European Review for Medical and Pharmacological Sciences 2024; 28: 863-878

Development of a clinical prediction model for diabetic kidney disease with glucose and lipid metabolism disorders based on machine learning and bioinformatics technology

Z. BI[1], L.-J. WANG[1], Y.-X. LIN[2], Y.-Y. ZHANG[1], S.-H. WANG[2], Z.-H. FANG[2,3]

[1]Anhui University of Chinese Medicine, Hefei, China
[2]The First Affiliated Hospital of Anhui, University of Chinese Medicine, Hefei, China,
[3]Xin'an Institute of Medicine and Modernization of Traditional Chinese Medicine, Hefei National Science Center, Hefei, China

Abstract. – OBJECTIVE: In this study, we investigated the internal relationship between the pathogenesis of diabetic kidney disease (DKD) and abnormal glucose and lipid metabolism to identify potential biomarkers for diagnosis and treatment and investigated the role of the immune microenvironment of glucose and lipid metabolism disorders in the occurrence and progression of DKD.

MATERIALS AND METHODS: The chip datasets GSE104948 and GSE96804 from the Gene Expression Common Database (GEO) were merged using the "lima" and "sva" software packages in R Software (4.2.3), and the merged dataset was used as the validation set. The intersection between the differential genes of DKD and the glucose and lipid metabolism genes in the MSigDB database was identified, and a nomogram of the incidence risk of DKD was built using three machine learning methods, namely LASSO regression, support vector machine (SVM), and random forest (RF), to validate the accuracy of the prediction model. Immune scores were conducted using the unsupervised clustering method, and patients were divided into two subgroups. The two subgroups were screened for differential genes for enrichment analysis. The differential genes of patients diagnosed with DKD were clustered into two gene subgroups for co-expression analysis. In this study, we utilized the Cytoscape software to construct a network of interactions among key genes.

RESULTS: Using machine learning, a diagnostic model was developed with G6PC and HSD17B14 as key factors. Enrichment analysis and immune scoring demonstrated that the development of DKD was related to the imbalance in the microenvironment brought about by glucose lipid metabolism disorders.

CONCLUSIONS: G6PC and HSD17B14 may be potential biomarkers for DKD, and the established predictive model is more helpful in predicting the incidence of DKD.

Key Words:
Bioinformatics, Diabetic kidney disease, Glucose metabolism, Immune microenvironment, Lipid metabolism, Machine learning.

Abbreviations

Diabetic Kidney Disease: DKD; Gene Expression Omnibus: GEO; Differential expression analysis: DEG; The Molecular Signatures Database: MSigDB; Support vector machines: SVM; Random Forest: RF; Protein-Protein Interaction Networks: PPI Networks ; Tumor microenvironment: TME; Human leukocyte antigen: HLA; Kyoto Encyclopedia of Genes and Genomes: KEGG; Weighted correlation network analysis: WGCNA; Natural killer cell: NK cell; Tricarboxylic Acid Cycle: TCA cycle; Principal Component Analysis: PCA; Reactive oxygen species: ROS.

Genes

CYP27B1: Cytochrome P450 family 27 subfamily B member 1; *HSD17B14*: Hydroxysteroid 17-beta dehydrogenase 14; *G6PC*: Glucose-6-phosphatase catalytic subunit; *TGFBI*: Transforming growth factor beta-induced; *FABP1*: Fatty acid binding protein 1; *EHHADH*: Enoyl-CoA hydratase and 3-hydroxyacyl CoA dehydrogenase; *ACOX1*: Acyl-CoA oxidase 1; *ECH1*: Enoyl CoA hydratase 1; *ECI2*: Enoyl-CoA delta isomerase 2; *ACOX2*: Acyl-CoA oxidase 2; *CRAT*: Carnitine O-acetyltransferase; CAT: Catalase; SLC27A2: Solute carrier family 27 member 2; *PECR*: Peroxisomal trans-2-enoyl-CoA reductase; *HAO2*: Hydroxyacid oxidase 2.

Introduction

Diabetic kidney disease (DKD) is one of the most prevalent microvascular complications of diabetes, and the kidney is a key target organ for microvascular injury in diabetes[1]. DKD is

Corresponding Author: Zhao-Hui Fang, MD; e-mail: fangzhaohuigh@126.com

the leading cause of end-stage renal disease. Relevant studies[2] revealed that the incidence of DKD differs among high-risk subgroups. The molecular mechanism of DKD is extremely complex and is characterized by multiple targets, stages, and genes. Several studies[3] revealed that changes in the kidney microenvironment, genetics, appearance, and other factors are involved in the pathogenesis of DKD. DKD is typically characterized by the interaction between genes and immune microenvironment[4]. Early diagnosis and targeted treatment of DKD have been extremely challenging due to its progressiveness and heterogeneity and the concealment of the origin of metabolic and genetic epigenetic disorders. Experimentation and clinical evidence[5] suggest that the activation of immune responses and chronic metabolic inflammation play a significant role in the progression of diseases. Over time, the accumulation of immune complexes may exacerbate kidney damage, whereas genetic/epigenetic factors may increase susceptibility to immune responses in specific populations. Poor glucose control (chronic hyperglycemia, transient hyperglycemia, hypoglycemia), hypertension, dyslipidemia, and diabetes family history are all established factors for identifying high-risk groups of DKD[6]. Kidney lipotoxicity caused by lipid metabolism disorders may be the pathogenesis of DKD and kidney insufficiency, and the development of lipomics has the potential for diagnosing and treating DKD[7,8]. Studies[9,10] have revealed that kidney tissues possess specific metabolic reprogramming and molecular pathological characteristics and that metabolic disorders exacerbate kidney stress and promote inflammatory responses. The interaction of these pathophysiological factors can trigger a series of chain reactions that result in glomerular filtration barrier damage[11]. Therefore, to explore a new vision for the diagnosis and treatment of DKD, metabolism was chosen as the target, and the immune microenvironment related to metabolism was analyzed.

In recent years, bioinformatics technology and microarray technology have been used extensively to identify pathogenic factors and disease mechanisms. The application of machine learning in bioinformatics has matured and enhanced over time, and the combination of these methods provides a solid foundation for disease diagnosis and prognosis. In this study, we developed a prediction model of diabetes related to glucose and lipid metabolism for establishing DKD, investigated the correlation between the pathogenesis of DKD and the immune microenvironment, and provided additional biomarkers for clinical diagnosis as well as new ideas for targeted immunotherapy.

Materials and Methods

Pre-Treatment of Gene Expression and Identification of Differential Genes

Gene Expression Omnibus (GEO) is an open gene expression database that stores chips, 2G sequencing, and other forms of high-throughput sequencing data. In this study, microarray datasets related to DKD were downloaded from the GEO (https://www.ncbi.nlm.nih.gov/geo/) database (the training set comprised a total of 71 patients, including 41 DKD patients and 20 non-DKD patients from GSE96804, as well as 7 DKD patients and 3 non-DKD patients from GSE104948. The validation set included 22 patients from GSE30528, with 9 DKD patients and 13 non-DKD patients). GSE104948 and GSE96804 were combined into one dataset for expression analysis. We used GSE30528 as the external validation dataset. The "limma" and "sva" software packages were used to eliminate batch effects and normalize the two sets of data. We used the "limma" package to identify differentially expressed genes (DEGs) between patients diagnosed with DKD and non-DKD patients.

Acquisition of Glucose and Lipid Metabolism Genes

In this study, 326 glycolysis-related genes (HALLMARK_GLYCOLYSIS, REACTOME_GLYCOLYSIS, BIOCARTA_GLYCOLYSIS-PATHWAY, KEGG_GLYCOLYSIS_GLUCONEOGENESIS, GO_GLYCOLYTIC-PROCESS) and 742 lipid metabolism-related genes (search term: "lipid metabolism" Select REACTOME_METABOLISM_OF_LIPIDS) were obtained and screened. Based on the "VennDiagram" software package, 1,050 genes related to glucose and lipid metabolism were taken and combined. Intersect genes related to glucose and lipid metabolism with DKD-DEGs were used as candidate genes for constructing the model.

Establishment and Validation of Candidate Genes and Risk Models Related to DKD Screening Based on Machine Learning

Support vector machine (SVM), LASSO regression, and random forest (RF) feature selection algorithms were used to screen key genes, and we used the intersection of the three machine learning genes to build a logistic regression model using the "glmnet" and "randomForest" packages. Using a column chart scoring system, the curve was calibrated to assess the accuracy of the column chart, and clinical utility was assessed using a clinical decision-making curve. The GSE30528 dataset was selected as the validation group for assessing the performance of the model on external datasets.

Analysis of Patient Subtypes and Immune Infiltration and Enrichment

DKD is a disease with biological heterogeneity, and accurate subtype identification and individualized prognosis are essential in preventing DKD from developing into end-stage kidney disease. Therefore, the "ConsensusClusterPlus" software package was used to perform unsupervised clustering for typing patients diagnosed with DKD based on the second step of obtaining DEGs in glucose and lipid metabolism. An ensemble clustering method was utilized, involving repeated subsampling of the dataset, application of the optimal clustering algorithm, and aggregation of results to identify consensus across iterations. Clustering analysis revealed that when samples were divided into two clusters, the samples within each cluster were relatively homogenous, leading to the formation of new molecular subtypes. To further assess the relationship between molecular subtypes and microenvironment, we used the ESTIMATE algorithm to calculate matrix scores, immune scores, and ESTIMATE scores, and to compare the differences among subtypes. Support vector regression analysis of gene expression data from mixed tissue samples was employed to estimate the relative abundance of various immune cell types, with the statistical significance of the estimates for each sample assessed using 1,000 permutation tests. DKD occurs in a microenvironment with a complex load system. We used the CIBERSORT algorithm to characterize immune cell composition based on gene expression profiles, and we used the Wilcoxon test to compare immune cell infiltration scores and tumor microenvironment (TME) scores across different subtypes. To examine the biological functional differences between subtypes, we used the Kyoto Encyclopedia of Genes and Genomes (KEGG, available at: https://www.genome.jp/kegg/) enrichment analysis to identify and analyze the biological functions of DEGs between the two subgroups, as well as the metabolic pathways and gene relationships.

Gene Subtypes and Immune Checkpoint Genes and Human Leukocyte Antigen Family Genes

Using the "ConsensusClusterPlus" software package, cluster searches were conducted for gene subtypes between the differential genes in the two subgroups, based on the subtypes of the patients. Human leukocyte antigen (HLA) is a gene family that regulates cell recognition and immune responses. Many studies have demonstrated that the HLA gene locus may interact genetically with the Type 1 and Type 2 diabetes regulatory genes. Immune checkpoints play crucial roles in regulating the expression of immune cells and immune activation. We used the Wilcoxon test to assess the expression differences between immune checkpoint genes and HLA family genes in glucose and lipid metabolism DEGs between subtypes.

Weighted Gene Co-Expression Network Analysis

We used the "WGCNA" software package to identify and cluster gene modules with high co-expression between subgroups. WGCNA analysis was performed between two gene subgroups, and the module genes were enriched with KEGG. The protein–protein interaction network was visualized using Cytoscape software (version 3.7.2, available at: https://cytoscape.org/index.html), and 10 hub genes were identified using the cytoHubba plugin.

Statistical Analysis

All statistical analyses and data processing were performed using R software (version 4.2.3, available at: https://www.r-project.org, The R Foundation for Statistical Computing, Vienna, Austria). The packages (available at: http://bioconductor.org/). A highly significant result, denoted by "***," indicates a p-value lower than 0.001 between the two groups. A result denoted by "**" signifies a p-value between 0.001 and 0.01, while "*" corresponds to a p-value between 0.01 and 0.05. These thresholds represent levels

of statistical significance, conferring a reasonable degree of confidence in the observed differences between groups. Significant differential gene expression in this study is defined by a corrected *p*-value (*p*-adjust) below 0.05 and an absolute log fold change (|logFC|) exceeding 1. Expression directionality, classified as upregulated or downregulated, is assigned for |logFC| values greater than 1 and lower than -1, respectively.

Results

Identification and Screening of DEGs

We used the merged datasets GSE96804 and GSE104948 and the COMBAT function in the sva package to eliminate batch effects, and principal component analysis (PCA) was performed to reduce compositional differences resulting from objective factors. Based on the significance criteria, as detailed in the volcanic and thermal maps, 148 DKD-DEGs were finally identified in this study (Figure 1). The genes associated with glucose and lipid metabolism were intersected with DKD-DEGs to obtain 20 genes for constructing the model. Box plots depicted the differential expression of related genes between patients diagnosed with DKD and non-DKD patients (Figure 2).

Machine Learning Construction of the DKD Predictive Diagnosis Model

To further narrow down the range of key immune cell-related genes, we used LASSO regression (Figure 3A1) to constrain the model by introducing penalty coefficients, and a 10-fold cross-validation was conducted. Eight candidate genes with a lamda coefficient of one standard error (lamada.1se) were selected (Figure 3A2). An RF model and an RF graph were constructed based on candidate genes (Figure 3B1). The feature significance of genes was assessed, and the top five genes were selected as candidate genes for the RF method, and a significant graph (Figure 3B2) was created. In binary classification problems, the SVM model is commonly used to convert low-dimensional linearly separable spaces into high-dimensional linearly separable spaces. We used the SVM recursive feature elimination algorithm (Figure 3C1) to assess gene significance. The 5 genes with the highest significance score were selected, 20 iterations of cross-validation were performed to obtain the corresponding errors, and graphs of cross-validation errors and cross-validation accuracy (Figure 3C2) were created. Finally, the candidate genes were intersected, and we screened out three genes (Figure 4A1). After stepwise regression, a logistic regression model was established with *HSD17B14* and *G6PC* as factors, and a column chart was drawn (Figure 4A2). Figure 4B1-3 demonstrates the corresponding receiver operating characteristic (ROC) curves, calibration curves, and clinical decision curves used to assess the discrimination, calibration, and clinical impact of the model. Figure 4C1-3 displays the good discrimination, calibration, and clinical validity of the model based on an external dataset.

Distinguishing Patient Subgroups from Immune and Stromal Cell Analysis

Unsupervised clustering was performed on patients diagnosed with DKD using the "ConsensusClusterPlus" software package, with a maximum clustering frequency of 10. Two subtypes were determined based on the consistency score and clustering results (Figure 5). Simultaneously, differential expression of genes involved in glucose and lipid metabolism was observed in the two subtypes (Figure 6A). Next, the immune cell composition of patients diagnosed with DKD was then analyzed using "CIBERSORT", and we used the "ESTIMATE" package to calculate the immune score and matrix score (Figure 6B1-3). We used the Wilcoxon test to examine the differences between the two subgroups. According to the results, there were significant differences in the composition of seven types of immune cells in the two subtypes of DKD, including macrophages M2, γδ T cells, regulatory T cells (Tregs), and other groups that can suppress immune responses. There were differences in stromal cells, immune cells, and overall scores between the two subgroups. We used secondary clustering to separate them into two new subtypes and conducted an immune infiltration analysis (Figure 7A). Based on the genes that distinguish the two subtypes (Figure 8), analyses were conducted on the expression differences between immune checkpoint genes and HLA family genes in DEGS subtypes (Figure 7B/C).

Co-Expression Analysis and Screening of Hub Genes

WGCNA analysis was conducted between two subgroups of DEGs with the aim of identifying gene modules with synergistic expression.

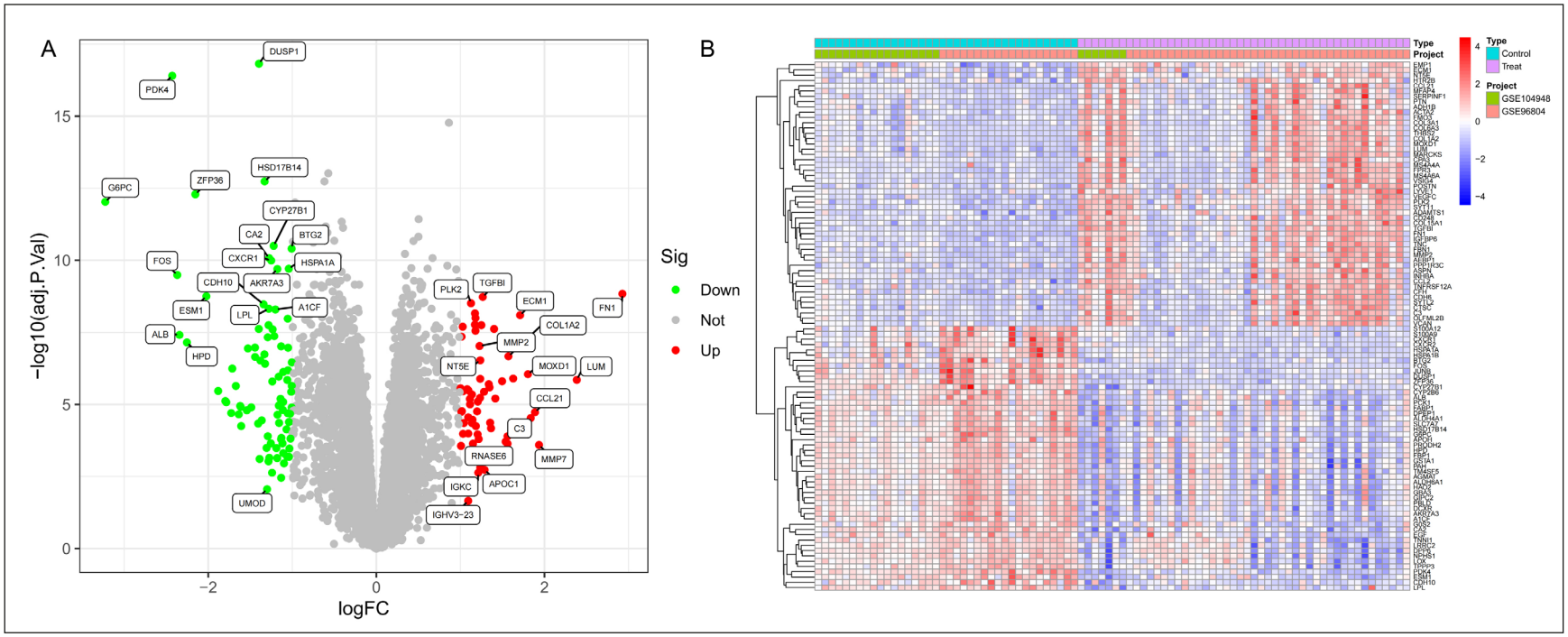

Figure 1. A, Differential genes with *p*-adjust < 0.05 and |logFC| > 1 were selected through the limma package, with red dots representing upregulated differential genes, green dots representing downregulated differential genes, and gray dots representing genes with no significant differences. **B**, DEGs heat maps of GSE104948 and GSE96804.

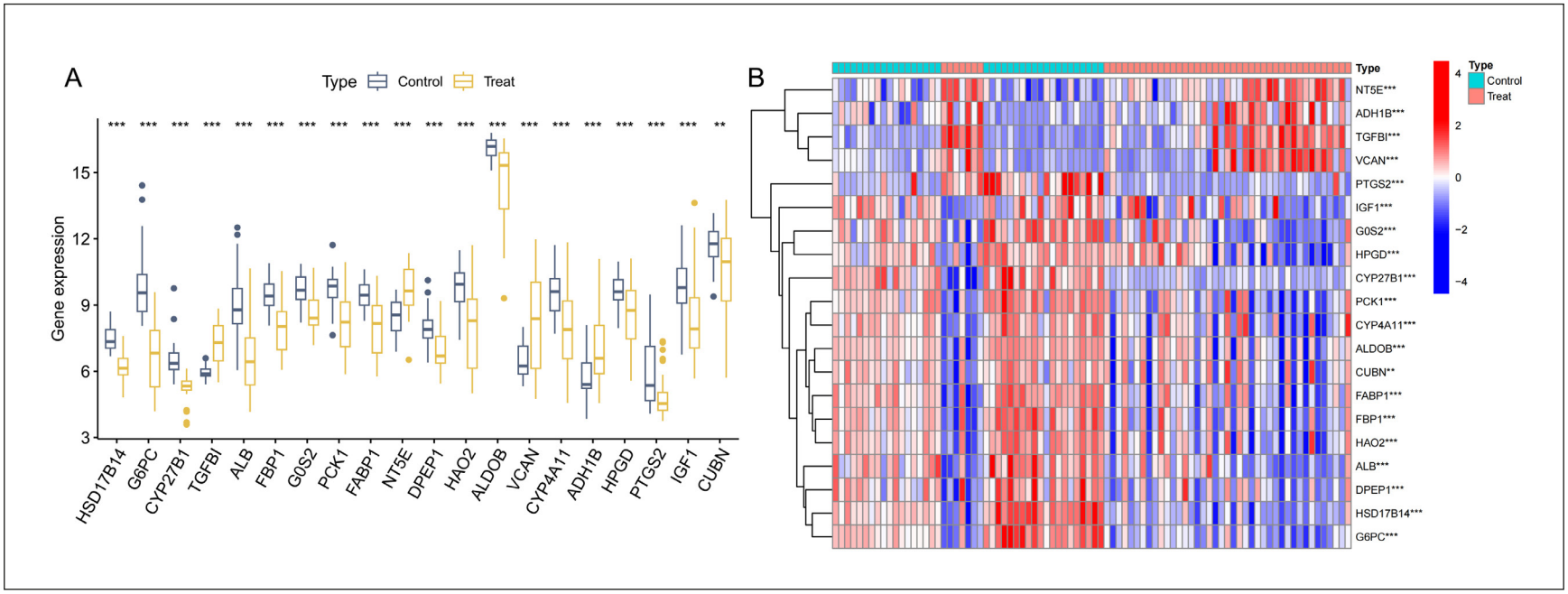

Figure 2. **A**, Differential expression heatmap of genes involved in glucose and lipid metabolism in the training set. **B**, Differential boxplot of gene expression. Significance of differences: ***represents $p < 0.001$, **represents $p < 0.01$, and *represents $p < 0.05$.

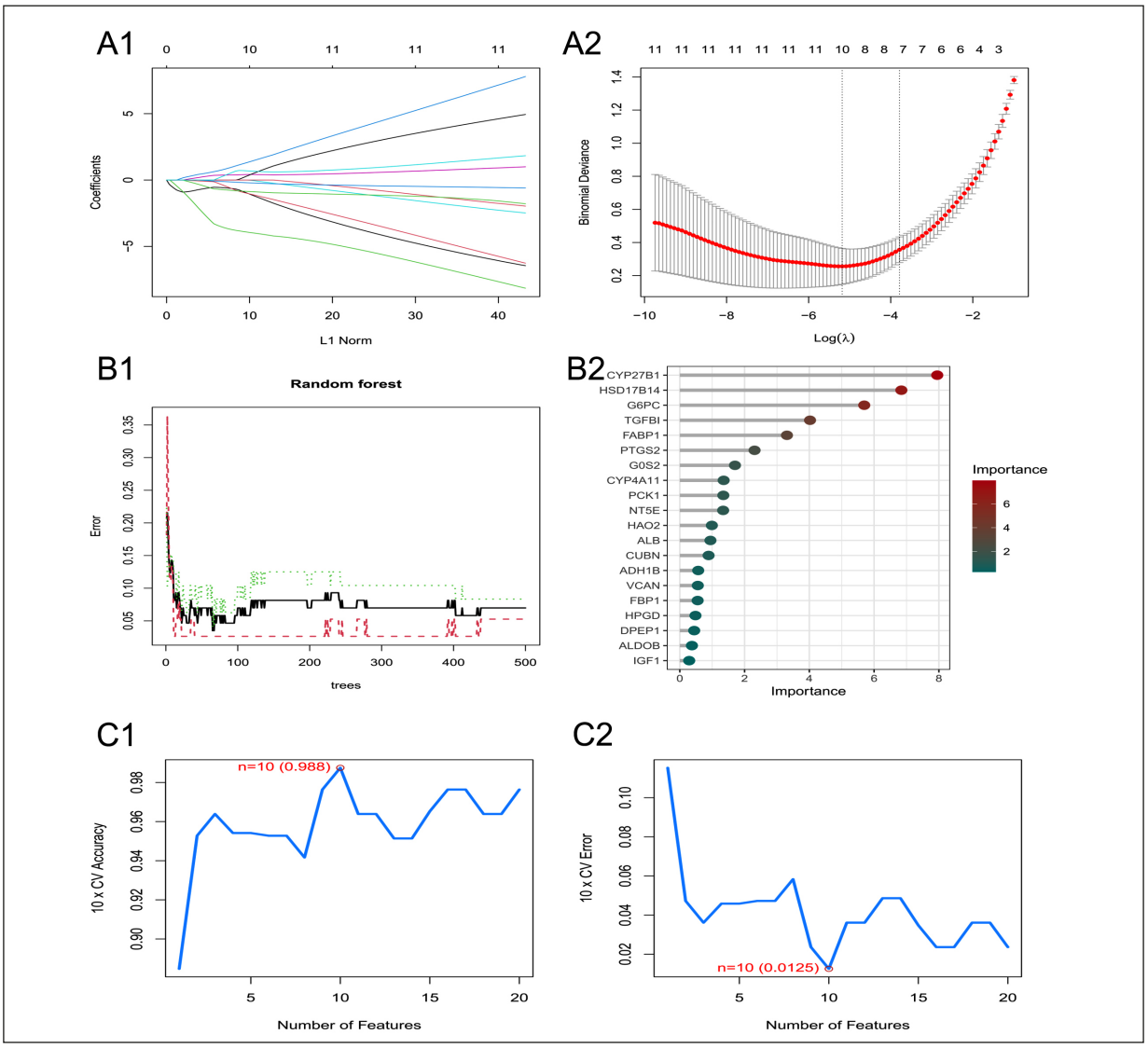

Figure 3. **A1**, LASSO regression model diagram to perform variable screening while fitting the generalized linear model. **A2**, LASSO regression 10-fold cross-validation, dashed lines are lambda.min and lambda.1se; the best λ was selected based on this graph. **B1**, The number of branches with the smallest error from the model graph was determined to establish an RF model. **B2**, A lollipop chart was created based on the significance of genes; as revealed in the figure, *CYP27B1*, *HSD17B14*, *G6PC*, *TGFBI*, and *FABP1* were the top five significant genes. **C1**, SVM machine learning accuracy curve. **C2**, SVM machine learning cross-validation error curve; the point with the smallest error and the highest accuracy was selected as the feature gene for SVM model screening.

Based on the correlation coefficient of genes, it was determined whether two genes had similar expression patterns, and a hierarchical clustering tree was established to represent different gene modules. After calculating and investigating the interaction between modules, it was determined that the expression significance of the MEblue module was relatively high (Figure 9A). Then, additional research was conducted on the MEblue key module and its subgroups, and enrichment analysis was performed on the 597 genes comprising the module (Figure 9B/C). A key module network was built with the Maximal Clique Centrality (MCC) algorithm for topology analysis in Cytoscape to identify 10 key genes (*EHHADH*, *ACOX1*, *ECH1*, *ECI2*, *ACOX2*, *CRAT*, *CAT*, *SLC27A2*, *PECR*, and *HAO2*) (Figure 10).

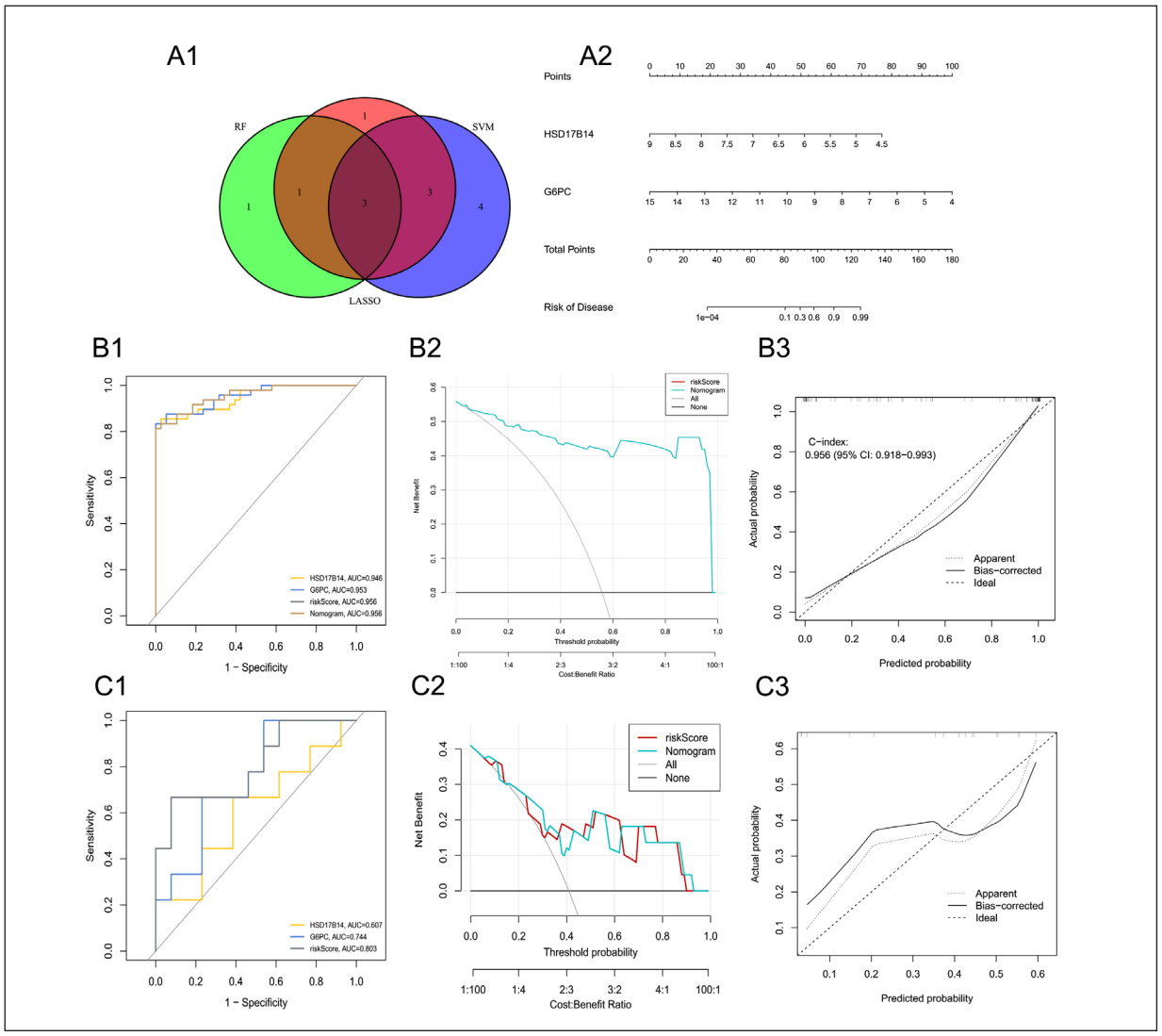

Figure 4. A1, Intersection of three machine learning screened genes, screening out three genes; (**A2**) A column chart with *HSD17B14* and *G6PC* as characteristic genes was finally established using the backward stepwise regression method. The ROC curves of the DKD column graph prediction model were compared in the training set (**B1**) and validation set (**C1**). The y-axis represents the true positive rate of risk prediction, whereas the x-axis represents the false positive rate of risk prediction. Analysis of the DKD patient decision curve using the training set (**B2**) and validation set (**C2**). **B3**, The training set and (**C3**) validation set. According to the calibration curve, the predicted probability of the model in the training and validation sets closely matches the actual probability.

Discussion

The pathogenesis of DKD is unclear, and its mechanism has not been fully elucidated. At present, there are no highly sensitive and specific biomarkers for early diagnosis of DKD. The determination of urine albumin and microalbuminuria can result in delayed diagnosis and treatment of the disease, which negatively impacts the prognosis of DKD and the quality of life of patients with DKD. Existing biomarkers like albuminuria and Glomerular Filtration Rate (GFR) no longer suffice for nuanced disease stratification and prognosis. Emerging markers like miRNAs, specifically miRNA-125b-5p and miRNA-181b-5p, have shown[12,13] promise as novel biomarkers and therapeutic targets in diseases such as Diabetic Kidney Disease and obesity, highlighting their critical role in metabolic regulation. We conducted this study to identify the potential changes in glucose metabolism and lipid metabolism in DKD and to identify reliable biomarkers. Two import-

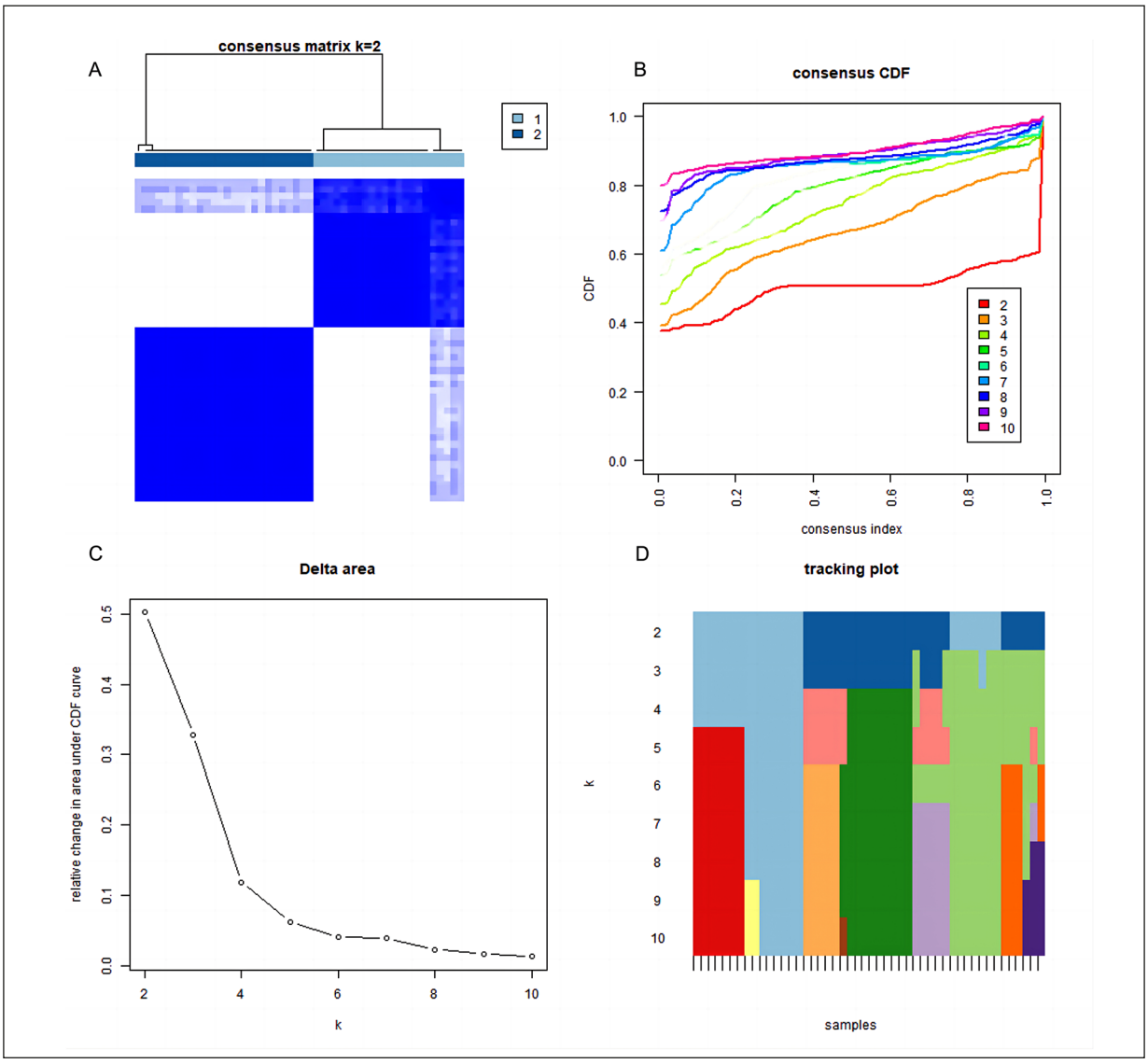

Figure 5. **A**, Consensus matrix for k=2 clusters, showing grouping consistency. **B**, Delta area plot with an elbow at k=2, indicating the optimal cluster count. **C**, CDF plot of clustering stability, plateauing at k=2. **D**, Tracking plot demonstrating sample assignment stability across varying k values.

ant genes, *HSD17B14* and *G6PC*, were identified by comprehensively analyzing the GEO database and by developing a logistic regression model using DKD-DEGs. Among them, *HSD17B14* was responsible for the metabolism of steroids and other substrates, including fatty acids, prostaglandins, and antibiotics. The gene and protein expression of 17-β dehydrogenase 14 (*HSD17B14*) of hydroxysteroids was diminished in proximal tubules of human diabetes and kidney-injured mice models despite *HSD17B14* being primarily a member of the enzyme family that regulates the relative balance of estrogen and androgen substrates, with secondary functions, such as fatty acid metabolism[14]. Studies[15] have revealed that the upregulation of *HSD17B14* expression increases intracellular estrogen levels to drive inflammation, causing an increase in ROS production and affecting steroid synthesis and metabolism. *G6PC* is one of the genes that mediates the encoding of glucose-6 phosphatase and is a key enzyme in maintaining glucose homeostasis, playing a crucial role in gluconeogenesis and glycogen breakdown. The expression of *G6PC* had

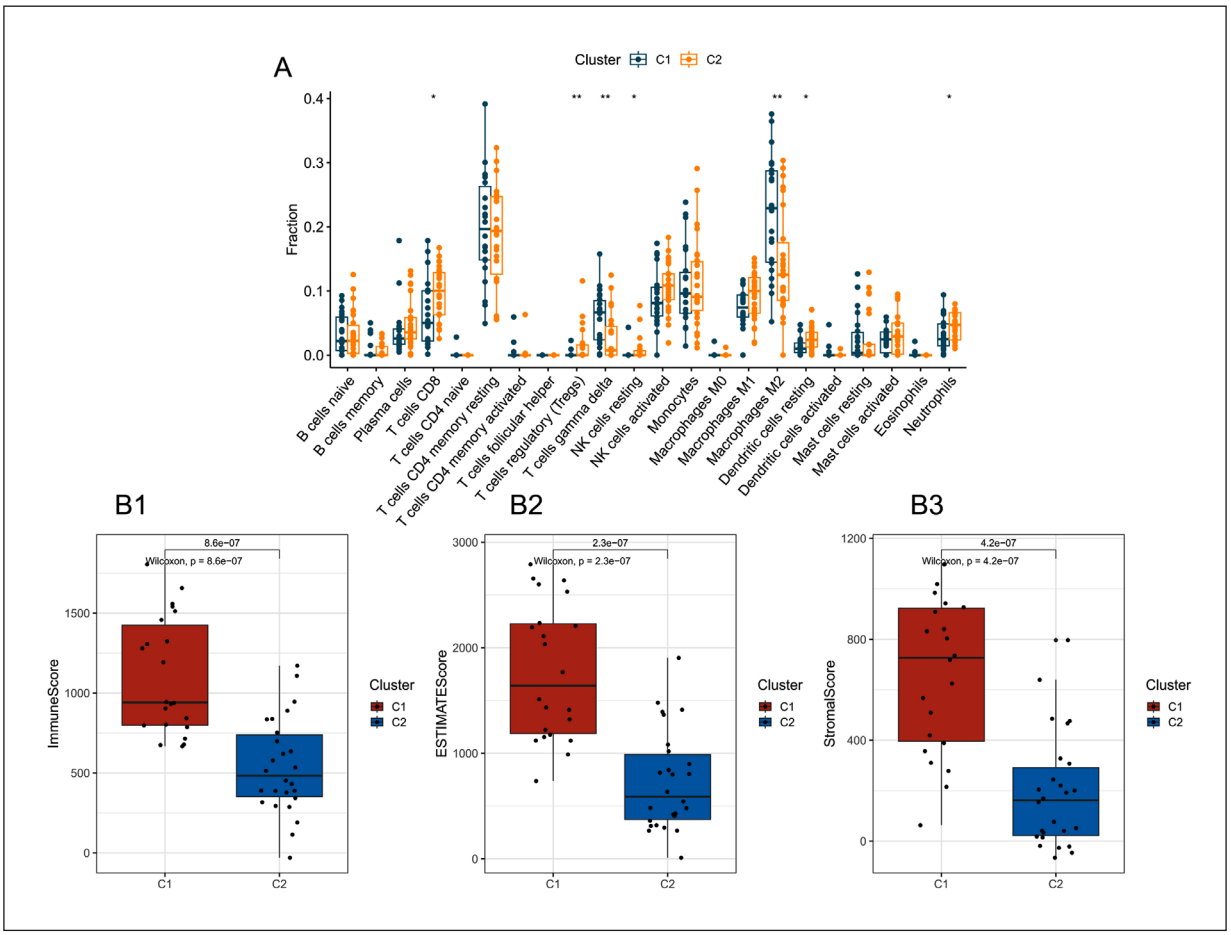

Figure 6. A, Immune cell distribution across clusters C1 and C2, highlighting significant differences in CD8 T cells, follicular helper T cells, Tregs, and M0 macrophages, reflecting immune heterogeneity. **B1**, Epidemic cell score; (**B2**) Stromal cell score; (**B3**) Comprehensive scoring.

an immediate impact on the expression level of glucose-6-phosphatase, thereby restricting glucose metabolism. Studies[16,17] suggest that down-regulation of *G6PC* expression increases glycogen storage, which leads to significant kidney enlargement and progressive glomerular hyperperfusion and ultrafiltration, followed by microalbuminuria. *G6PC* is predominantly expressed in the liver, but studies[18] have suggested its involvement in the immune infiltration of the microenvironment of clear kidney cell carcinoma. *G6PC* mutation causes glycogen storage disease type 1a, and *G6PC* overexpression affects glucose metabolism[19]. KEGG pathway analysis revealed that the differences between the two gene subgroups were closely associated with fatty acid metabolism and the degradation of branched-chain amino acids (BCAAs) such as valine, leucine, and isoleucine. Clinical trials[20] have demonstrated that serum BCAA levels gradually decrease as diabetic kidney disease (DKD) progresses in patients with Type 2 Diabetes Mellitus (T2DM).

Increasing evidence suggests that the occurrence of DKD may be associated with genetic, immune, and metabolic factors. Some evidence[21] suggests that abnormal immune response and cellular immune dysfunction in the kidneys are important factors promoting kidney function and structural enhancement. Mesenchymal stem cell (MSC) treatment in early diabetic nephropathy shows potential in preventing renal injury and restoring immune balance *via* modulation of inflammation and macrophage activity. It is necessary to investigate the pathogenesis from the perspective of the immune microenvironment and metabolic reprogramming.

Despite the fact that DKD is not typically classified as an inflammatory glomerular disease,

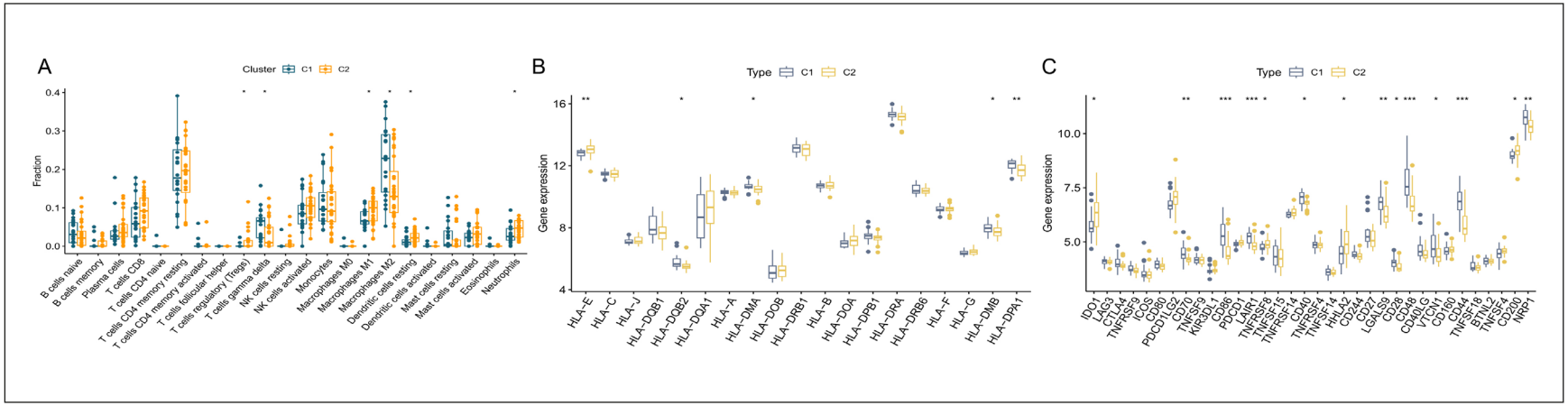

Figure 7. **A**, Differences in immune infiltrating cells between the two gene subtypes. **B**, Differential expression of HLA family genes among gene subtypes. **C**, Differential expression of immune checkpoint genes between the two gene subtypes.

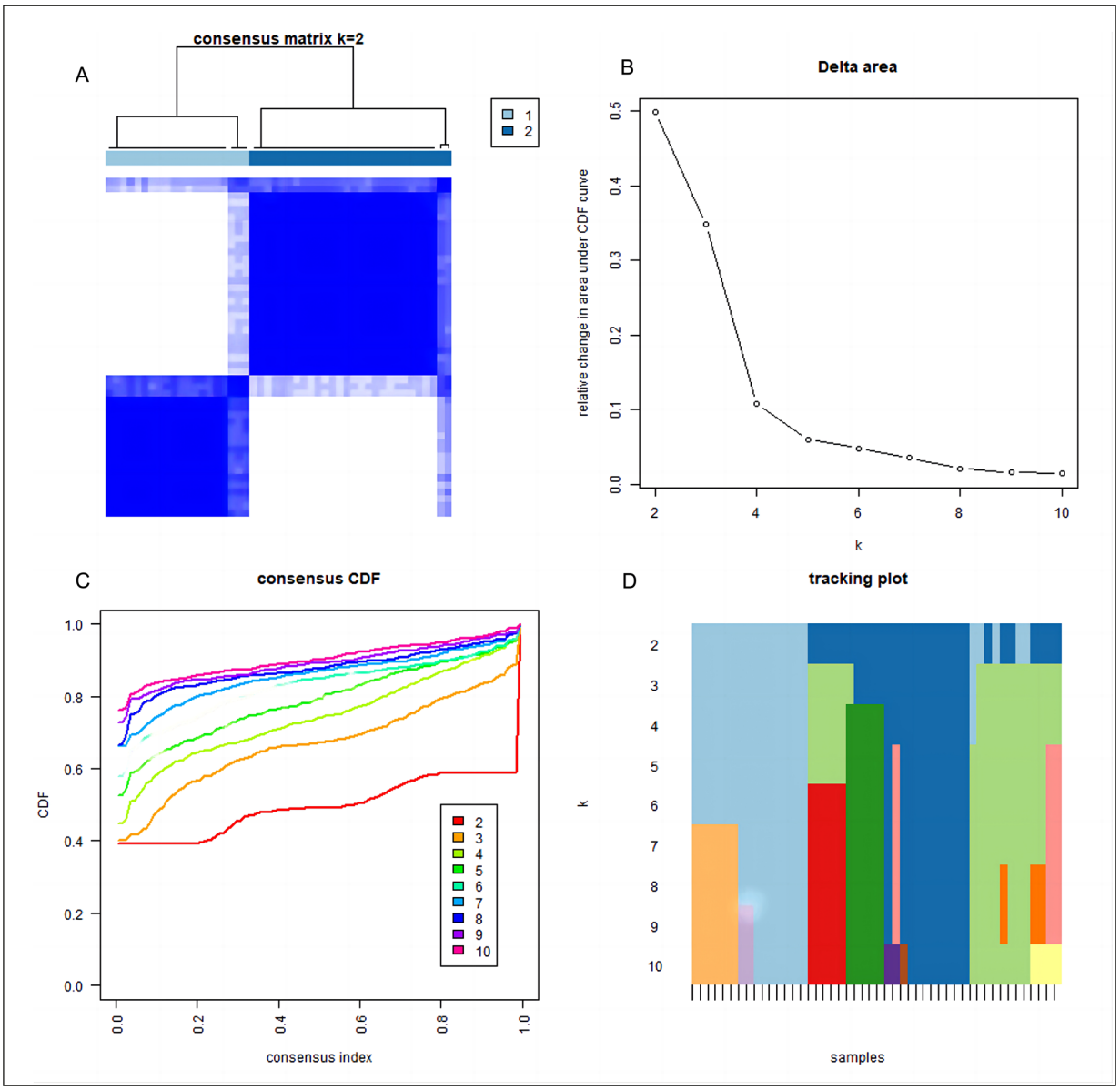

Figure 8. Gene subtype clustering post-patient subtyping: analysis indicates k=2 as the optimal gene subtype partitioning. **A**, Heatmap showcasing a strong consensus for bifurcating genetic subgroups, evidenced by the homogenous color density. **B**, The delta area plot, with a pronounced elbow at k=2 delineating the optimal bifurcation of the dataset into two discrete clusters. **C**, Array of cumulative distribution function curves for cluster counts one through ten, converging into a plateau at k=2, which substantiates the stability of the binary clustering solution. **D**, Tracking plot that delineates patient sample distributions across varying cluster counts, with consistent coloration within the bars underscoring the dependability of the dual-cluster.configuration.

increasing evidence[22] suggests that kidney inflammation is a crucial factor in the pathogenesis of DKD. Chronic, low-grade inflammation is one of the characteristic features of DKD. Chronic inflammation and chronic fibrosis are the primary causes of kidney function loss. Numerous studies[23,24] indicate that hyperglycemia significantly impairs the viability of human renal mesangial cells as well as the proliferation of pancreatic β-cells and insulin secretion, highlighting the detrimental effects of high glucose levels on renal and pancreatic function. Macrophages, which serve as antigen-presenting cells, interact with adaptive immune cells (such as T cells) during the inflammatory response, thereby shaping T cell responses and disrupting immune metabolism

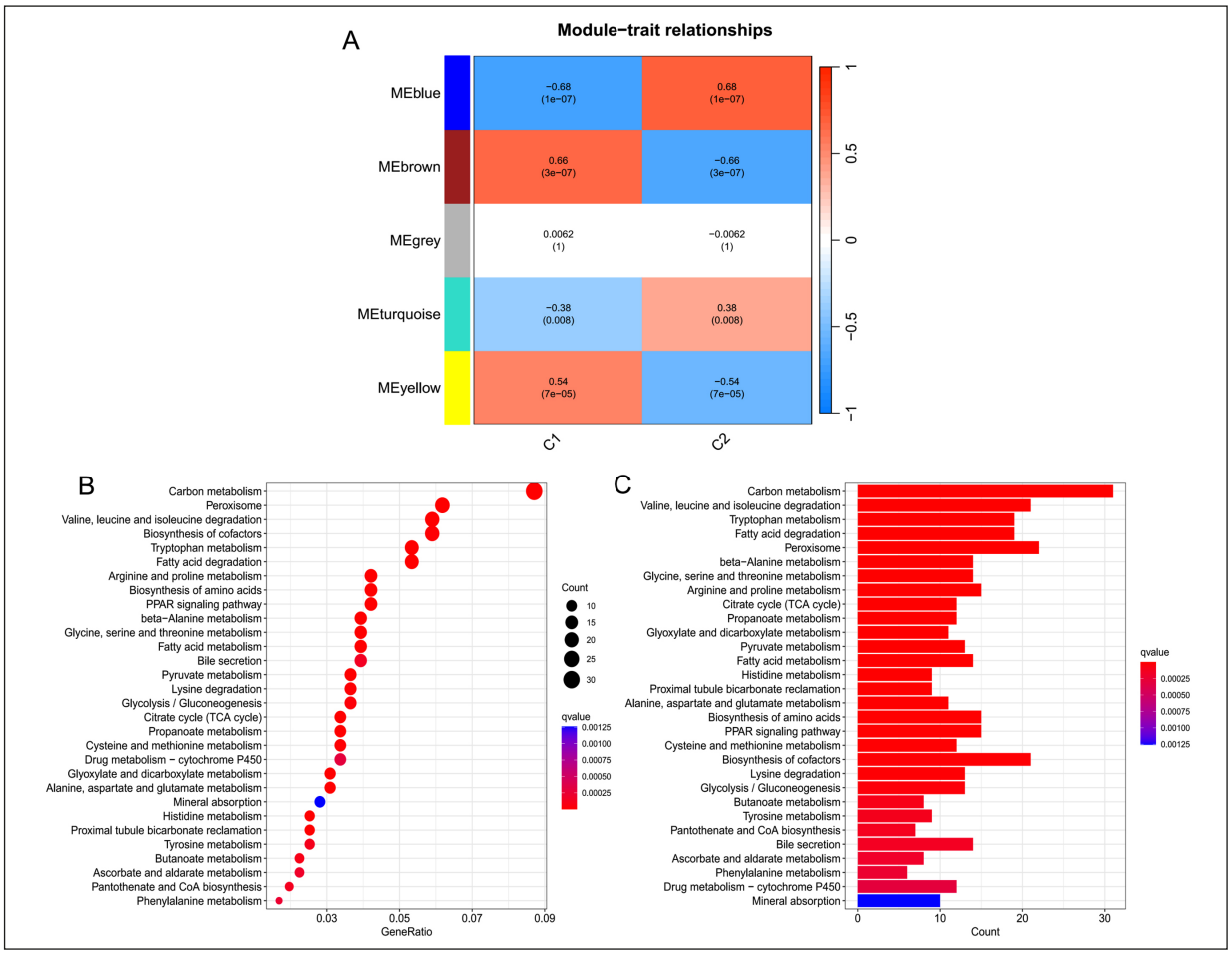

Figure 9. **A**, WGCNA analysis identified co-expressed modules; the heat map reveals that MEblue is the most significant co-expressed module. **B-C**, KEGG enrichment analysis was performed in the two gene subgroups, indicating a significant correlation between glucose metabolism (including TCA cycle) and lipid metabolism pathways in enrichment analysis.

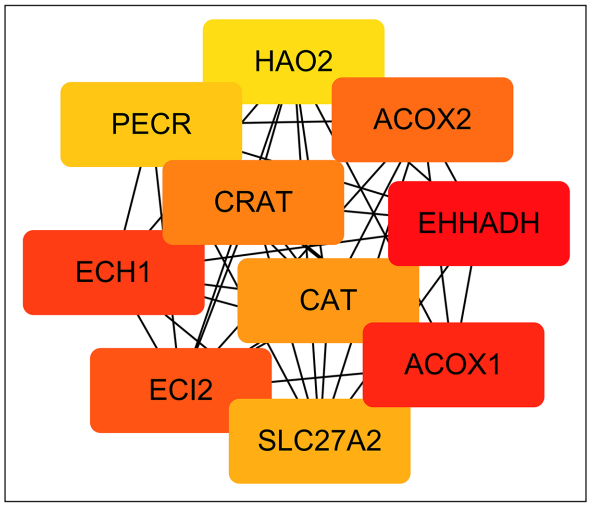

Figure 10. 10 Hub genes were identified using the cytoHub plugin.

homeostasis, leading to a pro-inflammatory environment[25]. In this study, we found that immune abnormalities can serve as an upstream mediator of inflammation and exacerbate metabolic imbalances. The condition and output of immune cells play a crucial role in the occurrence and development of DKD. The differences of immune cells between subgroups were compared, and six differential immune cells were obtained, namely, regulatory T cells (Tregs), γδT cells, M1 macrophages, M2 macrophages, resting dendritic cells, and neutrophils. High glucose-mediated metabolic disorders may interfere with macrophage and T-cell interactions[26]. In this condition, the composition of adaptive immune cells, including CD4+T, CD8+, and regulatory T cells (Tregs) changes[27]. The differential expression of regulatory T cells and γδT cells may result from the distinct protective mechanisms of gene sub-

groups against inflammatory responses. Multiple signaling pathways and local microenvironments can induce kidney macrophages to differentiate into M1 pro-inflammatory phenotype and M2 anti-inflammatory phenotypes[28]. In this study, the differential expression of M1 and M2 macrophages between the two subgroups suggests that differences in gene expression regulate macrophage polarization. The epigenetic phenomenon has been identified[29] as a risk factor for DKD, but its function and the consequences of epigenetic alterations are still unknown. Studies[30] have demonstrated that metabolic reprogramming is involved in innate and adaptive immune responses to regulate immune cell function, thereby establishing a connection between immune response and metabolic reprogramming.

Cellular fate decisions made by immune cells, including activation, proliferation, differentiation, and polarization, are influenced by alterations in cellular metabolism, according to cellular immunometabolism[31]. We discovered that immune cell subpopulations in disease states exhibit different metabolic pathways to promote cell survival and the formation of cell lineages. Due to the formation of numerous branches by intermediates in biochemical reactions, the presence of numerous enzyme subtypes, the reversibility of metabolic processes, the presence of numerous complement inputs into metabolic cycles, and the presence of numerous material sources, a complex metabolic network is formed[32-34]. The integration and reconstruction of biochemical reactions in material metabolism give the microenvironment of the body a great deal of vitality, which is also the physiological basis for the complex immune microenvironment of DKD. As the second-highest energy-consuming organ after the heart, the kidney must maintain homeostasis in energy metabolism, as abnormal energy metabolism can result in cell dysfunction and even death. In a study[35] involving 522 diabetic patients with concomitant hypertension, the importance of stable lipid levels in preventing poor glycemic control was particularly emphasized. Genome-wide association studies (GWAS) can be employed to identify genetic links associated with the accumulation of adipose tissue[36].

Studies[37] have revealed that pathological metabolic disturbances and gene reprogramming are frequently crucial to the occurrence and development of diseases, affecting not only the molecular pathways of kidney cells but also the immune system. Thus, we constructed a time-space line comprising of chronic low-grade inflammation, an abnormal immune microenvironment, and metabolic reprogramming of DKD progression.

Firstly, in this study, two diagnostic biomarkers of DKD were identified. Then, a diagnostic model was constructed, and the roles and mechanisms of related metabolic abnormal genes and immune microenvironments in DKD were investigated. There are limitations to this study, including its small sample size, and both the training set and the validation set should be expanded to increase the size of the data to achieve universality with a large sample size. Although the diagnostic prediction model performed well in this study, the relevant results were not further experimentally validated. Additionally, the study's reliance on retrospective data may introduce biases that could affect the findings. Prospective studies are needed to confirm these results and understand the model's real-world applicability and performance.

Conclusions

We screened two key genes and developed a highly accurate predictive model, providing novel references for the diagnosis, mechanism research, and treatment of this disease.

Conflict of Interest

The authors declare that they have no conflict of interests.

Acknowledgements

We are grateful to everyone who assisted us with our article.

Funding

This study was funded by the National Natural Science Foundation of China (No. 82174153) and Anhui University Collaborative Innovation Project (GXXT-2020-025).

Ethics Approval and Informed Consent

Ethical approval and consent were not required as this study was based on publicly available data.

Availability of Data and Materials

The relevant supporting data are available from the author upon request.

ORCID ID
Zheng Bi: 0009-0001-6642-1380
Lu-Jie Wang: 0009-0008-8887-1033
Yi-Xuan Lin: 0009-0004-7837-9927
Yin-Yu Zhang: 0009-0005-2914-0511
Si-Hai Wang: 0009-0001-3994-8988
Zhao-Hui Fang: 0000-0002-2174-0580

Authors' Contribution
Conception and design of the research:Zheng Bi, Zhao-Hui Fang. Acquisition of data: Lu-Jie Wang, Yi-Xuan Lin, Yin-Yu Zhang, Si-Hai Wang. Analysis and interpretation of the data: Zheng Bi, Lu-Jie Wang, Yi-Xuan Lin. Statistical analysis: Yin-Yu Zhang, Si-Hai Wang. Obtaining financing: Zhao-Hui Fang. Writing of the manuscript: Zheng Bi. Critical revision of the manuscript for intellectual content: Zhao-Hui Fang. All authors read and approved the final draft.

References

1) Thomas MC, Brownlee M, Susztak K, Sharma K, Jandeleit-Dahm KA, Zoungas S, Rossing P, Groop PH, Cooper ME. Diabetic kidney disease. Nat Rev Dis Primers 2015; 1: 15018.
2) Tuttle KR, Bakris GL, Bilous RW, Chiang JL, de Boer IH, Goldstein-Fuchs J, Hirsch IB, Kalantar-Zadeh K, Narva AS, Navaneethan SD, Neumiller JJ, Patel UD, Ratner RE, Whaley-Connell AT, Molitch ME. Diabetic kidney disease: a report from an ADA Consensus Conference. Diabetes Care 2014; 37: 2864-2883.
3) Fu H, Liu S, Bastacky SI, Wang X, Tian XJ, Zhou D. Diabetic kidney diseases revisited: A new perspective for a new era. Mol Metab 2019; 30: 250-263.
4) Reidy K, Kang HM, Hostetter T, Susztak K. Molecular mechanisms of diabetic kidney disease. J Clin Invest 2014; 124: 2333-2340.
5) Rayego-Mateos S, Rodrigues-Diez RR, Fernandez-Fernandez B, Mora-Fernández C, Marchant V, Donate-Correa J, Navarro-González JF, Ortiz A, Ruiz-Ortega M. Targeting inflammation to treat diabetic kidney disease: the road to 2030. Kidney Int 2023; 103: 282-296.
6) Macisaac RJ, Ekinci EI, Jerums G. Markers of and risk factors for the development and progression of diabetic kidney disease. Am J Kidney Dis 2014; 63: S39-S62.
7) Thongnak L, Pongchaidecha A, Lungkaphin A. Renal Lipid Metabolism and Lipotoxicity in Diabetes. Am J Med Sci 2020; 359: 84-99.
8) Xu T, Xu X, Zhang L, Zhang K, Wei Q, Zhu L, Yu Y, Xiao L, Lin L, Qian W, Wang J, Ke M, An X, Liu S. Lipidomics Reveals Serum Specific Lipid Alterations in Diabetic Nephropathy. Front Endocrinol (Lausanne) 2021; 12: 781417.
9) Wang Z, Fu W, Huo M, He B, Liu Y, Tian L, Li W, Zhou Z, Wang B, Xia J, Chen Y, Wei J, Abliz Z. Spatial-resolved metabolomics reveals tissue-specific metabolic reprogramming in diabetic nephropathy by using mass spectrometry imaging. Acta Pharm Sin B 2021; 11: 3665-3677.
10) Tesch GH. Diabetic nephropathy - is this an immune disorder? Clin Sci (Lond) 2017; 131: 2183-2199.
11) Mora-Fernández C, Domínguez-Pimentel V, de Fuentes MM, Górriz JL, Martínez-Castelao A, Navarro-González JF. Diabetic kidney disease: from physiology to therapeutics. J Physiol 2014; 592: 3997-4012.
12) Vettori A, Pompucci G, Paolini B, Del Ciondolo I, Bressan S, Dundar M, Kenanoğlu S, Unfer V, Bertelli M; Geneob Project. Genetic background, nutrition and obesity: a review. Eur Rev Med Pharmacol Sci 2019; 23: 1751-1761.
13) Ishii H, Kaneko S, Yanai K, Aomatsu A, Hirai K, Ookawara S, Morishita Y. MicroRNA Expression Profiling in Diabetic Kidney Disease. Transl Res 2021; 237: 31-52.
14) Mychaleckyj JC, Valo E, Ichimura T, Ahluwalia TS, Dina C, Miller RG, Shabalin IG, Gyorgy B, Cao J, Onengut-Gumuscu S, Satake E, Smiles AM, Haukka JK, Tregouet DA, Costacou T, O'Neil K, Paterson AD, Forsblom C, Keenan HA, Pezzolesi MG, Pragnell M, Galecki A, Rich SS, Sandholm N, Klein R, Klein BE, Susztak K, Orchard TJ, Korstanje R, King GL, Hadjadj S, Rossing P, Bonventre JV, Groop PH, Warram JH, Krolewski AS. Association of Coding Variants in Hydroxysteroid 17-beta Dehydrogenase 14 (HSD17B14) with Reduced Progression to End Stage Kidney Disease in Type 1 Diabetes. J Am Soc Nephrol 2021; 32: 2634-2651.
15) Qureshi R, Picon-Ruiz M, Aurrekoetxea-Rodriguez I, Nunes de Paiva V, D'Amico M, Yoon H, Radhakrishnan R, Morata-Tarifa C, Ince T, Lippman ME, Thaller SR, Rodgers SE, Kesmodel S, Vivanco MDM, Slingerland JM. The Major Pre- and Postmenopausal Estrogens Play Opposing Roles in Obesity-Driven Mammary Inflammation and Breast Cancer Development. Cell Metab 2020; 31: 1154-1172.e9.
16) Liu Q, Li J, Zhang W, Xiao C, Zhang S, Nian C, Li J, Su D, Chen L, Zhao Q, Shao H, Zhao H, Chen Q, Li Y, Geng J, Hong L, Lin S, Wu Q, Deng X, Ke R, Ding J, Johnson RL, Liu X, Chen L, Zhou D. Glycogen accumulation and phase separation drives liver tumor initiation. Cell 2021; 184: 5559-5576.e19.
17) Clar J, Gri B, Calderaro J, Birling MC, Hérault Y, Smit GP, Mithieux G, Rajas F. Targeted deletion of kidney glucose-6 phosphatase leads to nephropathy. Kidney Int 2014; 86: 747-756.
18) Xu WH, Xu Y, Tian X, Anwaier A, Liu WR, Wang J, Zhu WK, Cao DL, Wang HK, Shi GH, Qu YY, Zhang HL, Ye DW. Large-scale transcriptome profiles reveal robust 20-signatures metabolic prediction models and novel role of G6PC in clear cell renal cell carcinoma. J Cell Mol Med 2020; 24: 9012-9027.

19) Hutton JC, O'Brien RM. Glucose-6-phosphatase catalytic subunit gene family. J Biol Chem 2009; 284: 29241-29245.

20) Liu M, Yang Y, Liu Y, Peng X, Hou Y, Zhang X, Sun H, Shan C. Serum branched chain amino acids: an effective indicator of diabetic kidney disease. Front Endocrinol (Lausanne) 2023; 14: 1269633.

21) Li Y, Liu J, Liao G, Zhang J, Chen Y, Li L, Li L, Liu F, Chen B, Guo G, Wang C, Yang L, Cheng J, Lu Y. Early intervention with mesenchymal stem cells prevents nephropathy in diabetic rats by ameliorating the inflammatory microenvironment. Int J Mol Med 2018; 41: 2629-2639.

22) Zheng W, Guo J, Liu ZS. Effects of metabolic memory on inflammation and fibrosis associated with diabetic kidney disease: an epigenetic perspective. Clin Epigenetics 2021; 13: 87.

23) Zhu Y, Ruan CX, Wang J, Jiang FF, Xiong LS, Sheng X, Le J, Yu AQ, Wang Q, Liu YT, Qin SL. High glucose inhibits the survival of HRMCs and its mechanism. Eur Rev Med Pharmacol Sci 2022; 26: 5683-5688.

24) Zhang YN, Fu DX, Xu JX, Wang GY. The effect of SOX9 on islet β cells in high glucose environment through regulation of ERK/P38 signaling pathway. Eur Rev Med Pharmacol Sci 2019; 23: 8476-8484.

25) Schmidt V, Hogan AE, Fallon PG, Schwartz C. Obesity-Mediated Immune Modulation: One Step Forward, (Th)2 Steps Back. Front Immunol 2022; 13: 932893.

26) Peng HY, Lucavs J, Ballard D, Das JK, Kumar A, Wang L, Ren Y, Xiong X, Song J. Metabolic Reprogramming and Reactive Oxygen Species in T Cell Immunity. Front Immunol 2021; 12: 652687.

27) Savva C, Copson E, Johnson PWM, Cutress RI, Beers SA. Obesity Is Associated with Immunometabolic Changes in Adipose Tissue That May Drive Treatment Resistance in Breast Cancer: Immune-Metabolic Reprogramming and Novel Therapeutic Strategies. Cancers (Basel) 2023; 15: 2440.

28) Moratal C, Laurain A, Naïmi M, Florin T, Esnault V, Neels JG, Chevalier N, Chinetti G, Favre G. Regulation of Monocytes/Macrophages by the Renin-Angiotensin System in Diabetic Nephropathy: State of the Art and Results of a Pilot Study. Int J Mol Sci 2021; 22: 6009.

29) McKnight AJ, McKay GJ, Maxwell AP. Genetic and epigenetic risk factors for diabetic kidney disease. Adv Chronic Kidney Dis 2014; 21: 287-296.

30) Sun L, Yang X, Yuan Z, Wang H. Metabolic Reprogramming in Immune Response and Tissue Inflammation. Arterioscler Thromb Vasc Biol 2020; 40: 1990-2001.

31) Man K, Kutyavin VI, Chawla A. Tissue immunometabolism: development, physiology, and pathobiology. Cell Metab 2017; 25: 11-26.

32) Dey P, Kimmelman AC, DePinho RA. Metabolic Codependencies in the Tumor Microenvironment. Cancer Discov 2021; 11: 1067-1081.

33) Zhang YW, Nie F, Zheng XY, Zhao SJ. L-tyrosine metabolic pathway in microorganisms and its application in the biosynthesis of plant-derived natural products. World J Tradit Chin Med 2022; 8: 386-394.

34) Liu J, Zhang CC, Zhang SQ, Wang JH, Xu RR, Yang SL, Wang T, Liu QF, Wang HX, Tang XD. Clinical factors affecting platelet growth in the treatment of aplastic anemia by tonifying kidney and generating blood. World J Tradit Chin Med 2023; 9: 438-446.

35) Jarab AS, Al-Qerem W, Alqudah S, Abu Heshmeh SR, Mukattash TL, Beiram R, Aburuz S. Glycemic control and its associated factors in hypertensive patients with type 2 diabetes. Eur Rev Med Pharmacol Sci 2023; 27: 5775-5783.

36) Camilleri G, Kiani AK, Herbst KL, Kaftalli J, Bernini A, Dhuli K, Manara E, Bonetti G, Stuppia L, Paolacci S, Dautaj A, Bertelli M. Genetics of fat deposition. Eur Rev Med Pharmacol Sci 2021; 25: 14-22.

37) Kato M, Natarajan R. Epigenetics and epigenomics in diabetic kidney disease and metabolic memory. Nat Rev Nephrol 2019; 15: 327-345.