Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrated Bioanalysis and In Vivo Validation of Mitochondrial Energy Metabolism–Related Genes in Diabetic Nephropathy

Jinxiu Deng
Longyan First Affiliated Hospital of Fujian Medical University
Min Liao
Pudong New Area, Shanghai
Shuilian Zhang
Fujian Tingzhou Hospital
Xiuping Qiu
Longyan First Affiliated Hospital of Fujian Medical University
Meng Wu
doctorwum@yeah.net

Longyan First Affiliated Hospital of Fujian Medical University

Research Article

Keywords: Diabetic nephropathy, Mitochondrial energy metabolism, Diagnostic model, Immune infiltration, LASSO regression

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11232296/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Integrated Bioanalysis and In Vivo Validation of Mitochondrial Energy Metabolism–Related Genes in Diabetic Nephropathy

Jinxiu Deng [1#], Min Liao[2#], Shuilian Zhang [3], Xiuping Qiu[4*], Meng Wu [1*]

[1] Department of Nephrology, Longyan First Affiliated Hospital of Fujian Medical University, Longyan 364000, China.

[2] Sanlin Community Health Service Center, Pudong New Area, Shanghai，200129, China.

[3] Department of Nephrology, Fujian Tingzhou Hospital, Longyan, 366300, China.

[4]Department of Endocrinology,Longyan First Affiliated Hospital of Fujian Medical University，Longyan 364000, China.

[#]These authors contributed equally.

*Corresponding Author: Longyan First Affiliated Hospital of Fujian Medical University, No.105 Jiuyi North Road, Longyan, Fujian, China.

Xiuping Qiu;Tel: 86-15806017513; E-mail: qxp15806017513@163.com.

Meng Wu;Tel: 86-15160652288; E-mail: doctorwum@yeah.net.

Background: Diabetic nephropathy (DN) is a leading cause of end-stage renal disease, and early diagnostic biomarkers remain limited. Given the critical role of mitochondrial energy metabolism in DN pathogenesis, this study aimed to identify mitochondrial energy metabolism–related diagnostic signatures and associated immune features.

Methods: Three GEO microarray datasets (GSE30528, GSE96804, and GSE104948) were integrated after batch-effect correction. Differential expression analysis identified mitochondrial energy metabolism–related differentially expressed genes (MEMRDEGs). A diagnostic signature was developed using logistic regression for screening, SVM-RFE for feature refinement, and LASSO regression for final model construction, generating a seven-gene risk score model (PDK4, CKB, UCHL1, ADH1B, ALDH1L1, CR1, and PC). Diagnostic performance was evaluated using ROC analysis. Functional enrichment analyses (GO/KEGG) and immune infiltration estimation (ssGSEA) were performed to explore underlying biological pathways and immune characteristics. Regulatory networks (mRNA–TF and mRNA–miRNA) were further constructed for key genes, and immunohistochemistry was used to validate protein-level expression changes in DN tissues.

Results:A total of 20 MEMRDEGs were identified. LASSO regression yielded a diagnostic model comprising seven genes, which demonstrated excellent discriminative ability (AUC > 0.9). Functional enrichment suggested that DN was characterized by activation of inflammatory, TGF-β, and PI3K–AKT–mTOR signaling pathways. Immune infiltration analyses revealed prominent immune microenvironment remodeling, with notable involvement of CD8 T cells and regulatory T cells (Tregs). Immunohistochemistry confirmed the protein-level expression alterations of the seven signature genes in DN kidney tissues.

Conclusions: This study establishes a robust mitochondrial energy metabolism–based seven-gene diagnostic signature for DN and highlights metabolic–immune crosstalk associated with disease progression. The identified genes and immune-associated pathways may offer promising molecular targets for early diagnosis and therapeutic intervention in DN.

Keywords: Diabetic nephropathy, Mitochondrial energy metabolism, Diagnostic model, Immune infiltration, LASSO regression

Introduction

Diabetic nephropathy (DN) is a leading cause of end-stage renal disease (ESRD) worldwide, representing a major public health burden with profound implications for patient survival and healthcare systems(1). Despite advances in understanding its pathophysiology, clinical management remains hampered by the limitations of conventional diagnostic tools. Specifically, urinary albumin excretion rate and estimated glomerular filtration rate (eGFR) exhibit inadequate sensitivity and specificity for early detection and reliable prognosis prediction, often failing to identify renal injury until irreversible damage has occurred (2, 3). Furthermore, although sodium-glucose co-transporter 2 inhibitors and mineralocorticoid receptor antagonists have demonstrated renoprotective effects, a substantial proportion of patients continue to progress toward kidney failure, underscoring persistent therapeutic gaps(4, 5). The kidney's high energy demand renders it critically dependent on mitochondrial oxidative phosphorylation for normal function, and disruption of this metabolic equilibrium is increasingly implicated in the pathogenesis of chronic kidney disease, including DN (6, 7). Under sustained hyperglycemia, renal cells experience mitochondrial dysfunction characterized by impaired fatty acid β-oxidation (FAO), reduced ATP synthesis, and excessive reactive oxygen species (ROS) production(8-11). This bioenergetic deficit drives maladaptive metabolic reprogramming—particularly a shift from oxidative phosphorylation to glycolysis in podocytes—which promotes lipotoxicity, oxidative stress, inflammation, and fibrosis(12, 13). While these mechanisms are well-documented, their translation into clinically actionable biomarkers remains limited. Notably, systematic profiling of mitochondrial energy metabolism-related genes (MEMRGs) for diagnostic modeling and molecular subtyping in DN has been scarcely explored, creating a critical knowledge gap(14). To address this, we leveraged multi-cohort transcriptomic data from the Gene Expression Omnibus (GEO) and applied an integrative bioinformatics pipeline incorporating differential expression analysis, functional enrichment, and Least Absolute Shrinkage and Selection Operator (LASSO) regression to identify a robust MEMRG-based diagnostic signature. We further characterized the associated immune microenvironment and constructed regulatory networks involving transcription factors and microRNAs, with core findings validated at the protein level via immunohistochemistry in human renal tissues. This study thus aims to bridge mitochondrial pathobiology with precision diagnostics, offering a novel framework for early detection, risk stratification, and targeted therapeutic development in DN.

1 Materials and Methods:

1.1 Data Download

Three GEO microarray datasets were included: GSE30528, GSE96804, and GSE104948 (human kidney; Table1). The corresponding platforms were GPL571 (GSE30528), GPL17586 (GSE96804), and GPL22945/GPL24120 (GSE104948). The sample sizes were DN: 9, 41, 12 and controls: 13, 20, 21, yielding DN n = 62 and controls n = 54 after integration. the specific information is shown in Table 1.

Batch effects across datasets were corrected using the R package sva. Probe annotation and expression preprocessing were performed with limma, followed by principal component analysis (PCA) to confirm the effectiveness of batch removal.

1.2 Construction of mitochondrial energy metabolism gene sets

A mitochondrial energy metabolism gene set (MEMRGs) was compiled from GeneCards and PubMed. From GeneCards, genes matching "Mitochondrial Energy Metabolism" were retained as protein-

coding with relevance score > 0 (n = 220). From PubMed, additional mitochondria energy metabolism–related genes were identified (n = 188). After deduplication, 381 MEMRGs were obtained (Table S1).

1.3 Differential expression and identification of MEMRDEGs

DN and control samples were compared using limma. Differentially expressed genes (DEGs) were defined as:|log2FC| > 0.5 and adjusted P < 0.05 (Benjamini–Hochberg correction).Up and downregulated genes were classified based on the sign of log2FC. Mitochondrial energy metabolism–related DEGs (MEMRDEGs) were defined as the intersection between DEGs and MEMRGs.

1.4 Correlation analysis

Pairwise correlations among MEMRDEGs were evaluated using **Spearman's correlation**, and correlation heatmaps/scatter plots were generated in R. The correlation strength was interpreted as weak (<0.5), moderate (0.5–0.8), and strong (>0.8), consistent with the original analysis.

1.5 Functional enrichment analysis (GO and KEGG)

GO and KEGG enrichment analyses of MEMRDEGs were performed using clusterProfiler. Significance was set at：(1)P<0.05 and FDR (q value) < 0.25;(2)Multiple testing adjustment: Benjamini–Hochberg.Results were visualized with bubble plots and network-style diagrams.

1.6 Gene set enrichment analysis (GSEA)

GSEA was conducted using clusterProfiler on the full ranked gene list from the integrated dataset. Parameters were:seed = 2022;permutations = 1000;gene set size: min = 10, max = 500;gene sets: MSigDB C2 (All Canonical Pathways v2022.1; GMT).Screening thresholds were:adj. P < 0.05 and FDR q < 0.25;Benjamini–Hochberg for P adjustment.For risk stratification, DN samples were divided into high-risk and low-risk groups by the median LASSO RiskScore, followed by GSEA between the two groups.

1.7 Construction of the diagnostic signature and RiskScore

A three-step feature selection strategy was applied to derive a diagnostic model for DN：(1)Logistic regression on MEMRDEGs (features filtered by P < 0.05)；(2) SVM-RFE to select features with the best performance (lowest error / highest accuracy);(3)LASSO regression (binomial family, seed = 500) to obtain the final signature

The final key genes were selected by non-zero LASSO coefficients and used to compute the LASSO RiskScore:

$$RiskScore = \sum_{i} Coefficient\ (gene_i) * mRNA\ Expression\ (gene_i)$$

DN samples were classified into high- and low-risk groups based on the **median RiskScore**.

1.8 Model evaluation and visualization

ROC curves and AUC were calculated using pROC to assess diagnostic discrimination.Friends analysis (functional similarity) among key genes was performed using GOSemSim.A nomogram was constructed with rms.Calibration curves were generated using calibration analysis to evaluate prediction agreement.Decision curve analysis (DCA) was performed using ggDCA to estimate clinical utility.

1.9 Immune Infiltration Analysis (ssGSEA)

Immune cell infiltration was estimated using single-sample GSEA (ssGSEA). Enrichment scores

representing the relative abundance of 28 immune cell types were calculated for each sample. Differential immune infiltration between:control vs DN, and low-risk vs high-risk (DN) was evaluated using group comparison plots.

Correlation analyses between immune cell infiltration and key genes were performed using Spearman's correlation, with visualization through heatmaps and bubble plots.

1.10 Regulatory network construction

mRNA–TF network: transcription factors regulating key genes were retrieved from ChIPBase. Interactions were filtered by the number of upstream/downstream samples (> 9) and visualized using Cytoscape. The network included 6 key genes and 42 TFs.

mRNA–miRNA network: miRNA–mRNA interactions related to key genes were retrieved from starBase v3.0 and filtered to retain interactions supported by ≥ 3 sources. Visualization was performed in Cytoscape. The network included 4 key genes and 35 miRNAs.

1.11 Immunohistochemical staining

Formalin-fixed paraffin-embedded kidney tissues were used for IHC staining. Samples were provided by Longyan First Affiliated Hospital of Fujian Medical University. The study was approved by the ethics committee (Approval No: LYREC2024-k025-01), and informed consent was obtained.

Staining was performed on 5 μm sections. Primary antibodies were applied overnight at 4°C, followed by secondary antibody incubation and DAB visualization. Protein levels were quantified using ImageJ (IOD/area-based quantification). The assessed key genes were Rabbit Anti-CR1(Proteintech ,1/200)，Mouse Anti-ADH1B (Proteintech ,1/200)，Rabbit Anti-PDK4 (Proteintech，1/200)，Rabbit Anti- CKB(Proteintech,1/200) ， Rabbit Anti-UCHL1(Proteintech,1/200) ， Rabbit Anti-ALDH1L1(Proteintech,1/200)， Rabbit Anti-PC(Proteintech,1/200), using the antibody information and dilutions listed in the original manuscript.

1.12 Statistical analysis

All analyses were conducted in R (version 4.3.0). For two-group comparisons, Student's t-test was used for normally distributed variables; Mann–Whitney U test was used otherwise. For three or more groups, Kruskal–Wallis test was applied. Spearman correlation was used for association testing. Two-sided $P < 0.05$ was considered significant.

2 Results

2.1 Technology Road map

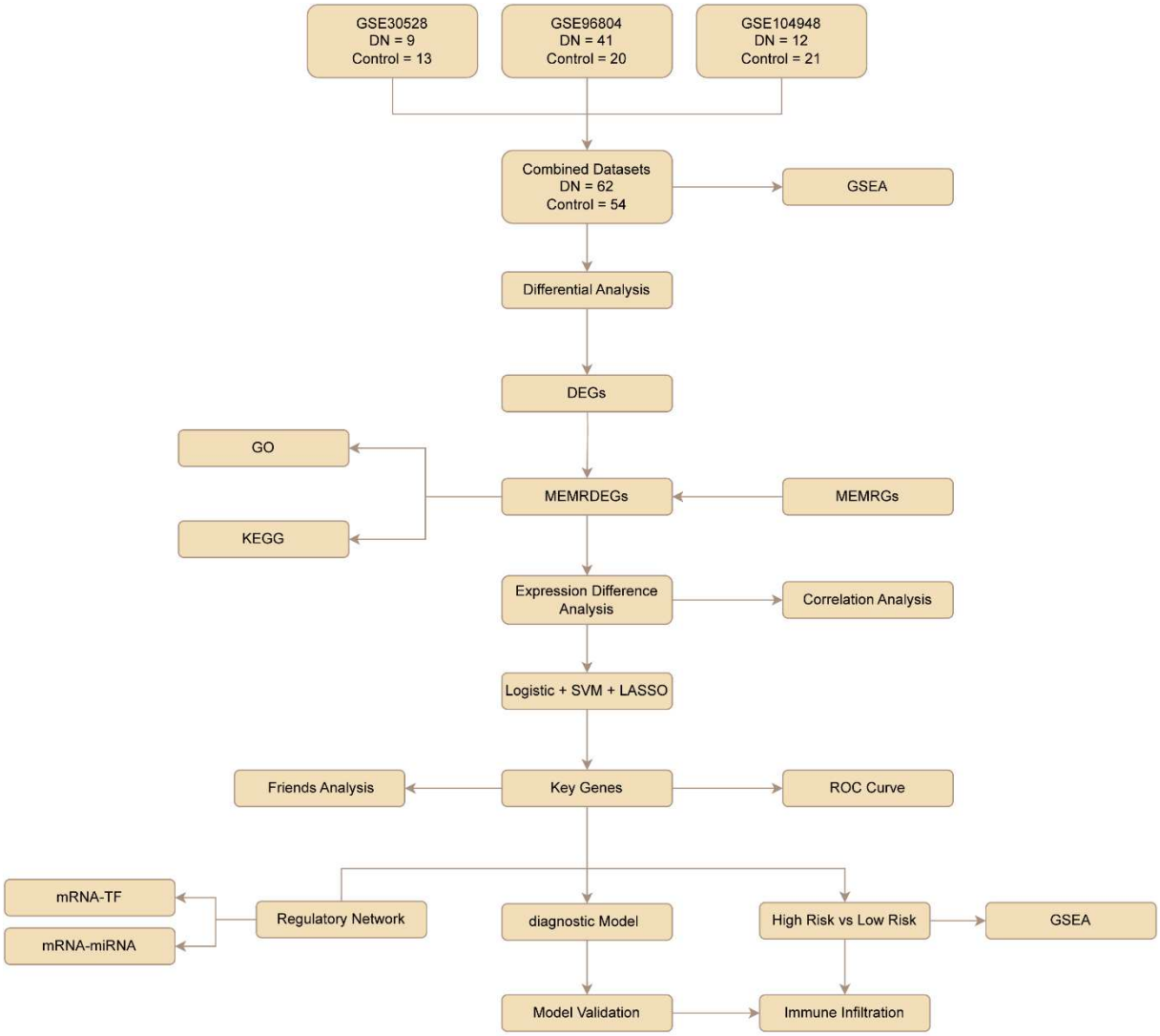


Fig.1 Flow Chart for the Comprehensive Analysis of MEMRDEGs

2.2 Batch-effect removal supports dataset integration

To integrate the three GEO microarray datasets, batch effects among GSE30528, GSE96804, and GSE104948 were removed using the sva package. Expression distributions (boxplots) and sample clustering patterns (PCA) were markedly improved after debatching, indicating effective batch correction and enabling downstream analyses (Fig. 2).

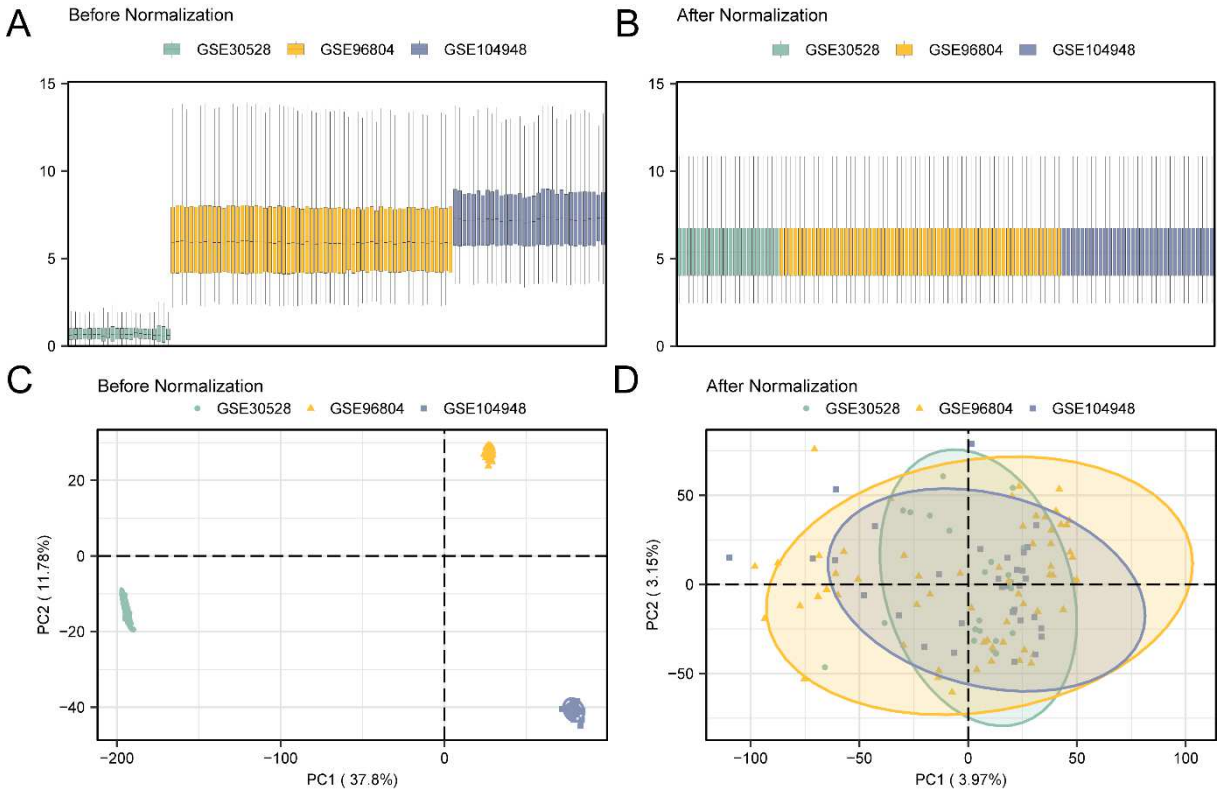


Fig.2 Batch Effects Removal of GSE30528, GSE96804 and GSE104948

(A) Boxplots of expression distributions across the combined GEO datasets before batch correction. (B) Boxplots after batch correction. (C) PCA of datasets prior to debatching. (D) PCA after debatching. GSE30528 (green), GSE96804 (yellow), and GSE104948 (blue).

2.3 Differentially expressed mitochondrial energy metabolism–related genes in DN

In the integrated cohort, differential expression between DN and controls was performed using limma. A total of 537 DEGs were identified using the criteria $|\log2FC| > 0.5$ and adjusted $P < 0.05$(Fig. 3A).A mitochondrial energy metabolism–related gene set (381 genes) was compiled, and DEGs were intersected with this set to obtain 20 MEMRDEGs (Fig. 3B–C).The 20 MEMRDEGs were: PDK4, CYP4A11, ALDH6A1, CR1, KNG1, EHHADH, ADH6, ACADSB, ALDH1L1, PC, ASS1, CKB, ECHS1, APOE, ALDH2, AHR, UCP2, UCHL1, HTR2B, and ADH1B. Chromosomal localization showed that these genes were enriched on chromosomes 3 and 4 (Fig. 3D).

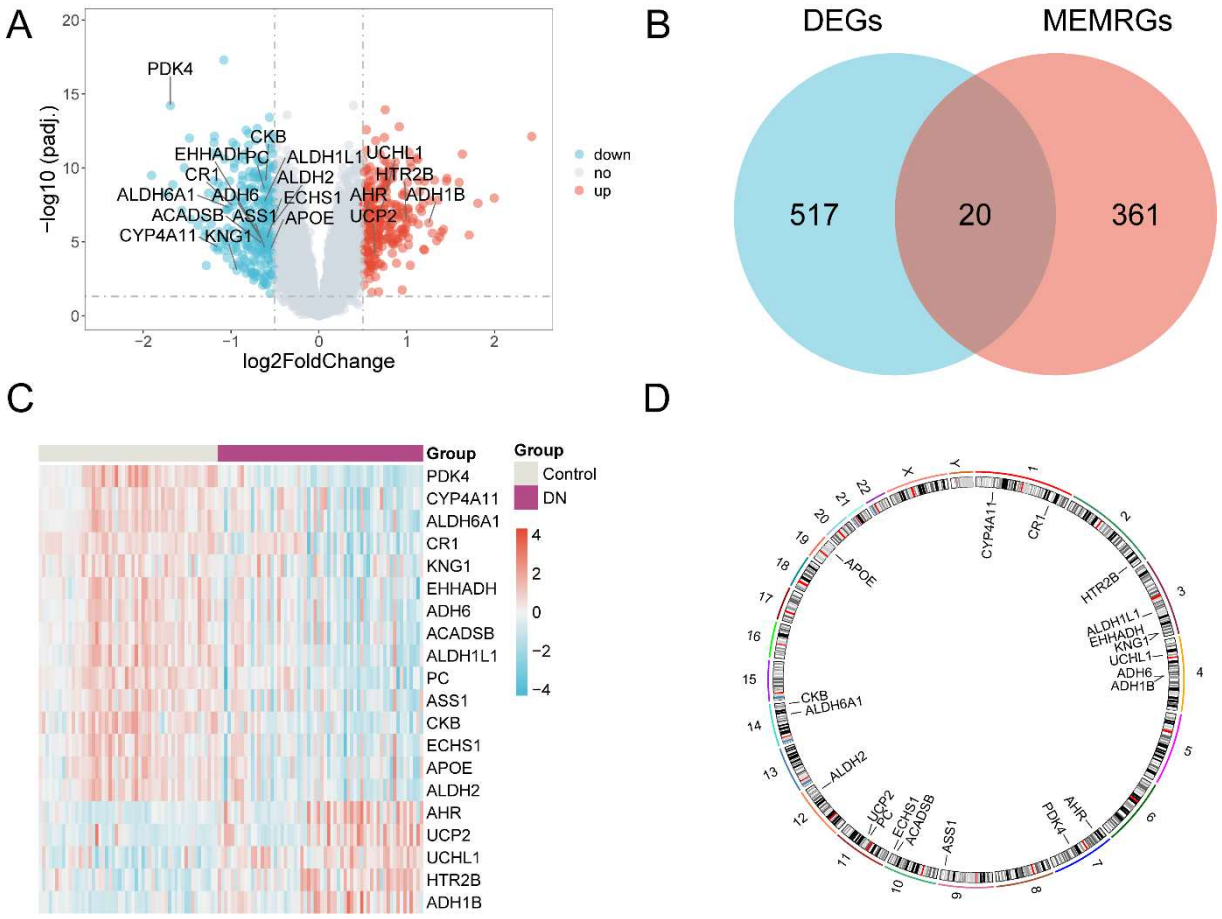


Fig.3 Differential Gene Expression Analysis

(A) Volcano plot of DEGs between DN and control samples in the integrated dataset. (B) Intersection between DEGs and MEMRGs. (C) Heatmap of MEMRDEGs. (D) Chromosomal distribution of MEMRDEGs. In panel C, red indicates higher expression and blue indicates lower expression

2.4 Expression changes and co-expression patterns of MEMRDEGs

The 20 MEMRDEGs exhibited robust differential expression between DN and controls ($P < 0.001$ for all genes) (Fig. 4A) .Spearman correlation analysis revealed coordinated expression among MEMRDEGs, including the strongest positive correlation between ALDH6A1 and ALDH1L1 ($r = 0.927$, $P < 0.001$) and a moderate negative correlation between PC and AHR ($r = −0.673$, $P < 0.001$). Scatter plots confirmed these strongest correlation pairs (Fig. 4B–D).

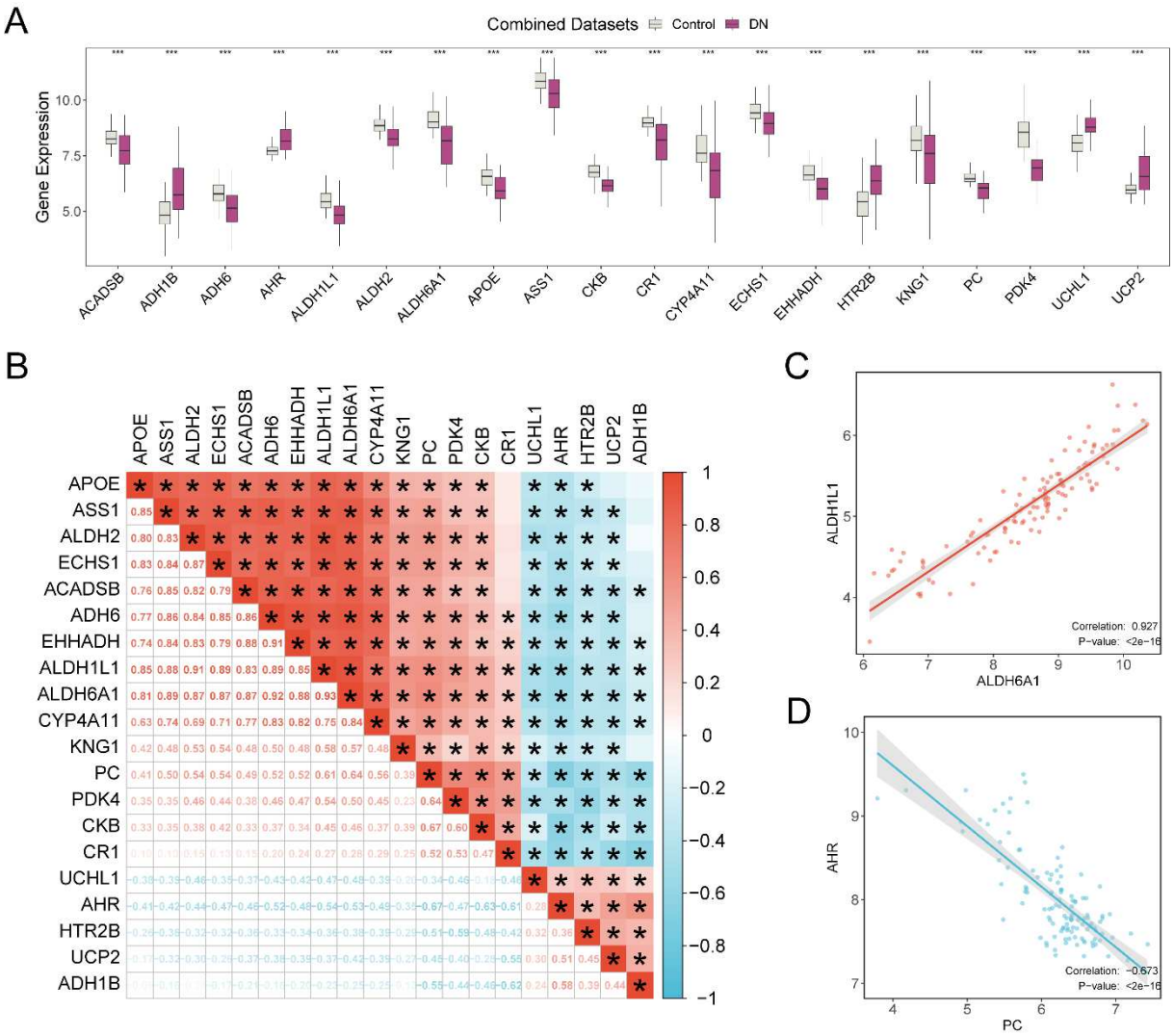


Fig.4 Differential Expression and Correlation Analysis

（A）Differential expression of MEMRDEGs between DN and control groups. (B) Heatmap of MEMRDEG expression patterns in the integrated dataset. (C) Spearman correlation between ALDH6A1 and ALDH1L1. (D) Spearman correlation between PC and AHR. Correlation coefficients (r) and P values are shown in the scatter plots.

2.5 Functional enrichment of MEMRDEGs

GO and KEGG enrichment analyses were conducted for the 20 MEMRDEGs. The results indicated enrichment in pathways related to small-molecule/organic acid catabolism, ethanol metabolism, and branched-chain amino acid catabolism at the biological process level, and mitochondrial matrix-associated cellular components and oxidoreductase-related functions at the molecular function level. KEGG enrichment further highlighted fatty acid degradation, valine/leucine/isoleucine degradation, β-alanine metabolism, pyruvate metabolism, and propanoate metabolism. Visual summaries were provided as bubble plots and gene–term networks (Fig. 5). the specific results are shown in Table 2.

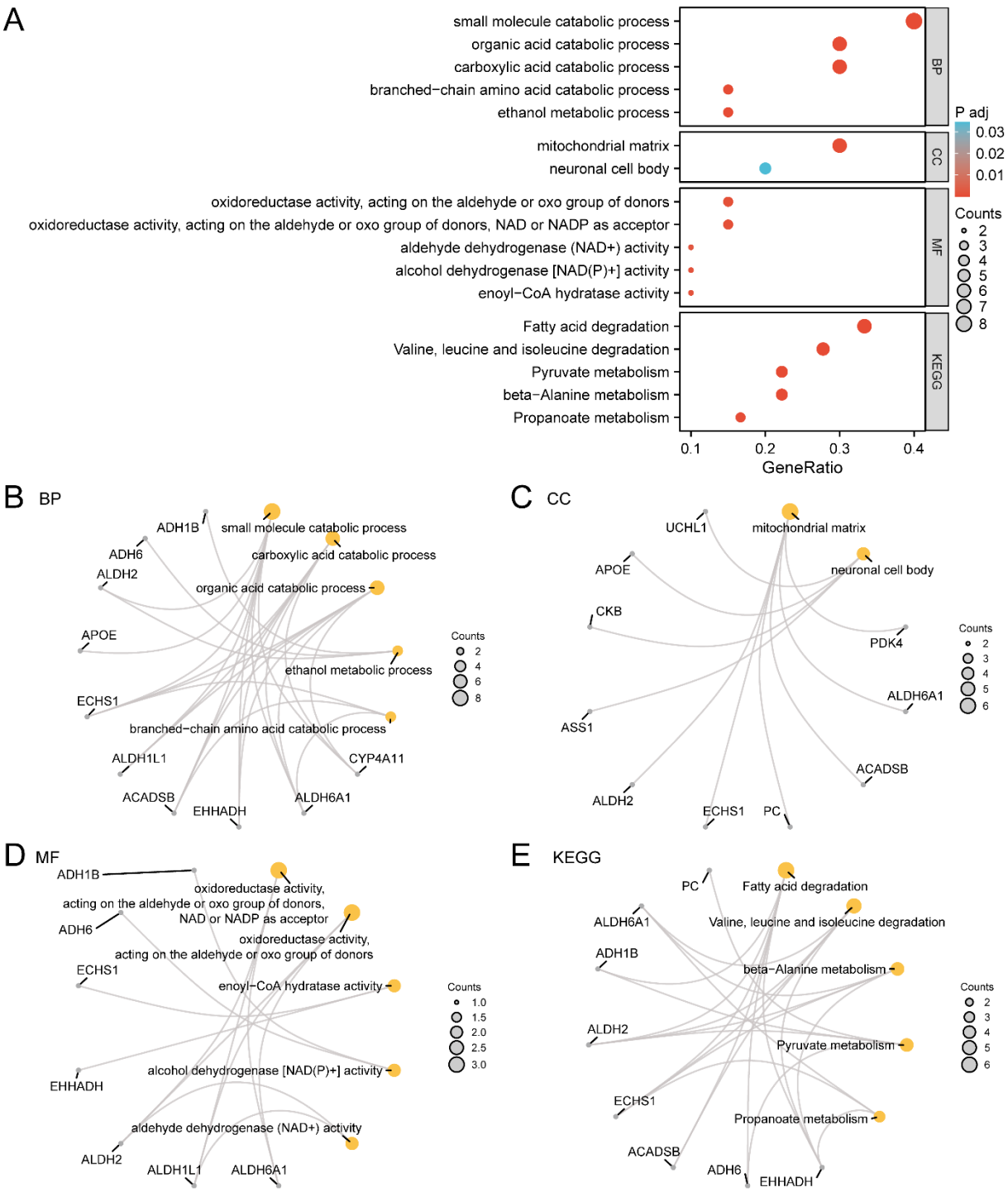


Fig.5 GO/KEGG enrichment of MEMRDEGs

(A)Bubble plot summarizing GO (BP/CC/MF) and KEGG terms. (B–E) Enrichment results of GO and KEGG displayed as separate panels for BP (B), CC (C), MF (D), and KEGG (E). Bubble size reflects gene ratio. Nodes in networks represent terms (yellow) and genes (gray); edges denote associations. Enrichment thresholds: adj. P<0.05P<0.05 and FDR (q value) < 0.25 (Benjamini–Hochberg).

2.6 Global GSEA highlights DN-associated inflammatory and injury programs

Gene set enrichment analysis (GSEA) was applied to the ranked gene list in the integrated dataset (all genes). DN samples showed significant enrichment in multiple inflammation- and injury-related pathways,

including inflammatory response, TGF-β signaling for EMT, FcεRI-mediated NF-κB activation, P130cas–MAPK signaling linked to integrins, IL-17 pathway, focal adhesion–PI3K–AKT–mTOR signaling, and apoptosis(Fig. 6). The link between the cellular components affected and the molecular functions shown in Table 3.

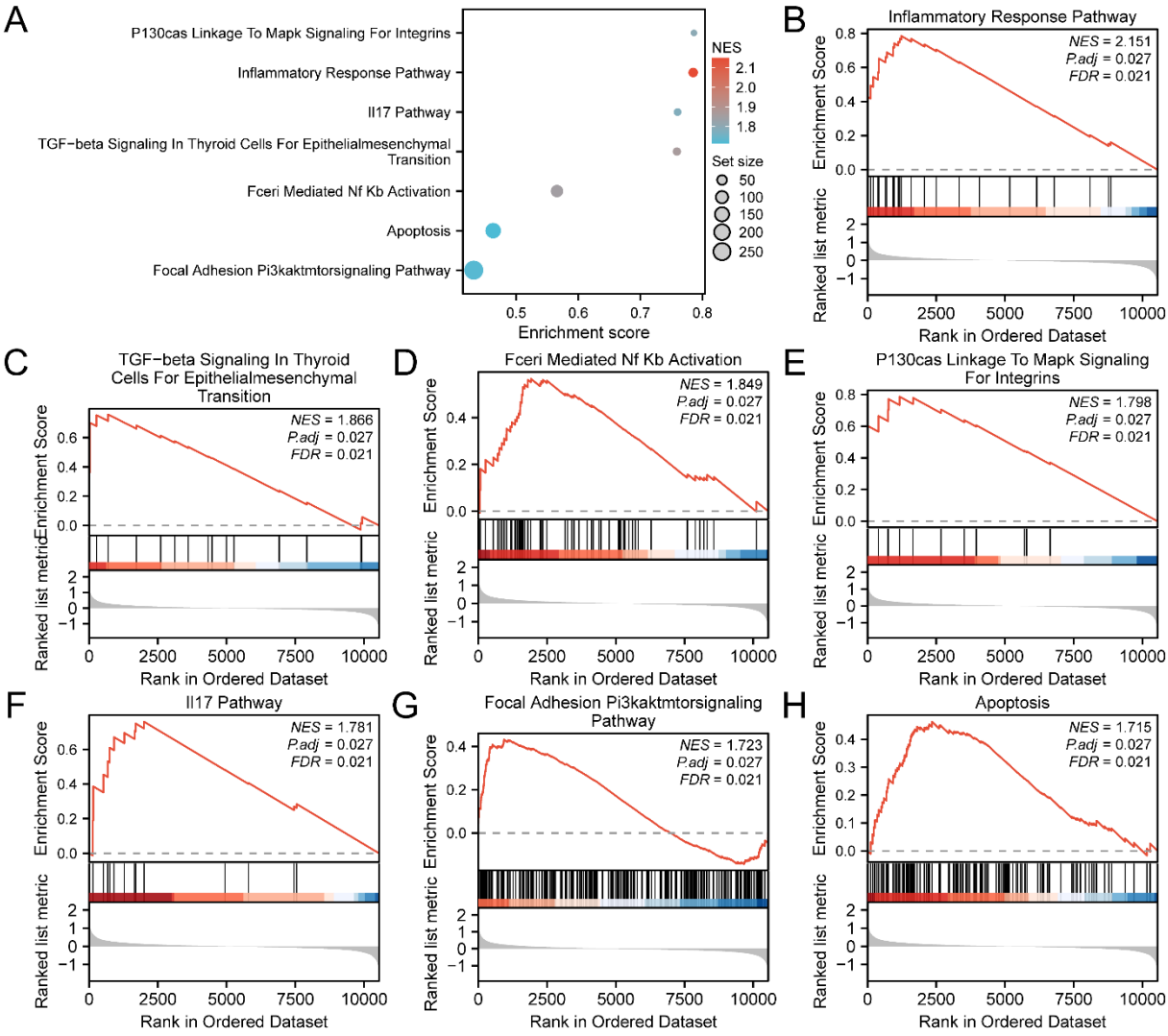


Fig.6 GSEA reveals DN-associated inflammatory and injury programs

(A)Bubble plot summarizing seven enriched functional gene sets. (B–H) GSEA enrichment plots for: Inflammatory Response Pathway (B), TGF-β signaling for EMT (C), FcεRI-mediated NF-κB activation (D), P130cas-linked MAPK signaling via integrins (E), IL-17 pathway (F), focal adhesion–PI3K–AKT–mTOR signaling (G), and apoptosis (H). NES and FDR are indicated for each pathway. GSEA cutoffs: adj. P<0.05*P*<0.05 and FDR q < 0.25 (Benjamini–Hochberg).

2.7 Construction of a 7-gene LASSO-based diagnostic signature

To establish a diagnostic model, MEMRDEGs were first assessed using logistic regression (forest plot). SVM-RFE was then used to estimate an optimal feature number, and LASSO regression was applied to derive the final sparse signature (Fig. 7A–E) . Ultimately , LASSO selected seven key genes: PDK4, CKB, UCHL1, ADH1B, ALDH1L1, CR1, and PC, which were used to calculate the RiskScore (Fig. 7D–E).

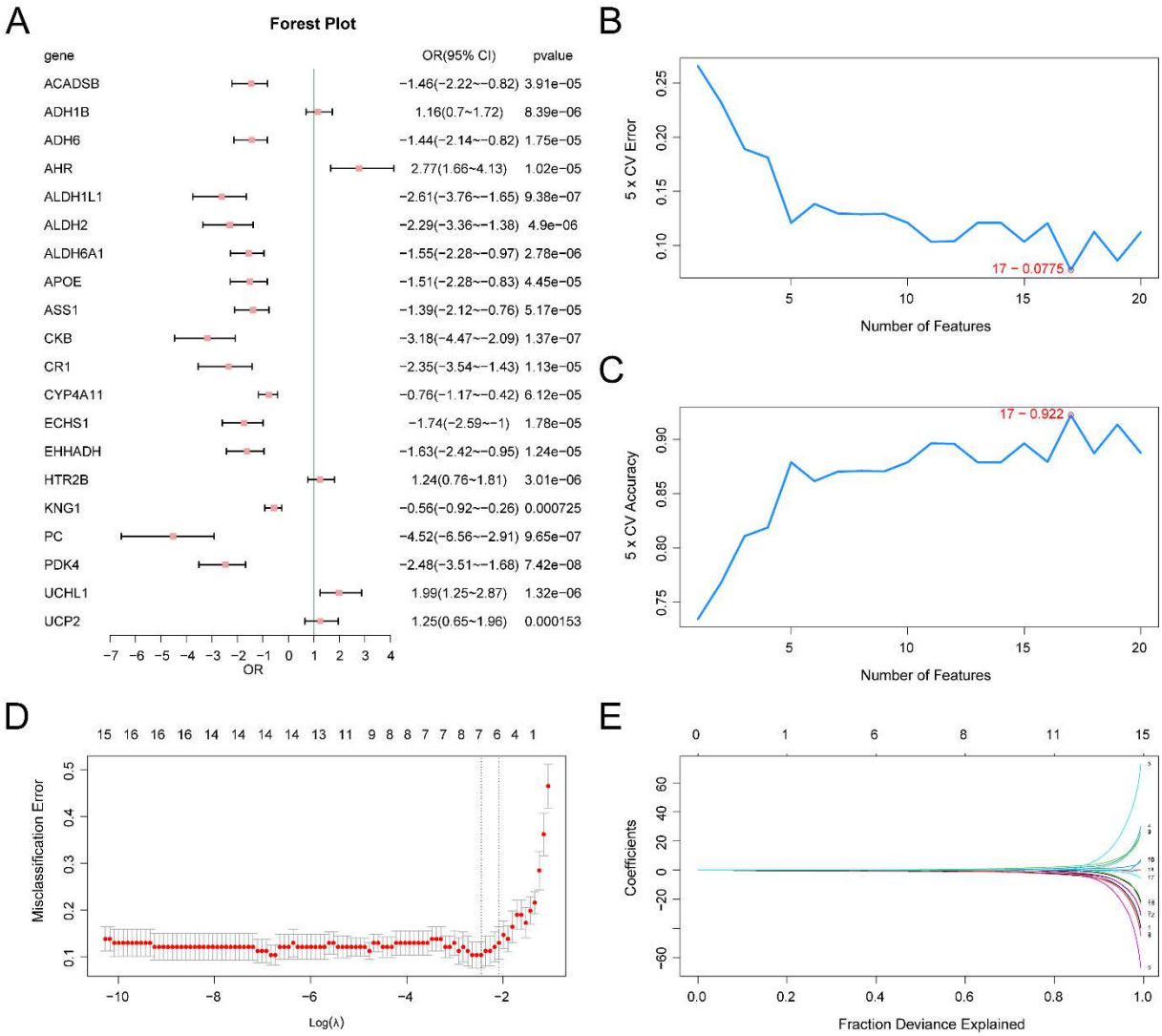


Fig.7 Construction of the 7-gene LASSO-based diagnostic signature

(A) Logistic regression forest plot for MEMRDEGs. (B–C) Feature selection performance using SVM-RFE (error rate and classification accuracy). (D) LASSO coefficient trajectories. (E) Final LASSO model variables. The LASSO-derived signature includes PDK4, CKB, UCHL1, ADH1B, ALDH1L1, CR1, and PC.

2.8 Diagnostic performance and model validation

ROC analysis based on the calculated RiskScore showed strong diagnostic discrimination between DN and controls (AUC > 0.9) in the integrated dataset (Fig. 8A). DN samples were further stratified into high-risk and low-risk groups using the median RiskScore. ROC analyses of individual key genes showed that PDK4 had the best diagnostic performance, while the remaining genes exhibited moderate discrimination (0.7 < AUC < 0.9) (Fig. 8B–D). Functional similarity analysis identified PC as a top-associated gene near the cutoff value (Fig. 8E). A nomogram model based on the seven key genes highlighted PDK4 as the strongest contributor to DN diagnosis. Calibration and decision curve analyses supported good agreement between predicted and observed probabilities and suggested potential clinical utility. (Fig. 8F–H).RiskScore formula:

$$RiskScore = \mathrm{PDK4} * (-0.7792) + \mathrm{CKB} * (-0.5616) + \mathrm{UCHL1} * (0.4806) + \mathrm{ADH1B} * (0.0637) + \mathrm{ALDH1L1} * (-0.1820) + \mathrm{CR1} * (-0.1195) + \mathrm{PC} * (-0.1382)$$

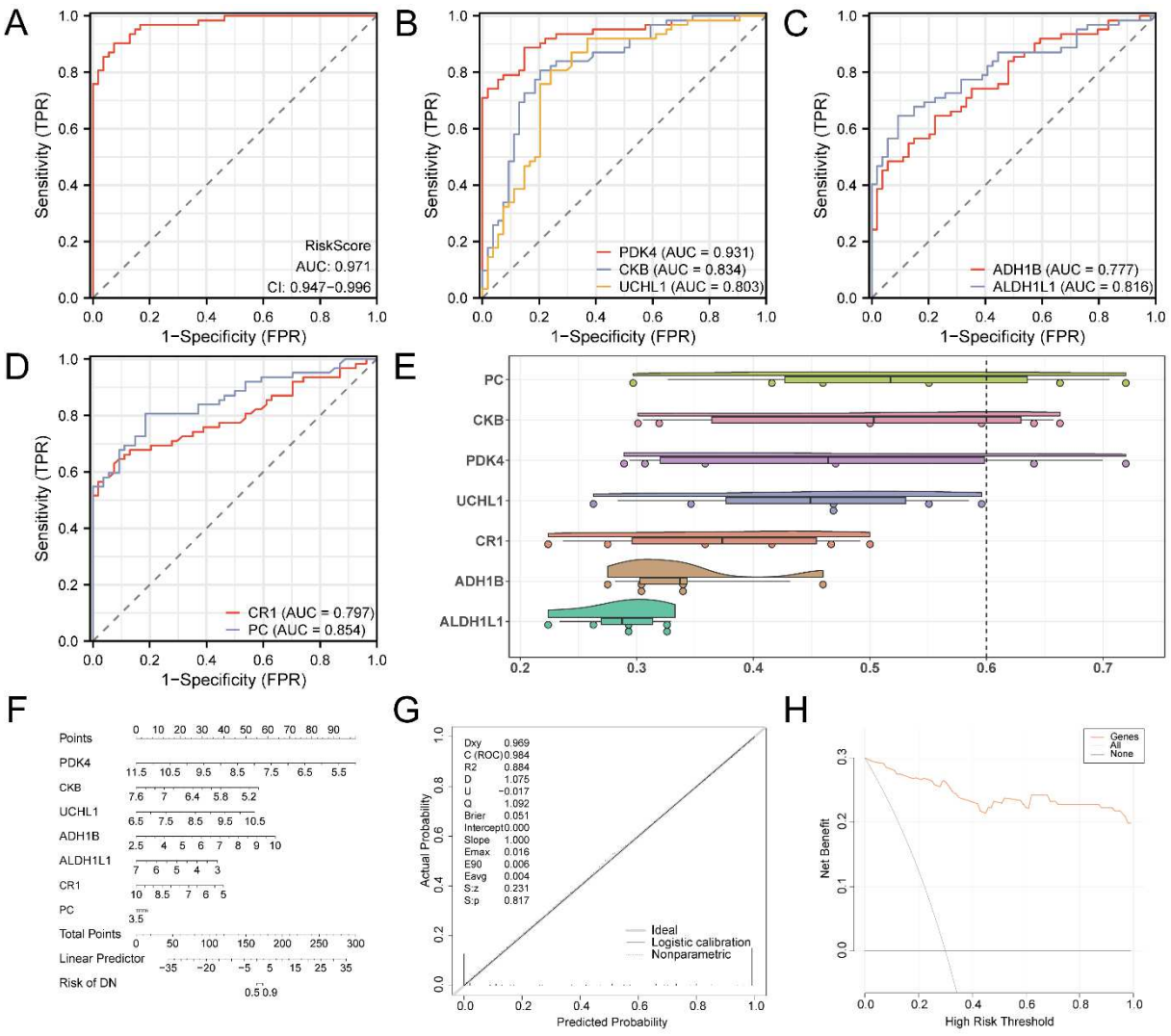


Fig.8 Diagnostic performance and validation of the risk model.

(A)ROC curve for the LASSO RiskScore in the integrated dataset. (B–D) ROC curves for individual signature genes. (E) Functional similarity ("Friends") analysis around the cutoff region. (F) Nomogram based on the seven signature genes. (G) Calibration curve of predicted versus observed probability. (H) Decision curve analysis (DCA) assessing net clinical benefit.

2.9 High-vs low-risk DN show distinct transcriptional and pathway profiles

DN samples were divided into high- and low-risk groups based on the median RiskScore. Differential expression analysis between high- and low-risk groups identified 263 DEGs ($|\log2FC| > 0.5$, adjusted $P < 0.05$), including 180 upregulated and 63 downregulated genes (Fig. 9A–B).

GSEA based on all genes showed that both high- and low-risk profiles were dominated by inflammatory and immune-related programs. In particular, high-risk DN was significantly enriched in pathways including inflammatory response, IL-1 and megakaryocytes in obesity, FcεRI-mediated NF-κB activation, focal adhesion–PI3K–AKT–mTOR signaling, TP53 network, pyroptosis, and TGF-β signaling associated with EMT (Fig. 9C–J). the detailed results are shown in Table 4.

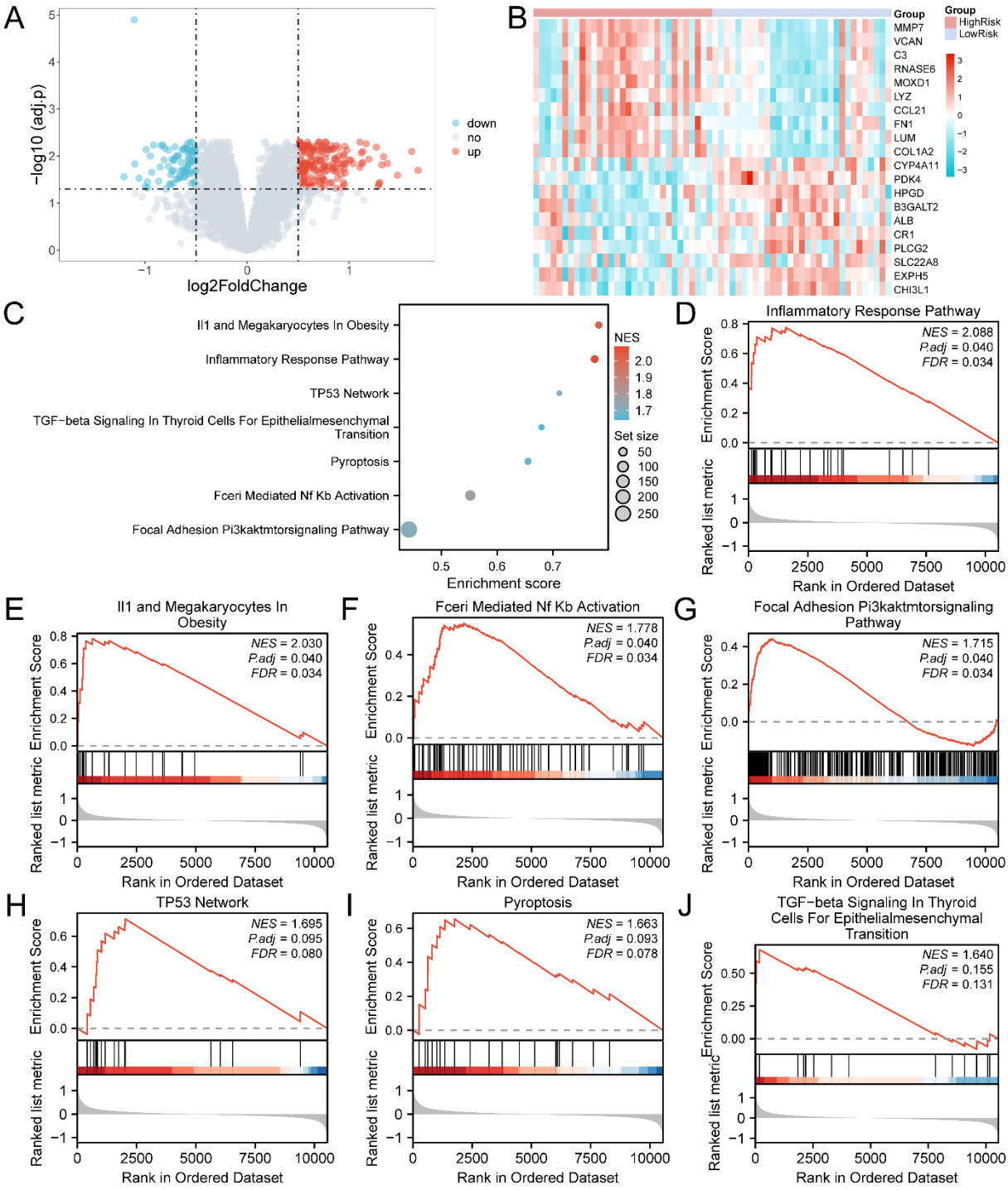


Fig.9 Risk stratification is associated with distinct transcriptional and pathway profiles

(A) Volcano plot of differential expression between high- and low-risk DN groups. (B) Heatmap of DEGs between the two risk groups. (C) Bubble plot summarizing enriched GSEA functional gene sets. (D–J) GSEA enrichment plots for: Inflammatory Response Pathway (D), IL-1 and megakaryocytes in obesity (E), FcεRI-mediated NF-κB activation (F), focal adhesion–PI3K–AKT–mTOR signaling (G), TP53 network (H), pyroptosis (I), and TGF-β signaling associated with EMT (J). GSEA screening: adj. P<0.05P<0.05 and FDR q < 0.25 (Benjamini–Hochberg).

2.10 Regulatory networks implicate multi-layer control of the signature genes

Transcription factors (TFS) combined with Key Genes were obtained through ChIPBase database, and

the screening conditions were Number of sample found (upstream + downstream) > 9. The mRNA-TF Regulatory Network was constructed and visualized by Cytoscape software (Fig10A). Among them, 6 Key Genes and 42 transcription factors (TFS) were included, and the specific information is shown in Table S2.

The mirnas related to the Key Genes were obtained through the StarBase database, and the screening criteria were mRNA-miRNA interactions recorded from at least three sources. The mRNA-miRNA Regulatory Network was constructed and visualized by Cytoscape software (Fig10B). Among them, there were 4 Key Genes and 35 mirnas, the specific information is shown in Table S3.

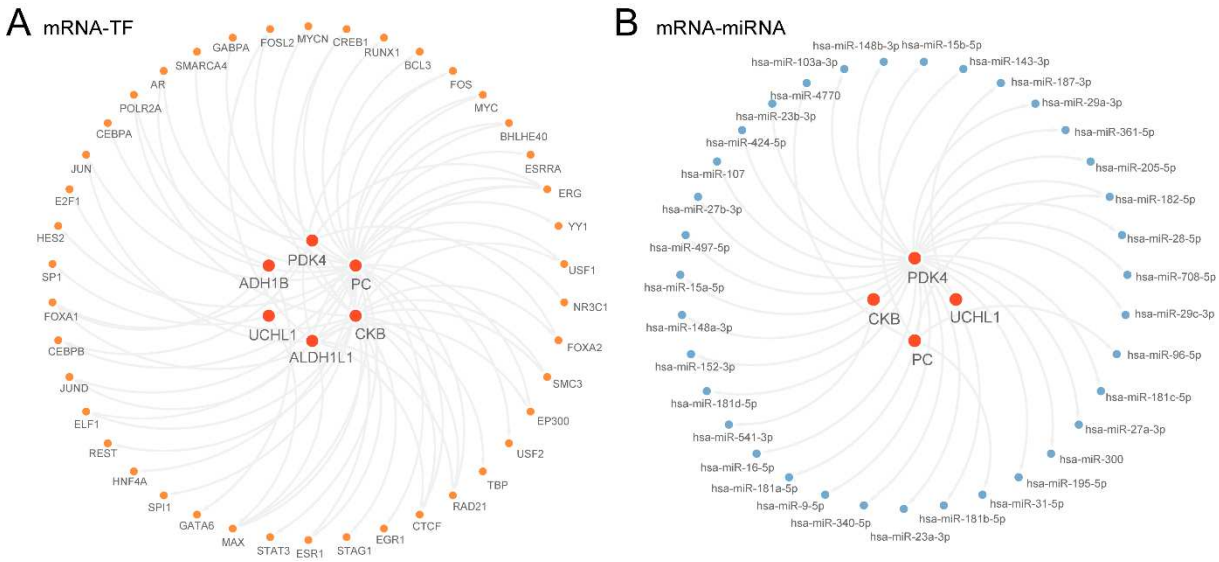


Fig.10 Regulatory networks suggest multi-layer control of the signature genes

(A) mRNA–TF regulatory network. (B) mRNA–miRNA regulatory network. Key genes are shown in red, transcription factors in orange, and miRNAs in blue..

2.11 Immune infiltration differs between high- and low-risk DN groups

Immune infiltration was estimated using ssGSEA for 28 immune cell types in DN samples. Compared with the low-risk group, 16 immune cell types showed significant abundance differences between high- and low-risk groups ($P < 0.05$) (Fig. 11A). Correlation analyses further suggested structured immune relationships: the strongest positive correlation in the low-risk group was between central memory CD4 T cells and regulatory T cells ($r = 0.825$, $P < 0.05$), while in the high-risk group the strongest positive correlation was between central memory CD8 T cells and regulatory T cells ($r = 0.933$, $P < 0.05$) (Fig. 11B–C). Gene–immune correlations indicated that ADH1B was positively associated with memory B cells in the low-risk group ($r = 0.808$, $P < 0.05$), whereas CR1 showed the strongest negative association with memory B cells in the high-risk group ($r = -0.891$, $P < 0.05$) (Fig. 11D–E).

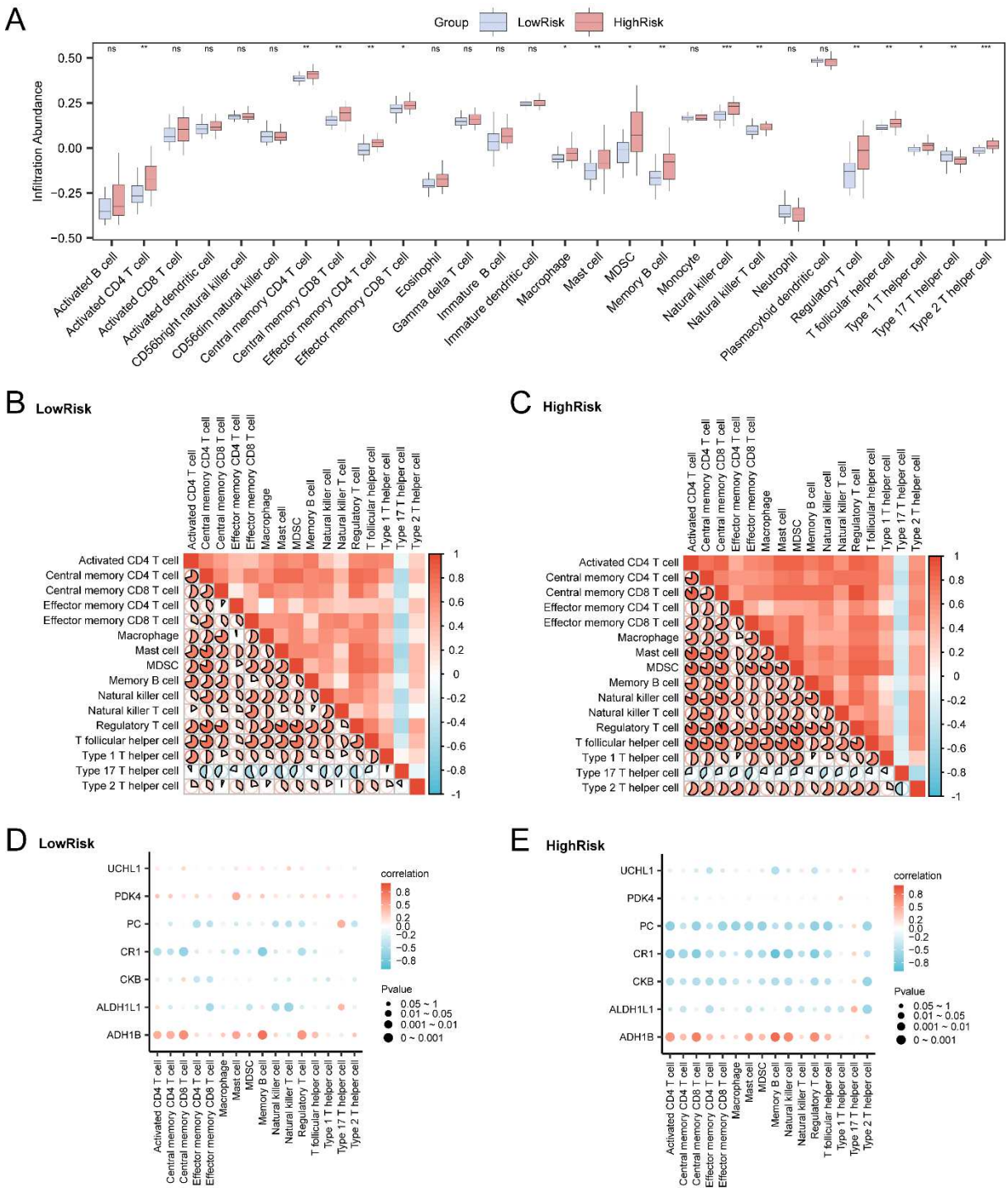


Fig.11 Immune infiltration differs between high- and low-risk DN groups

(A) ssGSEA-based comparison of immune cell infiltration between low- and high-risk groups. (B–C) Correlation analysis among immune cell infiltration levels in the low-risk (B) and high-risk (C) groups. (D–E) Gene–immune correlations between signature genes and immune cells in the low-risk (D) and high-risk (E) groups. Red indicates positive correlation and blue indicates negative correlation; color depth reflects correlation strength.

2.12 Immune infiltration differs between DN and controls

In the integrated dataset, ssGSEA showed significant immune infiltration differences between DN and controls for 20 immune cell types (P<0.05) (Fig. 12A). The strongest positive correlation in immune

infiltration was observed between regulatory T cells and T follicular helper cells (r = 0.834, P < 0.05) (Fig. 12B). Gene–immune association analysis showed the strongest negative correlation between CR1 and memory B cells (r = −0.836, P < 0.05) (Fig. 12C).

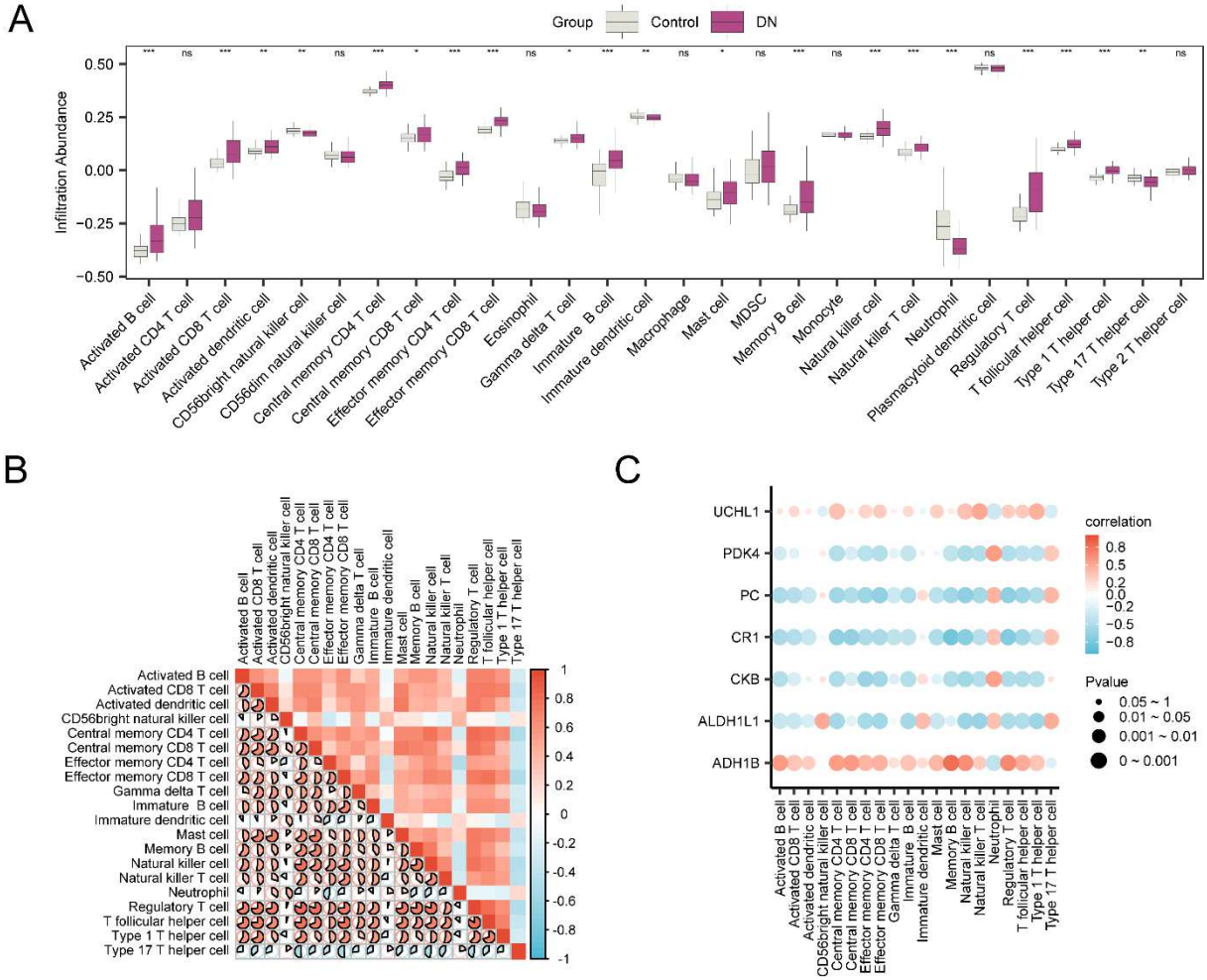


Fig.12 Immune infiltration differs between DN and controls

(A) ssGSEA-based comparison of immune cell infiltration between control and DN groups. (B) Heatmap of immune cell infiltration abundance in the integrated dataset. (C) Bubble plot of correlations between signature genes and immune infiltration. Control (gray) and DN (purple); red indicates positive correlation and blue indicates negative correlation.

2.13 Immunohistochemistry supports protein-level changes of the 7-key genes

To validate the diagnostic signature at the protein level, immunohistochemistry (IHC) was performed on formalin-fixed kidney tissues collected from biopsy-confirmed DN patients and matched adjacent normal tissues from tumor nephrectomy specimens. Compared with controls, DN tissues showed higher protein expression of ADH1B, CKB, PDK4, and UCHL1, and lower protein expression of ALDH1L1, CR1, and PC. Semi-quantitative analysis was performed using ImageJ by IOD/area (Fig. 13).

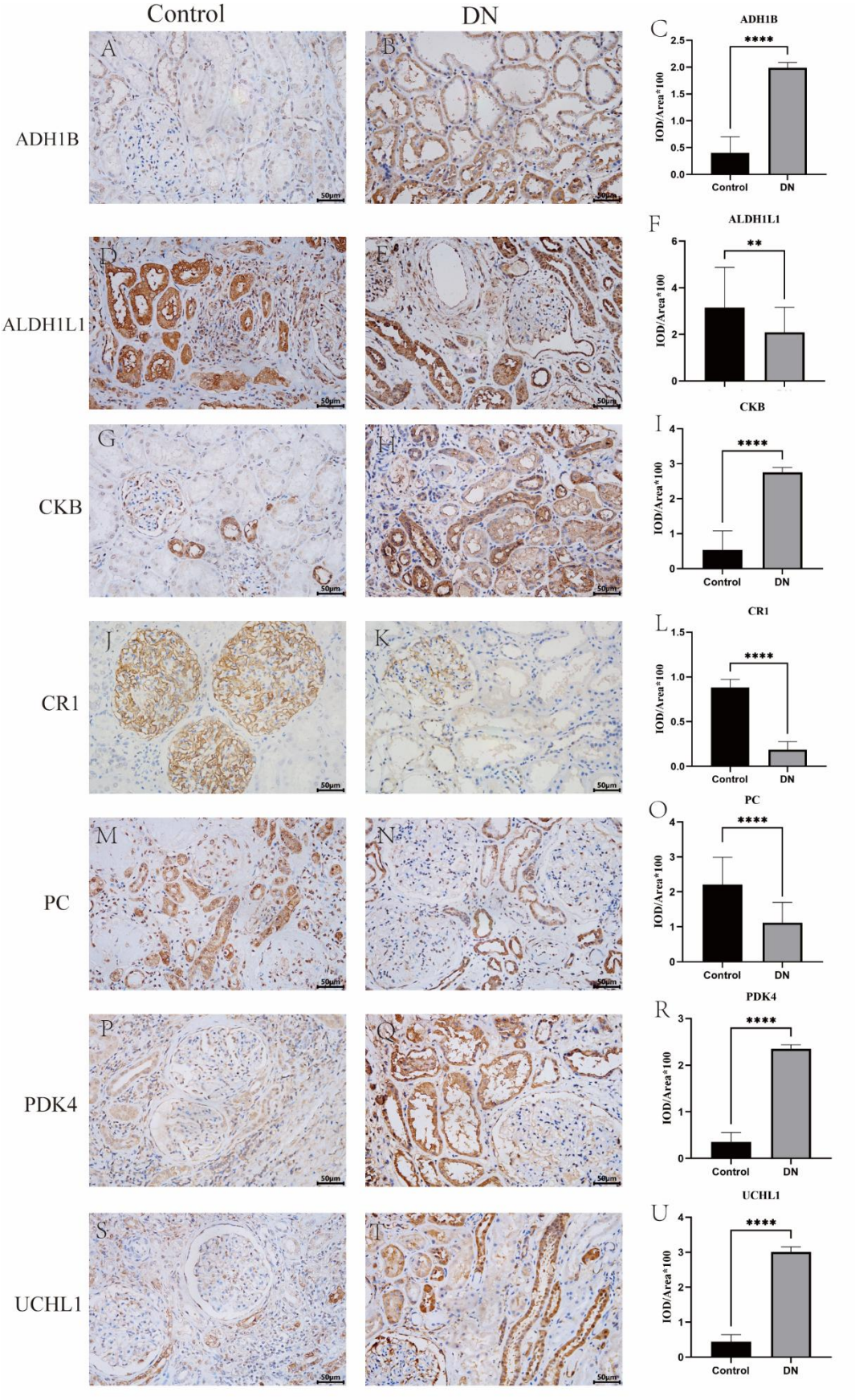


Figure 13 Immunohistochemistry of the seven signature genes in kidney tissues

(A, D, G, J, M, P, S) Representative IHC images of PDK4, CKB, UCHL1, ADH1B, ALDH1L1, CR1, and

PC in the Normal group; n=10*n*=10. (B, E, H, K, N, Q, T) Representative images for the same markers in the DN group; n=10*n*=10 (scale bar = 50 μm; magnification ×400). (C, F, I, L, O, R, U) Semi-quantitative analysis of staining intensity (IOD/area) for each marker. Statistical comparison versus the Normal group: ns P>0.05, * *P*<0.05, ****P*<0.001.

Discussion:

Our integrative multi-omics analysis identified a robust seven-gene diagnostic signature rooted in mitochondrial energy metabolism dysregulation, achieving an AUC of 0.971 in distinguishing DN from controls across three independent GEO datasets (62 DN vs. 54 controls). This model not only demonstrates superior discriminative performance compared to prior gene panels(15), but also captures the metabolic core of DN pathogenesis—extending beyond conventional risk factors to reflect intrinsic bioenergetic failure in renal cells.

Our findings suggest that PDK4 may hold notable diagnostic potential in predictive models for diabetic nephropathy DN, warranting further comparative evaluation against established variables. This observation aligns with emerging evidence highlighting the centrality of mitochondrial metabolic dysregulation in DN pathogenesis, where PDK4 serves not merely as a passive indicator but as an active driver of disease progression(16). In contrast to prior studies that primarily associate PDK4 with systemic metabolic disorders such as type 2 diabetes or fatty liver disease(17), our work specifically contextualizes its role within the renal microenvironment under diabetic conditions. While earlier literature has documented PDK4 upregulation in T2DM patients and high-fat diet models (17, 18), the present study extends this knowledge by demonstrating its direct mechanistic link to kidney injury, thereby refining its relevance from a general metabolic marker to a tissue-specific pathogenic mediator in DN. The mechanistic plausibility of PDK4's involvement in DN is supported by its canonical function as a mitochondrial kinase that phosphorylates and inactivates the E1α subunit of the pyruvate dehydrogenase complex (PDC), thereby suppressing pyruvate-to-acetyl-CoA conversion and redirecting carbon flux away from oxidative glucose metabolism(19, 20). Under high glucose (HG) conditions, this inhibition exacerbates mitochondrial dysfunction and, critically, suppresses Nrf2 activity—likely via Keap1 stabilization through impaired autophagy—leading to diminished expression of antioxidant enzymes and heightened oxidative stress in renal tubular cells(16).

The inclusion of genes such as PDK4 and PC directly implicates impaired pyruvate flux into the tricarboxylic acid cycle and compromised anaplerosis, aligning with emerging evidence that mitochondrial substrate inflexibility is a hallmark of early DN(21). In contrast to previous studies that emphasized inflammatory or fibrotic markers alone, our signature uniquely integrates metabolic regulators with immune-modulatory functions, revealing a coordinated shift from bioenergetic adaptation to maladaptive remodeling. For instance, while earlier work identified immune-related signatures involving T cells or monocytes(22), our model highlights CR1 and ADH1B—genes linking complement regulation and redox metabolism to memory B-cell infiltration—a connection recently corroborated in energy metabolism-focused DN research(15). Moreover, the observed enrichment of branched-chain amino acid catabolism and accumulation of toxic intermediates like acyl-carnitines and branched-chain ketoacids resonates with findings that metabolic infidelity drives oxidative stress and tubular injury in diabetes(23), thereby positioning our

MEMRDEGs within a mechanistically coherent framework. The RiskScore derived from this model stratifies patients into high- and low-risk subgroups that mirror a biological continuum from metabolic distress to fibro-inflammatory destruction. High-risk individuals exhibit pronounced activation of TGF-β signaling, extracellular matrix organization, IL-1–mediated inflammation, and pyroptosis—pathways consistently implicated in DN progression through mitochondrial damage-associated molecular patterns and NLRP3 inflammasome activation (24, 25). This transition aligns with experimental models where sustained mitochondrial stress triggers irreversible tissue remodeling, suggesting that our risk classification captures a critical threshold beyond which metabolic dysfunction begets structural damage. Notably, the concurrent downregulation of protective factors like PC and CR1 further supports a loss of metabolic resilience and immune homeostasis, reinforcing the model's biological plausibility.

Despite these insights, several limitations should be noted. The diagnostic model was developed and internally validated using only publicly available transcriptomic datasets, with no prospective external validation, which may limit generalizability across diverse populations and disease stages. Although immunohistochemistry confirmed protein-level changes of key targets (tubule-specific upregulation of PDK4 and CKB, and glomerular/tubular downregulation of ALDH1L1 and PC, the lack of functional experiments prevents causal inference. Moreover, immune cell infiltration was estimated computationally using ssGSEA, which cannot capture spatial architecture or functional states in the kidney microenvironment. Future studies with large multicenter, prospective cohorts, single-cell multi-omics, and mechanistic experiments are warranted to enable clinical translation of this signature.

Data Availability：The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article.

Ethics statement：This study was conducted with approval from the Ethics Committee of the Longyan First Affiliated Hospital of Fujian Medical University (Approval No: LYREC2024-k025-01). Clinical trial number: not applicable.

Author Contributions Statement:JD and ML drafted the manuscript. SZ performed specimen collection and data analysis. MW and XQ designed the study.

Funding:This study was supported by Longyan City Science and Technology Plan Project (No.2022LYF17103); Fujian Province Natural Science Foundation (No.2024J011614); Longyan Medical Science and Technology Team for Oncology and Metabolic Diseases；Excellent Community-Appropriate Talents Training Program of the Health System of Shanghai Pudong New Area (No.PWRs2025-04); Special Scientific Research Project of the ECHO Fund, Tongji University School of Medicine (No.TUSMCHC202405)

Conflicts of Interest：The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement: The core content of this study, including the study design, data collection, analysis, and conclusions, was carried out independently by the authors. During the preparation of the manuscript GPT-5.5 was used to assist with the following tasks: (1) optimization of English language

expression and grammar checking; (2) verification of conformity with academic writing standards; (3) standardization of reference formatting. All content generated with AI assistance has been carefully reviewed and revised by the authors to ensure accuracy and appropriate presentation. The authors assume full responsibility for all aspects of the article. This statement is intended to maintain academic transparency and integrity.

Publisher's note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Table 1 GEO Microarray Chip Information

	GSE30528	GSE96804	GSE104948
Platform	GPL571	GPL17586	GPL22945+GPL24120
Type	Array	Array	Array
Species	Homo sapiens	Homo sapiens	Homo sapiens
Tissue	Kidney	Kidney	Kidney
Samples in Disease group	9	41	12
Samples in Control group	13	20	21
Reference	PMID: 21752957	PMID: 29242313, 30511699	PMID: 29724730

Table 2 Result of GO and KEGG Enrichment Analysis for MEMRDEGs

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
BP	GO:0044282	small molecule catabolic process	8/20	376/18800	2.43 e-09	2.28 e-06	1.33 e-06
BP	GO:0046395	carboxylic acid catabolic process	6/20	238/18800	1.29 e-07	4.47 e-05	2.61 e-05
BP	GO:0016054	organic acid catabolic process	6/20	242/18800	1.43 e-07	4.47 e-05	2.61 e-05
BP	GO:0006067	ethanol metabolic process	3/20	21/18800	1.35 e-06	2.54 e-04	1.49 e-04
BP	GO:0009083	branched-chain amino acid catabolic process	3/20	21/18800	1.35 e-06	2.54 e-04	1.49 e-04
CC	GO:0005759	mitochondrial matrix	6/20	473/19594	5.57 e-06	3.01 e-04	2.23 e-04
CC	GO:0043025	neuronal cell body	4/20	482/19594	1.28 e-03	3.46 e-02	2.56 e-02
MF	GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	3/20	38/18410	9.03 e-06	9.45 e-04	4.85 e-04
MF	GO:0016903	oxidoreductase activity, acting on the aldehyde or oxo group of donors	3/20	46/18410	1.62 e-05	9.45 e-04	4.85 e-04
MF	GO:0004300	enoyl-CoA hydratase activity	2/20	10/18410	5.02 e-05	1.47 e-03	7.53 e-04
MF	GO:0018455	alcohol dehydrogenase [NAD(P)+] activity	2/20	10/18410	5.02 e-05	1.47 e-03	7.53 e-04
MF	GO:0004029	aldehyde dehydrogenase (NAD+) activity	2/20	15/18410	1.17 e-04	2.60 e-03	1.33 e-03
KEGG	hsa00071	Fatty acid degradation	6/18	43/8164	2.63 e-10	1.53 e-08	9.42 e-09
KEGG	hsa00280	Valine, leucine and isoleucine degradation	5/18	48/8164	4.59 e-08	1.33 e-06	8.21 e-07
KEGG	hsa00410	beta-Alanine metabolism	4/18	31/8164	5.02 e-07	9.70 e-06	5.98 e-06
KEGG	hsa00620	Pyruvate metabolism	4/18	47/8164	2.78 e-06	4.03 e-05	2.49 e-05
KEGG	hsa00640	Propanoate metabolism	3/18	32/8164	4.29 e-05	4.98 e-04	3.07 e-04

Table 3 Results of GSEA for Combined Datasets

ID	setSize	EnrichmentScore	NES	pvalue	p.adjust	qvalue
NABA_CORE_MATRISOME	176	0.67670	2.58850	1.92 e-03	2.70 e-02	2.13 e-02
REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION	234	0.64311	2.54717	1.87 e-03	2.70 e-02	2.13 e-02
REACTOME_ECM_PROTEOGLYCANS	67	0.75683	2.50250	2.01 e-03	2.70 e-02	2.13 e-02
KEGG_ECM_RECEPTOR_INTERACTION	70	0.73722	2.44703	2.03 e-03	2.70 e-02	2.13 e-02
REACTOME_DEGRADATION_OF_THE_EXTRACELLULAR_MATRIX	106	0.68579	2.43856	1.96 e-03	2.70 e-02	2.13 e-02
REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS	71	0.73152	2.43537	2.04 e-03	2.70 e-02	2.13 e-02
PID_INTEGRIN1_PATHWAY	54	0.75919	2.41172	2.01 e-03	2.70 e-02	2.13 e-02
NABA_ECM_GLYCOPROTEINS	116	0.66849	2.41088	1.94 e-03	2.70 e-02	2.13 e-02
REACTOME_COLLAGEN_DEGRADATION	52	0.74637	2.34923	2.03 e-03	2.70 e-02	2.13 e-02
NABA_COLLAGENS	30	0.82109	2.30941	1.98 e-03	2.70 e-02	2.13 e-02
REACTOME_COLLAGEN_CHAIN_TRIMERIZATION	30	0.82109	2.30941	1.98 e-03	2.70 e-02	2.13 e-02
REACTOME_MET_PROMOTES_CELL_MOTILITY	31	0.81082	2.30446	1.96 e-03	2.70 e-02	2.13 e-02
PID_SYNDECAN_1_PATHWAY	41	0.75241	2.28618	1.96 e-03	2.70 e-02	2.13 e-02
WP_INFLAMMATORY_RESPONSE_PATHWAY	26	0.78519	2.15082	1.99 e-03	2.70 e-02	2.13 e-02
WP_TGFBETA_SIGNALING_IN_THYROID_CELLS_FOR_EPITHELIALMESENCHYMAL_TRANSITION	17	0.75894	1.86619	1.97 e-03	2.70 e-02	2.13 e-02
REACTOME_FCERI_MEDIATED_NF_KB_ACTIVATION	65	0.56570	1.84926	2.04 e-03	2.70 e-02	2.13 e-02
REACTOME_P130CAS_LINKAGE_TO_MAPK_SIGNALING_FOR_INTEGRINS	12	0.78642	1.79839	1.93 e-03	2.70 e-02	2.13 e-02
BIOCARTA_IL17_PATHWAY	14	0.76004	1.78050	1.96 e-03	2.70 e-02	2.13 e-02
WP_FOCAL_ADHESION_PI3KAKTMTORSIGNALING_PATHWAY	255	0.43161	1.72312	1.92 e-03	2.70 e-02	2.13 e-02

REACTOME_APOPTOSIS	138	0.46287	1.71466	1.93 e-03	2.70 e-02	2.13 e-02

Table 4 Results of GSEA for Risk Group

ID	setSize	Enrichment Score	NES	pvalue	p.adjust	qvalue
NABA_CORE_MATRISOME	176	0.70422	2.63008	1.48 e-03	4.00 e-02	3.36 e-02
REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION	234	0.64162	2.46202	1.46 e-03	4.00 e-02	3.36 e-02
NABA_ECM_GLYCOPROTEINS	116	0.69471	2.44331	1.60 e-03	4.00 e-02	3.36 e-02
REACTOME_ECM_PROTEOGLYCANS	67	0.75142	2.44000	1.68 e-03	4.00 e-02	3.36 e-02
KEGG_ECM_RECEPTOR_INTERACTION	70	0.73710	2.41005	1.69 e-03	4.00 e-02	3.36 e-02
REACTOME_DEGRADATION_OF_THE_EXTRACELLULAR_MATRIX	106	0.69008	2.40533	1.58 e-03	4.00 e-02	3.36 e-02
PID_INTEGRIN1_PATHWAY	54	0.76553	2.38238	1.75 e-03	4.00 e-02	3.36 e-02
REACTOME_INTERLEUKIN_10_SIGNALING	40	0.80487	2.36473	1.81 e-03	4.00 e-02	3.36 e-02
REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS	71	0.71624	2.35005	1.69 e-03	4.00 e-02	3.36 e-02
WP_TYROBP_CAUSAL_NETWORK_IN_MICROGLIA	47	0.77573	2.34965	1.79 e-03	4.00 e-02	3.36 e-02
REACTOME_COLLAGEN_DEGRADATION	52	0.75546	2.33167	1.78 e-03	4.00 e-02	3.36 e-02
REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES	46	0.75756	2.30160	1.79 e-03	4.00 e-02	3.36 e-02
REACTOME_MET_ACTIVATES_PTK2_SIGNALING	22	0.84837	2.20131	1.84 e-03	4.00 e-02	3.36 e-02
WP_INFLAMMATORY_RESPONSE_PATHWAY	26	0.77480	2.08763	1.85 e-03	4.00 e-02	3.36 e-02
WP_IL1_AND_MEGAKARYOCYTES_IN_OBESITY	22	0.78219	2.02959	1.84 e-03	4.00 e-02	3.36 e-02
REACTOME_FCERI_MEDIATED_NF_KB_ACTIVATION	65	0.55207	1.77843	1.69 e-03	4.00 e-02	3.36 e-02
WP_FOCAL_ADHESION_PI3KAKTMTORSIGNALING_PATHWAY	255	0.44218	1.71540	1.46 e-03	4.00 e-02	3.36 e-02
WP_TP53_NETWORK	16	0.71151	1.69494	1.15 e-02	9.50 e-02	7.98 e-02
REACTOME_PYROPTOSIS	20	0.65535	1.66315	1.10 e-02	9.34 e-02	7.84 e-02
WP_TGFBETA_SIGNALING_IN_THYROID_CELLS_FOR_EPITHELIALMESENCHYMAL_TRANSITION	17	0.67954	1.64047	2.28 e-02	1.55 e-01	1.31 e-01

References:

1. Gupta S, Dominguez M, Golestaneh L. Diabetic Kidney Disease: An Update. Med Clin North Am. 2023;107(4):689-705.
2. Jung CY, Yoo TH. Pathophysiologic Mechanisms and Potential Biomarkers in Diabetic Kidney Disease. Diabetes Metab J. 2022;46(2):181-97.
3. Thipsawat S. Early detection of diabetic nephropathy in patient with type 2 diabetes mellitus: A review of the literature. Diab Vasc Dis Res. 2021;18(6):14791641211058856.
4. Wang N, Zhang C. Recent Advances in the Management of Diabetic Kidney Disease: Slowing Progression. Int J Mol Sci. 2024;25(6).
5. Barrera-Chimal J, Lima-Posada I, Bakris GL, Jaisser F. Mineralocorticoid receptor antagonists in diabetic kidney disease - mechanistic and therapeutic effects. Nat Rev Nephrol. 2022;18(1):56-70.
6. Fan X, Yang M, Lang Y, Lu S, Kong Z, Gao Y, et al. Mitochondrial metabolic reprogramming in diabetic kidney disease. Cell Death Dis. 2024;15(6):442.
7. Bhargava P, Schnellmann RG. Mitochondrial energetics in the kidney. Nat Rev Nephrol. 2017;13(10):629-46.
8. Yao L, Liang X, Qiao Y, Chen B, Wang P, Liu Z. Mitochondrial dysfunction in diabetic tubulopathy. Metabolism. 2022;131:155195.
9. Lee H, Lee H, Lim Y. Vitamin D(3) improves lipophagy-associated renal lipid metabolism and tissue damage in diabetic mice. Nutr Res. 2020;80:55-65.
10. Pavlovic N, Krizanac M, Kumric M, Vukojevic K, Bozic J. Mitochondrial Dysfunction: The Silent Catalyst of Kidney Disease Progression. Cells. 2025;14(11).
11. Yao L, Liang X, Liu Y, Li B, Hong M, Wang X, et al. Non-steroidal mineralocorticoid receptor antagonist finerenone ameliorates mitochondrial dysfunction via PI3K/Akt/eNOS signaling pathway in diabetic tubulopathy. Redox Biol. 2023;68:102946.
12. Zhao H, Sun D, Wang S, Liu Y, Zhao X, Tian W, et al. Branched-chain amino acids contribute to diabetic kidney disease progression via PKM2-mediated podocyte metabolic reprogramming and apoptosis. Nat Commun. 2025;16(1):7846.
13. Hao Y, Fan Y, Feng J, Zhu Z, Luo Z, Hu H, et al. ALCAT1-mediated abnormal cardiolipin remodelling promotes mitochondrial injury in podocytes in diabetic kidney disease. Cell Commun Signal. 2024;22(1):26.
14. Huang J, Pang D, Fan C, Yang G, Chen J, Chen S. Integrative transcriptomic and machine learning analyses identify HDAC9 as a key regulator of mitochondrial dysfunction and senescence-associated inflammation in diabetic nephropathy. Front Immunol. 2025;16:1627173.
15. He S, Ye J, Wang Y, Xie LY, Liu SY, Chen QK. Identification and functional analysis of energy metabolism and pyroptosis-related genes in diabetic nephropathy. Heliyon. 2025;11(3):e42201.
16. Tian S, Yang X, Lin Y, Li X, Zhou S, Yu P, et al. PDK4-mediated Nrf2 inactivation contributes to oxidative stress and diabetic kidney injury. Cell Signal. 2024;121:111282.
17. Bjune MS, Lawrence-Archer L, Laupsa-Borge J, Sommersten CH, McCann A, Glastad RC, et al. Metabolic role of the hepatic valine/3-hydroxyisobutyrate (3-HIB) pathway in fatty liver disease. EBioMedicine. 2023;91:104569.
18. Wang M, Ren B, Wu X, Guo J, Cao Y, Men L, et al. S100A9 inhibition ameliorates HFpEF by modulating mitochondrial fission and oxidative stress. Int

Immunopharmacol. 2025;163:115280.
19. Pradhan S, Mito T, Khan NA, Olander S, Zhaivoron A, McWilliams TG, et al. PDK4 and nutrient responses explain muscle specific manifestation in mitochondrial disease. Clin Transl Med. 2025;15(7):e70404.
20. Su H, Liu Z, Wei J, Liu Y, Zhong Y, Liu X, et al. PDK4 suppresses high glucose-induced microglial ferroptosis by restricting pro-ferroptotic PUFA biosynthesis. Neuroreport. 2026;37(1):1-10.
21. Khang AR, Kim DH, Kim MJ, Oh CJ, Jeon JH, Choi SH, et al. Reducing Oxidative Stress and Inflammation by Pyruvate Dehydrogenase Kinase 4 Inhibition Is Important in Prevention of Renal Ischemia-Reperfusion Injury in Diabetic Mice. Diabetes Metab J. 2024;48(3):405-17.
22. Wang J, Chen W, Chen S, Yue G, Hu Y, Xu J. Landscape of infiltrating immune cells and related genes in diabetic kidney disease. Clin Exp Nephrol. 2024;28(3):181-91.
23. Liu S, Li L, Lou P, Zhao M, Wang Y, Tang M, et al. Elevated branched-chain alpha-keto acids exacerbate macrophage oxidative stress and chronic inflammatory damage in type 2 diabetes mellitus. Free Radic Biol Med. 2021;175:141-54.
24. Dwivedi S, Sikarwar MS. Diabetic Nephropathy: Pathogenesis, Mechanisms, and Therapeutic Strategies. Horm Metab Res. 2025;57(1):7-17.
25. Tanase DM, Gosav EM, Anton MI, Floria M, Seritean Isac PN, Hurjui LL, et al. Oxidative Stress and NRF2/KEAP1/ARE Pathway in Diabetic Kidney Disease (DKD): New Perspectives. Biomolecules. 2022;12(9).