Original Article

Identification of the underlying mechanisms of pathogenesis in chronic kidney disease based on bioinformatics analysis

Chao Zhou [1,2]，Mengmeng Zhang[3]，Chundi Gao[1]，Yingying Li[1]，Yan Yao[1]，Farong Zhang[4*]

Abstract

Background：Chronic Kidney Disease (CKD) presents with a poor prognosis and limited treatment options. This study aimed to explore the key genes expressed in CKD, to identify new pathways and drug targets and to provide insight for subsequent molecular studies in terms of the potential mechanisms of CKD.

Methods: Four microarray data sets GSE15072, GSE41030, GSE66494, and GSE98603 were analyzed using the Expression Omnibus database (GEO). The SVA package in the R software was used to merge and correct batch effects in the four datasets and the limma package was used to screen the differentially expressed genes (DEGs) between CKD and normal samples. There were instances where the Gene Ontology (GO) and pathways indicated that the DEGs were associated with different pathways such as "platelet activation", "regulation of wound healing", "platelet degranulation", "focal adhesion", and the "PI3K-Akt signaling pathway". The GO plot package was used to perform the GO analysis of the DEGs and to generate a volcano map for DEGs and the Heml software was used to illustrate a heat map of the DEGs. The cluster Profiler package and the Kyoto Encyclopedia of Genes and Genomes（KEGG) website was used to perform pathway enrichment analysis and data on the differential genes were used to construct a protein-protein interaction (PPI) network to identify the central gene.

Results：Following removal of the batch effect, data on 10937 genes and 88 DEGs was obtained. A total of 53 differential genes were screened and the expression levels of 28 genes were upregulated and those of 25 genes were downregulated.

Discussions: GO analysis revealed that the response to potassium ion was most significant. KEGG pathway analysis helped identify four important pathways, namely the chemokine signaling pathway, pancreatic secretion, protein digestion and absorption and the regulation of actin cytoskeleton pathway. with the establishment of the PPI network, a central gene with high connectivity was selected, namely the Serum Amyloid A1 (SAA1) gene.

Conclusions：This study indicated that the SAA1 gene plays an extremely important role in the pathogenesis of CKD. The new pathway identified in this study may provide new insights into the underlying mechanism of CKD at the molecular level. [*Ethiop. J. Health Dev*. 2022; 36(1):000-000]

Keywords：Chronic Kidney Disease (CKD), Bioinformatics Analysis, Central Gene, Molecular Pathway, Pathogenesis.

Introduction

CKD is an essential potential risk factor for the development of cardiovascular diseases and cerebrovascular diseases (1，2) and has recently attracted considerable attention. Approximately 10.8% of adults in China present with CKD (3). Western medicine-based treatment methods, which include the use of angiotensin receptor blockers and angiotensin-converting enzyme inhibitors (ACEI) are commonly used to treat CKD; however, the efficacy is not satisfactory for application in clinical practice. As a disease involving multiple pathways, CKD treatment through the targeting of a single pathway does not result in a satisfactory therapeutic effect, causing

[1]Department of First Clinical College，Shandong University of Traditional Chinese medicine，Jinan，China;
[2]Department of Oncology，Weifang Hospital of Traditionl Chinese Medicine，Weifang，China;
[3]Department of Anesthesiology，Affliated Hospital of Weifang Medical University，Weifang，China;
[4]Department of Nephrology，Affiliated Hospital of Shandong University of TCM，Jinan，China;

considerable personal, social, and economic burdens. Therefore, exploration of the risk factors that affect the occurrence and development of CKD is of substantial interest for the nephrology community. With the application and development of epidemiological research, data on the induction and regulation of risk factors can be used in the early diagnosis and assessment of CKD in the clinic. A better understanding of the pathogenesis of this complex disease is essential for formulating successful treatment strategies.

In the present study, four original array datasets were selected from the GEO database, namely gene expression sequences GSE15072, GSE41030, GSE66494, and GSE98603. Bioinformatics methods were used to identify DEGs, and to perform the GO function annotation analysis and the KEGG pathway enrichment analysis. The purpose of these analyses was to study the pathogenesis of CKD at the gene level and to provide new options for exploring the ultimate drug treatment target. There were cases where the gene ontology and pathways indicated that the DEGs were associated with other different pathways, such as "platelet activation", "regulation of wound healing", "platelet degranulation", "focal adhesion", and "PI3K-Akt signaling pathway". Furthermore, there may be other significant transcription factors and microRNAs that may also be responsible for regulating the common DEGs symptoms. Additionally, a PPI network was established to reveal key genes. These results provide new insights into the pathogenesis of the disease, which may provide guidance for follow-up studies on CKD.

Methods

Acquisition of gene chip-based data

The gene expression profiles of GSE15072, GSE41030, GSE66494, and GSE98603 were derived from the GEO Database (https://www.ncbi.nlm.nih.gov/geo/). GSE15072 was based on the GPL96 platform (Affymetrix Human Genome U133A Array), GSE41030 was based on the GPL15950 platform (Agilent-028680 SurePrint G3 Human Exon 2x400K Microarray), GSE66494 was based on the GPL6480 platform (Agilent-014850 Whole Human Genome Microarray 4x44K G4112F), and GSE98603 was based on the GPL 13497 platform（Agilent-026652 Whole Human Genome Microarray 4x44K v2). This study used the SVA package in the R software to merge and correct batch effects in the four data sets and the limma package to screen the Differentially Expressed Genes (DEGs) between CKD and normal samples (1-5).

The GSE15072 file consisted of data on twenty-one CKD patients and eight normal samples, the GSE41030 file consisted of three CKD patients and three normal samples, the GSE66494 file consisted of fifty-three CKD patients and eight normal samples, and the GSE98603 file consisted of nine CKD patients and nine normal samples (Table 1). All samples were obtained from the kidney tissue of Homosapiens. All the data is freely available online, and ethical approval was not required for the present study.

Table 1. The four microarray databases derived from the GEO database

Dataset ID	Subjects	Organization name	country	Number of samples(CKD/NC)
GSE15072	kidney	University of Bari	Italy	21/8
GSE41030	kidney	U1076	France	3/3
GSE66494	kidney	Kyushu University Hospital	Japan	53/8
GSE98603	kidney	Lyon University Hospital	France	9/9

Abbreviations: GEO (Gene Expression Omnibus), **CKD** (Chronic Kidney Disease), NC (Normal Control).

Data consolidation and batch correction

The R software (version 4.0.1) was used for data analysis. The batch Normalize package was used to merge the GSE15072, GSE41030, GSE66494, and GSE98603 datasets, and the SVA package (4) was used to perform analysis of the merged datasets and batch

correction.

Screening of Differentially Expressed Genes (DEGs)

The limma package (5) was used to analyze GSE15072, GSE41030, GSE66494, and GSE98603, and the corrected P < 0.05, |logFC | > 1 was used as the screening criteria to select the DEGs. The ggplot2 package (6) and the Heml software were used to illustrate a volcano map and a heat map for the DEGs.

GO and KEGG pathway analyses

The GO plot package (7) was used to perform GO function annotation and visual analysis of the differential genes; the cluster Profiler package (8) and the KEGG website (david version 6.8, https://david.ncifcrf.gov/) were used to perform the differential gene pathway enrichment analysis (using p < 0.05, gene count ≥ 3 as screening conditions).

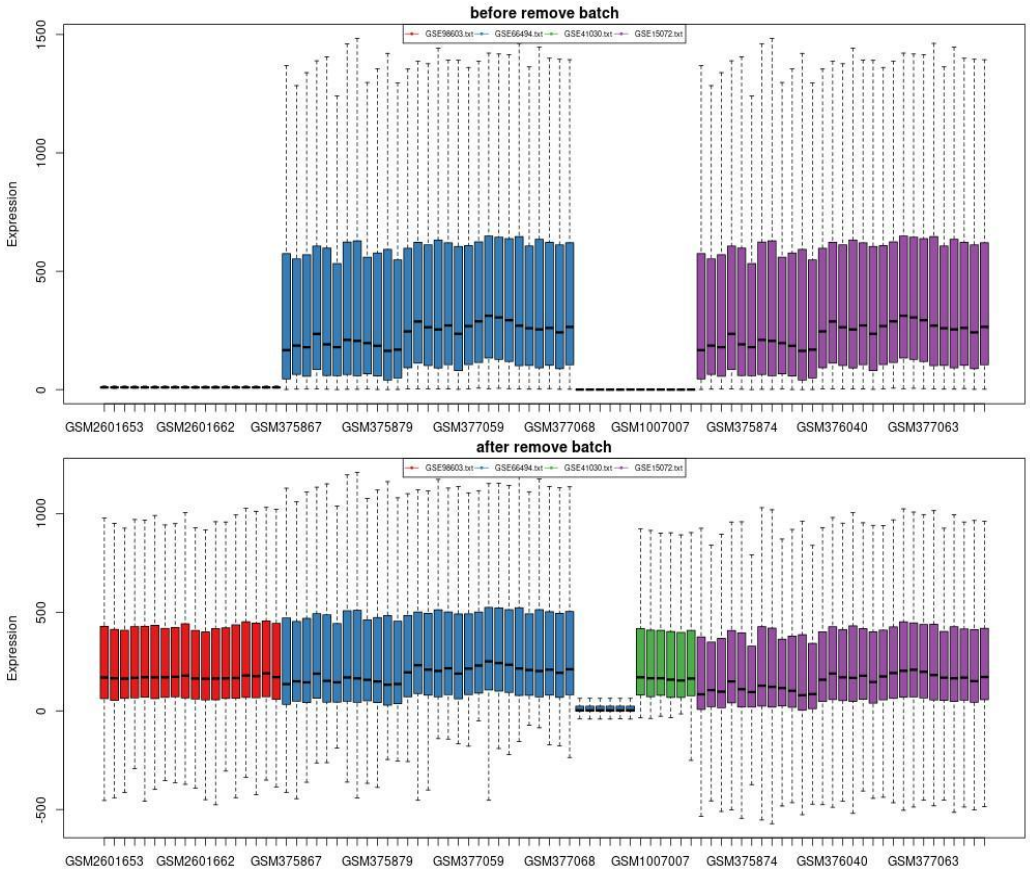
PPI network construction and central gene analysis

A search tool was used to explore the interactive gene (string, version 11.0, http://string-db.org/) online database (9), and the protein interaction network was predicted. The count package was used in R to determine PPI network-related protein pairs to obtain information on the central gene.

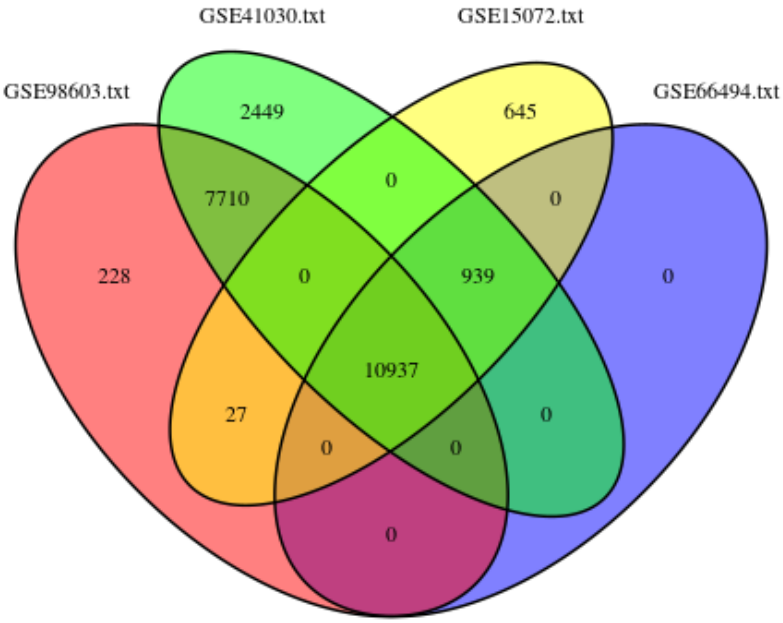
Results

Batch removal effect

According to the corrected cutoff criteria of p < 0.01 and | log2fc | ≥ 2.0, the R language package batch Normalize was used to merge GSE15072, GSE41030, GSE66494, and GSE98603 in order to obtain a large matrix, and the data was then merged according to the row name; subsequently, the R software package SVA was used after removal of the batch effect; data on the combat function of 10937 genes and 88 GSM were obtained.



(a)

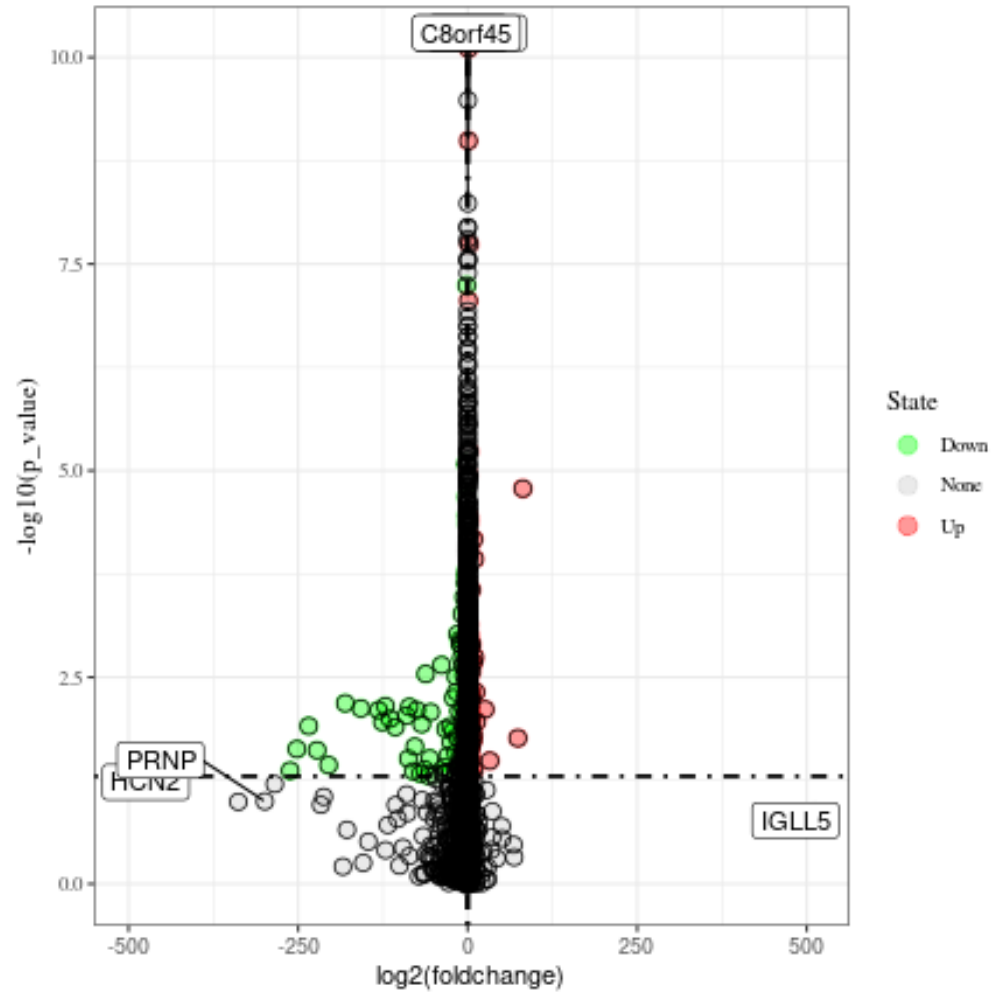


(b)

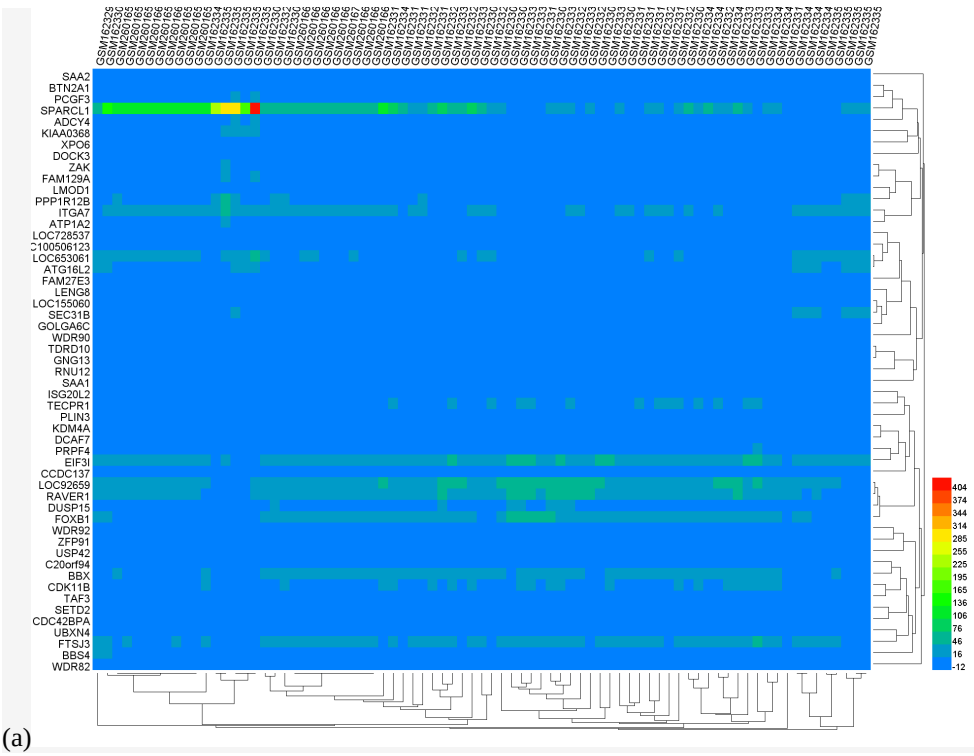
Figure 1. The effect diagram and Venn diagram illustrating data intersection after the batch removal effect. a) Diagram representing data after the batch removal effect; b) Venn diagram-based on intersection of each data set.

Differential gene expression and data visualization

After performing batch correction and standardization of the GSE15072, GSE41030, GSE66494, and GSE98603 datasets, a total of 53 differential genes were screened, of which the expression of 28 genes was found to be upregulated and that of 25 genes was found to be downregulated. A volcano map (Figure 2a) and a heat map (Figure 2b) were used to display the upregulated and downregulated Differentially Expressed Genes (see attachments up.xls and down.xls for specific gene names).



(a)

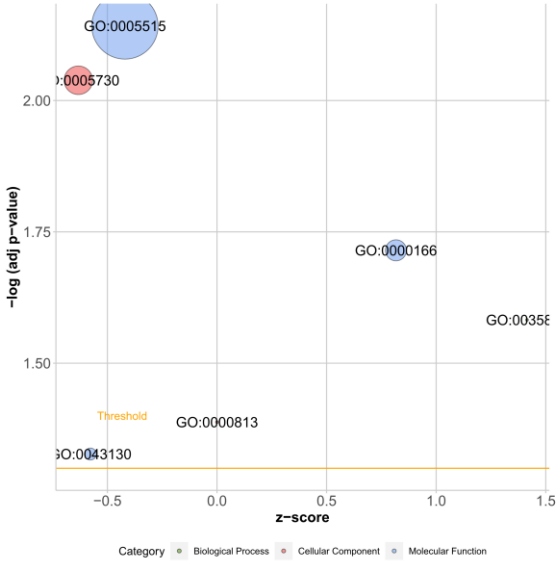


(b)

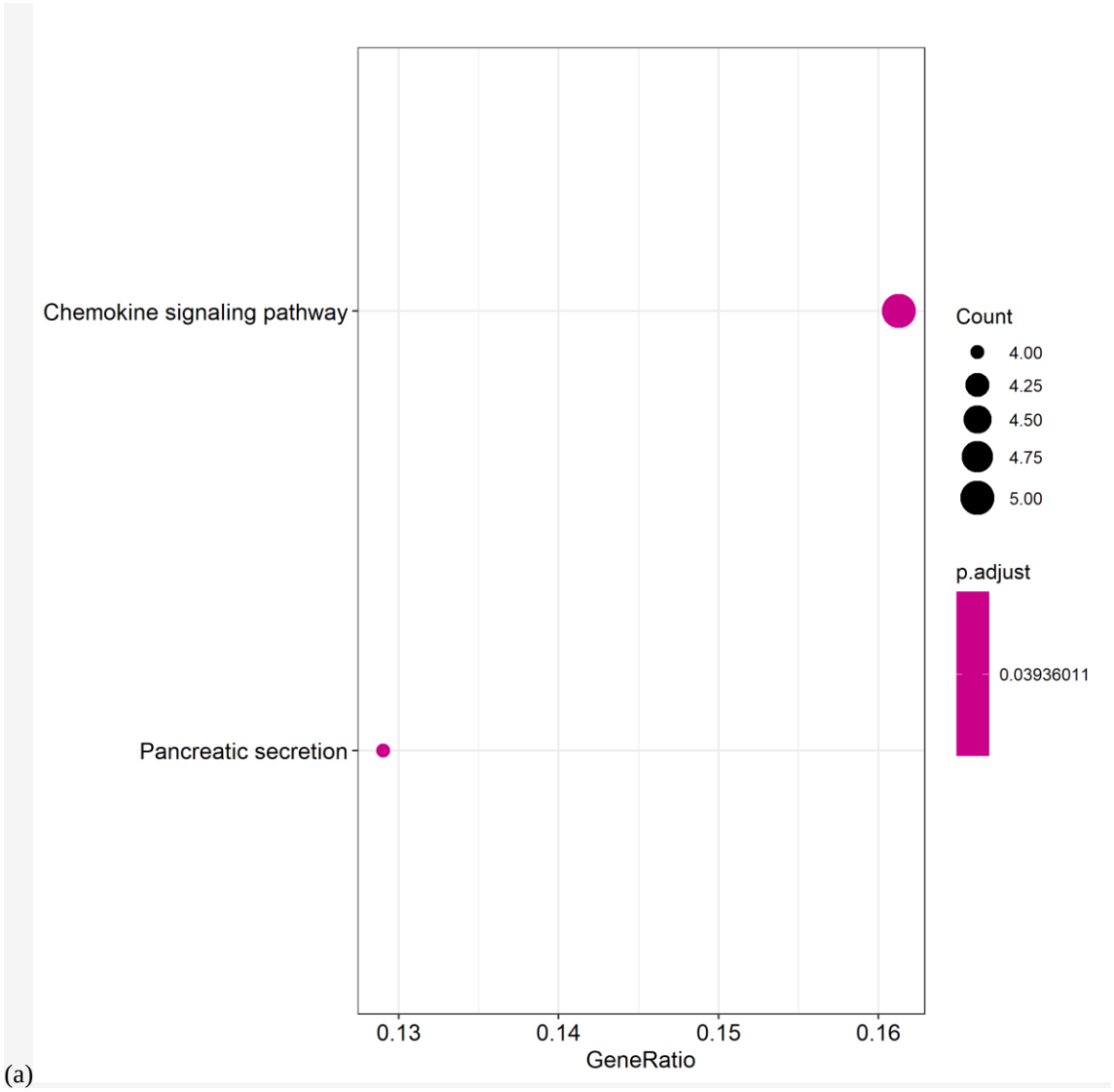
Figure 2. Volcano map and heat map of DEGs observed between the Chronic Kidney Disease group and the healthy control group. a) Volcano map of DEGs in the CKD group; b) heat map of DEGs in CKD group; (red indicates an upregulated gene and green indicates a downregulated gene.)

Investigation of DEGs based on the GO function and KEGG pathway enrichment analyses

The analysis of the GO function enrichment indicated that the DEGs were mainly enriched due to factors such as potassium ion, nucleolus, ESCRT I complex, protein binding, nucleotide binding, and ubiquitin binding (Figure 3a). The KEGG pathway enrichment analysis indicated that DEGs were mainly enriched in the chemokine signaling pathway, pancreatic secretion, protein digestion and absorption, and through the regulation of actin cytoskeleton and other pathways (Figure 3b、Table 2)



ID	Description
GO:0035864	response to potassium ion
GO:0005730	nucleolus
GO:0000813	ESCRT I complex
GO:0005515	protein binding
GO:0000166	nucleotide binding
GO:0043130	ubiquitin binding



(a)

(b)

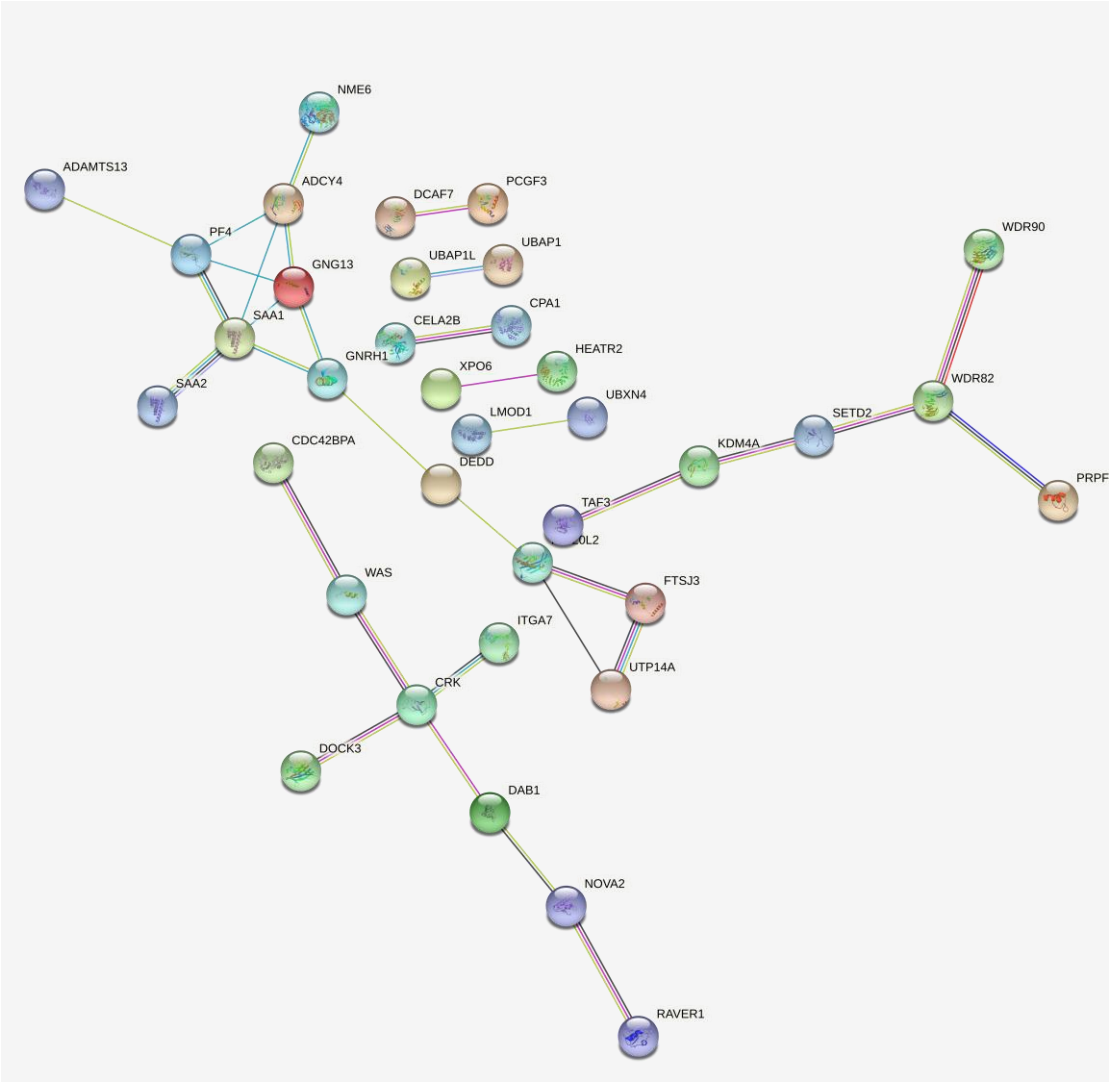
Figure 3a. Diagram illustrating GO functional enrichment analysis; (3b) Diagram illustrating KEGG pathway enrichment analysis

Table 2. KEGG pathway enrichment analysis

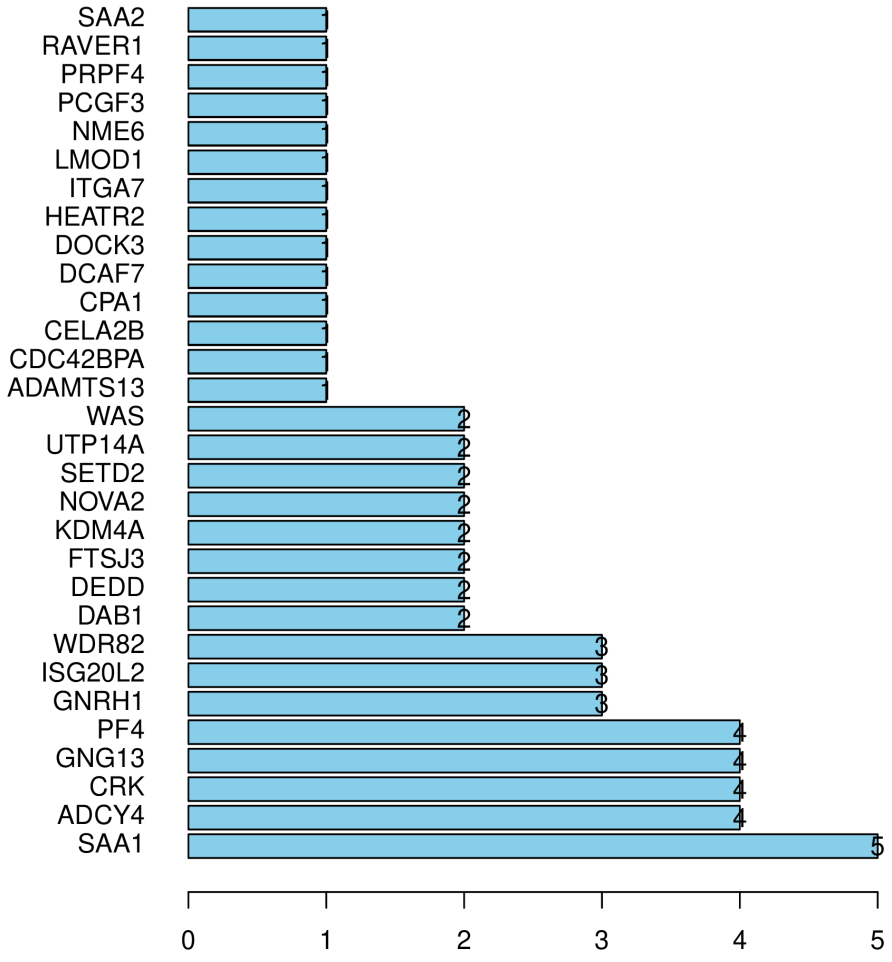
Category	Term	PValue	gene	count
KEGG_PATHWAY	hsa04062:Chemokine signaling pathway	0.004854184	ADCY4, WAS, CRK, GNG13, PF4	5
KEGG_PATHWAY	hsa04972:Pancreatic secretion	0.004965519	CPA1, CELA2B, ADCY4, ATP1A2	4
KEGG_PATHWAY	hsa04974:Protein digestion and absorption	0.043112467	CPA1, CELA2B, ATP1A2	3
KEGG_PATHWAY	hsa04810:Regulation of actin cytoskeleton	0.043517933	WAS, ITGA7, PPP1R12B, CRK	4

Construction of the PPI network of DEGs and analysis of central genes

STRING was used to construct a PPI network (Figure 4a), and the count package in R was used to conduct screening of the central genes (Figure 4b).


NME6
ADAMTS13
ADCY4
DCAF7
PCGF3
PF4
UBAP1L
UBAP1
GNG13
CPA1
SAA1
CELA2B
GNRH1
HEATR2
XPO6
SAA2
UBXN4
LMOD1
CDC42BPA
DEDD
KDM4A
SETD2
WDR90
WDR82
PRPF
TAF3
WAS
FTSJ3
ITGA7
UTP14A
CRK
DOCK3
DAB1
NOVA2
RAVER1

(a)



(b)

Figure 4. Diagram illustrating the PPI network constructed and central gene analysis of Differentially Expressed Genes. a) Construction of a PPI network of DEGs using STRING; b) data on central genes obtained from the PPI network.

Discussion

CKD pathogenesis is markedly complex. This study analyzed differential gene expression profiles and identified CKD-related DEGs. The purpose was to study the pathogenesis of CKD at the gene level and to provide new options for exploring the ultimate drug treatment target. Following the performance of GO enrichment analysis, the "response to potassium ion" was determined as the most significant enrichment GO term. The kidney is an essential organ and balances potassium ions in the body. Approximately 90% of the potassium content in the body is excreted through the kidneys (10). In patients with CKD, their renal function is impaired, glomerular filtration rate (GFR) is reduced, potassium ion excretion is blocked, and hyperkalemia occurs.

A previous study (11) indicated that the prevalence of hyperkalemia in CKD patients is noticeably higher than that in non-CKD patients, coupled with the deterioration of their renal function, the risk of seizures gradually increases. Approximately 50% of patients with CKD in stages 4 to 5 inevitably develop hyperkalemia. The incidence of hyperkalemia in CKD patients has increased significantly. The main reasons are:

1) decreased renal potassium excretion, GFR, and urine output or even anuria, resulting in the obstruction of potassium excretion; decreased secretion of hormones, such as angiotensin and renin, and long-term azotemia resulting in reduced potassium excretion via the renal tubules

2) potassium-sparing diuretics (spironolactone, etc), angiotensin-converting enzyme inhibitor

(ACEI)/angiotensin II receptor antagonist (ARB) (12), β-receptor blockers, tacrolimus, cyclosporine, vegetables, and Chinese herbal medicines which are rich in potassium ions can increase the levels of blood potassium

3) CKD patients often experience acidosis, wherein the pH decreases, and the potassium ions are released into the intracellular fluid, which leads to increased blood potassium levels

4) renal anemia leads to tissue hypoxia, thereafter NA+-K+-ATPase activity decreases, and the transfer of extracellular potassium to intracellular is blocked, which easily induces hyperkalemia

5) in CKD patients receiving hemodialysis, blood transfusions and other treatments, the tissue is destroyed, the intracellular potassium enters the extracellular fluid and increases potassium levels

6) the consumption of potassium-rich foods and drugs or excessive intravenous potassium can easily lead to high potassium levels in the body.

The hub gene selected through PPI network construction was identified as the serum amyloid A1 (SAA1) gene. This gene may play a key role in the development of CKD and demonstrates the highest degree of expression in the PPI network. SAA1, as the main component of the serum amyloid A family, is the most widely expressed, active, and sensitive subtype in the family (13). In various diseases and conditions, such as trauma, tumor, infection, and inflammation, the gene expression may increase by 1000-fold (14). Related studies have indicated that SAA also plays an important role in the occurrence and development of organ fibrosis. Piotti et al. (15) reported that the expression of SAA had significantly increased in early organ fibrosis. Xingcheng et al. (16) reported that the expression of SAA1 in mice with lupus nephritis was significantly higher than that of the normal group, suggesting that the increase in gene expression might be related to the occurrence of renal fibrosis; additionally, the tissue mRNA and protein levels were also noticeably increased. The test revealed that the expression of SAA1 had significantly increased, which was consistent with the anticipated results of this experiment. These results indicate that SAA1 can be used as a valuable laboratory index for renal interstitial fibrosis.

CKD patients exhibit a microinflammatory reaction state, and the patients present with abnormal expressions of various inflammatory factors such as TNF-a, IL-1B, and IL-6 (17). In patients with CKD, serum SAA levels are elevated, and regardless of the dialysis status of the patients, their serum SAA levels are significantly higher than those in healthy controls. Additionally, with the gradual decline of renal function, the serum SAA level gradually increases, and this increase is not related to age, sex, blood pressure, hemoglobin, and other factors. The increase in serum SAA levels in patients with CKD may be related to the expression of proinflammatory mediators such as TNF-a, IL-B, and IL-6 (18). In patients with CKD undergoing dialysis, the bio incompatibility of the dialysis membrane and the dialysate, contamination of endotoxins in the dialysate, and subclinical bacteremia caused by repeated fistula contribute further to the inflammatory reaction level in dialysis patients with CKD. The main reason for the increase is that the aforementioned factors can activate the complement system and inflammatory cells to induce the release of inflammatory mediators and to promote the inflammatory response (19). Thus, bioinformatics data analysis indicates that the SAA1 gene may be used as a marker for CKD treatment and detection, and the role of this gene warrants further investigation.

Acknowledgments

The authors wish to thank all those who contributed to this article. This study was funded by the National Natural Science Foundation of China (No.8177150812).

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of this research in ensuring that questions related to the accuracy or integrity of any part of this research are appropriately investigated and resolved.

References

1. Jha V，Garcia-Garcia G，Iseki K，et al．Chronic kidney disease: global dimension and perspectives. Lancet 2013;382:260-272．
2. Perco P，Wilflingseder J，Bernthaler A，et al．Biomarker candidates for cardiovascular disease and bone metabolism disorders in chronic kidney disease: a systems biology

perspective. J Cell Mol Med 2008;12:1177-1187.

3. Zhang L，Wang F，Wang L，et al．Prevalence of chronic kidney disease in China: a cross-sectional survey. Lancet 2012;379:815-822．
4. Benveniste O, Guiguet M, Freebody J, et al. Long-term observational study of sporadic inclusion body myositis.Brain 2011;134:3176-3184.
5. Chin CH, Chen SH, Wu HH,et al. Cytohubba: Identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014;8:S11.
6. Greenberg SA. Inclusion body myositis: clinical features and pathogenesis. Nat Rev Rheumatol 2019;15:257-272.
7. Keller CW, Schmidt J, Lunemann JD. Immune and myodegenerative pathomechanisms in inclusion body myositis. Ann Clin Transl Neurol 2017;4:422-445.
8. Ge JB,Xu YJ，Wang C. Internal Medicine. 9 editions. BEIJING: People's Medical Publishing House 2018:774-776．
9. KASHIHARA N，KOHSAKA S，KANDA E，et al． Hyperkalemia in real-world patients under continuous medical care in Japan．Kidney Int Rep 2019;4: 1248-1260．
10. Sun CH，Xie HM，Zhang YF，et al. Causative analysis of hyperkalemia in chronic kidney disease patients. International Journal of transplantation and blood purification.2019；17:26-29．
11. Yamada T，Kakihara T，Kamishima T，et al．Local expression of acute phase serum amyloid A are present in atherosclerotic lesions．Pathol International 1996;46:797-800.
12. Hansson GK. Inflammation, atherosclerosis, and coronary artery disease．N Engl J Med 2005;352: 1685-1695.
13. Piotti KC，Yantiss RK，Chen Z，et al．Serum amyloid immunohistochemical staining patterns in hepatitis．Histopathology 2016;69: 937-942.
14. Zhou XC,Zhang F，Yan Rui，et al．Expression of SAA1 in the kidney of Lupus mice and its effect on EMT of Renal Tubular epithelial cells. Chinese Journal of Immunology 2019；35:2561-2565.
15. Kaysen GA.The micminnammatory state in urerllia: causes and poences.J AM Soc Nephml 2001;12:1549.
16. Zhang HF,Wang JH, Wang YL,et al．Salvianolic Acid A Protects the Kidney against Oxidative Stress by Activating the Akt/GSK-3-β/Nrf2 Signaling Pathway and Inhibiting the NF-κB Signaling Pathway in 5/6 Nephrectomized Rats.Oxid Med Cell Longev 2019;18:2853534.
17. Zhang JB, Hua G, Zhang XG,et al．Regulatory T cells/T-helper cell 17 functional imbalance in uraemic patients on maintenance haemodialysis: a pivotal link between microinflammation and adverse cardiovascular events. Nephrology (Carlton) 2010;15:33-41.

18.Untersteller K, Meissl S, Trieb M, et al．HDL functionality and cardiovascular outcome among non-dialysis chronic kidney disease patients. Marsche G.J Lipid Res 2018;59:1256-1265.

18. R Auwul, C Zhang, Md R Rahman, Md. Shahjaman, S. A. Alyami, M A Moni, Network-based transcriptomic analysis identifies the genetic effect of COVID-19 to chronic kidney disease patients: A bioinformatics approach, Saudi J Biol Sci,28(10),2021,5647-5656,