Heliyon 11 (2025) e41872

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

# Machine learning-based identification and validation of amino acid metabolism related genes as novel biomarkers in chronic kidney disease

Guoqing Zhang [a,b,1], Hongyan Luo [a,b,1], Xiaohua Lu [a,b,1], Yonghua Liu [a,b], Mei Wang [a,b], Bo Li [a,c], Haixia Lu [a,b], Yali Zheng [a,b,*]

[a] Department of Nephrology, Ningxia Medical University Affiliated People's Hospital of Autonomous Region, Yinchuan, China
[b] The Third Clinical Medical College, Ningxia Medical University, Yinchuan, China
[c] Department of Nephrology Hospital, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, China

ARTICLE INFO

*Keywords:*
Chronic kidney disease
Amino acid metabolism
Immune cell infiltration
Machine
Learning
Biomarkers

ABSTRACT

*Objectives:* Chronic kidney disease (CKD) is a progressive illness with a high rate of morbidity and mortality with no proven therapy. Alterations of amino acid(AA) metabolism are associated with the incidence and progression of CKD. To characterize the potential value of AA metabolism related genes in the diagnosis and progression of CKD.

*Methods:* We filtered the key genes associated with AA metabolism based on the least absolute shrinkage and selection operator (LASSO) and SVM algorithm. Then, we constructed logistic regression models and evaluated the accuracy and specificity by nomogram analysis and DCA. Also, we mapped the ROC curves.Meanwhile, in order to determine the underlying mechanism and relevant biological features of CKD, we conducted differential analysis between high and low risk subgroups in CKD. Moreover,we employed ssGSEA algorithm to evaluate the infiltration abundance of immune cells and calculated the correlation among the immune cells with the key genes. Finally,we validated the expression and clinical relevance of amino acid metabolism key genes via cultured cells and clinical data. A total of six key genes related to amino acid metabolism were identified, including ALDH18A1, CENPF, CSAD, CTH, CYP27B1, HBB.

*Results:* All six genes exhibited promising diagnostic capabilities (AUC:0.7 to 0.9). Immune cells such as Activated $CD4^+$ T cells, Regulatory T cells, Immature B cells and MDSC,etc.infiltrated differentially in the high and low risk groups of CKD. There were correlations between immune cells abundance and the expression of key genes. All key genes correlated significantly with markers of kidney injury, such as eGFR and serum creatinine. The expression of ALDH18A1, CENPF were increased while CSAD, CTH and CYP27B1 were decreased in HK-2 cells cultured with indole sulfate.

*Conclusions:* Our study identified key genes involved in amino acid metabolism associated with immune cells infiltration and renal function in CKD, which may be potential biomarkers for the diagnosis and prognosis of CKD.

* Corresponding author. No.301 Zhengyuan North Street, Yinchuan, Ningxia Hui Autonomous Region, 750001, China.
*E-mail address:* zhengyali@nxmu.edu.cn (Y. Zheng).
[1] Equal contributors.

https://doi.org/10.1016/j.heliyon.2025.e41872
Received 1 August 2024; Received in revised form 3 January 2025; Accepted 9 January 2025
Available online 10 January 2025

## 1. Introduction

Chronic kidney disease (CKD) is a progressive illness with a high rate of morbidity and mortality with no proven therapy [1]. It is estimated that the approximate incidence of CKD is 8–16 % [2] and may ultimately lead to end-stage renal disease. The progression of CKD results in significant changes to the needs and metabolism of many nutrients [3]. However, there aren't effective therapy for patients with CKD, despite numerous advancements in the field of treatment. Furthermore,the diagnosis of CKD relies primarily on glomerular filtration rate (eGFR) and/or proteinuria, although eGFR is not detectable abnormally until kidney damage is evident. In order to manage the disease and prevent renal function loss at an early stage, identifying reliable biomarkers is crucial for CKD detection and prognosis.

It is well established that dietary components and their metabolites have a direct impact on the course of CKD. Amino acids (AA), commonly referred to as the building blocks of proteins, play a part in protein utilization and cell performance. The balance of the human body's amino acid levels is mostly dependent on the kidney [4],as the kidneys absorb substantially 97–98 % amino acids filtrated during one day [5]. However, as the mechanism of AA metabolism in CKD is still unclear, the identification of potent and trustworthy diagnostic and therapeutic biomarkers for chronic kidney disease on the basis of AA metabolism is imperative. The availability of amino acids, which serve as vital nutrients of immune cells, controls immune cell function. Immune cell differentiation is governed by a variety of factors, such as gene selection, cytokine stimulation, nutrition metabolism, etc [6]. The concentration of amino acids, transporters that are membrane-bound, and vital metabolic enzymes all play crucial roles in regulating immune cell development and function. However,the role of amino acids in protein synthesis is well recognized, but the mechanisms of metabolic regulation in CKD and immune cells remain poorly understood.

In this research,we investigated the features of critical genes and immune cell infiltration in amino acid metabolism of CKD. And we validated the expression and clinical relevance of amino acid metabolism related key genes via cultured cells in vitro and clinical data. In short, our study identified key genes involved in amino acid metabolism in CKD, which could be prospective indicators for CKD prediction and evaluation.

## 2. Material and methods

### 2.1. Flowchart of this research

The flow of this work is illustrated in Fig. 1 presented below.

### 2.2. CKD data process

Using the R package GEOquery [10], we retrieved the expression profiling datasets of patients with CKD GSE66494 [7], GSE116626 [8] from the GEO database [9]. The datasets GSE66494, GSE116626 were derived from Homo sapiens. The GSE66494 dataset consists of a total of 61 samples, including 53 kidney biopsies of CKD and 8 normal kidneys (Normal). The GSE116626 dataset

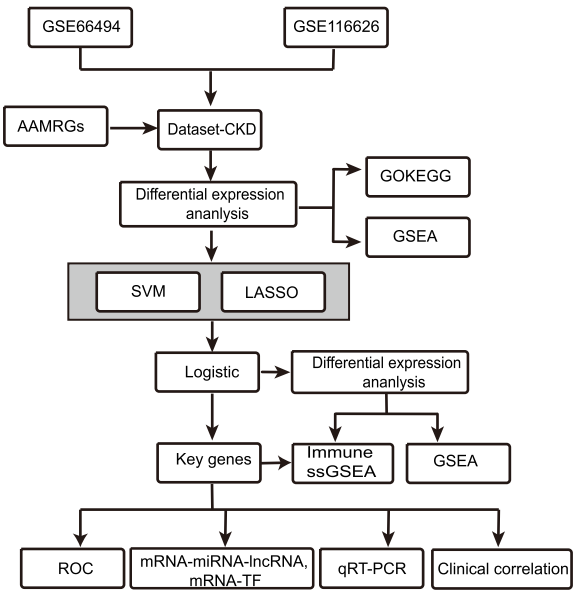



**Fig. 1.** Map of work flow.GO:Gene Ontology.KEGG:Kyoto Encyclopedia of Genes and Genomes.GSEA:Gene Set Enrichment Analysis.ROC:Receiver Operating Characteristic curve. AAMRGs:Amino Acid Metabolism related genes; AAMRDEGs:Amino Acid Metabolism related differentially expressed genes.

contains 52 patients with IgA nephropathy,22 patients with non-IgAN glomerulonephropathy and 7 biopsy samples from living kidney donors. 52 IgAN patients were scored by several pathologists according to the MEST-C classification and categorized into 4 groups according to differential renal pathologies: minor nephropathy 22, aggressive nephropathy 8, chronic nephropathy 12, and combined active and chronic pathologies group 10. Gene expression profiling data of 12 samples from IgAN patients diagnosed with chronic kidney disease (subgroup: CKD), and biopsy samples from 7 living kidney donors (subgroup: Normal) were selected for analysis. And specific information of datasets were listed in Table S1.

We obtained Amino Acid Metabolism related genes (AAMRGs) from the GeneCards [11]. "Amino Acid Metabolism" was the query word we used.A total of 238 AAMRGs were obtained with "Protein Coding" and "Relevance score >2″ as the screening criteria. Moreover, we performed a examine based on the MSigDB [12] according to "WP_AMINO_ACID_METABOLISM_IN_-TRIPLENEGATIVE_BREAST_CANCER_CELLS" and "WP_AMINO_ACID_BREAST_CANCER_CELLS" with "Amino Acid Metabolism" as search criteria. Subsequently, 13 AAMRGs were collected. Finally, we obtained altogether 241 AAMRGs by merging and de-duplicating, the specific details of which are presented in Table S2.

### 2.3. Differential genes analysis

In order to derive differentially expressed genes (DEGs) within the CKD and control patients(Normal), we firstly obtained the merged Dataset-CKD useing R package sva [13] with de-batch of GSE66494 and GSE116626, which comprised 65 CKD samples and 15 control (Normal) samples. To verify the effectiveness of debatch, we carried out Principal Component Analysis (PCA) [14] before and after the debatch process. Afterwards, to perform differential analysis and acquire the differentially expressed genes, we utilized the R package limma [15]. Finally, we intersected all DEGs based on the criteria of $|\log FC|>0.5$ and $Padj<0.05$ with the AAMRGs and plotted a Wayne diagram. Subsequently,we obtained the Amino Acid Metabolism related differentially expressed genes (AAMRDEGs). Moreover, the volcano plot is made by ggplot2 to show the difference analysis result. Then the correlation heatmap and group comparison plot of AAMRDEGs are also made by R package. Finally, we perform correlation heatmap between AAMRDEGs to check the correlation profiles between AAMRDEGs.

### 2.4. GOKEGG enrichment analysis

GO(Gene Ontology) [16] and KEGG (Kyoto Encyclopedia of Genes and Genomes) [17] are widely used for evaluation of genes biological pathways and features. With the R package clusterProfiler [18], we conducted GO and KEGG analysis of AAMRDEGS. Entries were filtered for $P.adjust<0.05$ and FDR value (qvalue) $< 0.05$.

### 2.5. Measurement of GSEA

To examine the variation of biological processes between two groups, we downloaded the collection of reference genes "c2.cp.v7.2.symbols.gmt" from the MSigDB [12] and used the GSEA [19] process by the package of clusterProfiler. The filtering criterion was P.adjust $<0.05$ and the FDR (qvalue) $< 0.25$.

### 2.6. Diagnostic models

LASSO (Least absolute shrinkage and selection operator) regression is currently used to construct diagnostic models for machine learning algorithms. To generate prognostic model in Dataset-CKD, we performed LASSO [20] regression to acquire the relevant AAMRDEGs with non-zero coefficients matching the assessment metrics' optimal lambda values. Subsequently, based on AAMRDEGs, the SVM model was built by the SVM (Support Vector Machine) [21] method. The number of genes with the best rate of accuracy and smallest error rate was selected for the AAMRDEGs. Finally,the key genes identified by combining SVM and LASSO analysis.

We subjected key genes to logistic regression analysis and constructed logistic regression models. After that, we built nomogram plots according to outcomes of the logistic regression analysis with R package rms. Next, we evaluated the calibration curve to assess the precision and distinction power of the logistic regression model through Calibration analysis. We created calibration curves with the "rms" algorithm in R. Decision curve analysis (DCA) was applied to estimate clinical prediction models. We evaluated the models and created DCA plots using the R package ggDCA [22]. Moreover, we demonstrated the correlation of key genes with the model using correlation scatter plots. The Spearman's approach was applied in order to investigate the relationship among genes.

### 2.7. ROC evaluation

The receiver operating characteristic curve, or ROC, is a positional graphical analyzer that could be employed to select the most suitable model or set the ideal threshold. The ROC curves for key genes in the Normal and CKD groups were mapped via the pROC software, and the area under the curve (AUC) was computed mathematically to evaluate the diagnostic efficacy.

### 2.8. Regulation network

miRDB database [25], which provides functional identification and estimation of miRNA target genes. We estimated the miRNAs interacting with key genes with ENCORI [24] and miRDB database, and took the intersection to obtain networks of mRNA-miRNA

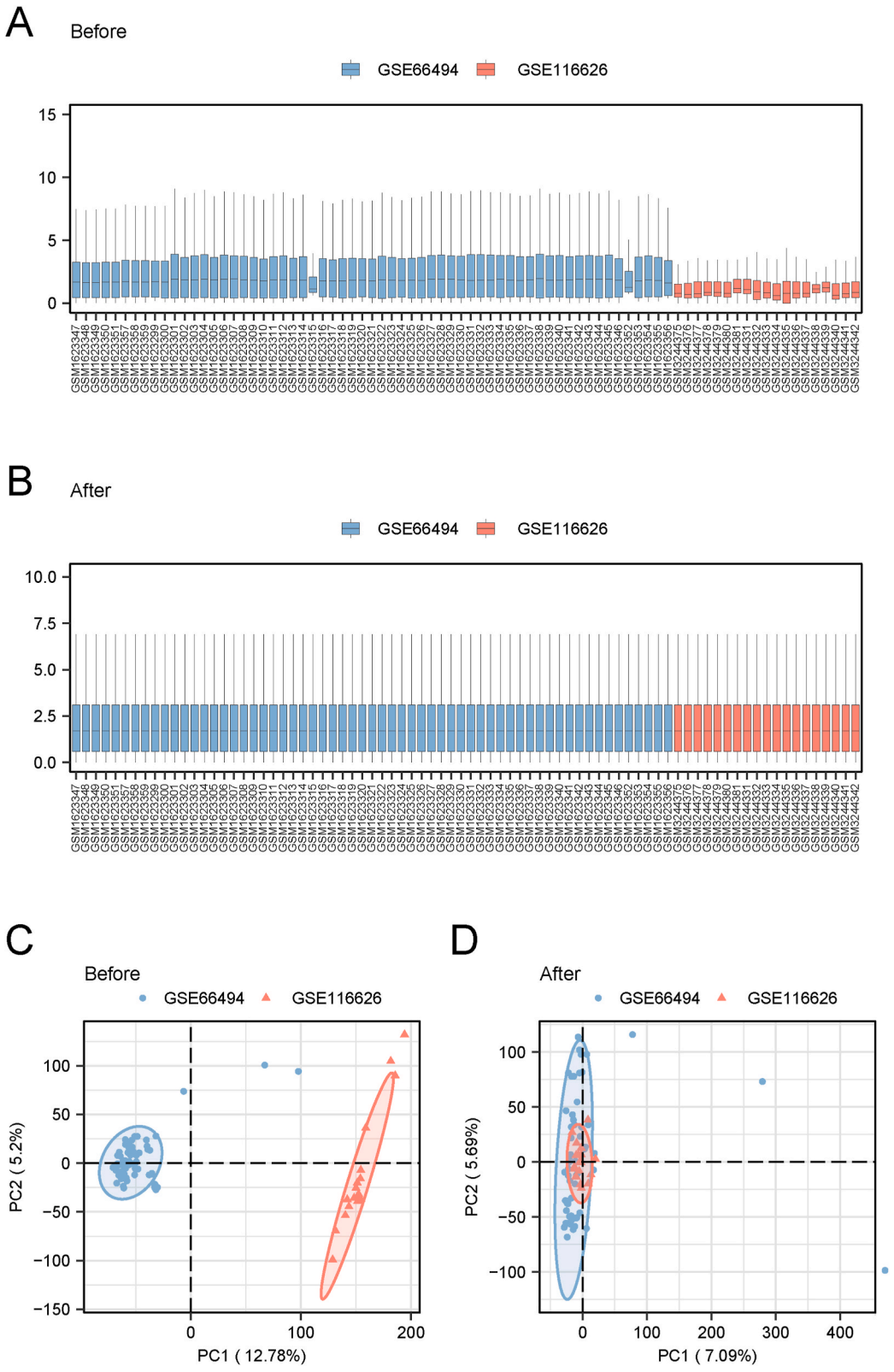



**Fig. 2.** Dataset and correction chart.(A) Box plot of the merged dataset before correction. (B) Box plot of the merged dataset after correction; (C) PCA plot of the merged dataset before correction; (D) PCA plot of the merged dataset after correction; PCA: Principal Component Analysis.

interactions. Following this, we have also predicted lncRNAs which interact with miRNAs using the ENCORI database as well as obtained miRNA-lncRNA interaction networks with clipExpNum ≥5 as a filtering criterion. Employing the CHIPBase [26] and hTFtarget [27] databases, we examined for transcription factors (TFs) that interact with significant genes. Transcription factors shared in both databases were selected to construct networks. Then, through Cytoscape software, the mRNA-miRNA-lncRNA and mRNA-TF networks were further displayed.

### 2.9. Analysis of DEGs and performed GSEA

To determine probable processes and related biology of different genes in the high- and low-risk subgroups of chronic kidney disease. We employed the limma package to conduct differentiation from groups of high and low risk, besides,volcano plots and heatmaps were presented to visualize the differentials. GESA was then implemented. The following were the criteria employed in this GSEA: 1000 calculations were made,100–500 genes contained in each gene set, and P.adjust $<0.05$ with FDR $<0.25$.

### 2.10. Immune cell analysis

The amounts of different immune cell types were indicated by the abundance scores generated by the ssGSEA [28] method from R software GSVA. Differences of immune cells observed in the groups at high and low risk were displayed by box plots. The ggplot2 was utilized to illustrate the association between various immune cells, which was determined using the spearman approach. After that, we computed relationships among immune cells and key genes in high- and low-risk groups based on the expression of genes and showed the relevance scatter plots by R package ggplot2.

### 2.11. Clinical features of key gene

Nephroseq v5 (http://v5.nephroseq.org) is a platform for integrated genotypic and phenotypic data of kidney diseases. We investigated the relationship among hub genes expression and clinical characteristics by querying the Nephroseq v5 database.

### 2.12. Culture of cells and analysis with RT-PCR

Human renal proximal tubular epithelium cells (HK-2,STCC10303P) obtained from servicebio, which were cultured in MEM with 10 % fetal bovine serum. The cells were then respectively incubated with Indole sulfate(IS,Sigma,Cat.,No.I3875) at different concentrations for 24h(1 mM and 2 mM). Extraction of experimental cells RNA using Total RNA Kit I (Omega, Cat. No., R6834-02). RT-qPCR was used to evaluate the expression of six key genes. Nanodrop procedure device was used to measure the concentration of RNA. Afterwards, Using the PrimeScript RT reagent Kit (Takara, Cat. No. RR037A), RNA was converted to cDNA. Using the SYBR Green detection method, the relative mRNA levels of key genes were determined. Applying a Takara RR820A reagent kit and an ABI 7500 Fast Real-Time PCR instrument, real-time quantitative PCR (RT-PCR) was conducted. Each sample was examined three times, and the cycle threshold (Ct) value was normalized to determine the relative level of each gene. Expression of key genes normalized by β-actin expression. Each primer sequences were available in Table S6 and raw Ct values of PCR were provided in Table S7.

### 2.13. Statistical analysis

All analysis in this study were performed with R software (Version 4.2.3). Comparisons between two groups of continuous variables were conducted with the Wilcoxon rank sum test. One-way ANOVA was used for comparison among three groups. The Chi-squared test or Fisher exact test was used to analyze the statistical significance between the two groups with categorical variables. Respondent Operating Characteristic Curve (ROC) was determined based on the R package pROC. If not specifically mentioned, the correlation coefficients between the different molecules were calculated by spearman correlation analysis. P-value statistics were two-sided, with P-value of less than 0.05 as the statistical criterion.

## 3. Results

### 3.1. CKD-dataset differential analysis

We employed genes expression from two datasets in our differential analysis, with errors due to different sequencing times and platform. To improve the comparability of the combined datasets and the reliability of the differential analysis, we performed batch effect analysis. Firstly, we processed the datasets GSE66494 and GSE116626 by removing batch effects using the R package sva to obtain the merged CKD-dataset. The two datasets are represented with the two colors. The two datasets are independent of one another and do not interact. Following the elimination of batch effect, sample distributions across datasets typically exhibit consistency. We demonstrated the before- and after-batch effect in the Dataset through box-and-line plots and PCA (Principal Component Analysis) (Fig. 2A–D). A total of 1254 differentially expressed genes(DEGs) were obtained, with P.adj $<0.05$ and a cutoff value of $|\log FC|>0.5$. In the group with CKD, there were 890 genes with decreased expression and 364 high-expression genes. The differentiation were indicated by volcano plotting with the R package ggplot2 (Fig. 3A). We mapped Wayne diagrams to demonstrate amino acid metabolism-related differentially expressed genes (AAMRDEGs)(Fig. 3B). There are 36 AAMRDEGs comprising these genes:ACAA1,

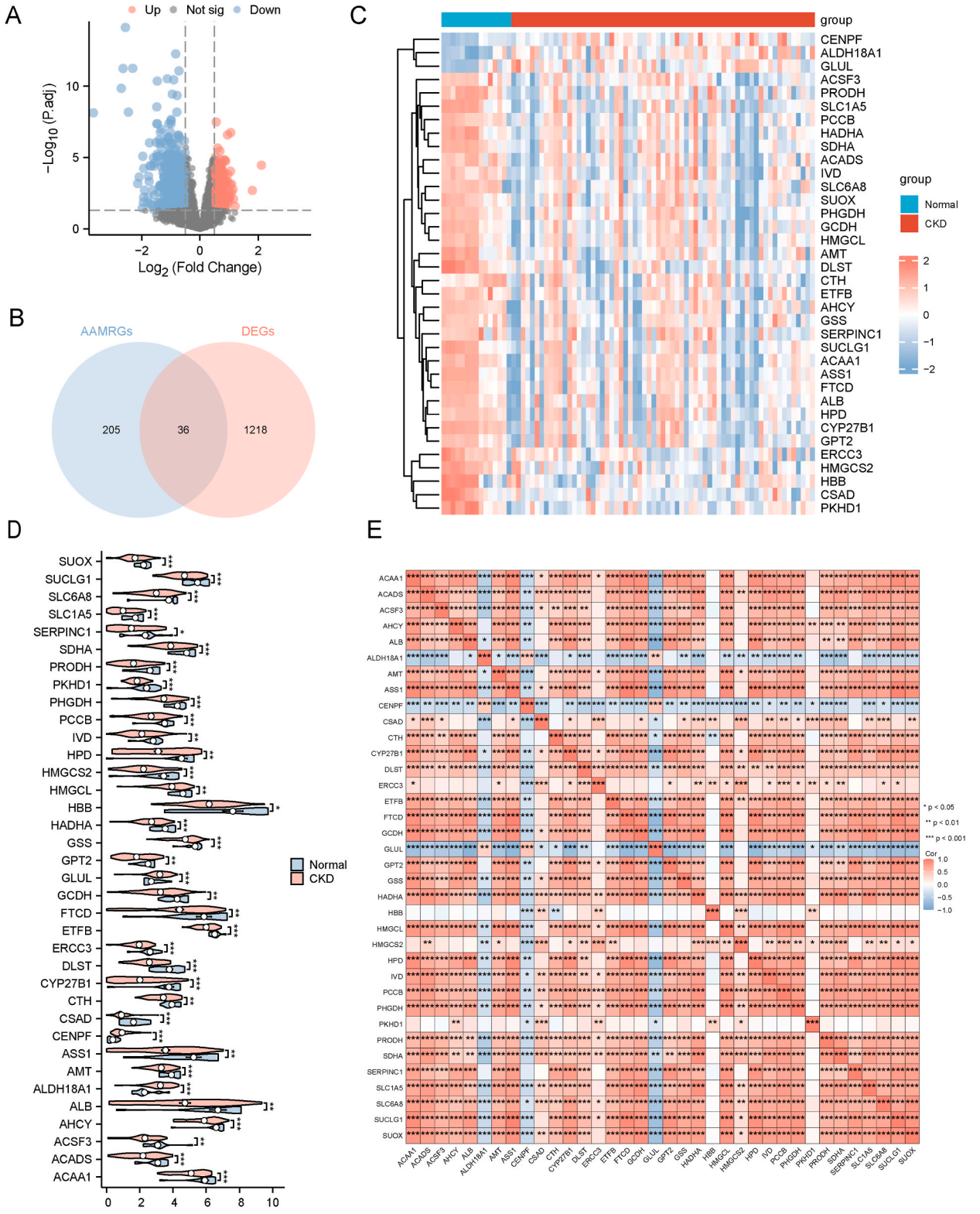



**Fig. 3.** Difference analysis of CKD-dataset.(A) Volcano plots of differentially expressed genes in the Normal and CKD groups; (B) Wayne plots of DEGs and AAMRGs; (C) Heatmap of AAMRDEGs; (D) Comparison of subgroups of AAMRDEGs; (E) Heatmap of correlation of among AAMRDEGs. DEGs: differentially expressed genes.AAMRGs: Amino Acid Metabolism related genes; AAMRDEGs: Amino Acid Metabolism related differentially expressed genes. *P < 0.05,**P < 0.01, ***P < 0.001.

ACADS, ACSF3,AHCY,ALB,ALDH18A1,AMT,ASS1,CENPF, CSAD,CTH,CYP27B1,DLST, ERCC3,ETFB, FTCD,GCDH, GLUL,GPT2,GSS, HADHA,HBB,HMGCL, HMGCS2,HPD,IVD,PCCB, PHGDH,PKHD1,PRODH, SDHA,SERPINC1,SLC1A5,SLC6A8,SUCLG1,SUOX. The expression of AAMRDEGs was presented as a heat map drawn by R package pheatmap (Fig. 3C). From the group comparison plot (Fig. 3D), we could observe that the expression differences of all AAMRDEGs were statistically significant, of which the majority of AAMRDEGs were up-regulated in control group. Lastly, we exhibited the correlation of AAMRDEGs with correlation heatmap (Fig. 3E), which revealed that most of the AAMRDEGs were associated with each other.

### 3.2. GO and KEGG

To characterize the biological processes of 36 AAMRDEGs in CKD. We performed GOKEGG functional enrichment analysis with the 36 AAMRDEGs. We demonstrated the findings by bar graphs (Fig. 4A). The analysis of GO suggested that AAMRDEGs were remarkably enriched in biological process (BP) (Fig. 4B),such as organic acid catabolic process, carboxylic acid catabolic process, cellular amino acid metabolic process, small molecule catabolic process and alpha-amino acid metabolic process; Cellular component(Fig. 4C), including mitochondrial matrix, tricarboxylic acid cycle enzyme complex, oxidoreductase complex; Molecular function (MF)(Fig. 4D), consisting of oxidoreductase activity, acting on the CH-CH group of donors, flavin adenine dinucleotide binding, ligase activity,acyl-CoA dehydrogenase activity and pyridoxal phosphate binding. This finding suggests that the majority of biological processes are related to the synthesis and catabolism of amino acids and may exert their biological features through oxidative cyclooxygenases and other enzymes which could play a role as one of the pathophysiological mechanisms in chronic kidney disease. The KEGG pathway (Fig. 4E) enriched predominantly in Valine, leucine and isoleucine degradation,Carbon metabolism, Biosynthesis of amino acids, Butanoate metabolism and Propanoate metabolism. Our data demonstrate clearly that amino acid metabolism, including essential amino acid metabolism, assumes an essential role in the progression of CKD. It is of great importance to investigate the genes and functions associated with amino acid metabolism.

### 3.3. GSEA

We examined the association between genes in CKD/Normal subgroup and the biological and pathogenesis processes by GSEA. The results displayed remarkable enrichment in KEGG_RETINOL_METABOLISM (Fig. 5B), KEGG_METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450 (Fig. 5C), REACTOME_METABOLISM_OF _AMINO_ACIDS_AND_DERIVATIVES (Fig. 5D), KEGG_DRUG_METABOLISM_CYTOCHROME_P450 (Fig. 5E), and other signal pathways (Fig. 5A–E). These biological functions are mainly

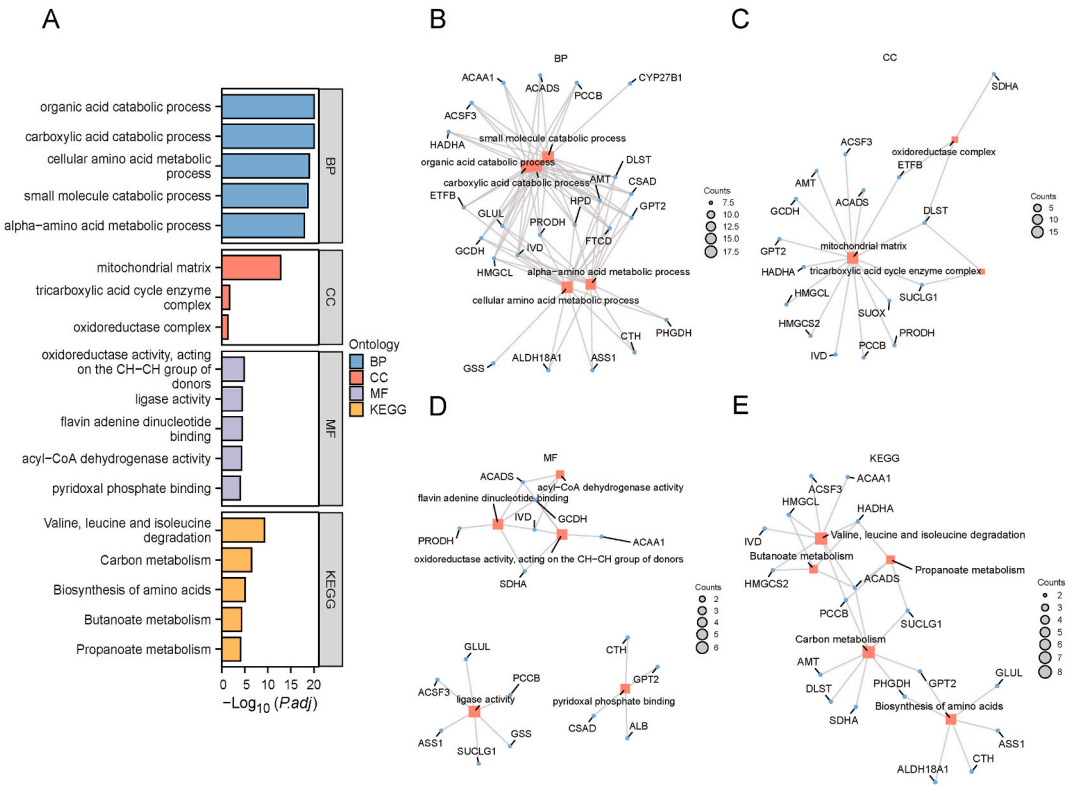



**Fig. 4.** GOKEGG enrichment analysis.(A) Bar graph of GO and KEGG enrichment analysis; (B–E): Biological process (BP) network (B); cellular component (CC) network (C); molecular function (MD) network (D); Network of KEGG enrichment analysis (E).

enriched in nutrient metabolism, metabolism of amino acids and their derivatives, as well as P450 enzymes also have an essential function in CKD. It further confirms the significance of amino acid metabolism in the progression of chronic kidney disease and suggests that the P450 enzyme metabolism-related pathway is a promising approach to study the pathogenesis of CKD.

### 3.4. Constructing the diagnostic models

We used LASSO regression analysis to create diagnostic models of 36 AAMRDEGs in order to determine their diagnostic function (Fig. 6A). In addition, we visualized the LASSO regression results and variable trajectory plots(Fig. 6B). Nine AAMRDEGs made up the LASSO diagnostic model we built, as shown in this figure:CTH, ALDH18A1, HBB, DLST, CYP27B1, CENPF, PKHD1, ERCC3, and CSAD. Meanwhile, we constructed SVM algorithm to acquire the gene numbers with highest accuracy (Fig. 6C) as well as lowest error rate (Fig. 6D).The achievement indicated that SVM model was the most precise when there were 15 genes:CTH, CYP27B1, FTCD, ASS1, CENPF, PHGDH, ACAA1, CSAD, SDHA, HBB, ALDH18A1, GPT2, SUCLG1, HPD, GSS. We combined the genes derived from the two algorithms and generate six overlapping genes:ALDH18A1, CENPF, CSAD, CTH, CYP27B1, HBB, for which we defined the six genes as key genes (Fig. 6E).

We performed logistic regression analysis and constructed logistic regression models for key genes (ALDH18A1, CENPF, CSAD, CTH, CYP27B1, HBB) and discovered that all six key genes fulfilled the screening requirements ($p < 0.05$). Next, To ascertain the model's diagnostic capabilities, we conducted nomogram analysis and created a nomogram plot(Fig. 6F). Furthermore, we plotted the calibration curves and performed prognostic calibration analysis on nomogram plots (Fig. 6G). We then evaluated the effectiveness of the logistic regression model in diagnosis with Decision Curve Analysis (DCA) and presented the results (Fig. 6H). As evident from our results (Fig. 6G and H), the model we constructed has a high accuracy in the diagnostic capability of pathogenesis of CKD.

Subsequently, CKD-dataset was grouped into high- and low-risk groups (High/Low) based on linear predictors. The logistic regression model's linear predictors score was computed using the following formula:

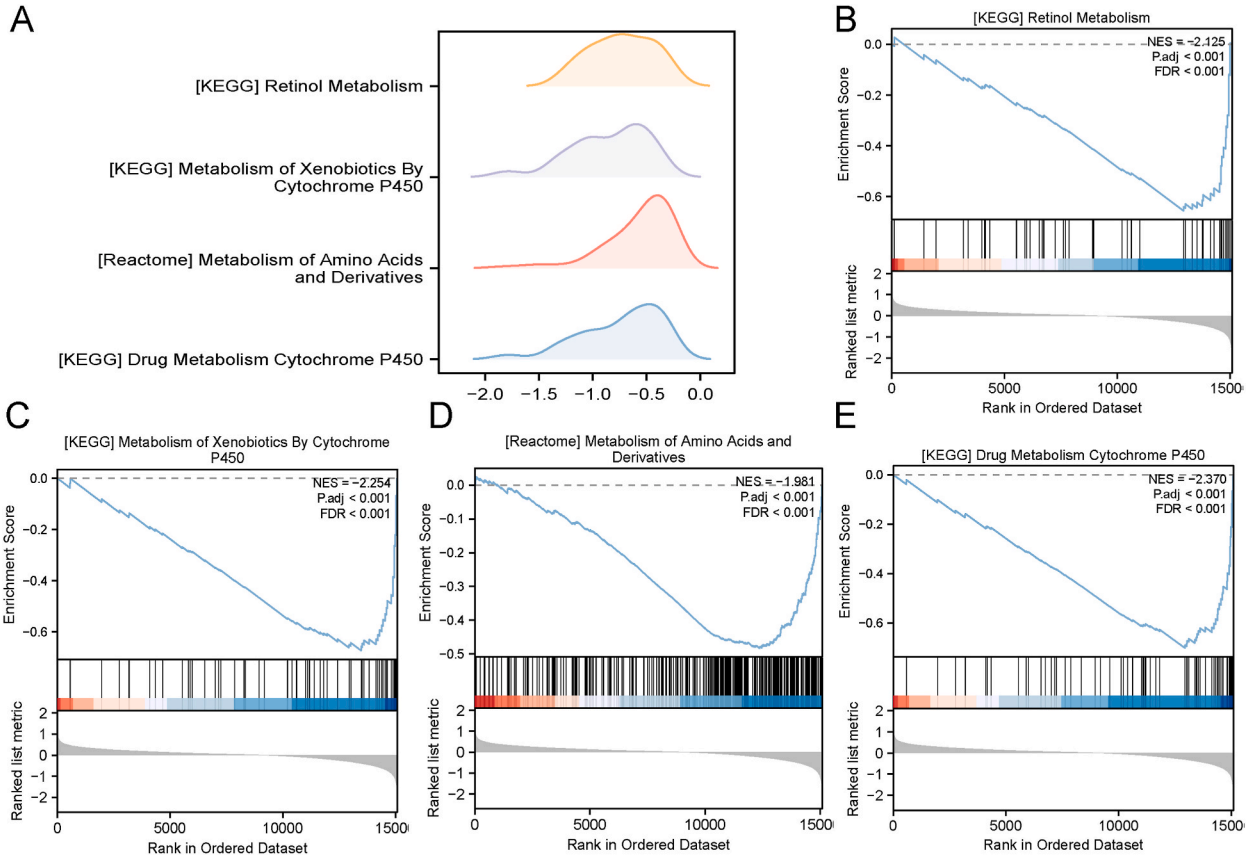



**Fig. 5.** GSEA of CKD-dataset.(A)The main four biological features of GSEA in CKD-datase; (B–E) CKD/Normal genes were significantly enriched in KEGG_RETINOL_METABOLISM (B); KEGG_METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450 (C); REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES (D); and KEGG_DRUG_METABOLISM_CYTOCHROME_P450 pathways (E).

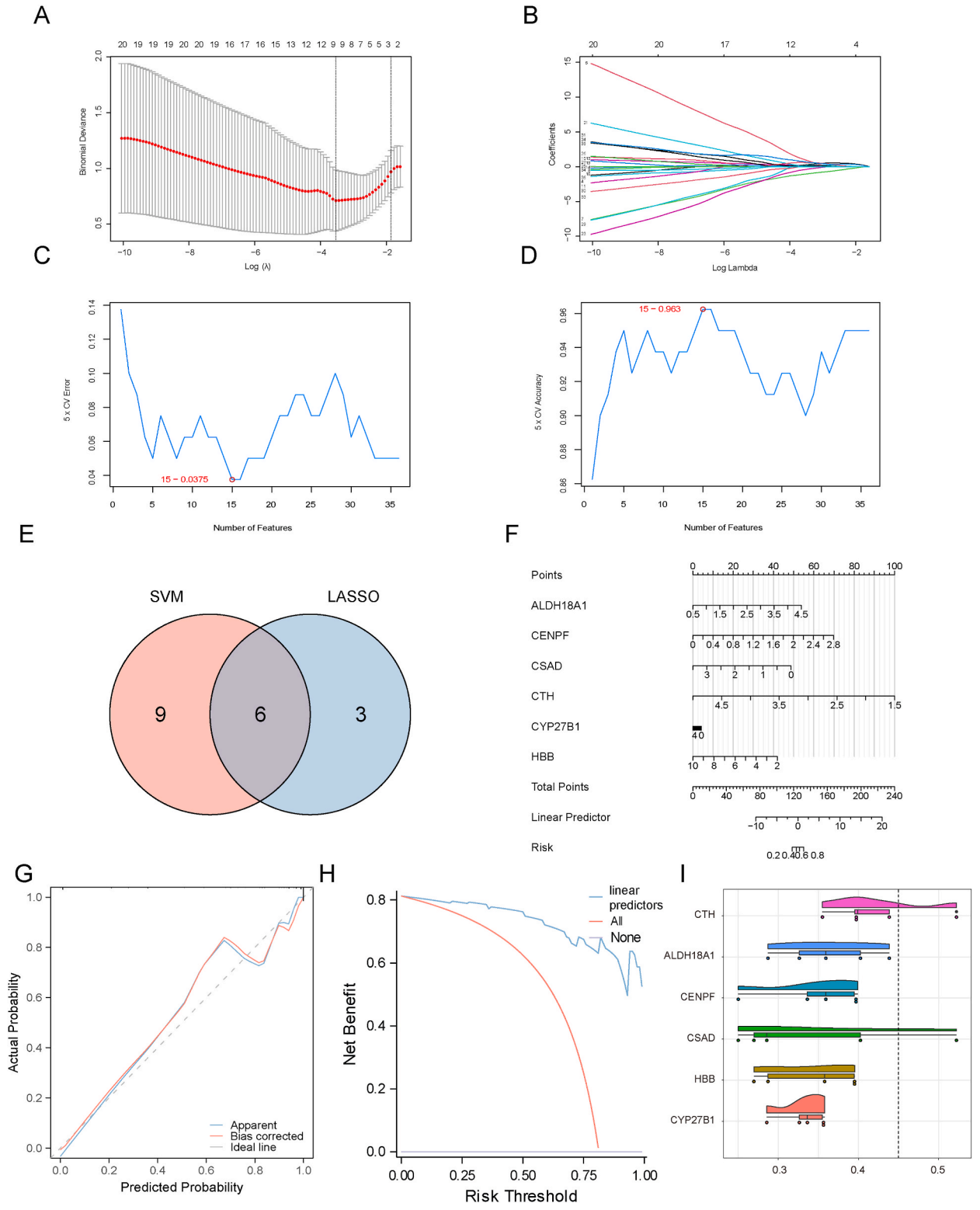



**Fig. 6.** LASSO model construction and SVM analysis of AAMRDEGs.(A) LASSO regression prognostic model plot for key genes; (B) Variable trajectory plot of LASSO regression diagnostic model; (C) Number of genes with the highest accuracy rate from SVM algorithm; (D) Number of genes with the lowest error rate from SVM algorithm; (E) Wayne's plot of the intersection of genes screened by LASSO regression and SVM algorithms; (F) Line plot illustrating logistics prediction values; (G) Calibration plot of the logistic predictive value; (H) DCA plot, the x-axis in the DCA plot represents the threshold probability, and the y-axis represents the net benefit; (I) Functional similarity analysis.

linear predictors = (2.683015623*ALDH18A1 expression) + (4.97955234*CENPF expression)
+ (−2.779012773* CSAD expression) + (−5.70791511*CTH expression) + ( − 0.169025009* CYP27B1 expression)
+ (−1.046509158*HBB expression)

These six major genes were the subjects for a functional similarity study, which we carried up with cloud and rain plots to illustrate the results (Fig. 6I). Our findings indicated that CTH had the highest functional similarity value with other key genes. In order to further explore the association of the six key genes with the model, we performed correlation analysis with the Spearman algorithm, and the outcomes were illustrated by correlation scatter plots (Fig. 7A–F). The findings indicated that vital genes: ALDH18A1, CENPF, CSAD, CTH, and CYP27B1 were moderately connected to the model, while HBB was weakly with the model.

### 3.5. Diagnostic ROC

Besides, we aimed to analyze the diagnostic value of the logistic regression model linear predictors and key genes (ALDH18A1, CENPF, CSAD, CTH, CYP27B1, HBB) (Fig. 8A). We combined linear predictors and 6 key genes to plot ROC. Linear predictors provided favorable diagnostic effect (AUC = 0.975); All the 6 key genes (Fig.B–G) offered better diagnostic accuracy (AUC 0.7–0.9).

### 3.6. Establishment of mRNA-miRNA-lncRNA, mRNA-TF interaction networks

We mapped mRNA-miRNA-lncRNA interaction networks with Cytoscape software (Fig. 9A). In the network, there were 4 mRNAs (ALDH18A1, CENPF, CTH, and CYP27B1), 33 miRNAs, and 33 lncRNAs composed. The particular mRNA-miRNA-lncRNA networks are summarized in the tables (Table S3, Table S4).

Transcription Factor (TF) monitors gene expression through interaction with target genes during the post-transcriptional period. We utilized the CHIPBase and hTFtarget databases to look for TFs that were correlated with key genes. Finally,we obtained 4 key genes (CENPF, CSAD, CTH, CYP27B1) and 96 transcription factors. Totally, there were 167 pairs of mRNA-TF interactions (Table S5) and visualized by Cytoscape software(Fig. 9B).

### 3.7. GSEA and differential analysis

In order to recognize the differential genes in high- and low-risk groups of CKD, we employed limma system to obtain DEGs in these two subgroups. The results were as follows: totally 1639 differentially expressed genes fulfilling $|\log FC| > 0.5$ with P.adj $<0.05$ were acquired. With this cutoff, there were 683 genes with high expression and 956 genes with low expression in high risk group. The

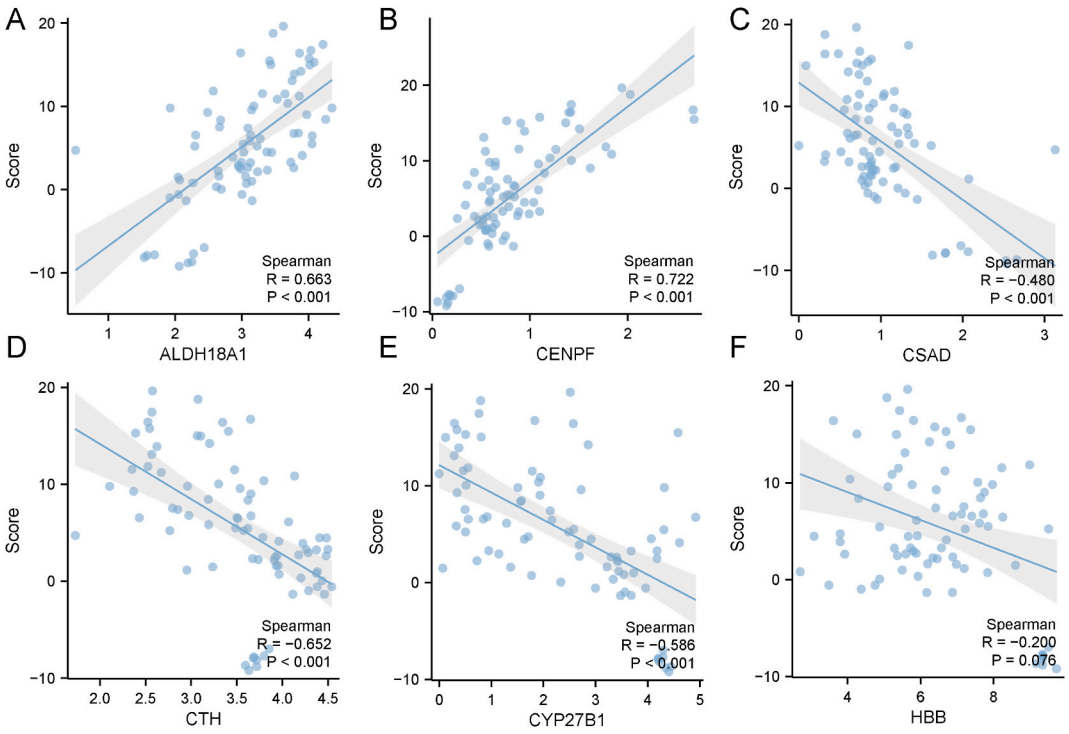



**Fig. 7.** Correlation scatter plot.(A–F) Scatter plot of correlation between key genes ALDH18A1, CENPF, CSAD, CTH, CYP27B1, HBB and the model.

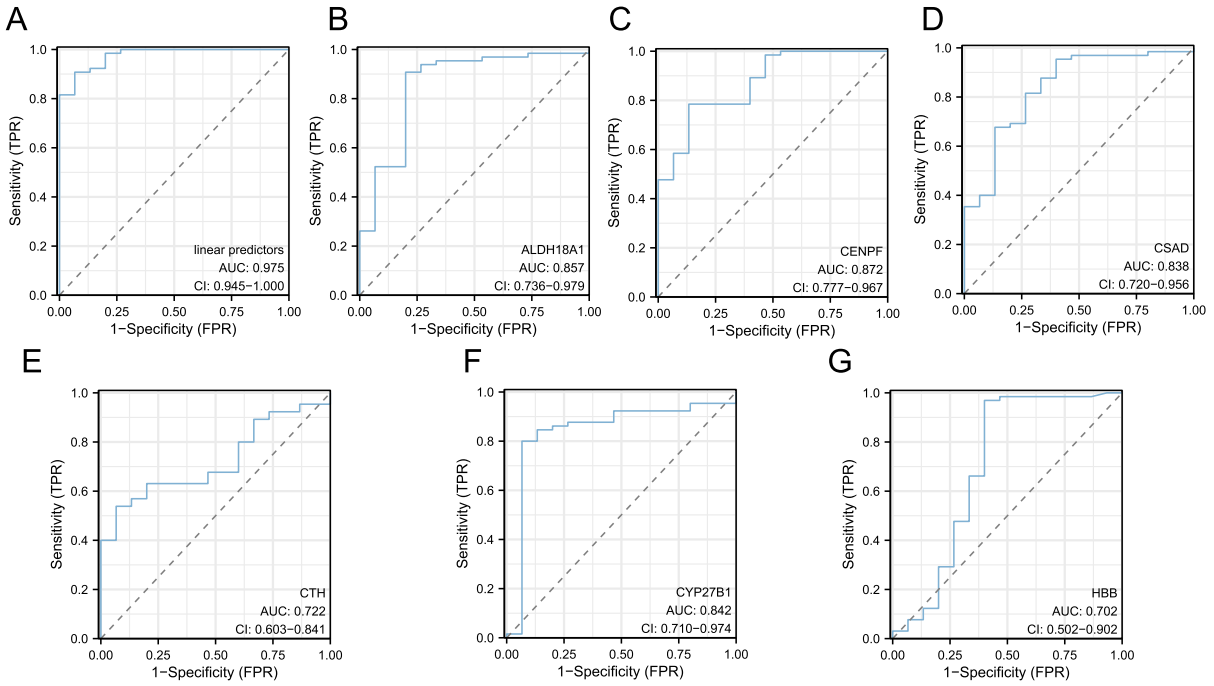


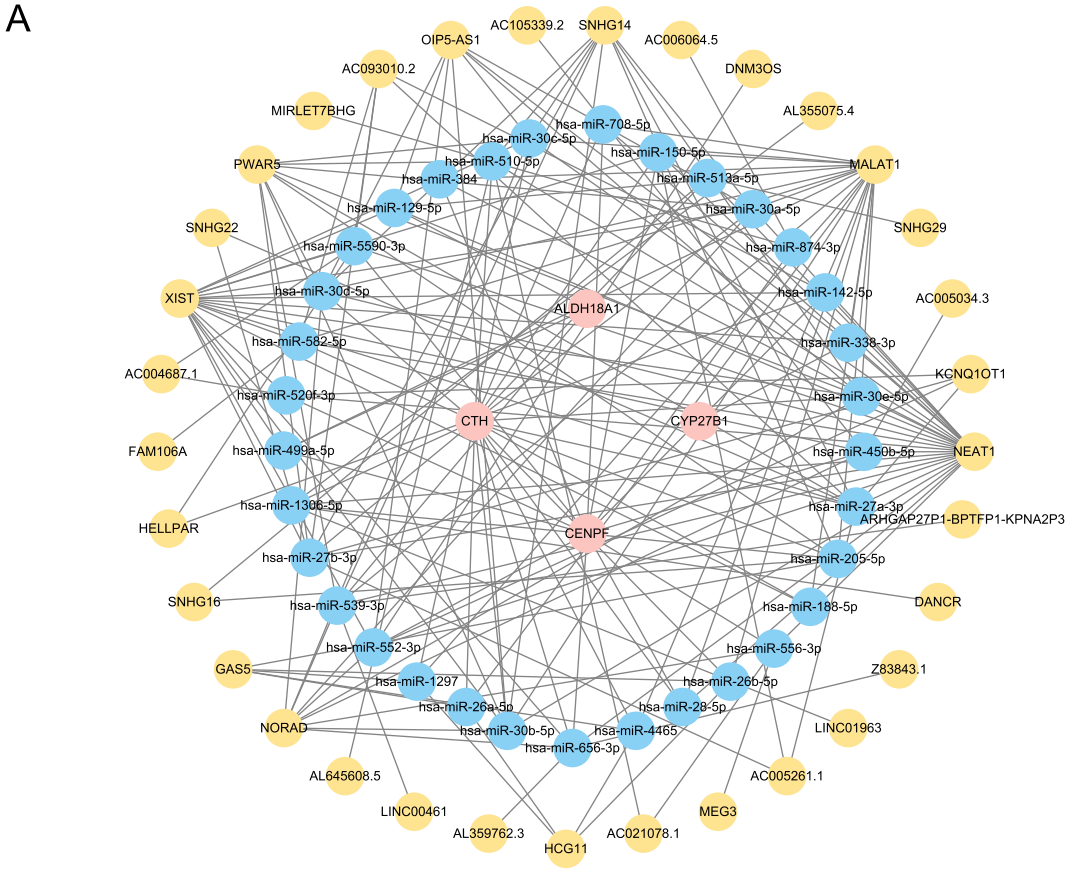

**Fig. 8.** ROC curves.(A) ROC profile outcomes of linear predictors; (B–G) ROC profile results of 6 key genes (ALDH18A1, CENPF, CSAD, CTH, CYP27B1, HBB) with Normal and CKD as the outcome variables.

findings of differential analysis were demonstrated by volcano and heatmap (Fig. 10A and B).

We organized the CKD-dataset by GSEA to identify the biological functions of genes among the groups at high and low risk.The experimental data indicated that genes in the high/low risk subgroup were dramatically enriched in REACTOME_ASSEMBLY_OF_COLLAGEN_FIBRILS_AND_OTHER_MULTIMERIC_STRUCTURES (Fig. 10D), WP_INTERACTIONS_BETWEEN_IMMUNE_CELLS_AND_MICRORNAS_IN_TUMOR_MICROENVIRONMENT (Fig. 10E), WP_CANONICAL_AND_NONCANONICAL_NOTCH_SIGNALING (Fig. 10F), KEGG_RETINOL_METABOLISM (Fig. 10G), KEGG_DRUG_METABOLISM_CYTOCHROME_P450 Fig. 10H), KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM (Fig. 10I), and in a variety of signaling pathways (Fig. 10C–I).

### 3.8. *Immune infiltration profiles*

To investigate the variations in immune cell infiltration, we computed the infiltration abundance of 28 types of immune cells with ssGSEA algorithm. Next, the level of infiltration differences of these 28 immune cells was evaluated and the outcomes were demonstrated through group comparison plots (Fig. 11A). The findings indicate that the infiltration abundance of 18 immune cells varied significantly between the high and low risk groups($P < 0.05$). These immune cells are:Type 2 T helper cell, Effector memory $CD4^+$ T cell, Activated $CD4^+$ T cell, Central memory $CD4^+$ T cell, Type 1 T helper cell, Type 17 T helper cell, Immature dendritic cell, Regulatory T cell, Natural killer cell, MDSC, CD56dim natural killer cell, Gamma delta T cell, Immature B cell, Effector memory $CD8^+$ T cell, Activated $CD8^+$ T cell, Activated B cell, Plasmacytoid dendritic cell, Mast cell. Afterwards,we then computed and presented the association between the infiltration levels of 18 immune cells (Fig. 11B and C). The results of the study showed that there was a positive relationship in the infiltration abundance of most immune cells of the high-risk group (Fig. 11C) and the low-risk group (Fig. 11B). Interestingly, the highest positive correlation between Activated $CD4^+$ T cells and Immature B cells was observed in the low risk groups (Fig. 11B). Meanwhile,the highest positive correlation between Regulatory T cells and MDSC was observed in the high risk groups (Fig. 11C). Immunometabolic changes in kidney disease involving a variety of cell types that contribute to the pathogenesis and progression of kidney disease [52]. Inflammation and fibrosis in chronic kidney disease are associated with aberrant activation of T cells, which are participated in tubular cell apoptosis and impairment of renal function [53]. Meanwhile, infiltration of T cells in glomeruli is correlated with the progression of proteinuria and worsens chronic kidney disease injury [54]. Antibodies generated by B cells are crucial in the development of glomerulonephritis. Inflammatory cell infiltration in kidney disease is exacerbated by B cells which are capable of producing pro-inflammatory cytokines [55]. Additionally, we calculated the connection, filtered at $P < 0.05$, between the expression of six essential genes and the number of these 18 different types of immune cells. The results were displayed using a correlation graph (Fig. 11D). As the achievements illustrated, there were a positive connection between CENPF and the majority of immune cells; there were negative relevance between CYP27B1 and CTH with immune cells. Our results confirm that amino acid metabolism related key genes are associated with immune cell infiltration in chronic kidney disease.It is an interesting finding to hypothesize that possibly abnormal amino acid metabolism and immune cells together promote the progression of kidney disease.




*(caption on next page)*

**Fig. 9.** mRNA-miRNA-lncRNA, mRNA-TF interaction network.(A) mRNA-miRNA-lncRNA interactions network of key genes; (B) mRNA-TF interactions network of key genes. mRNA-miRNA-lncRNA (A) interactions network in which pink round blocks are mRNAs; blue round blocks are miRNAs; yellow round blocks are lncRNAs. mRNA-TF (B) interactions network in which pink circles are mRNAs; green circles are transcription factors (TFs).TF, Transcription Factor.

### *3.9. Correlation analysis of clinical data for key genes*

To assess clinical significance of six key genes in CKD. We retrieved data for estimating glomerular filtration rate(eGFR, MDRD) and serum creatinine on Nephroseq v5 platform. Spearman correlation analysis was used to obtain the information, and gglot2 was used to show results(Fig. 12A–F). The results showed that ALDH18A1(Fig. 12A), CENPF(Fig. 12B) and HBB(Fig. 12F) were negatively correlated with eGFR, while CTH(Fig. 12D) and CYP27B1(Fig. 12E) were positively correlated with eGFR. We also performed correlation analysis between CSAD(Fig. 12C) and serum creatinine, which showed negative association. These findings demonstrate that the key genes we screened are correlative with kidney functional markers, indicating that these genes are potential and valuable biomarkers in the diagnosis and progression of CKD.

### *3.10. Validation expression of key genes*

Renal tubulointerstitial is an essential component of kidney, which is one of the major causes of the pathogenesis and progression in chronic kidney disease. We investigated the expression of six essential genes in HK-2 cells as determined by RT-PCR in order to validate the significance of these genes in the diagnosis and development of CKD. HK-2 cells were incubated with different levels of indole sulfate (uremic toxin) respectively, and RNA was collected to detect differences of key gene expression(Fig. 13A–E). The expression of ALDH18A1 and CENPF were increased compared with control following stimulation by indole sulfate, and the expression were raised with the indole sulfate concentration. Moreover,the expression of CSAD,CTH and CYP27B1 were decreased in response to the indole sulfate. The expression of these key genes were lower in the 2 mM IS compared to the 1 mM IS group. However,we discovered HBB exhibited weakly expression in HK-2. Therefore,we made no statistical analysis about the expression of this gene. On the basis of the above results,we hypothesized that the role of HBB in tubulointerstitial injury in CKD warrants additional experimental confirmation.

## 4. Discussion

Chronic kidney disease(CKD) is predicted to be the world's fifth greatest cause of death by 2040 [29]. Managing CKD and reducing its manifestations are crucial medical concerns. Nevertheless, there are no entirely successful interventions and CKD is still a serious and expanding health problem [30]. Therefore, the management of CKD requires innovative diagnostic and prevention approaches. In addition to being necessary for the production of proteins, amino acids are also involved in signaling and biomolecular metabolism [31]. For immune cells to sustain their basic metabolic processes and maintain their existence, they need constant accessibility to amino acids. Immune cells require amino acids urgently when they are triggered by inflammatory as well as antigen events [32]. Our main goal is to examine how amino acid metabolism related genes(AAMRGs) are altered in chronic kidney disease, with a focus on the potential utility of these genes in the diagnosis and prognosis of the condition. Additionally, we described the variations in immune cell infiltration and the significance of AAMRGs for immune cells.

However, which genes involved in amino acid metabolism are key genes in CKD and which ones have diagnostic value in CKD? We used two machine learning algorithms, LASSO and SVM, to acquire six overlapping genes in the CKD dataset and recognized these six genes as key genes. We constructed logistic regression models with key genes and evaluated the diagnostic capability and accuracy through nomogram and DCA analysis of the models. The models exhibit high diagnostic efficacy in CKD occurrence. In Addition, we plotted the ROC(receiver operating characteristic) curves of key genes. All of six key genes showed promising diagnostic reliability for CKD (AUC 0.7–0.9),including ALDH18A1,CENPF, CSAD,CTH, CYP27B1,HBB. To explore the regulatory factors of key genes, we established interaction networks of key genes with miRNAs, lncRNAs and transcription factors. Our results suggested that these genes have interplay relationships with certain non-coding RNAs and transcription factors, which provides research insights for future in-depth mechanism studies in CKD.

Several amino acids and metabolites are synthesized de novo by ALDH18A1, including proline, ornithine and arginine. Also, ALDH18A1 related metabolic pathways contribute to redox reactions in cells [33,34]. We selected ALDH18A1 via machine learning algorithm which was further confirmed by ROC curve that ALDH18A1 exhibited high diagnostic significance of CKD (AUC 0.857). Moreover, we downloaded transcriptomic data of CKD kidneys in Nephroseq v5, indicating that ALDH18A1 was negatively associated with eGFR. Indole sulfate, a uremic toxin metabolized from amino acids, accumulates with the progression of chronic renal failure. Further, ALDH18A1 gene expression was increased in HK-2 cells compared to controls when stimulated by indole sulfate. Previous studies have confirmed that the severity of CKD correlates with proline, arginine, and glutathione metabolism [35]. Shah VO et al. compared plasma metabolites of nondiabetic individuals and found altered arginine metabolism and increased inflammation with advanced CKD [36]. These analyses validated the potential role of ALDH18A1 in the detection of CKD and modulatory effects in amino acid metabolism. CENPF is known as serving a role in mitotic function [37]. Nevertheless, there are fewer studies on the diagnostic value of CENPF in CKD. Our findings confirm that CENPF may posses a favorable potential in diagnosis with CKD (AUC 0.872). With Nephroseq v5 platform data, we realized that CENPF was also negatively correlated with eGFR. CENPF is markedly overexpressed in patients with adrenocortical carcinoma and with high expression of CENPF predicting a poor prognosis [38]. In HK-2 cells, CENPF

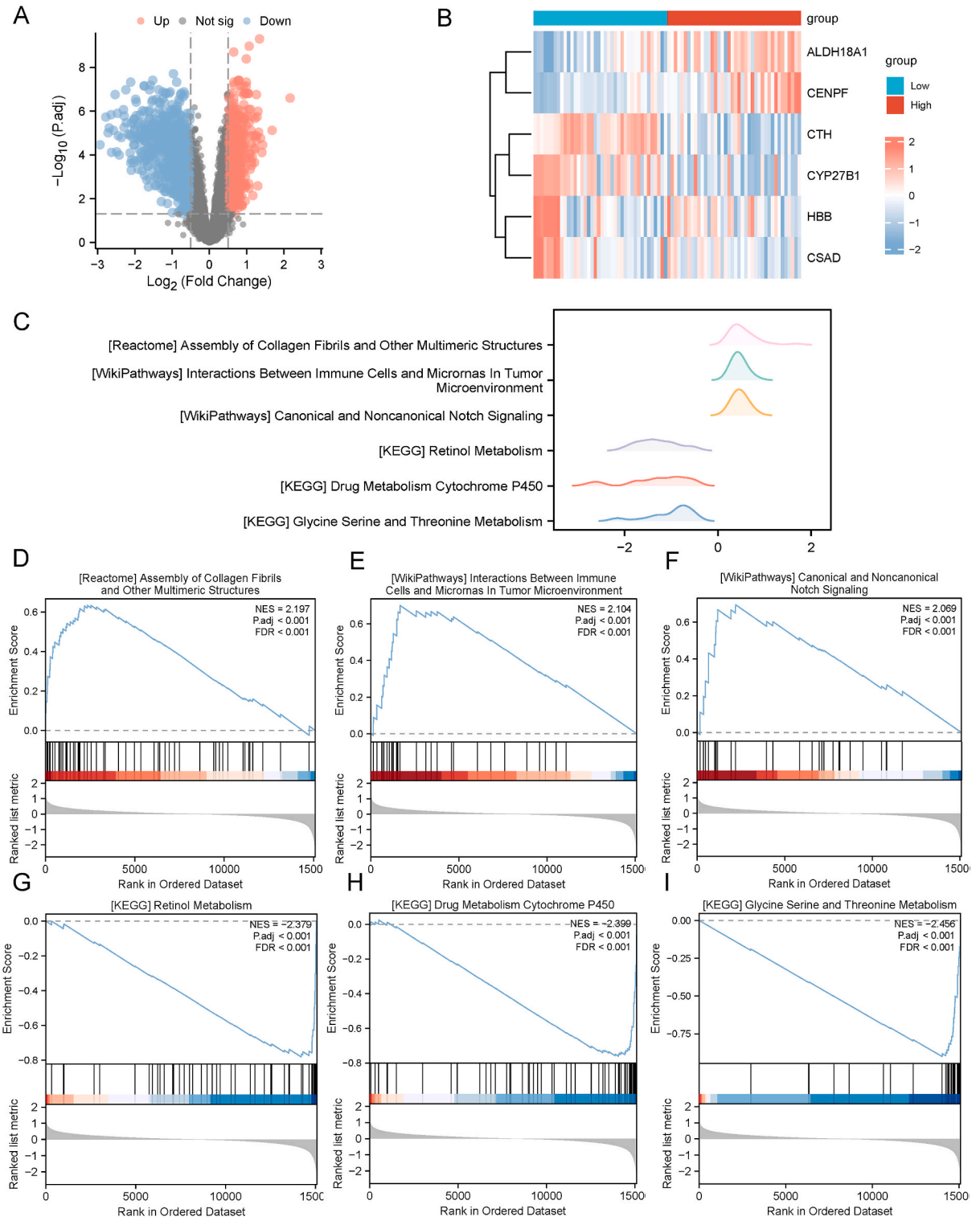



**Fig. 10.** Differential analysis of high and low risk groups and the GSEA.(A) Volcano plot of differential expression genes across subgroups (High/Low risk) in CKD-dataset; (B) Complex heatmap of key genes in high and low risk groups; (C) GSEA of high and low risk groups for the main 6 biological features; (D–I) High and low groups genes are prominently enriched in REACTOME_ASSEMBLY_OF_COLLAGEN_FIBRILS_AND_OTHER_MULTIMERIC_STRUCTURES (D), WP_INTERACTIONS_BETWEEN_ IMMUNE_CELLS_AND_MICRORNAS_IN_TUMOR_MICROENVIRONMENT (E), WP_CANONICAL_AND_NONCANONICAL_NOTCH_SIGNALING (F), KEGG_RETINOL_METABOLISM (G), KEGG_DRUG_METABOLISM_CYTOCHROME_P450 (H), and KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM pathways(I).

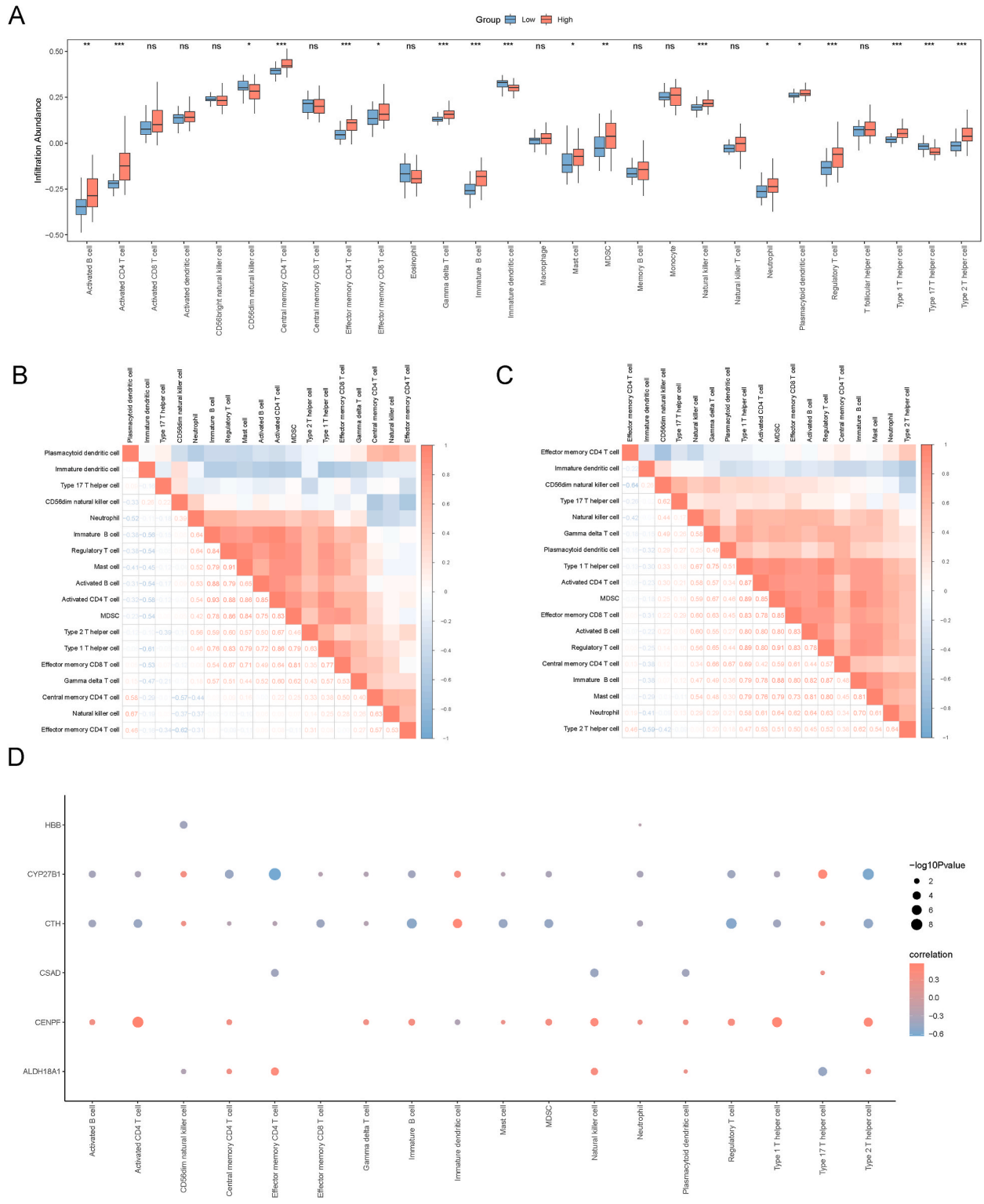



**Fig. 11.** Immune cell infiltration analysis.(A) Graphical presentation of grouped comparative results of ssGSEA immune infiltration analysis between high and low risk groups; (B–C) Correspondence analysis of immune cell infiltration abundance in low-risk (B), and high-risk (C) groups; (D) Correlation Heatmap of immune cells and key genes in CKD-dataset.

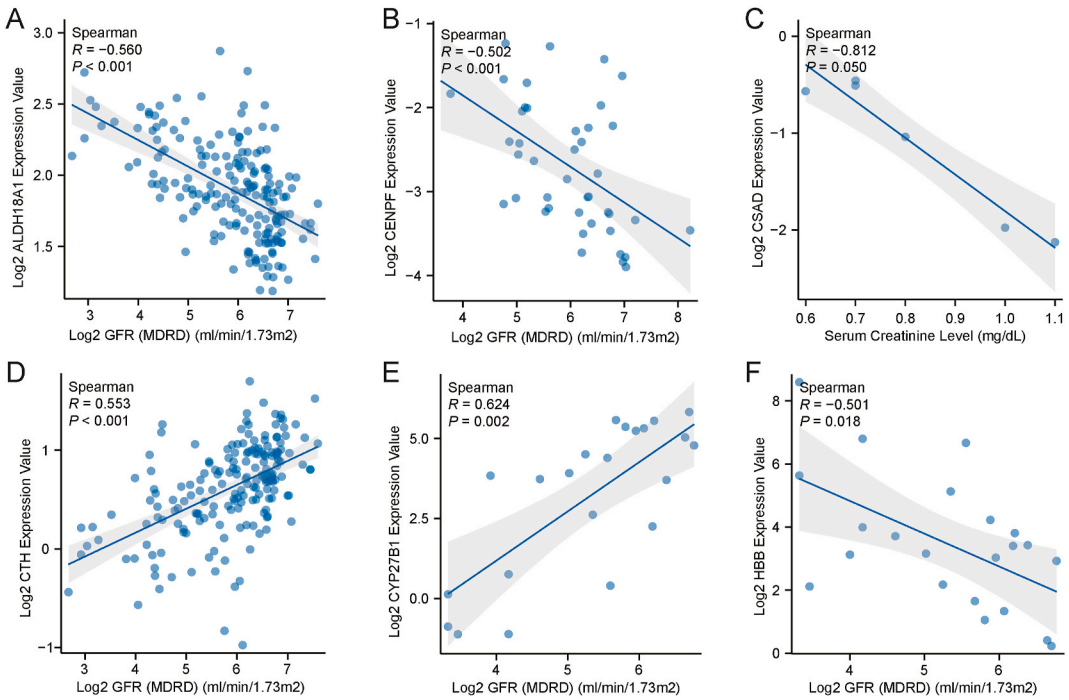



**Fig. 12.** Correlation analysis of clinical data for key genes.(A–B) Scatter plot of correlation between ALDH18A1,CENPF and GFR; (C) Scatter plot of correlation between CSAD and Serum Creatinine; (D–F) Correlation analysis between CTH, CYP27B1, HBB and the GFR.

gene expression was up-regulated with intervention of uremic toxin. The role played by CENPF in the pathologic injury of CKD deserves further study. Overall, we identified ALDH18A1 and CENPF as diagnostic features in CKD which could serve as biomarkers for recognizing CKD.

Cystathionine γ-lyase (CTH),a cytoplasmic enzyme involved in the trans-sulfuration process, which transforms cystathione from methionine into cysteine, is encoded by this gene. The synthesis of glutathione in the liver depends on the availability of cysteine. Methionine protects cells by regulating redox and binding oxidized metabolites. It has been reported that kidney CTH expression was decreased with older rats compared to younger animals [39]. Circulating levels of methionine are elevated in diabetes and diabetic nephropathy as well as predicting the progressive risk of diabetes [40]. The control of physiologically activated vitamin D levels is carried out by CYP27B1. The vitamin D form known as active, 1,25 hydroxyvitamin D [1,25(OH)2D], is created in the kidneys as the enzyme 1-a-hydroxylase(CYP27B1) further hydroxylates 25-hydroxyvitamin D. Active vitamin D and vitamin D receptor combination is capable of modulating the transcription of genes [41]. It has been shown that 1,25(OH)2D3 inhibits uPAR levels and decreased proteinuria in a 5/6 nephrectomized rat FSGS model [42]. We utilized Nephroseq v5 platform data to evaluate the correlation of CTH and CYP27B1 with estimated glomerular filtration(eGFR), which showed that both CTH and CYP27B1 were positively correlated with eGFR. By employing RT-PCR, it was also discovered that HK-2 cells cultivated with indoxyl sulfate treatment displayed considerably lower levels of CTH and CYP28B1 gene expression. Moreover, CTH and CYP28B1 present excellent diagnostic properties in CKD(AUC 0.722 and 0.842 respectively). In conclusion, CTH and CYP28B1 also represent novel diagnostic markers with effective efficiency in CKD.

Cysteine sulfinic acid decarboxylase (CSAD) is the rate-limiting enzyme of taurine bisynthesis.Taurine is considered a cytoprotective molecule with antioxidant properties [43]. In addition, taurine inhibits the expression of pro-inflammatory factors by decreasing ROS and modulates the expression of apoptosis markers via protective functions in the mitochondrial membrane [44]. According to Schaffer S,et al., taurine also decreases the proteinuria and prevents glomerular hypertrophy in diabetic nephropathy [45]. Meanwhile,we characterized the correlation between CSAD and serum creatinine and our results indicated that CSAD was negatively related with serum creatinine (R-0.968,p 0.002). Moreover, CSAD gene expression was reduced in HK-2 cells after uremic toxin intervention. Consistent with previous studies, our investigation also proved that CSAD may be a potential diagnostic marker for CKD(AUC 0.838). Hemoglobin subunit beta(HBB),however, fewer studies have been described on the diagnostics value and evaluation of HBB in CKD. In our work, we screened HBB differentially expressed in CKD. To evaluate its relevance with CKD, we found a negative correlation between HBB and eGFR using Nephroseq v5. However, we validated by RT-PCR and revealed that the expression of HBB was very low in HK-2 cells. The expression level of HBB in HK-2 cells was not analyzed and compared. Further research is required to demonstrate the diagnostic value and prediction capability of HBB in CKD progression in the future.

We divided the CKD dataset into two risk groups, high and low risk, based on the median of the linear predictors in order to better identify the genes that differ between high risk and low risk of CKD as well as the biological characteristics of the genes in the various risk groups. We further analyzed the biological processes in CKD by GESA and the results suggested remarkable enrichment in amino

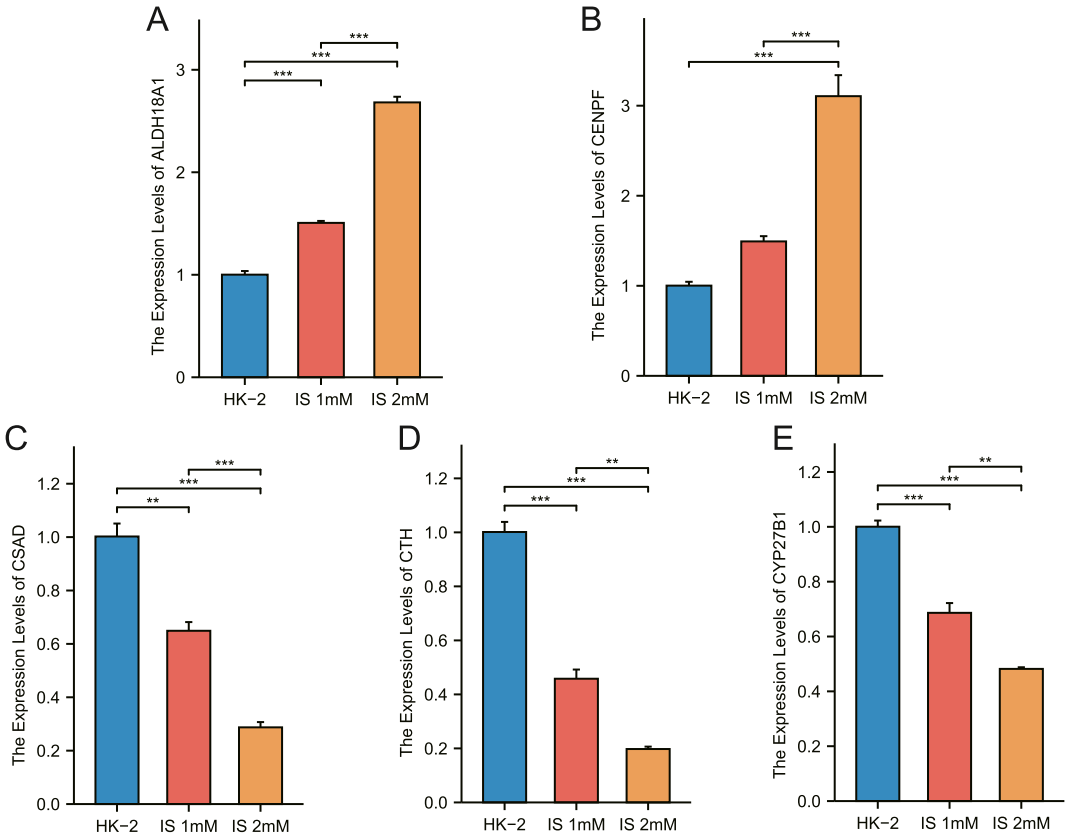



**Fig. 13.** Validation expression of key genes by RT-PCR.(A–E) Confirmed expression of key genes ALDH18A1(A),CENPF(B),CSAD(C),CTH(D) and CYP28B1(E) exhibited with histograms.Blue for HK-2(control) group,red for IS 1 mM group and yellow is IS 2 mM group.IS,Indole sulfate.*P $<$ 0.05,**P $<$ 0.01, ***P $<$ 0.001.

acid metabolism and immune cell pathways. So, whether diverse immune cell infiltration correlates with pathogenesis and progression of CKD? Immune cells are crucial in sustaining kidney homeostasis and processing kidney injury, with distinct immune cell infiltration profiles in chronic kidney disease linked to diverse pathologic damage. Our results confirmed that T cells, including CD4+T cells, CD8+T cells, and regulatory T cells, etc.; B cells, including immature B cells and activated B cells; as well as NK cells and dendritic cells indicates notable variations in infiltration abundance across CKD risk groups at high and low risk. T cells are crucial in the progression of CKD, where T cells activation and infiltration leading to long-term inflammation and kidney fibrosis. T cells infiltration in glomeruli correlates with the occurrence of proteinuria [46,47]. B cells are also of widespread concern in kidney injury. In glomerulonephritis, antibodies generated by B-cell activation are deposited in the glomerulus resulting in proteinuria and kidney injury [48]. Moreover, immune cells proliferation and activation is known to be influenced by metabolic pathways, including amino acid metabolism [49]. Interestingly, we revealed that the expression of key genes regarding amino acid metabolism were prominently correlated with immune cell infiltration. The relevance between cellular metabolism of immune cells and its function attracts extensive attention from researchers. Following T cells activation,T cells features experience dramatic alterations, including up-regulated expression of amino acid transporters, with rapid cell growth and catabolic metabolism of amino acids contributing to their energy and biological functional requirements [50]. Previous studies have implicated that leucine, methionine and tryptophan are crucial for T cell proliferation and clonal expansion, providing vital support in key metabolic events of T cells. In addition, glutamine deficiency in T cells is responsible for serious impairment of immune response to inflammation. Serine is an essential amino acid for purine synthesis and is implicated in T cells division activity [51]. Collectively, our findings established that immune cells varied in high and low risk groups of CKD. Furthermore, immune cells are relevant to amino acid metabolism-related key genes. Our research results provide a novel insight and perspective for future investigation on the influence of amino acid metabolism in immune cell responses in CKD.

Our investigation revealed that amino acid metabolism is strongly associated with the progression in chronic kidney disease. Immune cell infiltration and immune response are connected with glomerular and tubular injury in chronic kidney disease. This study also proved that amino acid metabolism-related genes are involved in immune cell infiltration of renal tissues, suggesting that amino acid metabolism together with immune cells contribute to the pathological injury in chronic kidney disease, which is an interesting finding and deserves further mechanistic studies. However, this study also has several limitations worth consideration. Firstly, This was an investigation that relied on previous CKD transcriptomic data. Further validation of the findings is required in future with larger clinical samples of CKD. Secondly, it remains inconclusive the specific biological functions of key genes in the development of CKD,

although the relative features of these genes have been analyzed using bioinformatics analyses. Besides, our analysis lacked examination of protein functions and mechanisms. Although we used indole sulfate (IS) in culturing HK-2 cells and verified the expression of key genes by qPCR, protein blotting (WB) experiments and immunohistochemistry (IHC) have not yet been performed to validate the protein expression levels. This deficiency may lead to an incomplete understanding of the functional relationship between genes and proteins, thus affecting the interpretation of the results. Meanwhile, whether amino acid metabolism in chronic kidney disease is involved in the pathological mechanism of renal tubulointerstitial fibosis is a very interesting study, but it was not confirmed in this study, which is also a limitation of our study. It is very well worth for us to investigate the gene and protein expression and alterations of amino acid metabolism in the pathological injury of renal tubulointerstitial fibrosis. Despite these limitations, we still believe that the present study provides fundamental data for the urgently needed field of chronic kidney disease and provides an important direction for further in-depth studies. In the future, we are planning to carry out further experiments to explore the influences of IS on other cells types and to complement studies related to protein expression in order to fully characterize their roles in chronic kidney disease.

## 5. Conclusions

In conclusion, our study screened amino acid metabolism key genes by machine learning algorithms and established diagnostic models of key genes as well as biological features. Moreover, we provided key genes and immune cells infiltration landscapes in high-risk and low-risk groups of CKD. Finally, we validated the expression and clinical relevance of amino acid metabolism key genes via cultured cells in vitro and clinical data. The present research provides novel diagnostic and intervention markers of the pathogenesis and progression in chronic kidney disease.

## CRediT authorship contribution statement

**Guoqing Zhang:** Writing – original draft. **Hongyan Luo:** Data curation, Conceptualization. **Xiaohua Lu:** Data curation, Conceptualization. **Yonghua Liu:** Investigation, Formal analysis. **Mei Wang:** Investigation, Formal analysis. **Bo Li:** Methodology, Investigation. **Haixia Lu:** Validation. **Yali Zheng:** Funding acquisition.

## Data availability

All data analyzed in this study are available in the supplementary files.The datasets were downloaded from GEO repository (numbers are GSE66494 and GSE116626).

## Ethics approval

The procedures and methods used in this study were in accordance with the ethical guidelines of the Helsinki and relevant regulations. All experimental procedures were approved by the ethics committee of Ningxia Medical University Affiliated People's Hospital of Autonomous Region. The results from this study are original, guided by ethical requirements. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

## Funding

This research was supported by Natural Science Foundation of Ningxia Province (No.2021AAC03311,No.2023AAC03458, No.2022AAC02059, No. 2022AAC03362),National Natural Science Foundation of China(No.81860136, No.81460161), and The Key Research and Development Program of Ningxia Province Region projects (No.2019BEG03017 and No.2022BEG03121).

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2025.e41872.

## References

[1] K. Kalantar-Zadeh, T.H. Jafar, D. Nitsch, B.L. Neuen, V. Perkovic, Chronic kidney disease, Lancet (London, England) 398 (10302) (2021) 786–802.

[2] L. Zhang, F. Wang, L. Wang, W. Wang, B. Liu, J. Liu, M. Chen, Q. He, Y. Liao, X. Yu, N. Chen, J.E. Zhang, Z. Hu, F. Liu, D. Hong, L. Ma, H. Liu, X. Zhou, J. Chen, L. Pan, W. Chen, W. Wang, X. Li, H. Wang, Prevalence of chronic kidney disease in China: a cross-sectional survey, Lancet (London, England) 379 (9818) (2012) 815–822.

[3] H.L. MacLaughlin, A.N. Friedman, T.A. Ikizler, Nutrition in kidney disease: core curriculum 2022, Am. J. Kidney Dis. : the official journal of the National Kidney Foundation 79 (3) (2022) 437–449.

[4] G. Garibotto, A. Sofia, S. Saffioti, A. Bonanni, I. Mannucci, D. Verzola, Amino acid and protein metabolism in the human kidney and in patients with chronic kidney disease, Clinical nutrition 29 (4) (2010) 424–433.

[5] K. Kalantar-Zadeh, D. Fouque, Nutritional management of chronic kidney disease, N. Engl. J. Med. 377 (18) (2017) 1765–1776.

[6] L. Yang, Z. Chu, M. Liu, Q. Zou, J. Li, Q. Liu, Y. Wang, T. Wang, J. Xiang, B. Wang, Amino acid metabolism in immune cells: essential regulators of the effector functions, and promising opportunities to enhance cancer immunotherapy, J. Hematol. Oncol. 16 (1) (2023) 59.

[7] Q.Q. Chen, K. Liu, N. Shi, G. Ma, P. Wang, H.M. Xie, S.J. Jin, T.T. Wei, X.Y. Yu, Y. Wang, J.Y. Zhang, P. Li, L.W. Qi, L. Zhang, Neuraminidase 1 promotes renal fibrosis development in male mice, Nat. Commun. 14 (1) (2023) 1713.

[8] S.N. Cox, S. Chiurlia, C. Divella, M. Rossini, G. Serino, M. Bonomini, V. Sirolli, F.B. Aiello, G. Zaza, I. Squarzoni, C. Gangemi, M. Stangou, A. Papagianni, M. Haas, F.P. Schena, Formalin-fixed paraffin-embedded renal biopsy tissues: an underexploited biospecimen resource for gene expression profiling in IgA nephropathy, Sci. Rep. 10 (1) (2020) 15164.

[9] T. Barrett, D.B. Troup, S.E. Wilhite, P. Ledoux, D. Rudnev, C. Evangelista, I.F. Kim, A. Soboleva, M. Tomashevsky, R. Edgar, NCBI GEO: mining tens of millions of expression profiles–database and tools update, Nucleic Acids Res. 35 (Database issue) (2007 Jan) D760–D765.

[10] S. Davis, P.S. Meltzer, GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor, Bioinformatics 23 (14) (2007) 1846–1847.

[11] G. Stelzer, N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T.I. Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, A. Kohn, N. Rappaport, M. Safran, D. Lancet, The GeneCards suite: from gene data mining to disease genome sequence analyses, Current protocols in bioinformatics 54 (2016) 1.30.1–1.30.33.

[12] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J.P. Mesirov, P. Tamayo, The Molecular Signatures Database (MSigDB) hallmark gene set collection, Cell systems 1 (6) (2015) 417–425.

[13] J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, Bioinformatics 28 (6) (2012) 882–883.

[14] M. Ringnér, What is principal component analysis? Nat. Biotechnol. 26 (3) (2008) 303–304.

[15] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (7) (2015) e47.

[16] Gene Ontology Consortium, Gene Ontology consortium: going forward, Nucleic Acids Res. 43 (Database issue) (2015 Jan) D1049–D1056.

[17] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (1) (2000) 27–30.

[18] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS A J. Integr. Biol. 16 (5) (2012) 284–287.

[19] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U.S.A. 102 (43) (2005) 15545–15550.

[20] W. Cai, M. van der Laan, Nonparametric bootstrap inference for the targeted highly adaptive least absolute shrinkage and selection operator (LASSO) estimator, Int. J. Biostat. (2020).

[21] H. Sanz, C. Valim, E. Vegas, J.M. Oller, F. Reverter, SVM-RFE: selection and visualization of the most relevant features through non-linear kernels, BMC Bioinf. 19 (1) (2018) 432.

[22] T. Tataranni, C. Piccoli, Dichloroacetate (DCA) and cancer: an overview towards clinical applications, Oxid. Med. Cell. Longev. 2019 (2019) 8201079.

[24] J.H. Li, S. Liu, H. Zhou, L.H. Qu, J.H. Yang, starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data, Nucleic Acids Res. 42 (Database issue) (2014 Jan) D92–D97.

[25] Y. Chen, X. Wang, miRDB: an online database for prediction of functional microRNA targets, Nucleic Acids Res. 48 (D1) (2020) D127–d131.

[26] K.R. Zhou, S. Liu, W.J. Sun, L.L. Zheng, H. Zhou, J.H. Yang, L.H. Qu, ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data, Nucleic Acids Res. 45 (D1) (2017) D43–d50.

[27] Q. Zhang, W. Liu, H.M. Zhang, G.Y. Xie, Y.R. Miao, M. Xia, A.Y. Guo, hTFtarget: a comprehensive database for regulations of human transcription factors and their targets, Dev. Reprod. Biol. 18 (2) (2020) 120–128.

[28] B. Xiao, L. Liu, A. Li, C. Xiang, P. Wang, H. Li, T. Xiao, Identification and verification of immune-related gene prognostic signature based on ssGSEA for osteosarcoma, Front. Oncol. 10 (2020) 607622.

[29] K.J. Foreman, N. Marquez, A. Dolgert, K. Fukutaki, N. Fullman, M. McGaughey, M.A. Pletcher, A.E. Smith, K. Tang, C.W. Yuan, J.C. Brown, J. Friedman, J. He, K.R. Heuton, M. Holmberg, D.J. Patel, P. Reidy, A. Carter, K. Cercy, A. Chapin, D. Douwes-Schultz, T. Frank, F. Goettsch, P.Y. Liu, V. Nandakumar, M.B. Reitsma, V. Reuter, N. Sadat, R.J.D. Sorensen, V. Srinivasan, R.L. Updike, H. York, A.D. Lopez, R. Lozano, S.S. Lim, A.H. Mokdad, S.E. Vollset, C.J.L. Murray, Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016-40 for 195 countries and territories, Lancet (London, England) 392 (10159) (2018) 2052–2090.

[30] J. Radhakrishnan, G. Remuzzi, R. Saran, D.E. Williams, N. Rios-Burrows, N. Powe, K. Brück, C. Wanner, V.S. Stel, S.K. Venuthurupalli, W.E. Hoy, H.G. Healy, A. Salisbury, R.G. Fassett, D. O'Donoghue, P. Roderick, S. Matsuo, A. Hishida, E. Imai, S. Iimuro, Taming the chronic kidney disease epidemic: a global view of surveillance efforts, Kidney Int. 86 (2) (2014) 246–250.

[31] N.S. Chandel, Amino acid metabolism, Cold Spring Harbor Perspect. Biol. 13 (4) (2021 Apr 1) a040584.

[32] T.L. McGaha, L. Huang, H. Lemos, R. Metz, M. Mautino, G.C. Prendergast, A.L. Mellor, Amino acid catabolism: a pivotal regulator of innate and adaptive immunity, Immunol. Rev. 249 (1) (2012) 135–157.

[33] C.A. Hu, S. Khalil, S. Zhaorigetu, Z. Liu, M. Tyler, G. Wan, D. Valle, Human Delta1-pyrroline-5-carboxylate synthase: function and regulation, Amino Acids 35 (4) (2008) 665–672.

[34] Z. Yang, X. Zhao, W. Shang, Y. Liu, J.F. Ji, J.P. Liu, C. Tong, Pyrroline-5-carboxylate synthase senses cellular stress and modulates metabolism by regulating mitochondrial respiration, Cell Death Differ. 28 (1) (2021) 303–319.

[35] H. Wang, A. Ainiwaer, Y. Song, L. Qin, A. Peng, H. Bao, H. Qin, Perturbed gut microbiome and fecal and serum metabolomes are associated with chronic kidney disease severity, Microbiome 11 (1) (2023) 3.

[36] V.O. Shah, R.R. Townsend, H.I. Feldman, K.L. Pappan, E. Kensicki, D.L. Vander Jagt, Plasma metabolomic profiles in different stages of CKD, Clin. J. Am. Soc. Nephrol. : CJASN 8 (3) (2013) 363–370.

[37] A. Varis, A.L. Salmela, M.J. Kallio, Cenp-F (mitosin) is more than a mitotic marker, Chromosoma 115 (4) (2006) 288–295.

[38] Y.G. Huang, D. Li, L. Wang, X.M. Su, X.B. Tang, CENPF/CDK1 signaling pathway enhances the progression of adrenocortical carcinoma by regulating the G2/M-phase cell cycle, J. Transl. Med. 20 (1) (2022) 78.

[39] D. Szlęzak, T. Hutsch, M. Ufnal, M. Wróbel, Heart and kidney H(2)S production is reduced in hypertensive and older rats, Biochimie 199 (2022) 130–138.

[40] S. Lamichhane, E. Kemppainen, K. Trošt, H. Siljander, H. Hyöty, J. Ilonen, J. Toppari, R. Veijola, T. Hyötyläinen, M. Knip, M. Orešič, Circulating metabolites in progression to islet autoimmunity and type 1 diabetes, Diabetologia 62 (12) (2019) 2287–2297.

[41] A.W. Norman, From vitamin D to hormone D: fundamentals of the vitamin D endocrine system essential for good health, Am. J. Clin. Nutr. 88 (2) (2008) 491s–499s.

[42] J. Ma, B. Zhang, S. Liu, S. Xie, Y. Yang, J. Ma, Y. Deng, W. Wang, L. Xu, R. Li, L. Zhang, C. Yu, W. Shi, 1,25-dihydroxyvitamin D(3) inhibits podocyte uPAR expression and reduces proteinuria, PLoS One 8 (5) (2013 May 31) e64912.

[43] M. Fakruddin, F.Y. Wei, T. Suzuki, K. Asano, T. Kaieda, A. Omori, R. Izumi, A. Fujimura, T. Kaitsuka, K. Miyata, K. Araki, Y. Oike, L. Scorrano, T. Suzuki, K. Tomizawa, Defective mitochondrial tRNA taurine modification activates global proteostress and leads to mitochondrial disease, Cell Rep. 22 (2) (2018) 482–496.
[44] J. Ma, Z. Yang, S. Jia, R. Yang, A systematic review of preclinical studies on the taurine role during diabetic nephropathy: focused on anti-oxidative, anti-inflammation, and anti-apoptotic effects, Toxicol. Mech. Methods 32 (6) (2022) 420–430.
[45] S. Schaffer, H.W. Kim, Effects and mechanisms of taurine as a therapeutic agent, Biomolecules & therapeutics 26 (3) (2018) 225–241.
[46] R. Sharma, G.R. Kinsey, Regulatory T cells in acute and chronic kidney diseases, Am. J. Physiol. Ren. Physiol. 314 (5) (2018) F679–f698.
[47] T.T. Tapmeier, A. Fearn, K. Brown, P. Chowdhury, S.H. Sacks, N.S. Sheerin, W. Wong, Pivotal role of CD4+ T cells in renal fibrosis following ureteric obstruction, Kidney Int. 78 (4) (2010) 351–362.
[48] E. Hoxha, L. Reinhard, R.A.K. Stahl, Membranous nephropathy: new pathogenic mechanisms and their clinical implications, Nat. Rev. Nephrol. 18 (7) (2022) 466–478.
[49] J. Li, Y. Yang, Y. Wang, Q. Li, F. He, Metabolic signatures of immune cells in chronic kidney disease, Expet Rev. Mol. Med. 24 (2022) e40.
[50] N.M. Chapman, M.R. Boothby, H. Chi, Metabolic coordination of T cell quiescence and activation, Nat. Rev. Immunol. 20 (1) (2020) 55–70.
[51] C. Han, M. Ge, P.C. Ho, L. Zhang, Fueling T-cell antitumor immunity: amino acid metabolism revisited, Cancer Immunol. Res. 9 (12) (2021) 1373–1382.
[52] P.J. Basso, V. Andrade-Oliveira, N.O.S. Camara, Targeting immune cell metabolism in kidney diseases, Nat. Rev. Nephrol. 17 (2021) 465–480.
[53] L. Xu, J. Guo, D.G. Moledina, L.G. Cantley, Immune-mediated tubule atrophy promotes acute kidney injury to chronic kidney disease transition, Nat. Commun. 13 (2022) 4892.
[54] S.Y. Hu, X.Y. Jia, J.N. Li, X. Zheng, J. Ao, G. Liu, Z. Cui, M.H. Zhao, T cell infiltration is associated with kidney injury in patients with anti-glomerular basement membrane disease, Sci. China Life Sci. 59 (2016) 1282–1289.
[55] M.J. Smith, K.M. Simmons, J.C. Cambier, B cells in type 1 diabetes mellitus and diabetic kidney disease, Nat. Rev. Nephrol. 13 (2017) 712–720.