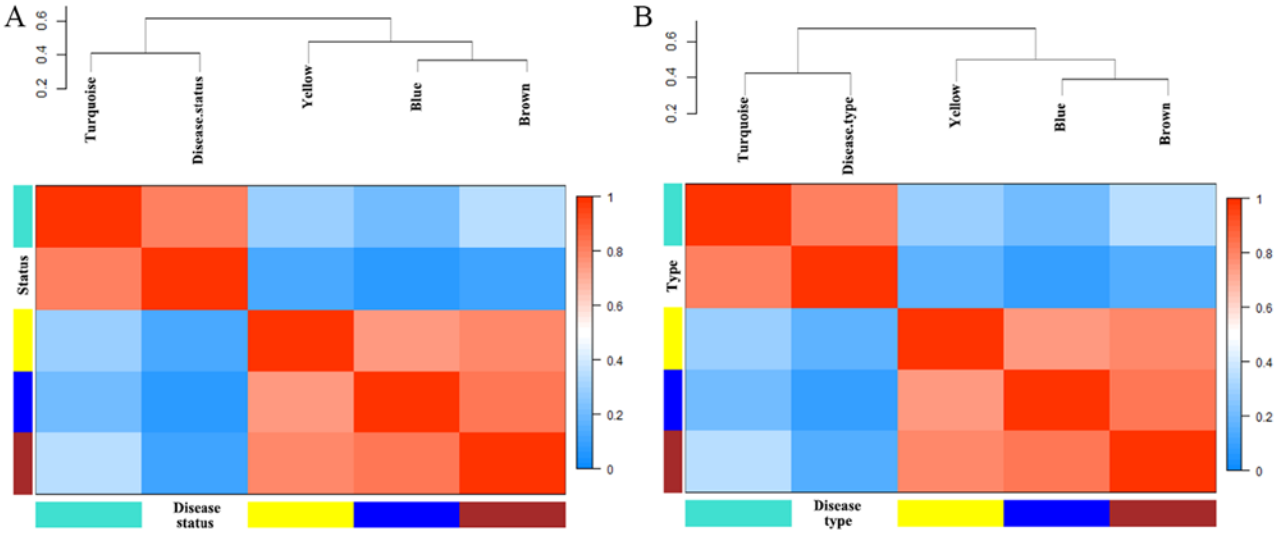

Figure 8. Eigengene dendrogram and heatmap identify groups of correlated eigengenes termed meta-modules. As a result, the upper dendrograms indicate that the Turquoise module is highly related to (A) CKD disease_status and (B) CKD disease_type. The heatmaps present eigengene adjacency.

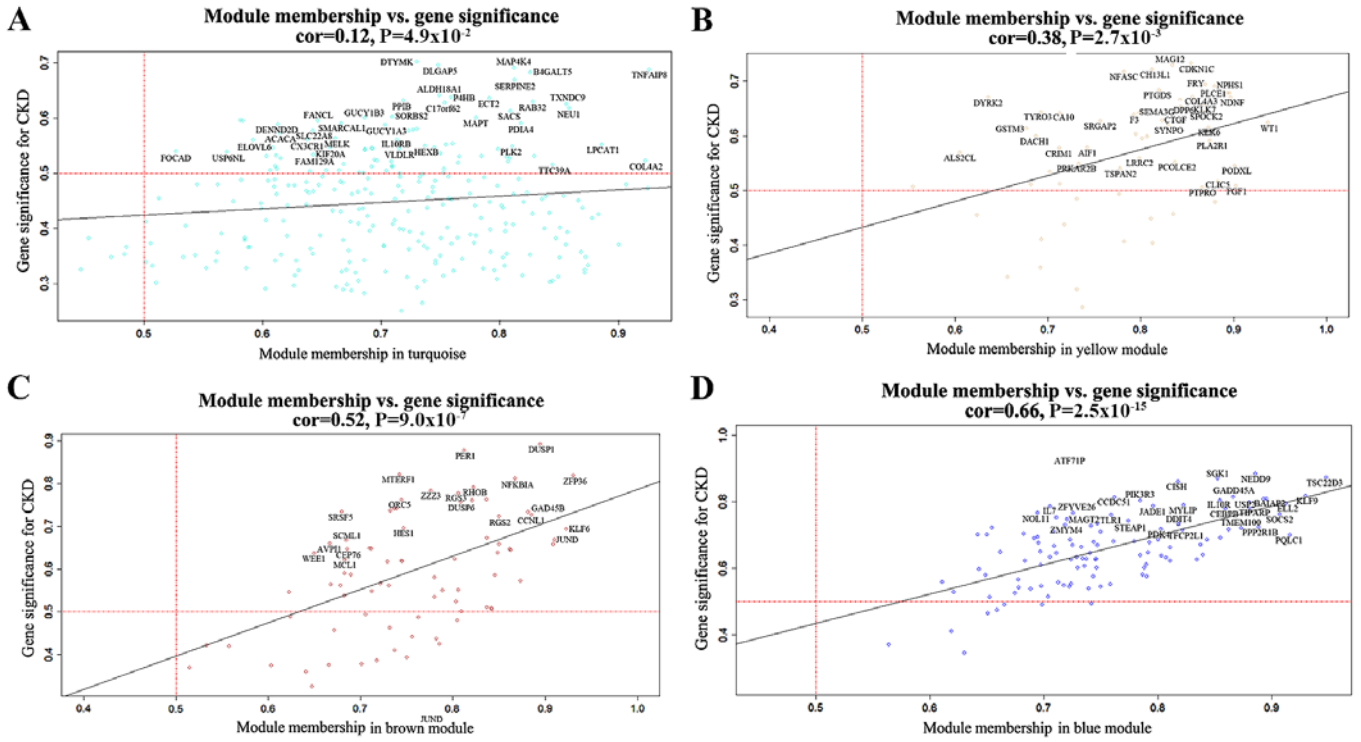

Figure 9. Scatterplot of gene significance for CKD vs. module membership in the (A) Turquoise module, (B) Yellow module, (C) Brown module and (D) Blue module. The correlation coefficient and P-value are listed above the scatterplots.

correlated with CKD clinical traits, whereas the other three modules (Blue, Brown and Yellow) were negatively correlated with CKD clinical traits and mainly contained genes that were downregulated in patients with CKD. The most central (hub) genes in these modules were ACACA, CDK1, WT1, NPHS2, JUNB, ATF3, FOXO1 and ABL1, which suggested direct/indirect regulation of the CKD-associated gene expression network; these genes may serve as potential biomarkers for detection and treatment of CKD.

CKD is a heterogeneous disease that arises from numerous diverse pathogenic mechanisms, including vascular, metabolic and immunological disorders (3). The glomeruli accumulate a large amount of extracellular matrix components, and the renal interstitium and periglomerular region become fibrotic (30,31). Histopathological analysis of end-stage kidney samples provides clues to the origin of the disease (32). These results have indicated that progression from the original occurrence to end-stage renal disease may have some common pathogenic mechanisms. High-throughput gene expression profile data have been used to identify various molecular mechanisms involved in distinctly original kidney diseases. In a previous study, analyses of peripheral blood lymphocyte cells from

Table II. GO enrichment analysis of genes in co-expression modules.

Module	Term ID	Term	Genes	Enrichment P-value	Term name
Turquoise	GO:0006952	BP	34	2.92×10^{-4}	Defense response
Turquoise	GO:0006955	BP	31	7.13×10^{-4}	Immune response
Turquoise	GO:0002376	BP	39	1.45×10^{-3}	Immune system process
Turquoise	GO:0050896	BP	86	9.04×10^{-3}	Response to stimulus
Turquoise	GO:0030574	BP	6	2.23×10^{-2}	Collagen catabolic process
Turquoise	GO:0002443	BP	8	2.48×10^{-2}	Leukocyte mediated immunity
Turquoise	GO:0001817	BP	15	2.56×10^{-2}	Regulation of cytokine production
Turquoise	GO:0000278	BP	19	4.13×10^{-2}	Mitotic cell cycle
Turquoise	GO:0016266	BP	5	4.74×10^{-2}	O-glycan processing
Turquoise	GO:0043067	BP	26	4.74×10^{-2}	Regulation of programmed cell death
Yellow	GO:0032835	BP	9	1.20×10^{-10}	Glomerulus development
Yellow	GO:0072006	BP	10	5.00×10^{-9}	Nephron development
Yellow	GO:0001822	BP	12	7.59×10^{-9}	Kidney development
Yellow	GO:0072001	BP	12	9.89×10^{-9}	Renal system development
Yellow	GO:0001655	BP	12	3.34×10^{-8}	Urogenital system development
Yellow	GO:0007275	BP	32	5.87×10^{-7}	Multicellular organismal development
Yellow	GO:0044767	BP	34	5.87×10^{-7}	Single-organism developmental process
Yellow	GO:0048731	BP	30	5.87×10^{-7}	System development
Yellow	GO:0032502	BP	34	6.59×10^{-7}	Developmental process
Yellow	GO:0072015	BP	4	3.37×10^{-6}	Glomerular visceral epithelial cell development
Brown	GO:0014070	BP	22	5.49×10^{-10}	Response to organic cyclic compound
Brown	GO:0051591	BP	11	8.16×10^{-10}	Response to cAMP
Brown	GO:0009605	BP	29	9.59×10^{-9}	Response to external stimulus
Brown	GO:0045944	BP	22	9.71×10^{-9}	Positive regulation of transcription from RNA polymerase II promoter
Brown	GO:0048518	BP	45	2.37×10^{-8}	Positive regulation of biological process
Brown	GO:0051254	BP	25	3.39×10^{-8}	Positive regulation of RNA metabolic process
Brown	GO:0009893	BP	37	5.48×10^{-8}	Positive regulation of metabolic process
Brown	GO:0045935	BP	26	5.61×10^{-8}	Positive regulation of nucleobase-containing compound metabolic process
Brown	GO:2000113	BP	23	6.13×10^{-8}	Negative regulation of cellular macromolecule biosynthetic process
Brown	GO:0048522	BP	41	6.30×10^{-8}	Positive regulation of cellular process
Blue	GO:0007167	BP	19	8.06×10^{-4}	Enzyme linked receptor protein signaling pathway
Blue	GO:0009719	BP	23	8.06×10^{-4}	Response to endogenous stimulus
Blue	GO:0009725	BP	18	8.06×10^{-4}	Response to hormone
Blue	GO:0010033	BP	32	8.06×10^{-4}	Response to organic substance
Blue	GO:0010941	BP	24	8.06×10^{-4}	Regulation of cell death
Blue	GO:0042221	BP	40	8.06×10^{-4}	Response to chemical
Blue	GO:0080090	BP	50	8.06×10^{-4}	Regulation of primary metabolic process
Blue	GO:0071495	BP	19	1.13×10^{-3}	Cellular response to endogenous stimulus
Blue	GO:0042981	BP	22	1.48×10^{-3}	Regulation of apoptotic process
Blue	GO:0019222	BP	54	1.84×10^{-3}	Regulation of metabolic process

BP, biological process; GO, Gene Ontology.

donors with membranous nephropathy and normal controls identified dysregulated microRNAs (miRNAs) that serve an important role in the pathogenesis of nephropathy, and may serve as reliable diagnostic markers and potential therapeutic targets (33). In addition, high-throughput sequencing has identified circulating miRNAs that may serve as potential biomarkers for kidney damage in patients with systemic lupus erythematosus (34). Notwithstanding these findings, the

Table III. AUC ROC values of different hub genes in patients with chronic kidney disease.

Hub genes	AUC	P-value	95% CI	Cut-off	Sensitivity	Specificity
ACACA	0.86	<0.0001	0.75-0.96	0.55	0.80	0.71
CDK1	0.83	<0.0001	0.72-0.94	0.51	0.77	0.67
NPHS2	0.83	<0.0001	0.71-0.95	0.57	0.77	0.67
JUNB	0.95	<0.0001	0.90-1.00	0.63	0.90	0.86
ATF3	0.90	<0.0001	0.81-0.99	0.72	0.83	0.76
WT1	0.87	<0.0001	0.77-0.96	0.57	0.77	0.67
FOXO1	0.89	<0.0001	0.80-0.98	0.59	0.83	0.76
ABL1	0.88	<0.0001	0.78-0.98	0.53	0.83	0.76
Combined	1.00	<0.0001	1.00-1.00	1.00	1.00	1.00

ABL1, cyclin-dependent kinase 1; ACACA, acetyl-CoA carboxylase α; ATF3, activating transcription factor 3; AUC, area under the curve; CDK1, cyclin-dependent kinase 1; FOXO1, forkhead box O1; JUNB, JunB proto-oncogene, AP-1 transcription factor subunit; NPHS2, NPHS2 stomatin family member, podocin; ROC, receiver operating characteristic; WT1, Wilm's tumor 1.

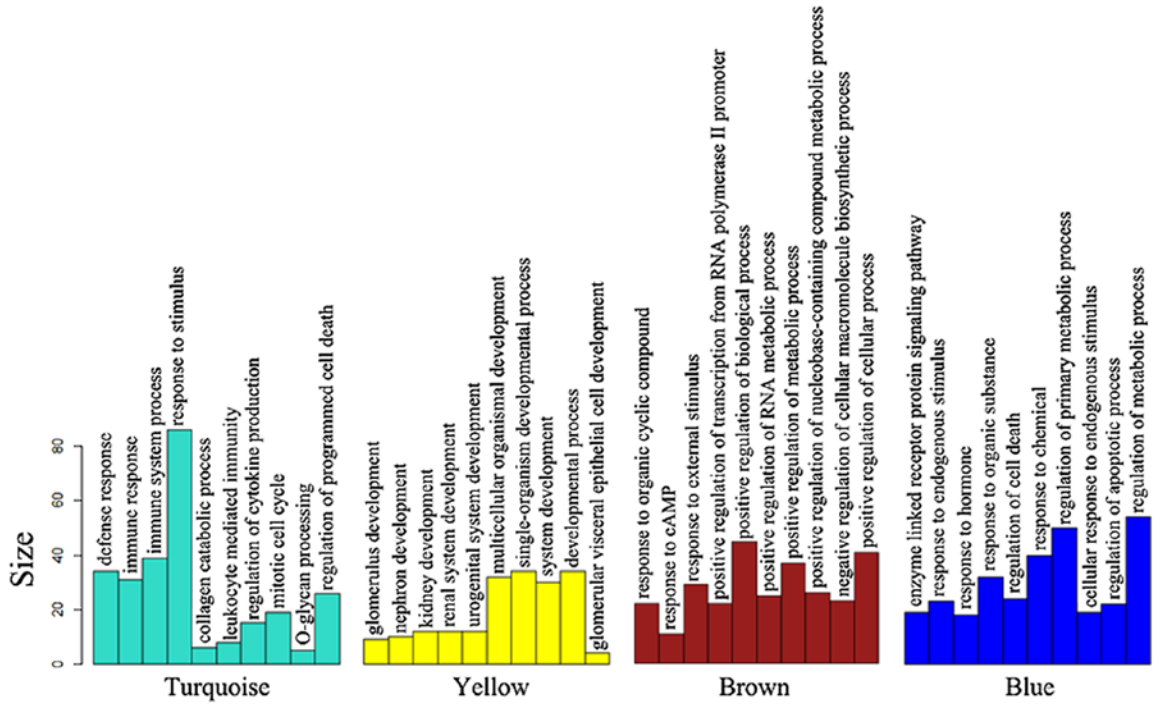

Figure 10. GO enrichment analysis of genes in the turquoise, yellow, brown, and blue modules.

molecular mechanisms underlying CKD remain poorly understood. The present results were somewhat consistent with these prior studies. A Turquoise module was identified, which contained genes that were mostly upregulated in patients with CKD, including genes that encoded proteins related to fibrosis ('collagen catabolic process' GO term, e.g., transforming growth factor β1, collagen type I α1 chain, collagen type III α1 chain, collagen type IV α1 chain, COL4A2 and collagen type IV α3 chain). In addition, a novel intriguing co-expression module (the Yellow module) was also detected in this study, which contained genes involved in 'glomerulus development', 'kidney development' and 'multicellular organismal development' GO terms [e.g. NPHS1 (35), NPHS2 (36), WT1 (37), podocalyxin-like and PLCE1]. Since the Yellow module mainly contained genes downregulated in patients with CKD and was negatively correlated with CKD clinical traits; these results indicated that the renal parenchyma may be damaged in CKD and the corresponding repair mechanisms may be suppressed.

Subsequently, the constructed modules were further visualized using a protein interaction network. The genes in the Turquoise module included ACACA, CDK1, cyclin-dependent kinases regulatory subunit 2 (CKS2), cyclin B1 (CCNB1), CDKN1C, COL4A2 and SERPINE2, which are mainly involved in the 'defense response', 'collagen catabolic process' and 'mitotic cell cycle' GO terms. ACACA encodes acetyl-CoA carboxylase α, which serves a critical role in the regulation and metabolism of fatty acid biosynthesis in

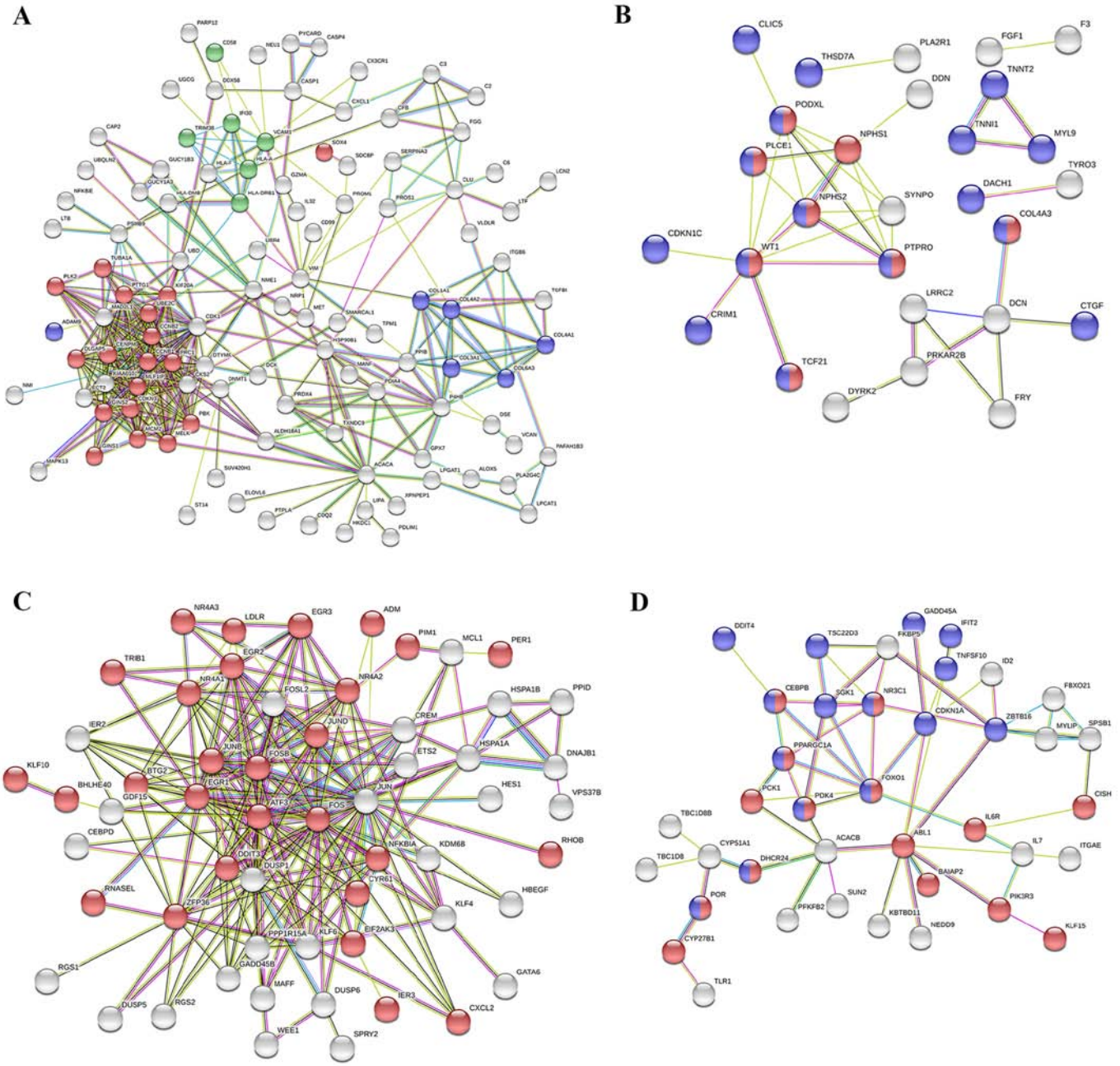

Figure 11. Visualization of co-expression of genes in the co-expression module. (A) Turquoise module, genes involved in the 'mitotic cell cycle', 'collagen catabolic process' and 'defense response' GO terms are marked in red, violet and green, respectively. (B) Yellow module, genes involved in 'glomerulus development' and 'multicellular organismal development' GO terms are marked in red and violet, respectively. (C) Brown module, genes involved in the 'response to external stimulus' GO term are marked in red. (D) Blue module, genes involved in the 'response to endogenous stimulus' and 'regulation of cell death' GO terms are marked in red and violet, respectively. GO, Gene Ontology.

mammals (38), although its role in CKD is currently unclear. CDK1, CKS2, CCNB1 and CDKN1C are genes involved in cell cycle regulation (39,40). A previous study reported that the expression levels of CDK1 and cyclin B2 are significantly upregulated in tubular epithelial cells from rats with chronic renal failure and that cycle arrest of tubular epithelial cells participates in kidney fibrogenesis (41). The other module of interest was the Yellow module. Genes in this module included WT1, NPHS1, NPHS2 and PLCE1, which mainly participate in the 'glomerulus development' GO term. WT1 is a pleiotropic transcription factor, and mutations in this gene lead to a set of clinical phenotypes that are caused by dysfunction of either renal podocytes or kidney progenitors (37). NPHS1, NPHS2 and PLCE1 are genes encoding proteins that maintain normal kidney development and function, and homozygous genetic variants in NPHS1, NPHS2 and PLCE1 have been reported to be associated with development of congenital nephrotic syndrome (42,43). According to the results of pathway enrichment analysis, the upregulated genes were mainly involved in cell cycle arrest and fibrogenesis, whereas the downregulated genes were mainly associated with kidney development and repair. Therefore, it was hypothesized that these hub genes may be potential diagnostic biomarkers and therapeutic targets for patients with CKD.

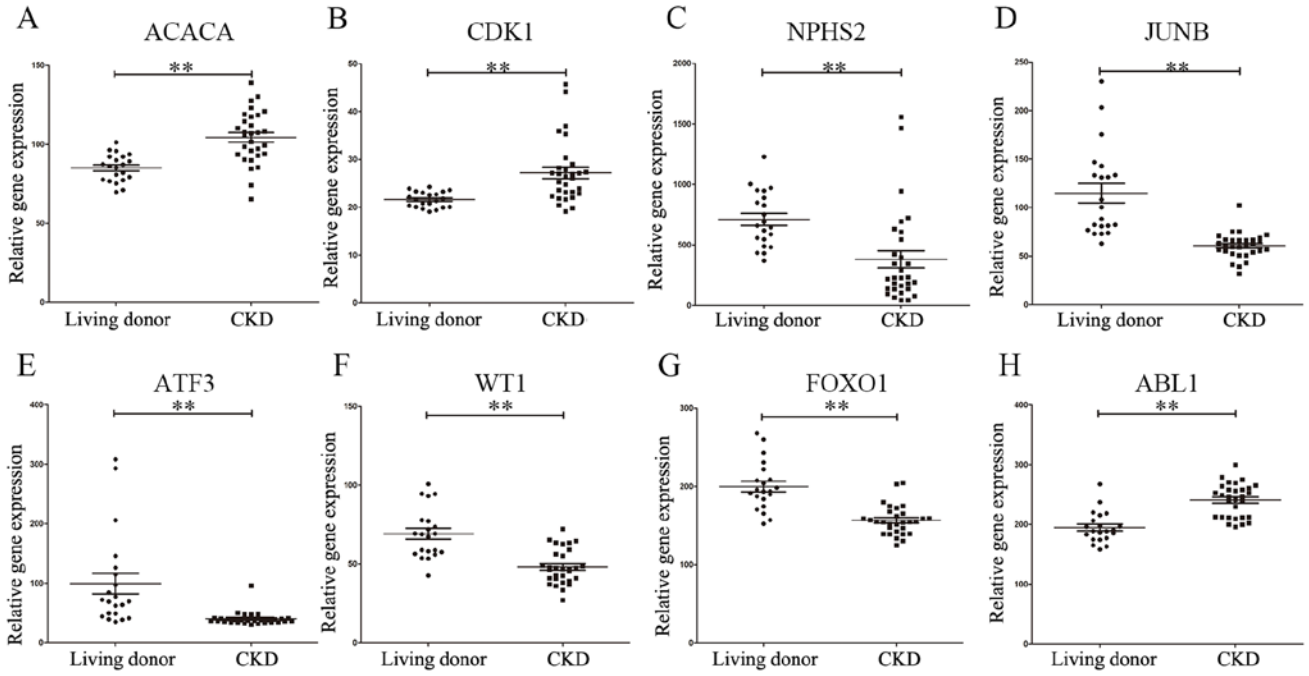

Figure 12. (A-H) Validation of the expression levels of hub genes in the renal biopsies. Mann-Whitney U test was used to analyze results. ABL1, cyclin-dependent kinase 1; ACACA, acetyl-CoA carboxylase α; ATF3, activating transcription factor 3; CDK1, cyclin-dependent kinase 1; CKD, chronic kidney disease; FOXO1, forkhead box O1; JUNB, JunB proto-oncogene, AP-1 transcription factor subunit; NPHS2, NPHS2 stomatin family member, podocin; WT1, Wilm's tumor 1.

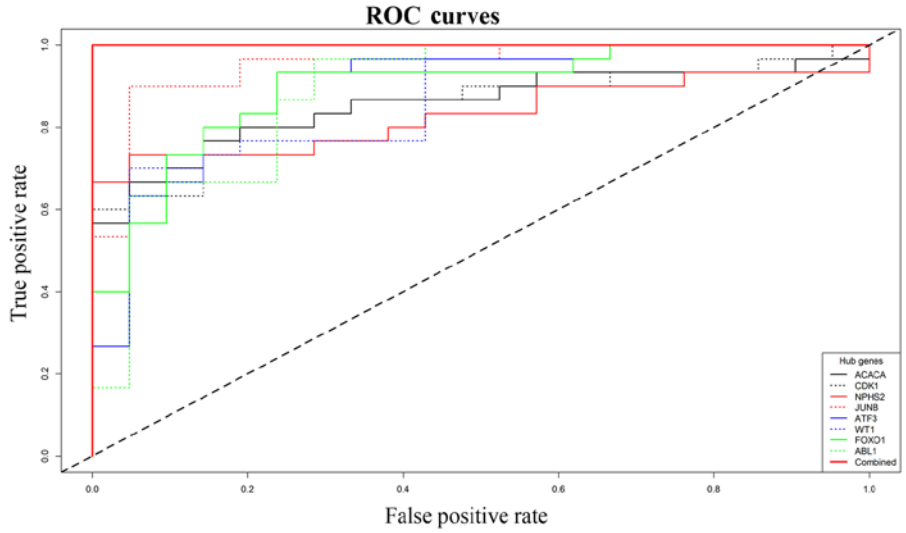

Figure 13. ROC curves of the hub genes. The AUC was determined for ACACA (AUC=0.86), CDK1 (0.83), NPHS2 (0.83), JUNB (0.95), ATF3 (0.90), WT1 (0.87), FOXO1 (0.89) and ABL1 (0.88), and for the combination of the eight hub genes (1.00). ABL1, cyclin-dependent kinase 1; ACACA, acetyl-CoA carboxylase α; ATF3, activating transcription factor 3; AUC, area under the curve; CDK1, cyclin-dependent kinase 1; FOXO1, forkhead box O1; JUNB, JunB proto-oncogene, AP-1 transcription factor subunit; NPHS2, NPHS2 stomatin family member, podocin; ROC, receiver operating characteristic; WT1, Wilm's tumor 1.

In conclusion, the Turquoise, Yellow, Brown and Blue modules were regarded as the most critical modules in patients with CKD based on gene expression data from renal biopsy samples, and the hub genes ACACA, CDK1, WT1, NPHS2, JUNB, ATF3, FOXO1 and ABL1 were significantly expressed in these modules. Further studies are underway to address the specific mechanisms of these hub genes in CKD. A detailed understanding of the roles served by these hub genes may provide insights into CKD, and lead to diagnostic and therapeutic opportunities for patients with CKD.

Acknowledgments

Not applicable.

Funding

This research program was financially sponsored by the Science and Technology Planning Project of Guangdong Province, China (grant nos. 2014A020212319 and 2015B020226002).

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the National Center of Biotechnology Information's GEO database (www.ncbi.nlm.nih.gov/gds/) with the accession no. GSE104954.

Authors' contributions

ZC and GP designed the study. LaX, JF and GL collected the data for analysis and drafted the manuscript. LZ, LuX and XL analyzed the data. YG and JM made substantial contributions to the design of the study and created the figures. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Klahr S, Schreiner G and Ichikawa I: The progression of renal disease. N Engl J Med 318: 1657-1666, 1988.
2. Carpenter AR and McHugh KM: Role of renal urothelium in the development and progression of kidney disease. Pediatr Nephrol 32: 557-564, 2017.
3. Webster AC, Nagler EV, Morton RL and Masson P: Chronic kidney disease. Lancet 389: 1238-1252, 2017.
4. Tonelli M, Wiebe N, Culleton B, House A, Rabbat C, Fok M, McAlister F and Garg AX: Chronic kidney disease and mortality risk: A systematic review. J Am Soc Nephrol 17: 2034-2047, 2006.
5. De Nicola L and Zoccali C: Chronic kidney disease prevalence in the general population: Heterogeneity and concerns. Nephrol Dial Transpl 31: 331-335, 2016.
6. Meguid El Nahas A and Bello AK: Chronic kidney disease: The global challenge. Lancet 365: 331-340, 2005.
7. Nugent RA, Fathima SF, Feigl AB and Chyung D: The burden of chronic kidney disease on developing nations: A 21st century challenge in global health. Nephron Clin Pract 118: c269-c277, 2011.
8. Lovisa S, Zeisberg M and Kalluri R: Partial epithelial-to-mesenchymal transition and other new mechanisms of kidney fibrosis. Trends Endocrin Met 27: 681-695, 2016.
9. Hewitson TD, Holt SG and Smith ER: Progression of tubulointerstitial fibrosis and the chronic kidney disease phenotype-role of risk factors and epigenetics. Front Pharmacol 8: 520, 2017.
10. Lajdova I, Okša A, Spustová A and Dzúrik R: Cellular calcium balance in chronic kidney disease. Vnitr Lek 58: 525-530, 2012.
11. Rossi GP, Seccia TM, Barton M, Danser AHJ, de Leeuw PW, Dhaun N, Rizzoni D, Rossignol P, Ruilope LM, van den Meiracker AH, *et al*: Endothelial factors in the pathogenesis and treatment of chronic kidney disease Part II: Role in disease conditions: A joint consensus statement from the european society of hypertension working group on endothelin and endothelial factors and the japanese society of hypertension. J Hypertens 36: 462-471, 2018.
12. Wan Q, Tang J, Han Y and Wang D: Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. Exp Eye Res 166: 13-20, 2018.
13. Liu R, Cheng Y, Yu J, Lv QL and Zhou HH: Identification and validation of gene module associated with lung cancer through coexpression network analysis. Gene 563: 56-62, 2015.
14. Zhai X, Xue Q, Liu Q, Guo Y and Chen Z: Colon cancer recurrenceassociated genes revealed by WGCNA coexpression network analysis. Mol Med Rep 16: 6499-6505, 2017.
15. Yan S: Integrative analysis of promising molecular biomarkers and pathways for coronary artery disease using WGCNA and MetaDE methods. Mol Med Rep 18: 2789-2797, 2018.
16. Guo N, Zhang N, Yan L, Lian Z, Wang J, Lv F, Wang Y and Cao X: Weighted gene coexpression network analysis in identification of key genes and networks for ischemicreperfusion remodeling myocardium. Mol Med Rep 18: 1955-1962, 2018.
17. Modena BD, Bleecker ER, Busse WW, Erzurum SC, Gaston BM, Jarjour NN, Meyer DA, Milosevic J, Tedrow JR, Wu W, *et al*: Gene expression correlated with severe asthma characteristics reveals heterogeneous mechanisms of severe disease. Am J Respir Crit Care Med 195: 1449-1463, 2017.
18. Liu Z, Li M, Fang X, Shen L, Yao W, Fang Z, Chen J, Feng X, Hu, Zeng Z, *et al*: Identification of surrogate prognostic biomarkers for allergic asthma in nasal epithelial brushing samples by WGCNA. J Cell Biochem 120: 5137-5150, 2018.
19. Yi G, Liang M, Li M, Fang X, Liu J, Lai Y, Chen J, Yao W, Feng X, Hu, *et al*: A large lung gene expression study identifying IL1B as a novel player in airway inflammation in COPD airway epithelial cells. Inflamm Res 67: 539-551, 2018.
20. Brandsma CA, van den Berge M, Postma DS, Jonker MR, Brouwer S, Paré PD, Sin DD, Bossé Y, Laviolette M, Karjalainen J, *et al*: A large lung gene expression study identifying fibulin-5 as a novel player in tissue repair in COPD. Thorax 70: 21-32, 2015.
21. Langfelder P and Horvath S: WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 9: 559, 2008.
22. Botia JA, Vandrovcova J, Forabosco P, Guelfi S, D'Sa K; United Kingdom Brain Expression Consortium, Hardy J, Lewis CM, Ryten M and Weale ME: An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. Bmc Syst Biol 11: 47, 2017.
23. Guo L, Zhang K and Bing Z: Application of a co-expression network for the analysis of aggressive and non-aggressive breast cancer cell lines to predict the clinical outcome of patients. Mol Med Rep 16: 7967-7978, 2017.
24. Grayson PC, Eddy S, Taroni JN, Lightfoot YL, Mariani L, Parikh H, Lindenmeyer MT, Ju W, Greene CS, Godfrey B, *et al*: Metabolic pathways and immunometabolism in rare kidney diseases. Ann Rheum Dis 77: 1226-1233, 2018.
25. Anders S and Huber W: Differential expression analysis for sequence count data. Genome Biol 11: R106, 2010.
26. Langfelder P and Horvath S: Fast R functions for robust correlations and hierarchical clustering. J Stat Softw 46: 1-17, 2012.
27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*: Gene ontology: Tool for the unification of biology. The gene ontology consortium. Nat Genet 25: 25-29, 2000.
28. The Gene Ontology Consortium: The gene ontology resource: 20 years and still going strong. Nucleic Acids Res 47: D330-D338, 2019.
29. Yu G, Wang LG, Han Y and He QY: clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS 16: 284-287, 2012.
30. Sharma K: Obesity, oxidative stress, and fibrosis in chronic kidney disease. Kidney Int Suppl 4: 113-117, 2014.
31. Boor P and Floege J: Chronic kidney disease growth factors in renal fibrosis. Clin Exp Pharmacol Physiol 38: 441-450, 2011.
32. Saborio P, Krieg RJ Jr, Chan W, Hahn S and Chan JC: Pathophysiology of growth retardation in chronic renal failure. Zhonghua Minguo Xiao Er Ke Yi Xue Hui Za Zhi 39: 21-27, 1998.
33. Chen W, Lin X, Huang J, Tan K, Chen Y, Peng W, Li W and Dai Y: Integrated profiling of microRNA expression in membranous nephropathy using high-throughput sequencing technology. Int J Mol Med 33: 25-34, 2014.
34. Navarro-Quiroz E, Pacheco-Lugo L, Lorenzi H, Díaz-Olmos Y, Almendrales L, Rico E, Navarro R, España-Puccini P, Iglesias A, Egea E and Aroca G: High-throughput sequencing reveals circulating miRNAs as potential biomarkers of kidney damage in patients with systemic lupus erythematosus. PLoS One 11: e0166202, 2016.
35. Eichinger A, Ponsel S, Bergmann C, Günthner R, Hoefele J, Amann K and Lange-Sperandio B: Cyclosporine A responsive congenital nephrotic syndrome with single heterozygous variants in NPHS1, NPHS2, and PLCE1. Pediatr Nephrol 33: 1269-1272, 2018.

36. Ramanathan AS, Vijayan M, Rajagopal S, Rajendiran P and Senguttuvan P: WT1 and NPHS2 gene mutation analysis and clinical management of steroid-resistant nephrotic syndrome. Mol Cell Biochem 426: 177-181, 2017.
37. Kaverina NV, Eng DG, Largent AD, Daehn I, Chang A, Gross KW, Pippin JW, Hohenstein P and Shankland SJ: WT1 is necessary for the proliferation and migration of cells of renin lineage following kidney podocyte depletion. Stem Cell Rep 9: 1152-1166, 2017.
38. Matsumoto H, Sasaki K, Bessho T, Kobayashi E, Abe T, Sasazaki S, Oyama K and Mannen H: The SNPs in the ACACA gene are effective on fatty acid composition in holstein milk. Mol Biol Rep 39: 8637-8644, 2012.
39. Enserink JM and Kolodner RD: An overview of Cdk1-controlled targets and processes. Cell Div 5: 11, 2010.
40. van den Heuvel S: Cell-cycle regulation. WormBook 1-16, 2005.
41. Nishihara K, Masuda S, Nakagawa S, Yonezawa A, Ichimura T, Bonventre JV and Inui K: Impact of cyclin B2 and cell division cycle 2 on tubular hyperplasia in progressive chronic renal failure rats. Am J Physiol Renal Physiol 298: F923-F934, 2010.
42. Hinkes BG, Mucha B, Vlangos CN, Gbadegesin R, Liu J, Hasselbacher K, Hangan D, Ozaltin F, Zenker M, Hildebrandt F and Arbeitsgemeinschaft für Paediatrische Nephrologie Study Group: Nephrotic syndrome in the first year of life: Two thirds of cases are caused by mutations in 4 genes (NPHS1, NPHS2, WT1, and LAMB2). Pediatrics 119: E907-E919, 2007.
43. Lozupone S, Tummolo A, Aceto G, Francioso G, Gigante M, Gesualdo L, Messina G, De Palo T and Penza R: Nphs1 and Nphs2 mutations in congenital and Late-Onset childhood nephrotic syndrome. Pediatr Nephrol 24: 1805-1805, 2009.