The Key Candidate Genes in Tubulointerstitial Injury of Chronic Kidney Diseases (CKD) Patients as Determined by Bioinformatic Analysis

Wanpeng Wang
Lianshui County People's Hospital,Kangda college of Nanjing Medical Univercity
https://orcid.org/0000-0002-4201-895X
Jianxiao Shen
Shanghai Jiao Tong University School of Medicine Affiliated Renji Hospital
Yan Pan
Lianshui County People's Hospital,Kangda College of Nanjing Medical Univercity
Chaojun Qi
Shanghai Jiao Tong University School of Medicine Affiliated Renji Hospital
Senlin Liang
Lianshui People's Hospital, Kangda college of Nanjing Medical Univercity
Qiaolin Kuai
Lianshui People's Hospital
Zhi Zuo (yehui0129@163.com)
https://orcid.org/0000-0002-3413-431X

Research article

Keywords: Bioinformatics, Weighted gene correlation network analysis (WGCNA), Chronic kidney disease (CKD), Tubulointerstitial injury (TIL), Molecular mechanism

Posted Date: November 11th, 2019

DOI: https://doi.org/10.21203/rs.2.12859/v2

Abstract

Background: Tubulointerstitial injury (TIL) is common in chronic kidney disease (CKD), which in turn, leads to loss of renal function. The aims of present study were to screen critical genes with tubulointerstitial lesion in CKD by weighted gene correlation network analysis (WGCNA).

Methods: GSE104954 gene expression data downloaded from GEO database were used for analysis for differential expression genes(DEGs); Meanwhile, 90 expression data of tubulointerstitial samples in GSE47185 combined with clinic information were applied to WGCNA. According to the enrichment analysis and relationship with estimated glomerular filtration rate (eGFR), the eGFR-associated modules were defined, and the hub gene is selected according to the average intramodular connectivity (Kwithin); Clinical data from Nephroseq were obtained to further validate the relationship between CKD and hub genes.

Results: Totally 294 DEGs were screened. Using the WGCNA, we identified 15 co-expressed modules, the blue, brown and yellow modules exhibited strongly association with eGFR, and were significantly enriched in several signaling pathways that have been reported involved in pathogenesis of CKD. Furthermore, it was found that the 4 genes (PLG, ITGB2, CTSS and CCL5) was one of the DEGs which also be identified as hub genes according to Kwithin. Finally, the Nephroseq online tool showed that the tubulointerstitial expression levels of PLG significantly positively correlated with the eGFR, while ITGB2, CTSS and CCL5 connected negatively to the eGFR.

Conclusion: WGCNA is an efficient approach to system biology. By this procedure, the present study improved the understanding of the transcriptome status of renal tubulointerstitial injury in CKD.

Background

Chronic kidney disease (CKD) is a world-wide disease with high morbidity and mortality[1]. Its effect on the function and structure of kidney is irreversible and it usually results in the end-stage renal disease (ESRD). Therefore, early diagnosis and management of the risk factors for CKD are important for developing more accurate diagnosis and optimal therapeutic design in CKD patients [2].

Tubulointerstitial injury is a common characteristic of CKD regardless of its cause and is associated with its progression to end-stage renal disease[3]. Generally, renal tubular epithelial cells (RTEC), which result from multiple pathological lesions (hypoxia, hyperglycemia, inflammation) lose their regenerative capacity, thus promoting apoptosis and fibrosis[4]. Nevertheless, the mechanisms underlying interstitial injury have not yet been clearly defined, which is why significant genetic biomarkers for interstitial injury remain uncovered. Due to important role of TIL in CKD, many studies have tried to identify the mechanisms underlying interstitial injury to determine the pathogenesis of CKD at molecular level. However, given its complex pathologic characteristics, it is hard to predict or evaluate the local part of the biological system just by using a single gene. Therefore, constructing a comprehensive genetic network to explore the whole process of biological system in interstitial injury is of high clinic relevance.

Currently, bioinformatics analysis, which has been successfully applied in multiple cancers, offers an alternative method to explore the genetic biomarkers for predicting the prognosis of CKD[5, 6]. Using the conduction of weighted gene correlation network analysis (WGCNA), significantly correlated genes and their co-expression modules can be constructed from expression data. In our previous study[7], the key node gene of MAGI2 was identified which is involved in the regulation of cytoskeletal rearrangement in podocytes by means of construction of glomerular gene co-expression module. However, there was no specific link was identified between each module and estimated glomerular filtration rate (eGFR), that may be the altered glomerular gene expression is one of early stage event in CKD. In present study, due to detection of eGFR is generally accepted as a classical index to evaluate the level of kidney function [2]. we re-analyzed publicly available tubulointerstitial section gene expression datasets using WGCNA methods, aimed to explored the eGFR-associated gene modules or hub genes, which furthers the understanding of the transcriptome status of renal tubulointerstitial injury in CKD.

Methods

Microarray data information

Microarray data of the two gene expression profiles (GSE104954 and GSE47185) were downloaded from Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/)[8]. Samples from these two datasets were RNA extracted from the tubulointerstitial section and processed for hybridization on Affymetrix Human Genome U133 Plus 2.0 Array or Affymetrix Human Genome U133A Array, annotated using Human Entrez Gene ID custom CDF version 19. 190 samples from GSE104954 were used for differential expressed genes (DEGs) analysis, including32 lupus nephritis (LN), 25 IgA nephropathy (IgAN), 21 rapidly progressive glomerulonephritis (RPGN), 20 hypertensive nephropathy (HT), 18 membranous glomerulonephritis (MGN), 17 Diabetic Nephropathy (DN), 13 minimal change disease (MCD), 13 focal segmental glomerulosclerosis (FSGS), 6 thin basement membrane nephropathy (TMD), 4 focal segmental glomerulosclerosis combined with minimal change disease (FSGS-MCD) and 21 living donors (LD). 90 CKD samples from GSE47185 were used for WGCNA, including 17 DN, 12 FSGS, 4 FSGS-MCD, 12 MCD, 18 MGN, 21 RPGN and 6 TMD samples. Clinical characteristics of sample donors in GSE47185 (age, sex, serum creatinine and eGFR) were obtained from nephroseq online tools (https://www.nephroseq.org/resource/), which is a free platform used by the academic and non-profit community for integrative data mining of genotype/phenotype data[9].

Preprocessing of raw datasets

The raw expression data was preprocessed with R v3.4.1 (https://www.r-project.org/). Batches and platforms from different datasets were processed individually with their corresponding BrainArray CDF. Normalization was performed by Robust Muti-array Average (RMA) algorithmin using the Affy package, including background correction, quantile normalization and pro-summarization. After normalization, sample data were merged by common EntrezGene IDs, and the empirical Bayes algorithm (sva package) was used to correct the batch effect of the gene expression value[10].

Identification of DEGs

To compare expression value between each group, DEGs were performed by using the Linear Models for Microarray Data (LIMMA) package in Bioconductor. The corresponding p value of genes after T-test was obtained using the Benjamini-Hochberg procedure that generated the adjusted p-value (FDR), FDR < 0.05 and fold change (FC) ≥ 1.5 were defined as the criteria for DEGs. Heat map and volcano plot were generated by using the "heatmap.2" function of the R package "ggplots" [11].

Protein-protein interaction (PPI) network construction and hub module selection

Since proteins cannot work alone, it is necessary to study the interactions among proteins[12]. In this study, the protein-protein interaction (PPI) network of the DEGs or genes assigned to specific WGCNA modules was constructed with a confidence score > 0.4 by using the Search Tool for the Retrieval of Interacting Genes (STRING) database[13] (STRING, version 10.5, https://string-db.org/). Then, the plug-in Molecular Complex Detection (MCODE) of cytoscape was used to identify the most significant module in the PPI network. Cut-off criteria were as follows: degree = 5, k-core = 2, node score =0.2 and max. depth = 100.

Construction of weighted gene correlation network

To explore the relationship between the genes and to construct genetic co-expression network, WGCNA was performed to convert co-expression measures into connections weight or topology overlap measure [14], which are typically used for exploring correlations at transcription levels. It is widely accepted that genes involved in the same pathway or functionality tend to exhibit similar expression pattern [15, 16]. Therefore, the construction of a gene co-expression or correlation network facilitates the identification of genes with similar biological functions[17]. In this study, all genes of GSE47185 after preprocessing were inputted to construct weighted co-expression modules by using the WGCNA package in R language. Parameters were set as follows: TOMType = "unsigned" minModuleSize = 30, reassignThreshold = 0, mergeCutHeight = 0.25, numericLabels = TRUE, pamRespectsDendro = FALSE.

Enrichment analyses

GO (Gene Ontology) function and pathway enrichment analyses of DEGs or genes in each module were performed based on DAVID (Database for Annotation, Visualization and Integrated Discovery) (http://david.abcc.ncifcrf.gov/)[18]. FDR < 0.05 was considered significantly different.

Results

Recognition of DEGs in CKD and the enrichment of these genes

The DEGs of GSE104954 were analyzed after preprocessing and removing batch effects. A total of 294 DEGs containing 180 upregulated genes and 114 down regulated genes were identified (**Figure 1A, B**) (**Table. S1**). The DEGs are shown in the volcano map (**Figure 1A**), and are also visualized on a heatmap (**Figure 1B**). To further determine the biological functions of the DEGs, DAVID was used to explore the DEGs' biological significance including GO terms and KEGG pathways. Statistically significant enriched GO terms of

DEGs (FDR < 0.05) were identified (**Figure 1C** and **Table S2**). The results showed that the DEGs were mainly enriched in immune response, type I interferon signaling pathway, and innate immune response (**Figure 1C**). Meanwhile in the KEGG analysis, the DEGs were mainly enriched in the Staphylococcus aureus infection, Rheumatoid arthritis, and Type I diabetes mellitus. Identification of significant terms and pathways, where the altered genes in tubulointerstitial section of CKD patients were enriched, may further contribute to our understanding of the role of DEGs in tubulointerstitial injury of CKD patients.

Integration of protein–protein interaction (PPI) network and clusters analysis

To explore the expressive relationships among DEGs, we inputed the DEGs to STRING. Then, PPI networks were visualized using the cytoscape software. As a result, a PPI network with 230 nodes and 1135 edges was constructed (**Figure S1**). Then, plug-in MCODE was used to screen the clusters inside the PPI network. 8 clusters were finally identified in the PPI network. The top 4 clusters are shown in **Figure 1E-H**, while the functional annotation of the genes involved in each module were analyzed, respectively. Enrichment analysis showed that the genes in module 1-4 were mainly associated with type I interferon signaling pathway, fibrinolysis, and extracellular matrix organization, specifically inflammatory response and immune response in cluster rank 4 (**Table 1**). Previous studies have reported these GO terms to be related to kidney or tubular epithelial cell damage [19, 20]. Therefore, genes included in the four clusters were more suspectable to participate in the triggering of injury mechanism of renal tubule-interstitium compared to other DEGs. Also, construction of DEGs regulatory networks further promotes the understanding of the molecular mechanisms underlying CKD progression.

Construction of weighted gene correlation network analysis (WGCNA)

Weighted gene correlation network analysis (WGCNA) has been widely applied in cancer-related studies [21, 22]. In the present study, we performed WGCNA to explore the possible mechanisms of tubulointerstitium damage using genetic expression profile data of GSE47185, which included a total of 107 tubulointerstitium samples from kidney biopsy of various CKD patients. RMA method and Human Entrez Gene custom CDF annotation version 19 were used for the normalization. Each of the hybridization was separately normalized and the batch-corrected data were processed using Combat. 17 samples were removed because the lack of patients' clinical characteristics and the remaining samples were kept for further analysis. Consequently, 90 tubulointerstitial samples from CKD patients containing 12031 genes were used for the construction of WGCNA. **Figure 2A** shows the clusters of the 90 tubulointerstitial samples and clinical characteristics of the patients. No abnormal sample was found. The correlation network was constructed with WGCNA package in the R software. $\beta = 6$ was set in order to achieve a scale-free topology ($R^2 > 0.9$) (**Figure 2B**). Then the pairwise correlation was converted into an adjacency matrix of connection strengths using soft-thresholding approach (connection strength = $|\text{correlation}|^{\beta}$)[23] . A dissimilarity matrix based on topological overlap measure (TOM) was used to identify gene modules through a dynamic tree-cutting algorithm[22]. All modules were assigned to a different color (**Figure 2C**). Finally, these DGEs were categorized into 15 modules with sizes ranging from 36 to 3711 genes, and each module was assigned to a different color: turquoise, blue, brown, yellow, green, red, black, pink, magenta, purple, green-yellow, tan, salmon, cyan and midnight-blue, representing 3711,

2269, 1746, 1133, 214, 209, 129, 117, 107, 105, 86, 80, 75, 53 and 36 genes, respectively. Meanwhile, there were 1961 genes independent of any of 15 modules, which were classified as a module and assigned to grey color (**Table 2**, **Figure 2C**).

Weighted gene correlation network analysis (WGCNA) identifies critical modules correlating with estimated glomerular filtration rate (eGFR)

Using the WGCNA analysis, we identified 15 co-expressed modules in this study (**Figure 2C**). There were four very large modules (turquoise, blue, brown and yellow) indicating that the expression levels of a large number of genes were strongly correlated. The module eigengene (ME), which was calculated by the first principal component, was often used to represent each module. To explore the relationship between each module, the clustering analysis for ME was applied. We examined the association between ME and eGFR or CKD stage. Both indexes were crucial clinical measurements for CKD patients. The results showed that CKD stage cluster together with brown, and the distance between eGFR and blue module were very close (**Figure 2D**). Next, we examined the association between each of the modules and demographic and clinical parameter traits, including age, sex, eGFR, serum creatinine and disease status (**Figure 2F**). Interestingly, the turquoise module with the most genes exhibited a weak correlation with eGFR ($r = 0.220$, $p = 0.038$). The blue, brown and yellow modules were identified as strongly associated with eGFR ($|r|>0.5$, $P<0.001$) (**Table 2**). These results indicate that genes included in these three modules are more likely to participate in the progression of CKD and accelerate the reduction of renal function compared to other modules.

To further explore the biological functions and pathway of the genes in specific modules, enrichment analyses of GO and KEGG were performed. Whole enriched GO BP terms and KEGG pathway terms (FDR < 0.05) involved in each module were shown, respectively (Supplementary table S3 and supplementary table S4). Since DAVID does not allow lists of > 3000 symbols to be uploaded in order to maintain system stability, the top 3000 genes in turquoise with the highest intramodule connectivity were used for GO analysis. Top 8 GO BP terms and KEGG pathways for blue, brown and yellow module were shown, respectively (**Figure 3B-G**). The blue module, which contained 2269 genes where 55 of them were DEGs, exhibited significantly positive correlation with the eGFR. Bio-functions of this module were primarily involved in the oxidation-reduction process, fatty acid beta-oxidation, metabolic process and fatty acid beta-oxidation using acyl-CoA dehydrogenase, and so on (**Figure 3B**, **Table S3**). Pathway analysis exhibited a similar result according to which the genes in the module were mainly enriched in metabolic pathways, biosynthesis of antibiotics and carbon metabolism, and so forth. (**Figure 3D**, **Table S4**). The other two important modules, brown and yellow, had significantly negative correlation with eGFR. Genes in these two modules were mainly enriched in biological processes like cell division, mitotic nuclear division, cell proliferation, response to lipopolysaccharide, chemokine-mediated signaling pathway, adaptive immune response (**Figure 3C, E, F**). Interestingly, the KEGG pathways of yellow module also exhibited significant relevance to specific signaling pathways relevant to inflammation such as Chemokine signaling pathway or T cell receptor signaling pathway (**Figure 3G**, **Table S4**). Numerous previous studies have identified these GO terms or pathways as related to renal tubular mesenchymal damage.

Hub-based analysis

Since blue, brown and yellow modules exhibit strong association with eGFR, the gene significance (GS) and module membership (MM) were conducted. The GS and MM showed a very significant correlation (**Figure 4A-C**), indicating that genes involved in the blue, brown and yellow module are more likely to be highly correlated with eGFR. Highly connected hub nodes are central to the network's architecture[22] and some studies suggested that genes with more centralized position in the network are more likely to be key drivers to proper cellular function than peripheral genes[24]. In this study, the top 5% nodes with the highest intramodular connectivity (Kwithin) were defined as hub genes of each module [23]. Complete list of hub genes for each module is shown in supplementary table S5. Weighted co-expression networks of genes in blue, brown and yellow module with a weight (w) > a threshold of 0.2, 0.15, 0.2 are shown in **Figure 4E-G**, respectively (**Figure 4E-G**). It can be seen that the relative expression of most hub genes (**Figure 4E-F**, diamond nodes) has consistent trend and can exhibit high connectivity with neighboring genes whose functions are consistent with the analysis results of GO and KEGG. Taking the blue module, which shows functional enrichment in metabolic process and fatty acid metabolism as an example some of the hub genes have also been reported to participate in similar process. Some of these are SLC13 Family (SLC13A3)[25] or MSRA, where the latter one has a protective role in the progression of UUO-induced kidney fibrosis via suppression of fibrotic responses caused by oxidative stress[26]. Meanwhile, though some famous genes were not defined as hub genes in this study, they have evident high connectivity with other hub-genes, such as FABP1 [27], also known as L-FABP. Numerous studies have demonstrated that FABP1 is a promising biomarker for several kidney diseases, that can also attenuate renal injury [28, 29].

In order to further identify which genes may be potentially important for the development of CKD and to refine the hub genes in blue, brown and yellow module, the genes in these modules were mapped to STRING with a combined score > 0.4. After visualization of PPI networks via cytoscape software, top 10% genes with the highest degree of each module PPI network or DEGs PPI network were examined and intersected with hub genes for each module respectively (**Figure 4D**, **Figure S2-4**). As a result, four genes appeared as follows: PLG (blue module) and ITGB2, CTSS, CCL5 (yellow module). These four genes were also defined in the core clusters (**Figure 1F** and **Figure 1H**). Considering the core position in all networks and statistically significant expression in the tubulointerstitial of CKD patients, it is reasonable to assume that these four genes could be candidate biomarkers in the diagnosis of CKD and could serve as gene targets in the treatment of renal tubular mesenchymal damage.

The relationship between expression of candidate genes and clinical parameters of CKD patients

To further validate the relationship between CKD and these four candidate genes in patients, clinical data of CKD patients from Nephroseq (www.nephroseq.org, Apr 2018, University of Michigan, Ann Arbor, MI) were obtained for further analysis. The tubulointerstitial expression levels of these four genes were significantly correlated with eGFR of CKD patients (**Figure 5**). Generally, the expression levels of tubulointerstitial or renal PLG had significant positive correlation with the eGFR of patients receiving kidney transplants ($r = 0.691, p = 9.7e-10$, **Figure 5A**), patients with nephritic syndrome (NS) ($r = 0.559, p = 2.96e-5$, **Figure 5C**), IgAN ($r = 0.798, p = 2.99e-6$, **Figure 5D**) or with other CKD ($r = 0.593, p = 5.09e-19$, **Figure 5B**). The expression levels of tubulointerstitial or renal

ITGB2, CTSS or CCL5 had significant negative correlation with the eGFR of patients receiving kidney transplants, IgAN, Diabetic Nephropathy (DN) or other CKD affirmed in a series of different datasets (**Figure 5E-P**). Expression trend of these candidate genes was in accordance with each ME involved in different modules. Besides, the glomerulus expression levels of PLG were down-regulated in patients with IgAN, vasculitis, FSGS, LN, HT, DN (**Figure 6A**), while ITGB2, CTSS, CCL5 were up-regulated in these diseases (**Figure 6B-D**). This expression trend was in line with that of tubulointerstitial section. Moreover, the expression level of tubulointerstitial PLG had significant negative correlation with the urinary protein content of renal transplant patients ($r = -0.926$, $p = 0.008$, **Figure 6E**), while the tubulointerstitial ITGB2 and CTSS or blood ITGB2 expression were positively related to the urinary protein content of IgAN or DN patients. The expression level of PLG was also significantly down-regulated in patients with kidney transplant and renal dysfunction compared with patients with no rejection ($P = 2.60E-02$, $FC = -1.692$, **Table S6**). The integrated significant clinical parameters for PLG, ITGB2, CTSS, CCL5 are listed in Supplementary Table S6-9 respectively, with $p < 0.05$ and $|FC| \geq 1.5$ (or $r \geq 0.5$) set as cut-off criteria. All the reported results suggest that these four candidate genes do not merely serve as biomarkers for tubulointerstitial lesions, but also participate in the pathogenesis of kidney damage and thus are associated with progression of CKD.

Discussion

A total of 10-16% of the adult population in Europe, Asia and the USA are affected by chronic kidney disease (CKD) [30-32]. Physiological changes in the CKD can lead to a complex series of outcomes, including loss of kidney function, and end-stage renal failure (ESRD). In CKD, the irreversible and progressive loss of nephrons cause glomerular sclerosis, tubular atrophy, and tubulointerstitial injury. Pathological changes of cardinal features such as tubular cell atrophy and interstitial injury are common cardinal features of CKD as well as progressive renal disease, irrespective of the initial etiology [33, 34]. However, the molecular characteristics of these pathological events remain to be fully elucidated. Therefore, in the current study we employed various bioinformatics methods to explore the regulatory network of gene expression profiling in renal tubulointerstitial lesions in disease states.

The gene expression profile data that came from 169 tubulointerstitial samples of CKD (including IgAN, LN, FSGS and so on) patients and 21 living donors who served as controls were downloaded from the GEO dataset (GSE104954). As a result, a total of 294 DEGs were identified. Furthermore, GO biological processes (BP) enrichment analyses were performed followed by cluster analysis. The obtained results revealed that the biological effects of genes in top 4 clusters were mainly related to type I interferon signaling pathway, interferon-gamma-mediated signaling pathway, extracellular matrix organization, and inflammatory response, respectively. Consistent previous studies, these BP terms have been reported to be involved in a variety of renal diseases. For example, type I interferon signaling may represent the molecular signature of chronic antibody-mediated rejection (CAMR)[35], and it is also widely accepted that inflammatory response as well as extracellular matrix organization are at the core of renal fibrosis in CKD patients [33, 34]. At the same time, some of DEGs in these top four clusters have been reported to be associated with the development of CKD and could serve as promising diagnostic and prognostic molecular biomarkers and therapeutic targets. For example, epidermal growth

factor (EGF), which was significantly down-regulated in our analysis (log2FC = -0.893, FDR < 0.001) and was identified in core of rank 2 cluster, has been reported to be responsible for modulation of tissue response to injury in the kidney after tubulointerstitial damage[36], and can be used as a predictor of CKD progression in patients with kidney disease as well as urinary EGF (uEGF)[37]. While still some of other DEGs of the clusters have been rarely reported s related to CKD, such as IFITM1, IFIT1, IFIT2 or IFIT3, identification of DEGs in these 4 top clusters provides a set of useful targets for future investigations of the molecular mechanisms and biomarkers.

Gene expression analysis with microarrays has become one of the most widely used preaches for genome-wide functional analysis. Also, it is well accepted that co-expressed genes are more likely to be co-regulated and functionally related[38]. Therefore, identifying co-expressed protein-coding genes can help assign the functions to disease-causing gene[39]. However, few studies have used this approach to investigate the genes related to development of CKD. Objectives of this study were to gain molecular insights into the progression of CKD and to further the understanding of the relationship between genes expression data and eGFR, which most previous studies have used as the primary outcome measure, since they are considered to reflect the pathological process in the kidney. As a result, a total of 15 modules with sizes ranging from 36 to 3711 genes were obtained through weighted gene expression network analysis (WGCNA). The key findings in the current study revealed that the blue, brown and yellow modules had a stronger correlation with eGFR. Moreover, the GO and pathway enrichment analysis indicated that genes in blue module were mainly enriched in oxidation-reduction process and metabolic process, while genes in yellow module were mostly enriched in immune response or various signal pathways related to inflammatory response. These results are very much in line with multiple previous research indicating the importance of energy metabolism in chronic kidney disease[40].

As the genes in the blue, brown and yellow modules had the highest correlation with eGFR, the genes in this module were further selected for hub genes with a cut off of top 5% intramodule connectivity. The key candidate genes were identified from the intersection with the hub genes, top 10% degrees genes of DEG PPI network and module genes PPI network. Briefly, 4 key genes were excavated and chosen for further analysis, as follows: PLG, ITGB2, CTSS and CCL5. The protein encoded by PLG, also known as plasminogen, is a secreted blood zymogen that is activated by proteolysis and converted to plasmin and angiostatin. Plasmin dissolves fibrin in blood clots and is an important protease in many other cellular processes while angiostatin inhibits angiogenesis [41]. Recent studies on specific expression network analysis have shown that PLG had a key role in regulating the incidence of DN, while previous study has confirmed that Plg deficiency significantly attenuates tubular EMT and renal fibrosis in mice responding to unilateral ureteral obstruction (UUO). Similarly, key genes in yellow module have been identified as pro-inflammatory factors in a variety of human diseases, like CTSS, ITGB2 and CCL5 [42], which is consistent with the BP term or KEGG pathway enriched in yellow module. Meanwhile, some “famous” inflammatory factors have also been found to have high connectivity to these hub genes, such as CSF1R, TLR7, CASP1, etc. These observations further support the premise that proposed roles for hub genes of unknown function may be inferred from clusters of genes similarity expressed across many biological conditions.

It is important to note that since the four genes in our study were selected by calculating the degree in PPI network with a combined score > 0.4, which was based on experimental verification, there are many studies that have addressed the role of these key genes. Nonetheless, a lot of hub genes in the blue, brown and yellow module have received little attention, but may have an important role in CKD progression. For example, FBP1, TMSB10 and CD53 which were identified as DEGs with the maximum intramodular connectivity compared to other DEGs for each module. The validation of these three genes in Nephroseq also showed a strong correlation between tubulointerstitial expression levels of these three genes and eGFR in series of datasets (Figure S5). All these unreported hub genes could provide helpful information for potential biomarkers in further experiments due to their significant dysregulation in the CKD patients and since they share a high connectivity in eGFR-associated modules.

Admittedly, there were also some limitations to this study, since only the networks between hub genes and signaling pathways were explored in TIL while relevant molecular mechanisms and protein regulation were not. Moreover, our results are mostly based on literature searches or bioinformatics predictions; thus, further validation is required and necessary.

Conclusions

In short, the current study gave an explicit elucidation of dysregulated tubulointerstitial genes in CKD by analyzing the microarray datasets in GEO database. We presented a novel approach using WGCNA to explore the gene changes during the CKD process, especially with the occurrence of tubulointerstitial lesions. According to the network constructed by WGCNA, 15 modules were identified, and three modules (blue, brown and yellow) were found to be strongly correlated with eGFR and thus were selected for detailed analysis. The GO and pathway enrichment analysis indicated that genes in blue module were mainly enriched in oxidation-reduction process and metabolic process, which may further the understanding of pathways and genes underlying tubulointerstitial lesion of CKD. In the meantime, as DEGs and hub genes of these eGFR-associated modules, PLG, ITGB2, CTSS and CCL5 were found to be closely associated with eGFR. In sum, WGCNA results as an efficient approach to system biology. The present study improves the understanding of the transcriptome status of CKD and provides directions for the further investigations of the mechanisms of CKD.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable.

Availability of supporting data

The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests

The authors declare that they have no conflicts of interest.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 81600549), the Jiangsu Provincial Commission of Health and Family Planning (CN) (Z2018026), National Natural Science Foundation of China (Grant No. 81700586), the Huai'an Municipal Science and Technology Bureau (CN) (Grant No. HAB201737).

Authors' contributions

Zhi Zuo and Wanpeng Wang conceived and designed the study. Wanpeng Wang, Yan Pan, and Chaojun Qi performed the analysis procedures. Wanpeng Wang, Yan Pan, Senlin Liang and Qiaolin Kuai analyzed the results. Jianxiao Shen, Senlin Liang and Chaojun contributed analysis tools. Wanpeng Wang, Yan Pan and Chaojun Qi contributed to the writing of the manuscript. All authors reviewed the manuscript.

Acknowledgments

We would like to acknowledge the KEGG database developed by Kanehisa Laboratories. We would like to acknowledge the GEO and Nephroseq databases for free use.

References

1. Zaza G, Granata S, Rascio F, Pontrelli P, Dell'Oglio MP, Cox SN, Pertosa G, Grandaliano G, Lupo A: **A specific immune transcriptomic profile discriminates chronic kidney disease patients in predialysis from hemodialyzed patients.** *BMC Med Genomics* 2013, **6:**17.
2. Levey AS, Becker C, Inker LA: **Glomerular Filtration Rate and Albuminuria for Detection and Staging of Acute and Chronic Kidney Disease in Adults: A Systematic Review.** *Jama* 2015, **313:**837.
3. Liu BC, Tang TT, Lv LL, Lan HY: **Renal tubule injury: a driving force toward chronic kidney disease.** *Kidney Int* 2018, **93:**568-579.
4. Garrett SH, Clarke K, Sens DA, Deng Y, Somji S, Zhang KK: **Short and long term gene expression variation and networking in human proximal tubule cells when exposed to cadmium.** *BMC Med Genomics* 2013, **6 Suppl 1:**S2.
5. Wang Z, Lyu Z, Pan L, Zeng G, Randhawa P: **Defining housekeeping genes suitable for RNA-seq analysis of the human allograft kidney biopsy tissue.** *BMC Med Genomics* 2019, **12:**86.
6. Xia L, Su X, Shen J, Meng Q, Yan J, Zhang C, Chen Y, Wang H, Xu M: **ANLN functions as a key candidate gene in cervical cancer as determined by integrated bioinformatic analysis.** *Cancer Manag Res* 2018, **10:**663-670.

7. Zuo Z, Shen JX, Pan Y, Pu J, Li YG, Shao XH, Wang WP: **Weighted Gene Correlation Network Analysis (WGCNA) Detected Loss of MAGI2 Promotes Chronic Kidney Disease (CKD) by Podocyte Damage.** *Cell Physiol Biochem* 2018, **51:**244-261.
8. Olbryt M, Habryka A, Student S, Jarzab M, Tyszkiewicz T, Lisowska KM: **Global gene expression profiling in three tumor cell lines subjected to experimental cycling and chronic hypoxia.** *PLoS One* 2014, **9:**e105104.
9. Krochmal M, Kontostathi G, Magalhaes P, Makridakis M, Klein J, Husi H, Leierer J, Mayer G, Bascands JL, Denis C, et al: **Urinary peptidomics analysis reveals proteases involved in diabetic nephropathy.** *Sci Rep* 2017, **7:**15160.
10. Kupfer P, Guthke R, Pohlers D, Huber R, Koczan D, Kinne RW: **Batch correction of microarray data substantially improves the identification of genes differentially expressed in rheumatoid arthritis and osteoarthritis.** *BMC Med Genomics* 2012, **5:**23.
11. Reimers M, Carey VJ: **Bioconductor: an open source framework for bioinformatics and computational biology.** *Methods Enzymol* 2006, **411:**119-134.
12. Zhang K, Kong X, Feng G, Xiang W, Chen L, Yang F, Cao C, Ding Y, Chen H, Chu M, et al: **Investigation of hypoxia networks in ovarian cancer via bioinformatics analysis.** *J Ovarian Res* 2018, **11:**16.
13. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ: **STRING v9.1: protein-protein interaction networks, with increased coverage and integration.** *Nucleic Acids Res* 2013, **41:**D808-815.
14. Lue HW, Yang X, Wang R, Qian W, Xu RZ, Lyles R, Osunkoya AO, Zhou BP, Vessella RL, Zayzafoon M, et al: **LIV-1 promotes prostate cancer epithelial-to-mesenchymal transition and metastasis through HB-EGF shedding and EGFR-mediated ERK signaling.** *PLoS One* 2011, **6:**e27720.
15. Jiang J, Sun X, Wu W, Li L, Wu H, Zhang L, Yu G, Li Y: **Construction and application of a co-expression network in Mycobacterium tuberculosis.** *Sci Rep* 2016, **6:**28422.
16. Wan X, Kong Z, Chu K, Yi C, Hu J, Qin R, Zhao C, Fu F, Wu H, Li Y, Huang Y: **Co-expression analysis revealed PTCH1-3'UTR promoted cell migration and invasion by activating miR-101-3p/SLC39A6 axis in non-small cell lung cancer: implicating the novel function of PTCH1.** *Oncotarget* 2018, **9:**4798-4813.
17. Miller JA, Cai C, Langfelder P, Geschwind DH, Kurian SM, Salomon DR, Horvath S: **Strategies for aggregating gene expression data: the collapseRows R function.** *BMC Bioinformatics* 2011, **12:**322.
18. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA: **The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists.** *Genome Biol* 2007, **8:**R183.
19. Freitas MC, Uchida Y, Lassman C, Danovitch GM, Busuttil RW, Kupiec-Weglinski JW: **Type I interferon pathway mediates renal ischemia/reperfusion injury.** *Transplantation* 2011, **92:**131-138.
20. Zhang X, Ritter JK, Li N: **Sphingosine-1-Phosphate Pathway in Renal Fibrosis.** *Am J Physiol Renal Physiol* 2018.

21. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis.** *Stat Appl Genet Mol Biol* 2005, **4:**Article17.
22. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC Bioinformatics* 2008, **9:**559.
23. Li S, Li B, Zheng Y, Li M, Shi L, Pu X: **Exploring functions of long noncoding RNAs across multiple cancers through co-expression network.** *Sci Rep* 2017, **7:**754.
24. Yang Y, Han L, Yuan Y, Li J, Hei N, Liang H: **Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types.** *Nat Commun* 2014, **5:**3231.
25. Colas C, Pajor AM, Schlessinger A: **Structure-Based Identification of Inhibitors for the SLC13 Family of Na(+)/Dicarboxylate Cotransporters.** *Biochemistry* 2015, **54:**4900-4908.
26. Kim JI, Noh MR, Kim KY, Jang HS, Kim HY, Park KM: **Methionine sulfoxide reductase A deficiency exacerbates progression of kidney fibrosis induced by unilateral ureteral obstruction.** *Free Radic Biol Med* 2015, **89:**201-208.
27. Katagiri D, Hamasaki Y, Doi K, Negishi K, Sugaya T, Nangaku M, Noiri E: **Interstitial renal fibrosis due to multiple cisplatin treatments is ameliorated by semicarbazide-sensitive amine oxidase inhibition.** *Kidney Int* 2016, **89:**374-385.
28. Dihazi H, Koziolek MJ, Datta RR, Wallbach M, Jung K, Heise D, Dihazi GH, Markovic I, Asif AR, Muller GA: **FABP1 and FABP3 Have High Predictive Values for Renal Replacement Therapy in Patients with Acute Kidney Injury.** *Blood Purif* 2016, **42:**202-213.
29. Xu Y, Xie Y, Shao X, Ni Z, Mou S: **L-FABP: A novel biomarker of kidney disease.** *Clin Chim Acta* 2015, **445:**85-90.
30. Wen CP, Cheng TY, Tsai MK, Chang YC, Chan HT, Tsai SP, Chiang PH, Hsu CC, Sung PK, Hsu YH, Wen SF: **All-cause mortality attributable to chronic kidney disease: a prospective cohort study based on 462 293 adults in Taiwan.** *Lancet* 2008, **371:**2173-2182.
31. Murphy D, McCulloch CE, Lin F, Banerjee T, Bragg-Gresham JL, Eberhardt MS, Morgenstern H, Pavkov ME, Saran R, Powe NR, et al: **Trends in Prevalence of Chronic Kidney Disease in the United States.** *Ann Intern Med* 2016, **165:**473-481.
32. Hallan SI, Coresh J, Astor BC, Asberg A, Powe NR, Romundstad S, Hallan HA, Lydersen S, Holmen J: **International comparison of the relationship of chronic kidney disease prevalence and ESRD risk.** *J Am Soc Nephrol* 2006, **17:**2275-2284.
33. Nath KA: **Tubulointerstitial changes as a major determinant in the progression of renal damage.** *Am J Kidney Dis* 1992, **20:**1-17.
34. Remuzzi G, Bertani T: **Pathophysiology of progressive nephropathies.** *N Engl J Med* 1998, **339:**1448-1456.
35. Rascio F, Pontrelli P, Accetturo M, Oranger A, Gigante M, Castellano G, Gigante M, Zito A, Zaza G, Lupo A, et al: **A type I interferon signature characterizes chronic antibody-mediated rejection in kidney transplantation.** *J Pathol* 2015, **237:**72-84.

36. Torres DD, Rossini M, Manno C, Mattace-Raso F, D'Altri C, Ranieri E, Pontrelli P, Grandaliano G, Gesualdo L, Schena FP: **The ratio of epidermal growth factor to monocyte chemotactic peptide-1 in the urine predicts renal prognosis in IgA nephropathy.** *Kidney Int* 2008, **73:**327-333.
37. Ju W, Nair V, Smith S, Zhu L, Shedden K, Song PXK, Mariani LH, Eichinger FH, Berthier CC, Randolph A, et al: **Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker.** *Sci Transl Med* 2015, **7:**316ra193.
38. Stuart JM, Segal E, Koller D, Kim SK: **A gene-coexpression network for global discovery of conserved genetic modules.** *Science* 2003, **302:**249-255.
39. Signal B, Gloss BS, Dinger ME: **Computational Approaches for Functional Prediction and Characterisation of Long Noncoding RNAs.** *Trends Genet* 2016, **32:**620-637.
40. Mount PF, Power DA: **Balancing the energy equation for healthy kidneys.** *J Pathol* 2015, **237:**407-410.
41. Zhao S, Dorn J, Napieralski R, Walch A, Diersch S, Kotzsch M, Ahmed N, Hooper JD, Kiechle M, Schmitt M, Magdolen V: **Plasmin(ogen) serves as a favorable biomarker for prediction of survival in advanced high-grade serous ovarian cancer.** *Biol Chem* 2017, **398:**765-773.
42. Nair S, Lee YH, Rousseau E, Cam M, Tataranni PA, Baier LJ, Bogardus C, Permana PA: **Increased expression of inflammation-related genes in cultured preadipocytes/stromal vascular cells from obese compared with non-obese Pima Indians.** *Diabetologia* 2005, **48:**1784-1788.

Tables

Table 1. The top three Gene Ontology (GO) biological process (BP) terms in the enrichment analysis of DEGs for each cluster.

Module	BP Term	Count	FDR	Genes
cluster 1	type I interferon signaling pathway	15	2.98E-26	EGR1, IFITM1, OAS1, HLA-B, STAT1, PSMB8, HLA-F, ISG20, IFIT3, IFIT2, IFIT1, IFI27, ISG15, MX1, GBP2
	defense response to virus	13	8.35E-16	IFIT3, PLSCR1, IFIT2, IFIT1, ISG15, IFITM1, OAS1, IFI16, MX1, STAT1, TRIM22, ISG20, GBP1
	response to virus	11	1.44E-13	DDX58, IFIT3, IFIT2, IFIT1, IFIH1, IFITM1, OAS1, IFI44, MX1, TRIM22, ISG20
cluster 2	interferon-gamma-mediated signaling pathway	7	5.55E-07	OAS1, HLA-B, STAT1, TRIM22, GBP2, GBP1, HLA-F
	platelet degranulation	11	4.88E-16	FGG, ALB, FGB, CLU, SERPINA3, APOH, ACTN1, SERPING1, HRG, EGF, PLG
	fibrinolysis	5	4.26E-06	FGG, FGB, SERPING1, HRG, PLG
cluster 3	extracellular matrix organization	8	3.48E-09	COL4A2, COL4A1, LUM, COL3A1, TGFBI, COL1A2, POSTN, VCAN
	collagen fibril organization	4	1.55E-03	P4HA1, LUM, COL3A1, COL1A2
	collagen catabolic process	4	7.01E-03	COL4A2, COL4A1, COL3A1, COL1A2
cluster 4	inflammatory response	7	4.31E-04	CCL20, C3, ANXA1, CCL19, ITGB2, CXCL6, CCL5
	immune response	4	6.84E-03	CCL20, ANXA1, CCL19, CCL5
	chemokine-mediated signaling pathway	6	2.12E-02	CCL20, C3, CCL19, CXCL6, CTSS, CCL5

Table 2. Overview of 15 modules and grey module (including genes that cannot be assigned to any modules) constructed by Weighted gene correlation network analysis (WGCNA). (DEG: the differentially expressed genes (DEGs) through the above analysis; hub: the 5% of genes with the highest connectivity was defined as hub genes in each module; Hub-DEG: the intersection of DEGs and hub; Bold represent the top 10% genes with highest degree in PPI network of DEGs; Underline represent the top 10% genes with highest degree in PPI network of genes included in each module)

Module	Size	DEG	Hub-DEG	Hub-DEG	eGFR
	N	N	N	Gene Symbol	r (p-value)
turquoise	3711	5	0		0.220 (3.75E-02)
blue	2269	55	9	FBP1 DPYS PLG MME AZGP1 XPNPEP2 MARC2 UPB1 AFM	**0.569 (4.92E-09)**
grey	1961	18	0		0.203 (5.48E-02)
brown	1746	110	36	TMSB10 HLA-B VIM ANXA2 TUBA1A WFDC2 MMP7 S100A11 S100A13 TIMP1 TPM1 PLSCR1 COL4A1 SERPING1 ARPC1B SERPINA3 CEBPB COL4A2 C1S MYOF PSMB8 NNMT PLP2 HN1 TNFAIP8 ANXA1 NMI DSE CKLF ACTN1 PSMB10 LPCAT1 ADAMTS1 QPCT ITGB6 GBP2	**-0.640 (1.11E-11)**
yellow	1133	79	24	CD53 LAPTM5 FCER1G MS4A6A TYROBP CD48 IFI16 RGS10 HCLS1 TRAC CD52 PYCARD ITGB2 CD3D EVI2A LY96 CTSS CCL5 GZMA CASP1 EVI2B MNDA PLAC8 CSTA	**-0.527 (9.72E-08)**
green	214	0	0		-0.002 (9.84E-01)
red	209	0	0		-0.267 (1.10E-02)
black	129	2	0		**0.366 (3.83E-04)**
pink	117	1	0		-0.222 (3.57E-02)
magenta	107	2	0		0.146 (1.71E-01)
purple	105	0	0		0.249 (1.78E-02)
greenyellow	86	3	1	TSPAN8	-0.151 (1.55E-01)
tan	80	3	1	MOXD1	**-0.347 (8.12E-04)**
salmon	75	1	0		**0.427 (2.67E-05)**
cyan	53	3	0		0.282 (7.07E-03)
midnightblue	36	12	2	LIC5 FGF1	0.164 (1.23E-01)

Figures

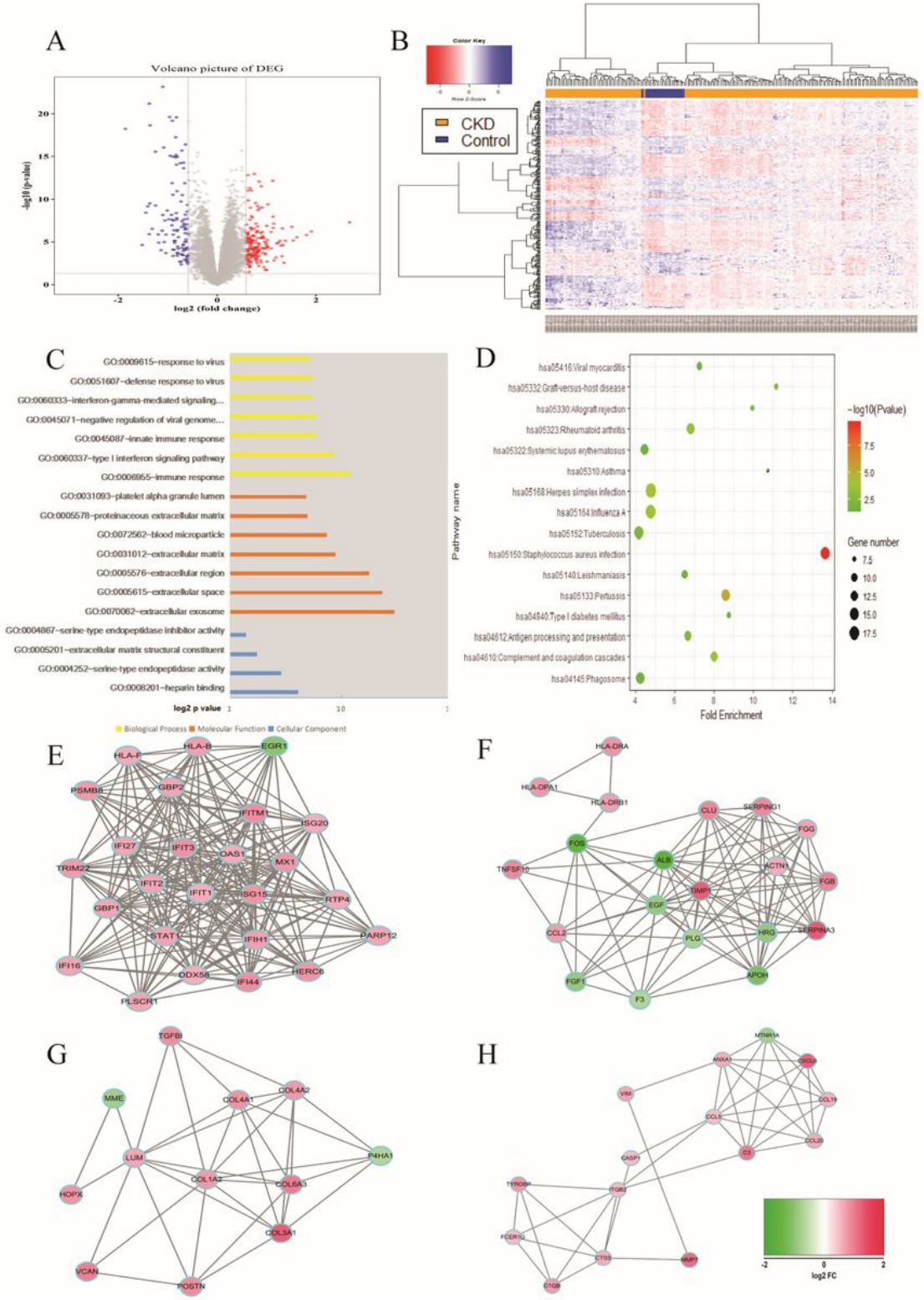

Figure 1

Differentially expressed genes (DEGs) identified in GSE104954 dataset and construction of protein–protein interaction (PPI) network. (A) Volcano map of tubulointerstitial DEGs between Chronic kidney disease (CKD) and normal living donors (LD). (B) Heatmap of the DEGs. (C) Top 8 Gene Ontology (GO) terms in the enrichment analysis of DEGs. (D) Top 8 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in the enrichment analysis of DEGs. (E-H) Construction of protein–protein interaction (PPI)

network of DEGs; Top 4 clusters from the PPI network: (E) cluster rank 1, (F) cluster rank 2, (G) cluster rank 3, (H) cluster rank 4.

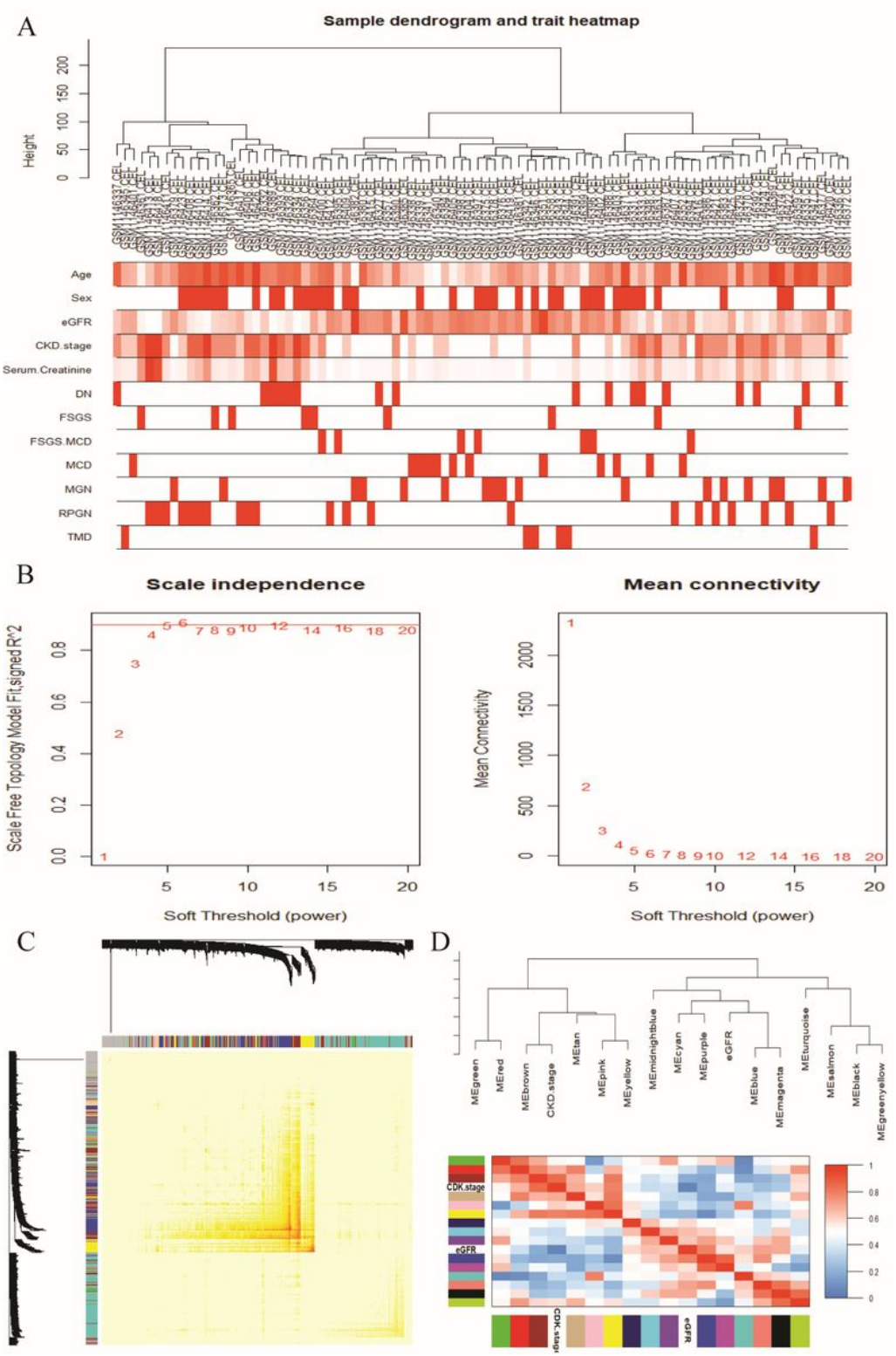

Figure 2

Weighted gene correlation network analysis (WGCNA). (A) Cluster results and clinic information on data samples. (B) Determination of parameter β of the adjacency function in the WGCNA algorithm. The left panel shows the scale-free topology fitting index (R2, y-axis) as a function of the soft-thresholding power

(x-axis). The right panel displays the mean connectivity (degree, y-axis) as a function of the soft-thresholding power (x-axis). Red Arabic numerals in the panels denote different soft-thresholds. The red line in left panel means R2 = 0.9. There is a trade-off between maximizing scale-free topology model fit (R2) and maintaining a high mean number of connections. Thus, we set β = 6. (C) Construction of the gene co-expression network. Each color represents a certain gene module. (D) Clustering tree based on the module eigengenes of modules.

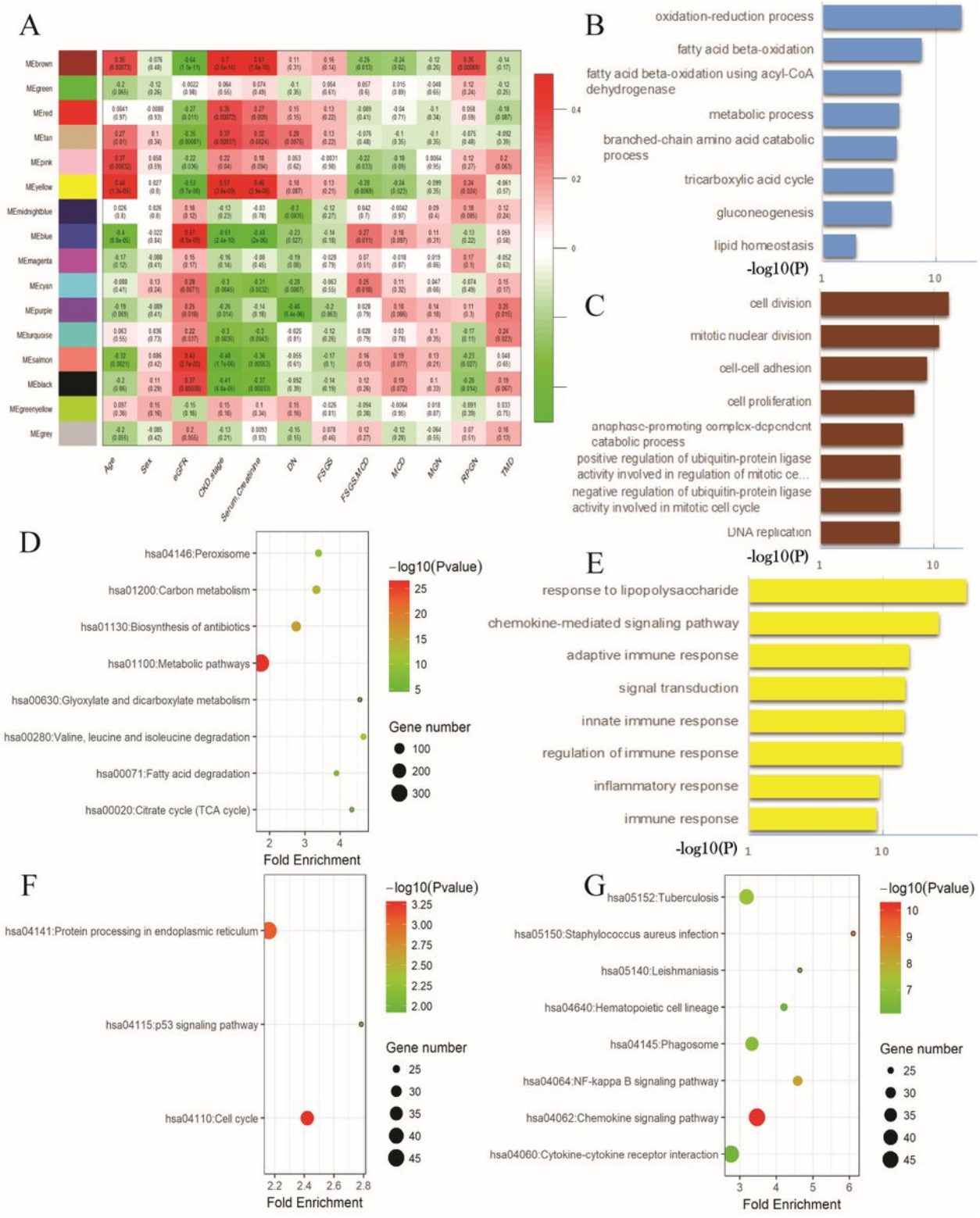

Figure 3

Weighted gene correlation network analysis (WGCNA) Identifies Critical Modules Correlating with estimated glomerular filtration rate (eGFR). (A) Correlation between modules and traits. The upper number in each cell refers to the correlation efficiency of each module in the trait, and the lower number is the corresponding p-value. Among them, blue, brown and yellow modules were the most relevant modules with eGFR. (B, C, E) Top 8 Gene Ontology (GO) biological process (BP) terms in the enrichment analysis of genes in blue (B), brown (C), yellow (E) modules. (D, F, H) The top 8 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in the enrichment analysis of genes in blue (D), brown (F), yellow (H) modules.

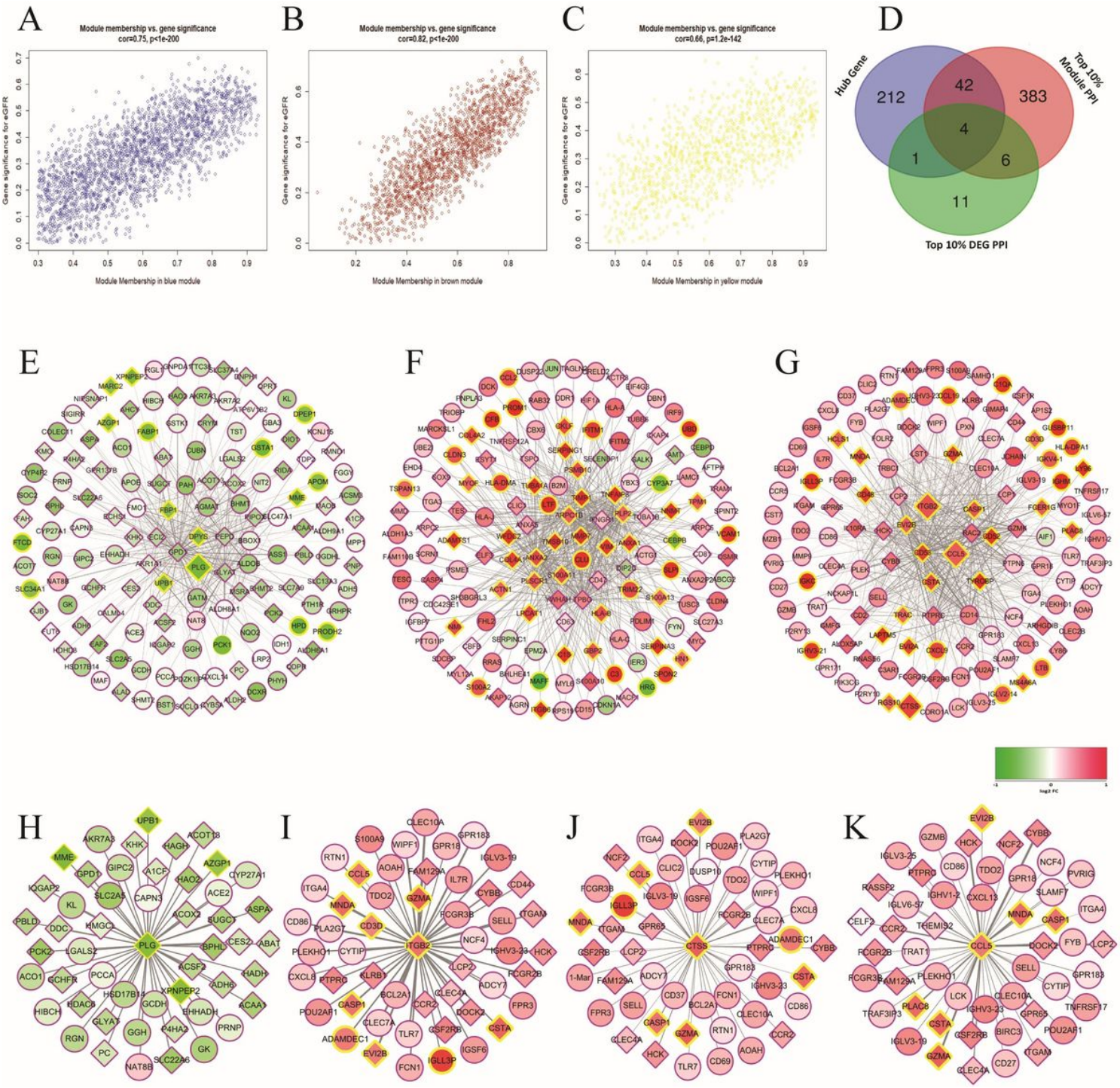

Figure 4

Hub-based analysis. (A-C) scatter plots of gene significance (GS) for estimated glomerular filtration rate (eGFR) versus the module membership (MM) in the blue (A), brown (B) and yellow (C) module. (D) The Venn plot of hub genes, top 10 genes with the highest degree in DEGs and module genes PPI network. (E-G) The cytoscape network visualization of genes in blue (E), brown (F), yellow (G) module, in which only edges with weight (w) above a threshold of 0.2, 0.15, 0.2 are displayed, respectively. The nodes with yellow border denote the hub genes which were the 5% of genes with the highest connectivity; The diamond nodes denote the differentially expressed genes (DEG), The color of nodes represent the log2Fold Chang (FC) value. (H-K) The top 50 adjacent genes with highest intramodule connectivity to candidate genes: PLG (H), ITGB2 (I), CTSS (J), CCL5 (K).

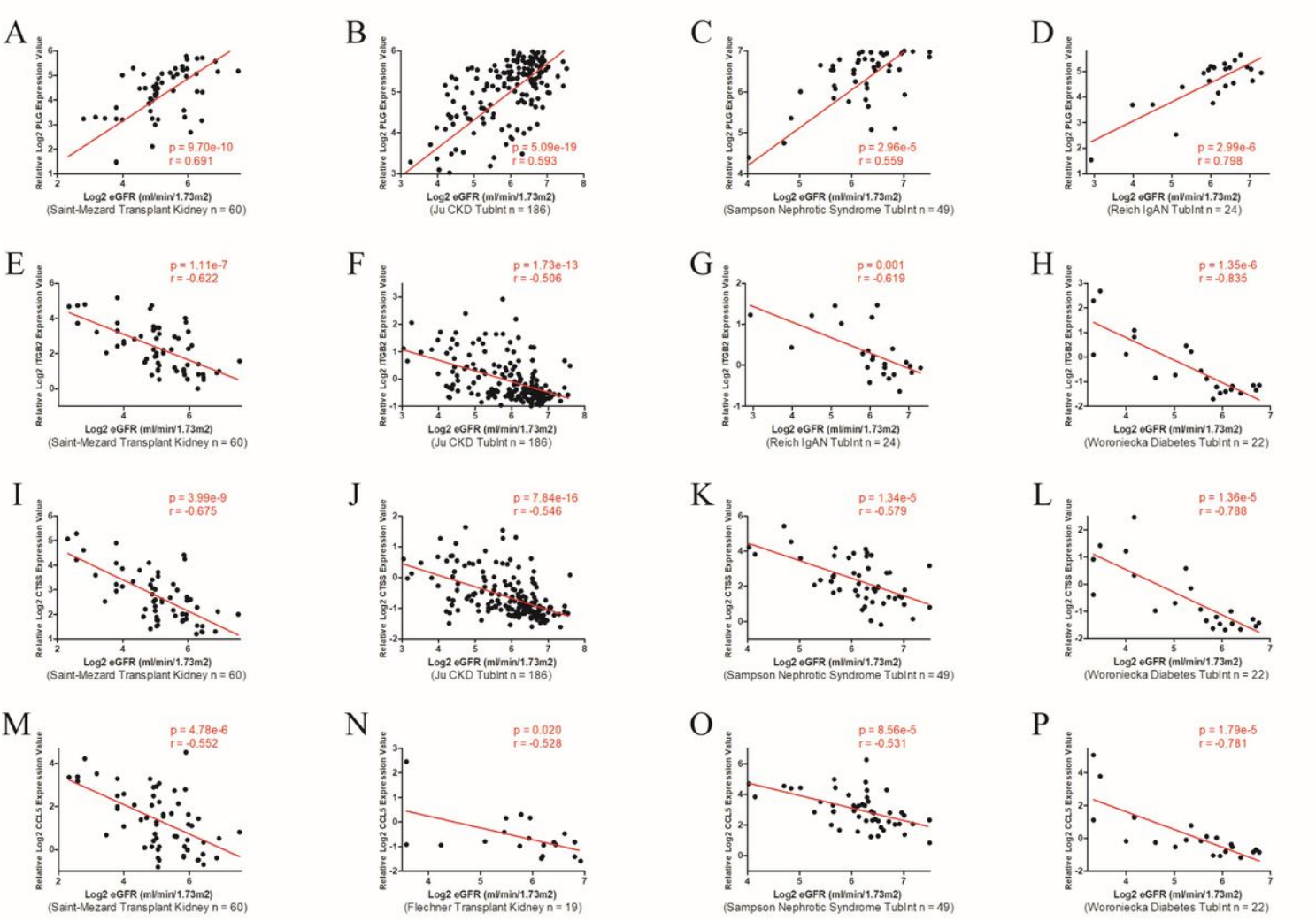

Figure 5

The correlation between estimated glomerular filtration rate (eGFR) and the expression of candidate genes validated in series of datasets. (A-D) The expression level of tubulointerstitial PLG is significantly positively correlated with eGFR of patients with receiving kidney transplants (A), nephritic syndrome (NS) (C), IgA nephritis (IgAN) (D) or multiple Chronic kidney disease (CKD) (B); (E-P) the expression level of tubulointerstitial ITGB2 (E-H), CTSS (I-L) or CCL5 (M-P) is significantly nagetively correlated with the eGFR

of patients with receiving kidney transplants (E, I, M, N), IgAN (G), Diabetic Nephropathy (DN) (H, L, P), NS (K,O) other CKD (F, J).

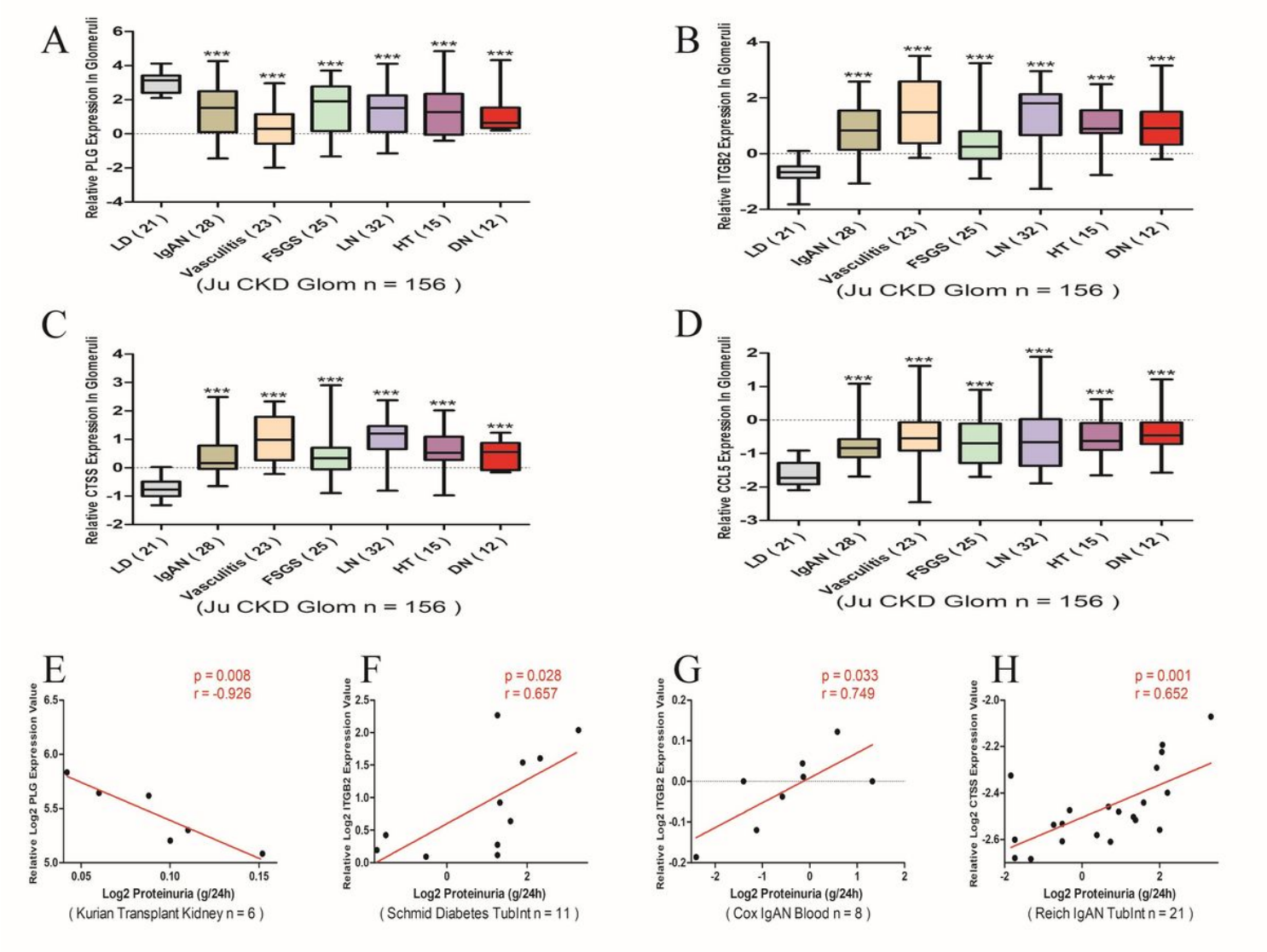

Figure 6

The expression levels and clinical significance of candidate genes in Chronic kidney diseases (CKD). (A-D) The relative expression levels of glomerulus PLG (A), ITGB2 (B), CTSS (C) and CCL5 (D) in the various Chronic kidney diseases (CKD), ***p<0.001. (E) the expression level of tubulointerstitial PLG is significantly negatively correlated with the urinary protein content of renal transplant patients. (F-G) The expression of the tubulointerstitial ITGB2 (F) and CTSS (H) or blood ITGB2 (G) expression positively to the urinary protein content of IgA nephritis (IgAN) (G,H) or Diabetic Nephropathy (DN) (F) patients.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx