RESEARCH

Check for updates

Uncovering glycolysis-driven molecular subtypes in diabetic nephropathy: a WGCNA and machine learning approach for diagnostic precision

Chenglong Fan[1†], Guanglin Yang[1,2†], Cheng Li[1†], Jiwen Cheng[1*], Shaohua Chen[2*] and Hua Mi[1*]

Abstract

Introduction Diabetic nephropathy (DN) is a common diabetes-related complication with unclear underlying pathological mechanisms. Although recent studies have linked glycolysis to various pathological states, its role in DN remains largely underexplored.

Methods In this study, the expression patterns of glycolysis-related genes (GRGs) were first analyzed using the GSE30122, GSE30528, and GSE96804 datasets, followed by an evaluation of the immune landscape in DN. An unsupervised consensus clustering of DN samples from the same dataset was conducted based on differentially expressed GRGs. The hub genes associated with DN and glycolysis-related clusters were identified via weighted gene co-expression network analysis (WGCNA) and machine learning algorithms. Finally, the expression patterns of these hub genes were validated using single-cell sequencing data and quantitative real-time polymerase chain reaction (qRT-PCR).

Results Eleven GRGs showed abnormal expression in DN samples, leading to the identification of two distinct glycolysis clusters, each with its own immune profile and functional pathways. The analysis of the GSE142153 dataset showed that these clusters had specific immune characteristics. Furthermore, the Extreme Gradient Boosting (XGB) model was the most effective in diagnosing DN. The five most significant variables, including *GATM*, *PCBD1*, *F11*, *HRSP12,* and *G6PC*, were identified as hub genes for further investigation. Single-cell sequencing data showed that the hub genes were predominantly expressed in proximal tubular epithelial cells. In vitro experiments confirmed the expression pattern in NC.

Conclusion Our study provides valuable insights into the molecular mechanisms underlying DN, highlighting the involvement of GRGs and immune cell infiltration.

Keywords Diabetic nephropathy, Glycolysis, Glycolysis-related genes, Machine learning algorithm, Hub genes

†Chenglong Fan, Guanglin Yang and Cheng Li these authors contributed equally to this work.

*Correspondence:
Jiwen Cheng
chengjiwen@stu.gxmu.edu.cn
Shaohua Chen
silverchan1994@163.com
Hua Mi
mihua2019@163.com

BMC

Fan *et al. Biology Direct*   (2025) 20:10

Page 2 of 16

Background

DN is a serious diabetes-related complication and the leading cause of end-stage renal disease (ESRD), accounting for about 40% of all ESRD cases in the United States [1]. The prevalence of DN increases with the increasing prevalence of diabetes and may worsen if treatment strategies to prevent DN are not developed. About one-third of diabetics develop DN after the incubation period, which can last for several years [2]. The occurrence and frequency of DN in China have markedly increased over the last ten years, with about 24.3 million diabetes patients suffering from chronic kidney disease [3]. To date, the pathogenesis of DN is unclear due to its complexity. Research has indicated that even with conventional therapy, encompassing rigorous management of glucose levels and blood pressure, DN can progress to ESRD and increase mortality [4]. Therefore, understanding the pathophysiological mechanisms underlying DN, critical risk factors, and effective management approaches is crucial for DN treatment.

Endothelial cells of blood vessels produce energy through glycolysis. Abnormal glycolysis occurs in endothelial cells in patients with diabetes, atherosclerosis, pulmonary hypertension, and arthritis [5]. Recent studies have shown that glycolysis occurs in the proximal tubules based on multi-photon microscopy [6]. Nevertheless, the relationship between glycolysis-related genes (GRGs) and DN is unclear. Therefore, it is crucial to explore the molecular classification and genomic diversity within the DN cohort, particularly regarding glycolysis and its driver genes, to enhance our understanding of the fundamental pathogenic mechanisms that promote the progression and development of DN.

In this study, the expression profiles of GRGs were first assessed to identify differences between patients with DN and normal controls (NC), followed by a detailed analysis of the immune cell infiltration in these samples. Furthermore, DN samples were extracted from the training cohort. Consensus clustering was conducted utilizing the differentially expressed GRGs mentioned above. The results indicated that DN samples can be categorized into two unique clusters related to glycolysis, each displaying a distinct immune profile, functional classification, and diverse pathways. Hub genes related to glycolytic clusters were identified via Weighted Gene Co-expression Network Analysis (WGCNA) algorithm. Shared genes between DN and those associated with the glycolytic modules were identified by intersecting these hub genes with GRGs. A diagnostic model for DN was then developed by evaluating and comparing different machine-learning techniques. Nomogram, calibration plot, decision curve analysis (DCA), and the independent verification dataset GSE142153 were used to verify the identification and stability of the model. By utilizing a high-glucose induced cell model, which was assessed through quantitative real-time polymerase chain reaction (qRT-PCR). Additionally, the scRNA-seq dataset GSE183276 served as supplementary validation for single-cell sequencing. The research flow chart is shown in Fig. 1.

Materials and methods

Data collection and sample details

Five unprocessed datasets were obtained from the Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/). A DN diagnostic model was established using GSE30122, GSE30528, and GSE96804 as training datasets, including tissue samples from 60 DN patients and 70 NC individuals. The model was verified using the scRNA-seq dataset GSE183276 via single-cell sequencing. An independently validated dataset GSE142153 was used to evaluate the forecasting capability of the model. The dataset included blood samples from 30 DN patients and 10 NC individuals.

Identification and analysis of GRGs

First, a comprehensive review of previous studies related to glycolysis was conducted, followed by a comprehensive search in the GeneCards database (https://www.genecards.org/) (relevant score threshold: 4.0). Finally, a total of 69 GRGs were identified. The differential expression of GRGs in DN and NC samples was analyzed using R package "limma" [7]. The results were visualized using the R package "ggpubr" and the R package "pheatmap". The relationship representation of diverse glycolysis was created utilizing the R package "circlize" [8].

Consensus clustering and Glycolysis patterns in the training set

The R package "ConsensusClusterPlus" [9] was employed for cluster analysis of the DN sample training dataset based on GRGs expression levels. The ideal number of clusters was identified using a consensus matrix plot, consensus cumulative distribution function (CDF) plot, and trace plots. Principal component analysis (PCA) was used to visually describe the distribution of glycolysis-related patterns in a sample, concentrating on the first two principal components following the clustering process.

Gene set variation analysis (GSVA)

"GSVA" package in R is widely used for enrichment analysis to investigate biological functions and pathways across different clusters [10]. Two gene sets, "c5.go.symbols" and "c2.cp.Kegg.v7.2.symbols" were extracted from the Molecular Signature Database (https://www.
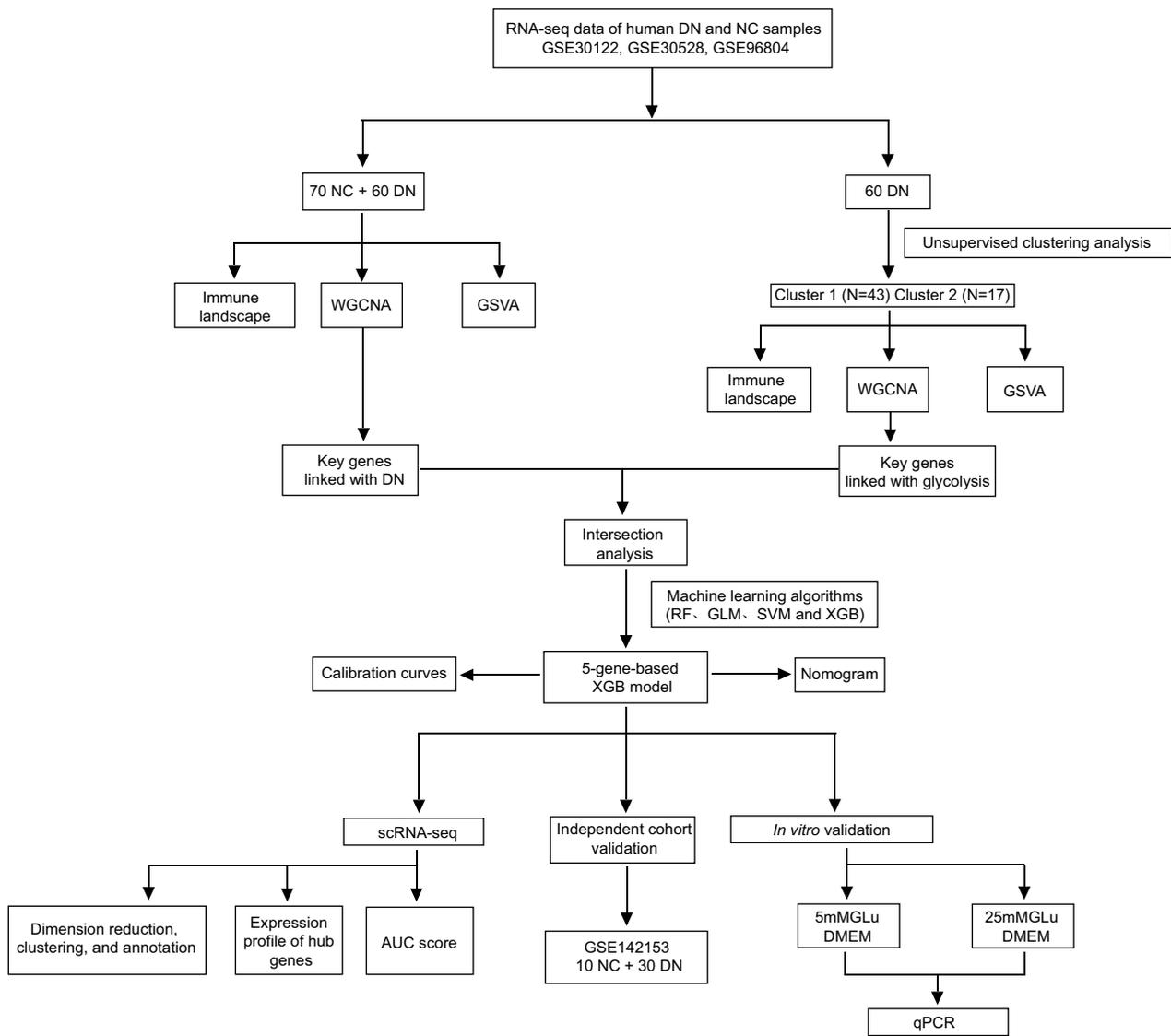
Fan *et al. Biology Direct*   (2025) 20:10

Page 3 of 16

Fig. 1 A flow chart of the study

gsea-msigdb.org/gsea/msigdb). The significant items (*P* < 0.05) determined by the student T-test were represented using a barplot, with orange and cyan colors representing up-regulated and down-regulated pathways, respectively.

Identification of key genes and their relationship with disease traits through WGCNA analysis of gene modules

The relationship between gene modules and disease traits was studied using the WGCNA algorithm to identify the key genes closely related to DN. The process included extracting 25% of genes with the highest variation rate from GSE30122, GSE30528, and GSE96804 datasets, hierarchical clustering of DN samples. Pearson correlation coefficient was used to support the establishment of a similarity matrix. This matrix was then converted to an adjacency matrix and topological overlay matrix using an appropriate soft threshold power. The genes were grouped into modules using a dynamic tree-cutting algorithm, identifying hub genes with gene significance (GS) ˃0.2 and module membership (MM) ˃0.6. The threshold for minimum module size was established at 100 genes. Each module was assigned a random color. Every module's eigengene profile represented global gene expression.

Fan *et al. Biology Direct*   (2025) 20:10

Page 4 of 16

Immune cell infiltration analysis

The "CIBERSORT" R package is widely used to estimate the proportion of specific cell types by leveraging a reference gene expression signature matrix. This package relies on established gene expression profiles to provide insights into the cellular composition of samples. A unique machine learning method that resists noise, "CIBERSORT", utilizes a specific selection process that adaptively chooses genes from a defined matrix using linear support vector regression, which requires an input matrix of transcriptome data [11]. This approach allows for the effective deconvolution of a given mixture. An overall P value for the deconvolution process is calculated based on empirical determination [12]. In this study, CIBERSORT was used to determine the relative composition of 22 immune cells according to their expression profiles. The relative composition of these immune cell types in different groups and their correlation with glycolysis were also examined. The"ggplot2" and "ggpubr" package were used for visualization.

Establishment and validation of the diagnostic model for DN using various machine learning algorithms

Cross-profiling of genes in the most important modules using WGCNA led to the identification of key genes with significant potential for diagnosing specific aspects of glycolysis. To determine the significance of these genes, we utilized four separate machine learning algorithms. For this purpose, the R package "kernlab," "randomForest," and "xgboost" were employed. In subsequent analyses, we selected the disease phenotype as the dependent variable and the gene set identified by WGCNA as the primary variable. Next, we utilized the model of machine learning algorithms to construct the "caret" R package. Subsequently, exploratory analysis was conducted on the model using the "DALEX" R package explanation function. The plot function was utilized to generate a cumulative residual distribution map and a residual box plot to facilitate the determination of the best diagnostic model. The performance of the model was evaluated based on the "pROC" R package. The analysis was refined by identifying the five most crucial features of the model. In addition, validation of the diagnostic model was performed on another validation dataset.

Single-cell RNA sequencing

To investigate single-cell characteristics, the R package "Seurat" was utilized to preprocess and analyze the scRNA-seq dataset GSE183276. Cells were rigorously filtered, excluding those with fewer than 400 genes, more than 5000 total genes, or more than 30% mitochondrial genes. Data that met these criteria were analyzed using the R package "harmony" to mitigate batch effects

between samples. Subsequently, cell cluster annotation was performed based on previous research, and visualized using Uniform Manifold Approximation and Projection (UMAP). The R package "AUCell" was applied to score the feasibility of the diagnostic model gene set of a single cell.

Cell culture and treatment

The HK2 cell was purchased from Whelab (ShangHai, China, Cat.No:C1116) and cultured with DMEM-normal glucose (Gibco, China, Cat.No:11885084) and DMEM-high glucose (Gibco, China, Cat.No:11965118) at 37 °C and 5% CO_2. The medium was changed daily. When the cell confluence reached 80–90%, the cells were passed in a 1:2 ratio by using 0.25% trypsin–EDTA (Gibco, China, REF:C25200-072). The cells were inoculated in a 6-well plate at a density of 5×10^5 cells per milliliter and attached to the wall overnight. The cells were inoculated to 90% confluent and treated with medium DMEM-high glucose (25 mMD-glucose) and DMEM-normal glucose (5.5 mMD-glucose) for 24 h.

qRT-PCR

Total RNA was extracted from treated HK-2 cells using the HiPureUniversal RNA Kit (Magen, Shanghai, China, REF:R4130-02). The purity and concentration of the RNA was determined and then reverse transcribed to cDNA using the PrimeScript™ RT kit (Takara, Dalian, China, REF:RR092A). It was then processed using the Start-up reagent: PowerUp™SYBR™Green Master MIX (Thermo Fisher, USA, REF:A25742). Finally, PCR was conducted on the LightCycler®96 instrument (Roche Diagnostics Gmbh, Switzerland). The β-Actin primer pairs was used as the internal control. The primer sequences used are shown in Additional file 1: Table S1.

Statistical analysis

The non-parametric Wilcoxon test was used to compare two sets of data with smaller sample sizes, while the student t-test was used for normally distributed data. Spearman correlation test was used to show the correlation. Statistical analyses were conducted using R software 4.3.3 and Graphpad Prism 10.1.2., $P < 0.05$ was considered statistically significant.

Results

Identification of multiple glycolysis expression patterns in DN

First, 69 GRGs were obtained from public databases. The roles of the 69 GRGs in the glycolysis pathway are listed in Additional file 2: Table S2. Gene expression data from 60 DN samples and 70 NC samples from the GSE30122, GSE30528, and GSE96804 datasets were

Fan *et al. Biology Direct* (2025) 20:10

Page 5 of 16

examined to investigate the expression patterns of these GRGs in DN. Batch effects were removed from the training sets using "SVA" R package. The PCA cluster plot was used to visualize the effects of removing inter-batch differences (Additional file 3: Figure S1A-B). The distribution of the three datasets on PC1 and PC2 was closer after batch correction, indicating that the batch effect was effectively eliminated, thus improving comparability between the data. A total of 32 of 69 GRGs showed differential expression in DN samples. Specifically, 11 significantly differentially expressed genes with *P < 0.001* were selected, and results showed that the expression levels of *PFKL, MPC1, PC, PKLR, ALDOB,*

FBP1, and PCK1 genes were significantly reduced in DN samples, while the expression levels of *PFKP, TPP2, HIF1A, and TP53* genes were elevated (Fig. 2A). The differential expression of GRGs in tissue samples of DN and NC individuals was visually represented in Heatmaps (Fig. 2B). In addition, the chromosomal locations of these Glycolysis genes are shown in Fig. 2C. Pro-inflammatory cytokine release and systemic and local low-grade inflammation (primarily due to innate immune system-driven inflammation) are associated with the onset and progression of DN [13]. The risk of developing DN is associated with systemic and local activation of inflammatory processes. Some studies
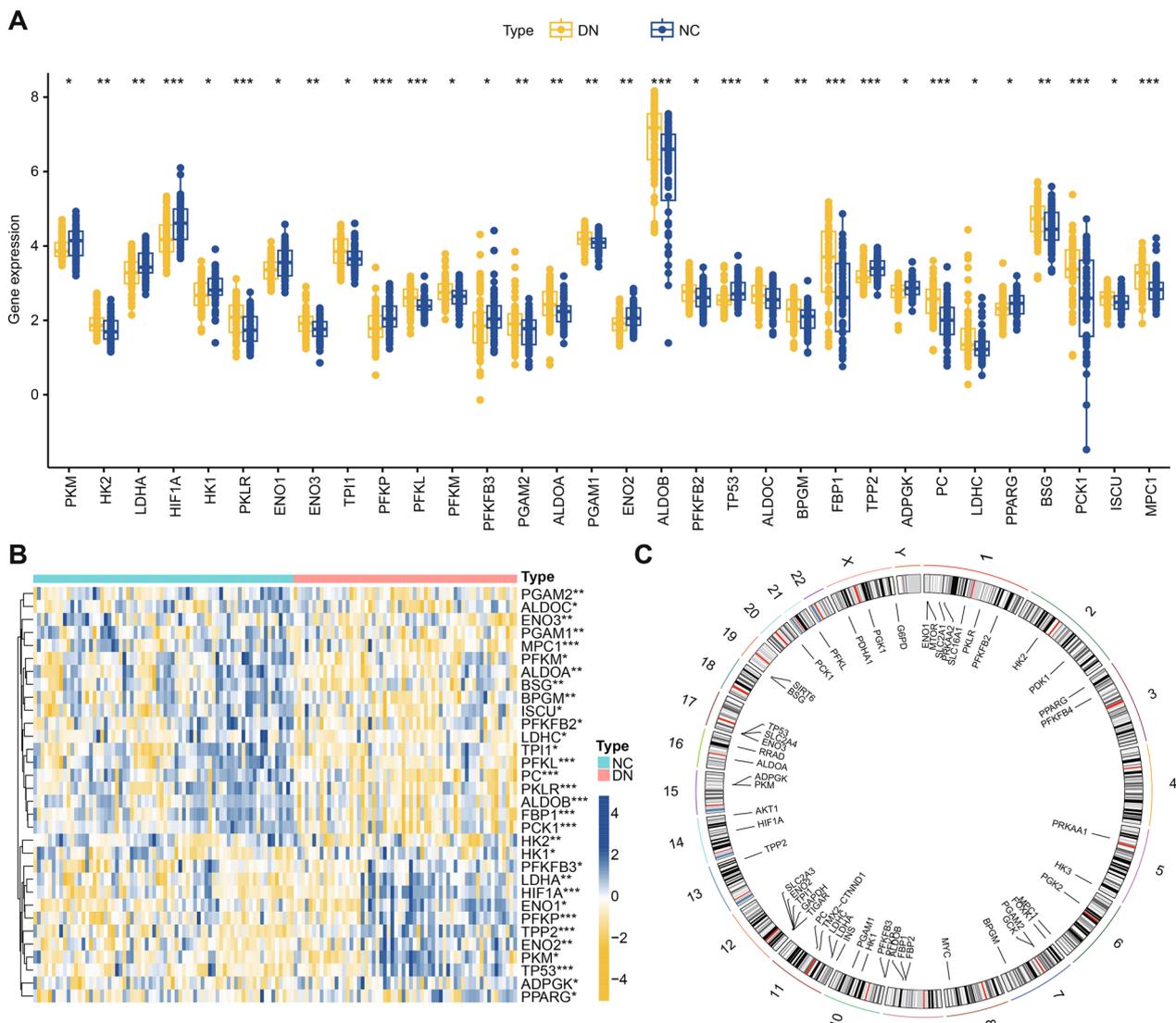

Fig. 2 The mode of GRGs expression in DN. **A** Box plot showing the differential expression of 32 GRGs between NC and DN samples. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. **B** The relative expression calorigrams of 32 differentially expressed GRGs, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. **C** The chromosomal locations of 32 differentially expressed GRGs

Fan *et al. Biology Direct*   (2025) 20:10

Page 6 of 16

have shown that macrophages, T cells, B cells, ILC2, and other cells participate in DN pathogenesis[14], suggesting that immune cells are potential therapeutic targets. In this study, a relative analysis of immune cell abundance in GSE30122, GSE30528, and GSE96804 samples was detected using CIBERSORT algorithm, combined with visualization using a heatmap (Additional file 4: Fig. S2A). Correlation analysis (Additional file 4: Fig. S2B) showed that GRGs were strongly associated with different populations of immune cells in the local environment. Additionally, we performed a correlation analysis between the relative abundance of immune cells and the expression of differential genes. The results indicate a significant association between GRGs and distinct immune cell subpopulations in the local environment, with a notable correlation between M2 macrophages and GRGs (Additional file 4: Fig. S2C). This concurrent occurrence of glycolysis and immune cell subpopulations in the immune microenvironment suggests a potential link between glycolysis and the development of DN. Moreover, the distribution of various infiltrating immune cells varied across the cohort, highlighting the complex interplay between glycolysis and immune responses in DN.

These findings underscore the role of GRGs in DN development and their potential impact on the immune microenvironment.

Unsupervised cluster analysis and machine learning algorithm for analysis of differential expression of glycolysis genes in DN samples

Sixty DN samples were selected from the training datasets to investigate the different expression patterns of GRGs in DN. Consensus clustering methods showed two different clusters based on the consensus matrix graph (k = 2), indicating clear differences (Fig. 3A). The minimum fluctuation of the consensus CDF curve at different consensus indices confirmed the stability of the cluster (Fig. 3B). The trace plot also confirmed the cluster's stability (Fig. 3C). Additionally, each cluster had a consistency score ﹥ 0.8 when k = 2 (Fig. 3D). As a result, the 60 DN samples were categorized into two distinct clusters: Cluster 1 (C1), 43 samples) and Cluster 2 (C2), 17 samples). PCA delineated these clusters (Fig. 3E).

A systematic analysis was conducted to fully understand the molecular characteristics of different glycolysis clusters. Eleven significantly different genes were selected (P < 0.001, Fig. 2A). Differential expression of multiple
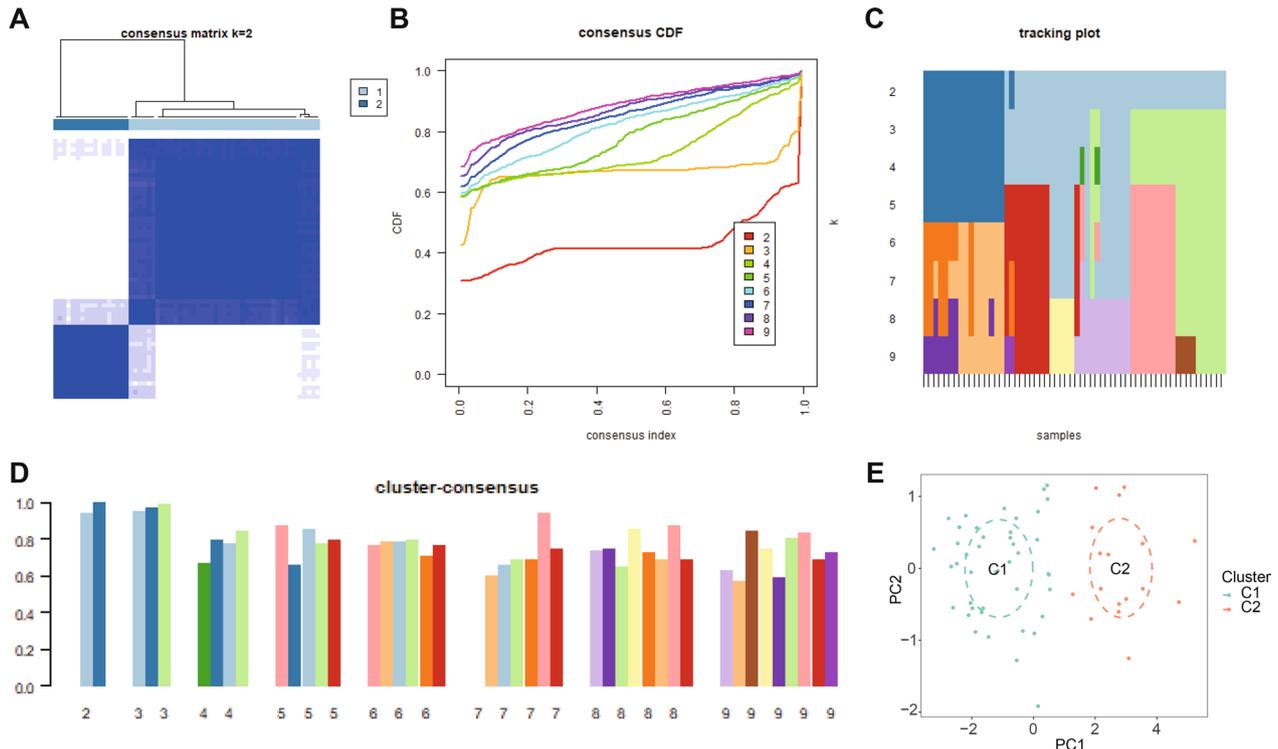

Fig. 3 Cluster analysis of differentially expressed GRGs in DN samples. **A** When k = 2, the sample was divided into 2 distinct clusters. **B** Consensus clustering CDF when k = 2 ~ 9. **C** A tracer showing the clustering results for each sample at different k values. (2–9) **D** Calculate a consistent clustering score when the value of k varies systematically from 2 to 9. (**E**) PCA analysis visually illustrates the distribution of two identified unsupervised consensus clusters of glycolytic clusters

GRGs was observed between C1 and C2, with 9 out of the 11 glycolysis genes showing differential expression (Fig. 4A). Heatmap was used to describe the relative expression patterns of these 11 glycolytic genes in DN samples (Fig. 4B). In addition, GSVA highlighted the upregulation of metabolic-related pathway and carbohydrate metabolism pathways in C2, including alanine aspartate and glutamate metabolism, peroxisome, citrate cycle, and PPAR signaling pathways. In contrast, C1 was enriched in immune signaling pathways, such as RIG-i-like receptor signaling pathways (Fig. 4C). The CIBERSORT algorithm was used to estimate the proportion of infiltrating immune cells in the two clusters. The relative abundance of the immune cells was expressed using barplot (Additional file 5: Fig. S3A), while different types of infiltrating immune cells were expressed using the box plot (Additional file 5: Fig. S3B). The relative abundance was significantly different between the two infiltrating immune cell types. This comprehensive analysis provided detailed insights into the differences between the two glycolysis clusters, further improving the understanding of the underlying mechanisms.

Identification of key genes related to DN and glycolysis using WGCNA

WGCNA algorithm was used to identify key genes associated with DN. The scale-free network was established after selecting the top 25% of the variance genes and removing abnormal samples in the GSE30122, GSE30528, and GSE96804 datasets. The soft threshold and the scale-free R^2 value were 12 and 0.85, respectively (Fig. 5A). Four distinct co-expression modules were identified (Fig. 5B). Notably, the brown module had the highest correlation with DN (r = 0.49) and significant *P*-value (P = 3e^{-09}) (Fig. 5C). Further
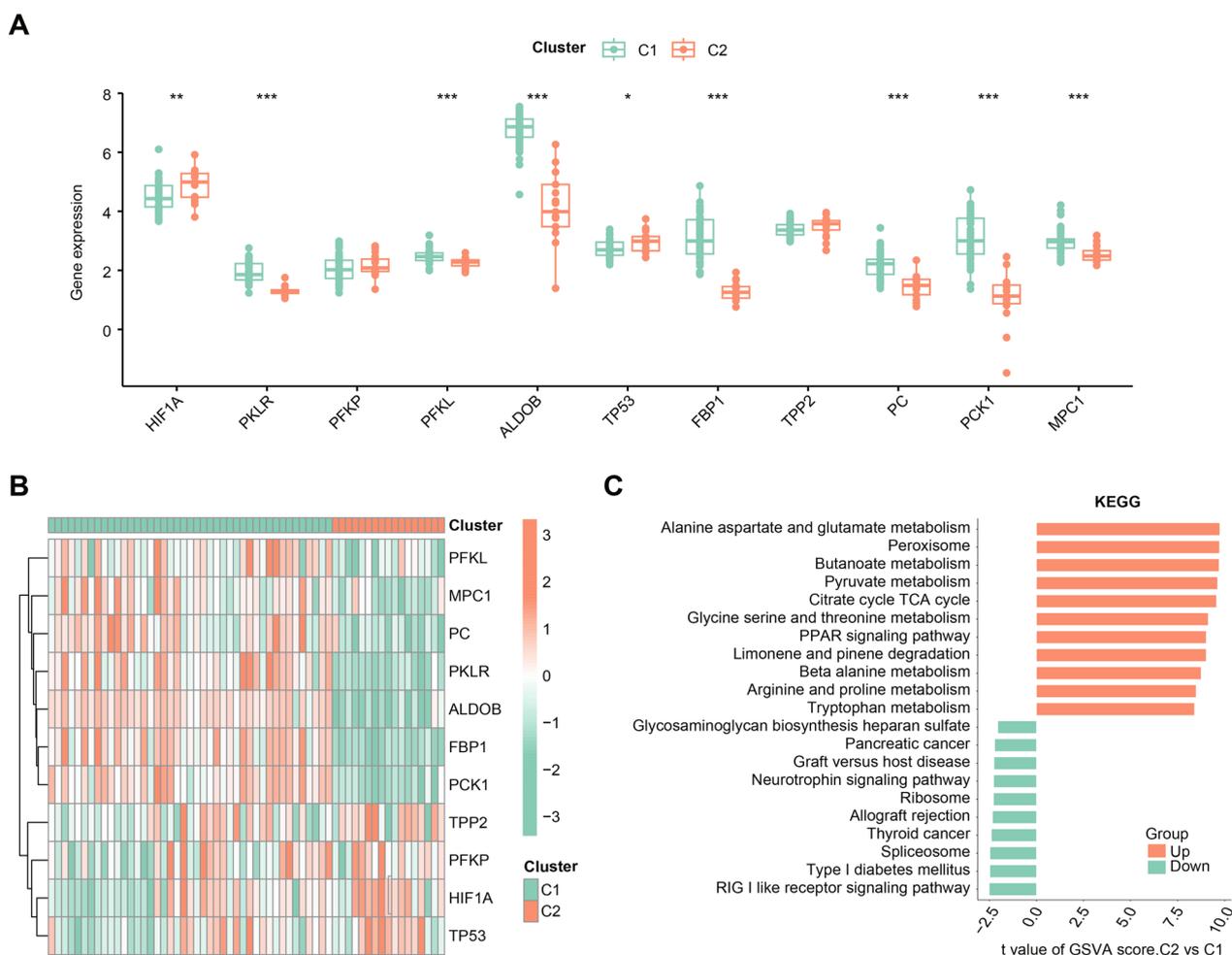

Fig. 4 Differences in expression patterns of GRGs in two unsupervised consensus clusters. **A** Box plots displaying GRGs with differential expression between two glycolytic groups. **B** Heat maps showing the relative expression levels of 11 CRGs in glycolytic clusters C1 and C2. **C** GSVA enrichment analysis based on the HALLMARK pathway among samples of glycolytic clusters C1 and C2, sorted by T-value. *P < 0.05, **P < 0.01, ***P < 0.001
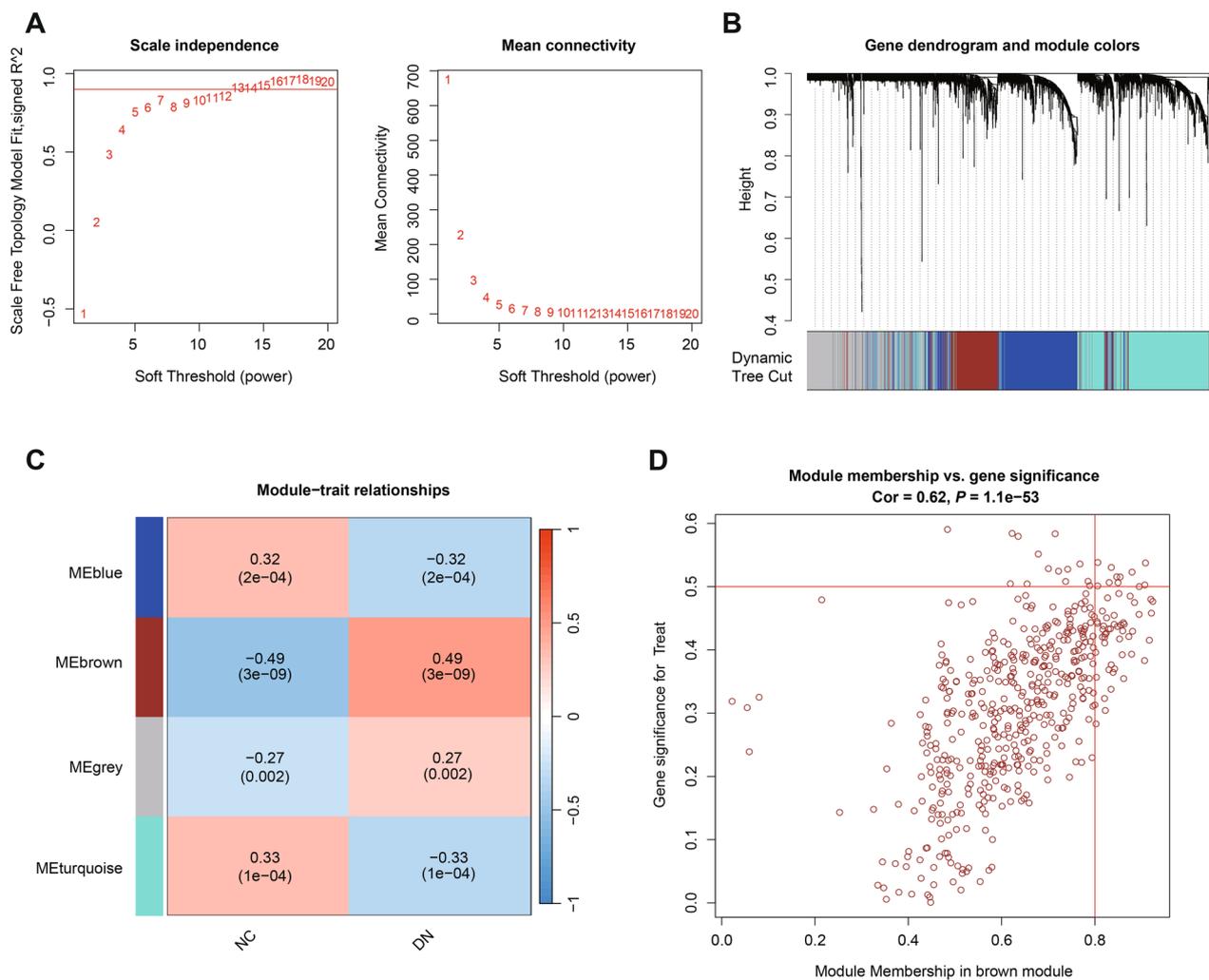
Fig. 5 Construction and module analysis of WGCNA. **A** Network topology analysis under different soft threshold powers. **B** Clustering Dendrogram, illustrating the hierarchical grouping of genes by topological overlap, with the specified module colors representing different gene clusters. **C** Correlation analysis for the relationship between different coexpression modules and clinical features. **D** Correlation between brown module members and DN

examination of the 282 genes in the module demonstrated a significant positive correlation (Fig. 5D).

Important genes related to glycolysis clusters in DN patients within the GSE30122, GSE30528, and GSE96804 dataset were identified using WGCNA algorithm. A scale-free network was constructed using a soft threshold $\beta = 8$ and the R^2 value of 0.86 (Additional file 6: Fig. S4A-B). Similarly, turquoise module showed the strongest correlation ($r = 0.69$) with glycolysis clusters and significant *P*-values ($P = 8e^{-10}$) (Additional file 6: Fig. S4C). Further analysis of 482 genes in the module showed a significant correlation (Additional file 6: Fig. S4D).

Cross-analysis of key genes obtained through WGCNA showed that there were 261 shared genes

related to DN patients and NC individuals, as well as module-related genes in glycolysis clusters (Additional file 7: Fig. S5A). Further Gene Ontology (GO) functional enrichment analysis revealed the main role of shared genes in regulating oxidative stress signaling pathways and metabolic responses (Additional file 7: Fig. S5B). Similarly, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway signal enrichment analysis highlighted that the shared genes were enriched in carbon metabolism, fatty acid degradation, and tryptophan metabolism (Additional file 7: Fig. 5C).

Fan *et al. Biology Direct*   (2025) 20:10

Page 9 of 16

Construction of DN diagnostic model and external cohort validation 5 gene-diagnostic model via multiple machine learning methods

About 70% of the samples from GSE30122, GSE30528, and GSE96804 datasets were used to screen out the hub genes that can diagnose DN from 261 shared genes obtained from WGCNA operation. A diagnostic model was then built using four machine-learning methods (RF, SVM, XGB, and GLM). The analysis of the cumulative residual distribution maps (Fig. 6A) and the residual boxplots (Fig. 6B) for the four algorithms showed that the residual values for both XGB and RF were small. The top ten variables for each model are displayed in Fig. 6C, ranked according to their root-mean-square error. The diagnostic performance of the models was assessed using receiver operating characteristic (ROC) curves, focusing on the remaining 30% of the samples from GSE30122, GSE30528, and GSE96804 (Fig. 6D). The models showed excellent recognition abilities, and the area under the curve (AUC) was more than 0.97. However, XGB was considered to be the best diagnostic model for DN based on its predictive power and reliability. The five most vital variables in the model, *GATM, PCBD1, F11, HRSP12,* and *G6PC*, were identified as hub genes for further analysis. The predictive potential of the 5-gene diagnostic model was evaluated using the independent validation dataset GSE142153. ROC analysis showed a high AUC value of 0.722 (Fig. 6E).

Moreover, the nomogram was constructed based on the five hub genes: *GATM*, *PCBD1*, *F11*, *HRSP12*, and *G6PC* (Additional file 8: Fig. S6A). The correction plot indicated that the model's predictions closely matched the actual outcomes, demonstrating strong predictive accuracy (Additional file 8: Fig. S6B). Additionally, the DCA supported the model's usefulness in clinical decision-making (Additional file 8: Fig. S6C), highlighting its relevance and practicality in real-world scenarios.

GRGs are associated with PT cells in human DN model

scRNA-seq analysis of the diabetic kidney of the human model was conducted to better characterize the relationship between glycolysis and DN at the single-cell level. Gene expression profiles of 9235 cells from the NC sample and 27,929 cells from DN samples were obtained after data screening and integration as described in the Methods (Fig. 7A). Thirteen types of cell clusters were annotated and visualized, including proximal tubule cells (PT), thick ascending branch cells (LOH), and distal convoluted tubule cells (DCT) (Fig. 7B). A bar chart (Fig. 7C) was used to show the different cell composition ratios in DN and NC samples, suggesting differences in PT, LOH, and other cell types. Moreover, the DN samples had significantly lower proportions of PT than

the NC samples. Interestingly, the hub genes (*GATM, PCBD1, F11, HRSP12, G6PC*) of 13 major cell types were verified. The results suggested that *GATM* and *PCBD1* were highly expressed in PT, and *PCBD1* was differentially expressed in LOH, DCT, and other cells (Fig. 7D-H). In addition, the expression of hub genes sets in the DN was evaluated using AUCell score (Fig. 7I-J). The results showed that the AUC score was highest in PT cells, indicating its specific expression in PT. In conclusion, cell composition and gene expression were significantly altered in DN samples, particularly a reduction in PT cells. The identification of hub genes, such as *GATM* and *PCBD1*, and their high expression in PT cells suggests that these genes may play a pivotal role in DN progression. Therefore, these genes offer potential targets for therapeutic intervention of DN in the future.

Construction of a cell model to verify the hub genes

A high-glucose-induced cell model was established to further evaluate the expression of the hub genes in HK-2 cells. Briefly, HK-2 cells were treated with 25 mMD-glucose, and control HK-2 cells were treated with 5.5 mMD-glucose. The corresponding cells were collected after 24 h to extract cDNA. qRT-PCR was used to verify the differential expression of hub genes (*GATM, PCBD1, F11, HRSP12*) in the constructed high-sugar-induced HK2 cell model (Fig. 8). Notably, *G6PC* and part of the *F11* in the samples had a large Cq value, and its data did not have clear confidence, thus no comparison was added to the analysis. The results showed that the gene expressions of *GATM, PCBD1, F11,* and *HRSP12* in the DN group were significantly decreased compared with the NC group. This finding indicates that the hub genes play a pivotal role in GRGs-related function in DN. Nonetheless, further studies should explore the mechanism and mode of action of the hub genes.

Discussion

DN is a diabetes-related complication and a major cause of ESRD [15], affecting up to 40% of patients with type 1 and type 2 diabetes [16]. Biomarkers, such as urinary albumin and creatinine ratios, have improved early detection and monitoring of DN [17]. Nonetheless, the burden of the disease remains high, necessitating continued research and development of new treatment strategies to improve patient outcomes [15]. Besides, understanding the underlying mechanisms, determining the severity of DN, and developing targeted interventions are critical in improving patient outcomes. Therefore, addressing knowledge gaps and classifying DN subtypes are necessary priorities in both research and clinical practice. This study aimed to investigate the related factors and pathological mechanisms of DN. The incidence of DN is
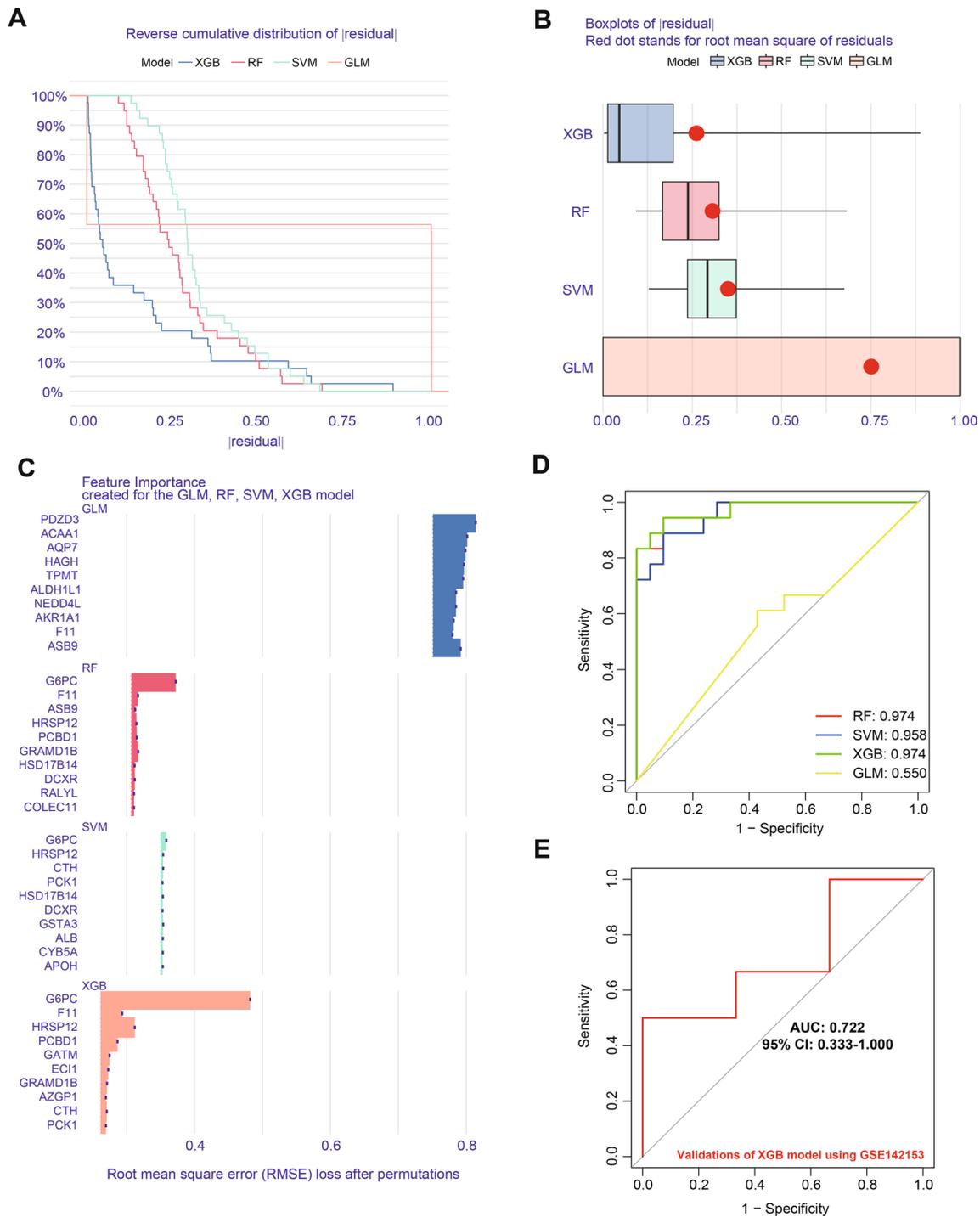
Fan *et al. Biology Direct*   (2025) 20:10

Page 10 of 16

Fig. 6 Residual and performance assessment of machine learning models on different feature sets. **A** Cumulative residual Distribution: The reverse cumulative distribution of residual for four machine learning models (XGB, RF, SVM, and GLM). The curve displays differences in accuracy of the diverse models in fitting the data. **B** Residual box plot: Comparison of the residual distribution of the four models. The red dots represent the residual root-mean-square error (RMSE) of each model. **C** The top ten variables in the RMSE ranking used for evaluating the feature importance of the models (GLM, RF, SVM, XGB), and the significance contribution of each model to the input features was analyzed. **D** The ROC curves for RF, SVM, XGB, and GLM models and their corresponding AUC values. **E** ROC curve and AUC values obtained by the XGB model were verified using the GSE142153 dataset
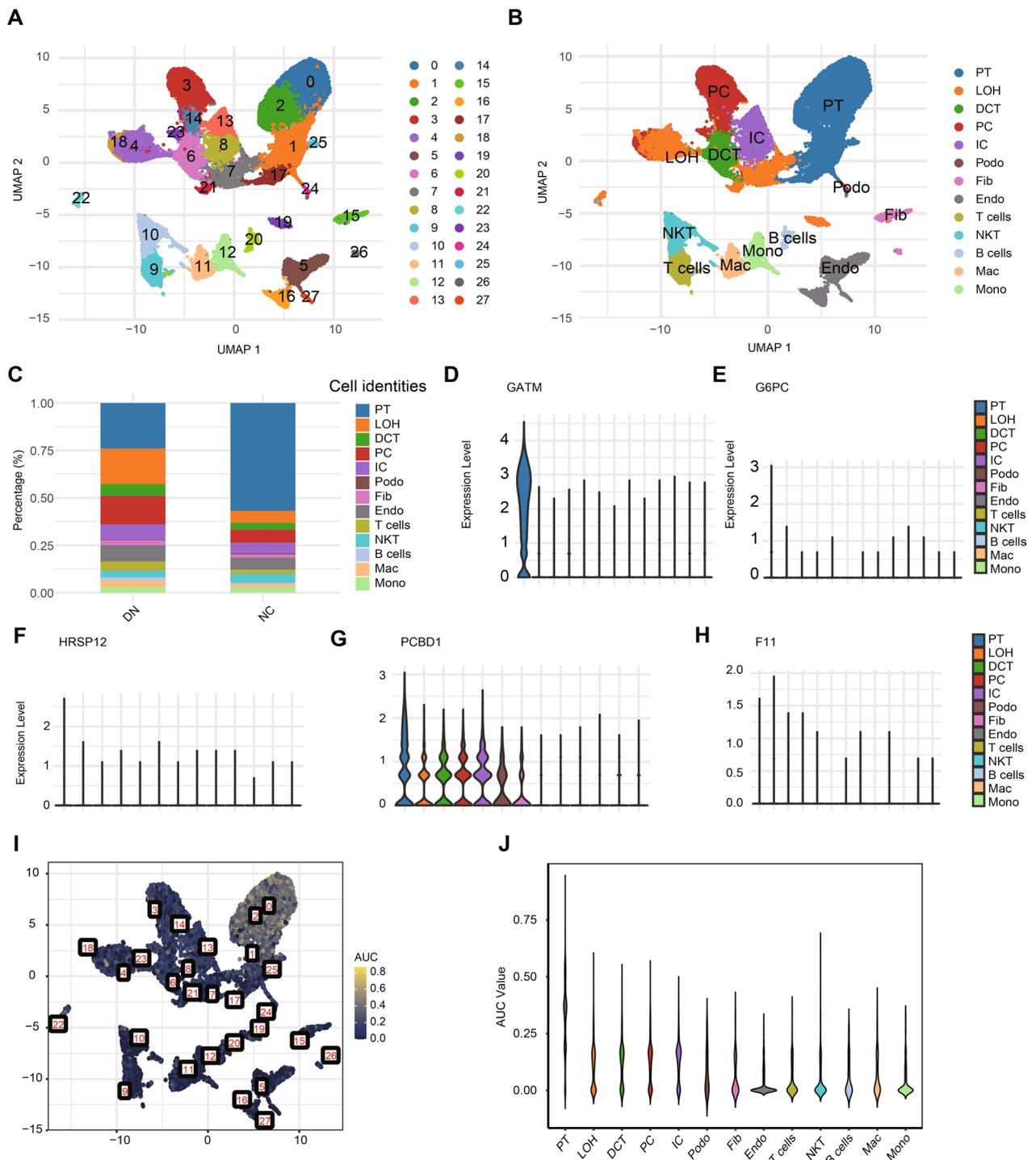
Fig. 7 Characterization of cell populations and gene expression patterns in DN and NC samples through scRNA-seq data. **A** UMAP displaying the cellular gene expression profiles of NC and DN samples in the dataset GSE183276. **B**. Annotations and visualizations illustrating the cell clusters based on the expression profiles. **C** Histogram showing the proportion of major cell types in DN and NC samples. **D** Differential expression of *GATM* in 13 cell clusters (E) Differential expression of *G6PC* in 13 cell clusters (**F**) Differential expression of *HRSP12* in 13 cell clusters (**G**) Differential expression of *PCBD1* in 13 cell clusters (**H**) Differential expression of *F11* in 13 cell clusters (**I**) illustrating the spatial coordinate system, the regional distribution of different cell clusters and their corresponding AUC values. **J** Violin plot showing the distribution density of AUC values by cell type
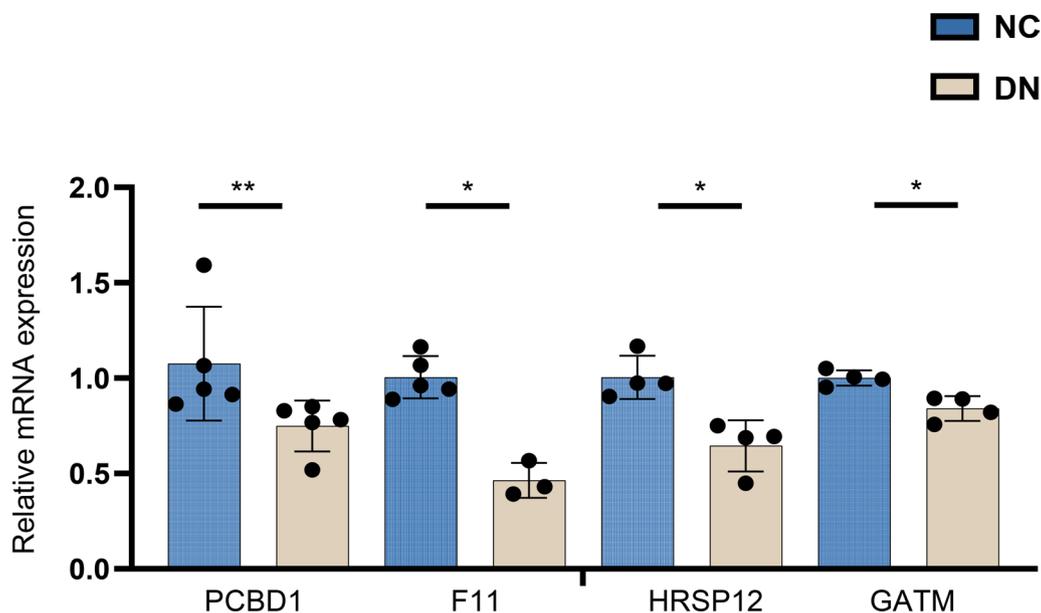
Fan *et al. Biology Direct*   (2025) 20:10

Page 12 of 16

Fig. 8 Validation of hub genes in in vitro hyperglycemic cell models: qRT-PCR validation of *GATM*, *PCBD1*, *F11*, *HRSP12*, and *G6PC* expression between DN patients and NC individuals. *$P < 0.05$, **$P < 0.01$

significantly increased in patients with diabetes. Besides, DN is significantly associated with multiple clinical biomarkers, such as blood glucose levels, blood pressure, and serum creatinine [18, 19]. In addition, a key gap in literature was assessed to conduct a comprehensive and systematic exploration of glycolysis between DN patients and healthy individuals.

This research uncovers notable discrepancies in the expression patterns of glycolysis within the DN domain, confirming a profound interaction between glycolysis and DN pathogenesis. Interestingly, a different immune landscape was revealed in the DN microenvironment, emphasizing the diverse subtypes of macrophage and T cells. These discoveries show a significant propensity for heterogeneity compared with NC individuals and highlight the close involvement of immune cells in DN progression. Specifically, relative T cell abundance, especially T cells CD4 memory resting [20], T cells gamma delta [21], M2 macrophages [22]. Monocytes [23] and neutrophils [14] were more abundant in NC individuals than in DN samples. Both in vitro and in vivo studies have shown that chronic hyperglycemia increases the polarization of M2 macrophages [24]. M1 macrophages produce large amounts of pro-inflammatory cytokines iNOS, TNF-α, MCP-1, and other pro-inflammatory mediators that amplify inflammation, resulting in further damages during DN pathogenesis. M2 macrophages, on the other hand, suppress kidney inflammation and reduce damage by secreting anti-inflammatory cytokines such as IL-10 and Arg-1 [25]. Therefore, regulation of

M1/M2 macrophage phenotypes has anti-proteinuria and renal protective effects on DN progression [26–28]. You et al. suggested that renal tissue can be better protected by clearing macrophages from mice in DN models molded using streptozotocin (STZ) [29]. In this study, two distinct glycolytic clusters were identified in DN patients using consensus clustering methods, revealing a unique innate immune environment, particularly involving T cells. Moon et al. also showed that activated T cells are associated with abnormal diabetic kidney damage and hyperglycemia in a mouse model of STZ [30]. Taken together, these observations suggest that multiple immune cells in the microenvironment promote highly complex interactions, with innate and adaptive immunity playing a conductive role. This phenomenon creates a connection between the atypical immune response within the immune microenvironment and the clinical manifestations of DN.

Machine learning-based biological image analysis is promising in the field of nephrology, including diagnosis of kidney pathology. As a result, it is considered the ultimate standard for identifying kidney disease. This diagnostic approach directly affects the range of treatment options and patient outcomes [31]. Diabetes is the leading cause of kidney failure in the Western Hemisphere [32–34]. The initial clinical sign of DN is usually the presence of microalbuminuria, defined as excretion ≥ 30 mg/ day or 20 μg/min. However, kidney biopsy studies have shown that microalbuminuria is not a complete indicator of type 2 DN because only

Fan *et al. Biology Direct* (2025) 20:10

Page 13 of 16

20–40% of patients progress to significant kidney disease without targeted therapy. In contrast, about 20% of patients with type 2 diabetes maintain normal urinary albumin levels when they progress to stage 3 CKD, characterized by a GFR of less than 60 mL/min/1.73 m^2 [32]. Therefore, new non-invasive biomarkers that can more accurately detect the early stages of DN and predict progression to kidney damage or kidney failure are necessary. Previous studies had various limitations, such as small sample size, cohort size, and fewer learning algorithms [35]. Therefore, a new approach is needed to break through these constraints that hinder research goals. Genetic diagnostic models provide new insights for the prediction of multiple diseases in the clinic [36, 37], We combined the expression matrix of DN patients in the training sets, extracted the characteristic gene expression of all DN samples as the original expression matrix for consensus clustering, determined the optimal cluster number to enhance the stability of the model and avoid overfitting, and explored potential related GRGs. However, a powerful DN diagnostic model should be constructed according to different machine learning algorithms, such as XGB and Lasso Cox regression analysis [38], to provide sufficient diagnostic feasibility. In this study, a unique DN diagnostic model was built using four different machine-learning algorithms. IN our dataset, there may be nonlinear relationships or complex feature interactions that affect GLM performance, leading to poor feasibility of GLM model. Notably, XGB was the most reliable and precise algorithm, demonstrating superior predictive capability compared with other algorithms. The diagnostic model based on five genes demonstrated exceptional recognition performance and stability when evaluated with an independently validated dataset. These findings confirm that the constructed model had good clinical value. Mutations in trexate 4α methylamine dehydrogenase 1 (*PCBD1*) cause hyperphenylalaninemia, hypomagnesemia, and diabetes. *PCBD1* is mainly expressed in the kidney and liver [39]. Silvia Ferre et al. noted that hepatic nuclear factor-1 β (HNF-1β, vHNF1) is a development-regulated transcription factor required for tissue-specific gene expression in epithelial cells of many organs, including the kidney. HNF-1β forms a heterotetrameric complex with dimeric cofactors of protein trexin-4α-methylamine dehydrogenase/hepatic nuclear factor 1 homeobox A (*PCBD 1* [MIM 126090]) [40]. HNF-1β-related disorder, Renal Cysts and Diabetes (RCAD; MIM:137,920) is syndrome characterized by autosomal dominant inheritance, renal cystic abnormalities, maturity-onset diabetes of the young type 5 (MODY5) [41, 42]. Notably, homozygous or complex heterozygous

PCBD1 mutations in humans are associated with MODY diabetes and renal Mg2 + consumption, with different penetrance. These findings are consistent with functional deficiencies of *PCBD1* as a cofactor of HNF-1β dimerization [43]. These findings indicate that *PCBD1* may participate in metabolic adaptation during DN. However, further studies are needed to clarify the potential relationship between *PCBD1* and DN pathogenesis. The gene *HRSP12* encodes human heat response protein, and its expression level is positively correlated with HbA1c, indicating that it may participate in renal stress response induced by hyperglycemia. The expression level of *HRSP12* in urinary extracellular vesicles can reflect the changes in renal function in diabetic patients. The DN candidate marker *HRSP12* has protective effects on cells or cell proteins under stress, such as *GPX3, HRSP12, MSRA, MSRB1,* and *CRYAB. GPX3, GPX1*, and *GPX4,* belonging to the *GPX* family, are involved in reducing oxidative stress damage in cells, indicating that they may play a role in protecting the kidneys by reducing kidney stress caused by hyperglycemia [44]. The *GATM* gene encodes L-arginine, glycine amidinotransferase [45]. *GATM* may be involved in creatinine production rather than renal function since it encodes glycine aminotransferase, an enzyme involved in creatine biosynthesis [46], which may be related to DN severity. The level of *GATM* is negatively correlated with the degree of type 2 diabetes. The PI3K-AKT and AMPK pathways may be potential targets for IR-related glucose metabolism regulation in type 2 diabetes and obese patients. Furthermore, the expression of *GATM* is down-regulated in the liver tissues of mouse models and may affect the liver AMPK pathway, induce glucose metabolism disorders, and further affect the development of type 2 diabetes [47]. The mRNA of *GATM* shows immune cell specificity [46], indicating the correlation between DN and immune cells. The glucose-6-phosphatase catalytic (*G6PC*) subunit is associated with the severity of complications in diabetic patients, indicating that *G6PC* may be involved in metabolic adaptation during DN. *G6PC* catalyzes the hydrolysis of glucose-6-phosphate to glucose, which is the final step in gluconeogenesis and glycogen degradation. Glucose catalyzed by *G6PC* leaves the liver through glucose transporter 2 [48]. *G6PC* plays a key role in maintaining normal blood glucose. *G6PC* gene is up-regulated in diabetic patients due to insulin tolerance or hypoinsulinemia [49]. However, further studies are needed to clarify the potential relationship between *G6PC* and the pathogenesis of the disease. The gene *F11* encodes coagulation factor XI, which is involved in the endogenous coagulation pathway of humans. *F11* is also expressed in Langerhans islands of pancreas and

Fan *et al. Biology Direct*   (2025) 20:10

Page 14 of 16

renal tubular cells in humans [50]. Several studies have shown that the blood system of diabetic patients is often in a hypercoagulable state. Blood clots are more likely to form in diabetic patients than in healthy people [51]. The thrombus activation in diabetic patients is stronger than that in normal people. Multiple plasma coagulation factors are increased in diabetic patients, further aggravating kidney injury [52]. Sun et al. showed that the level of renal function-related indicators is significantly higher in DN patients than in diabetic patients without kidney damage. DN patients also have shorter APTT than diabetic patients, suggesting that endogenous coagulation function is enhanced in DN patients [53]. In summary, *F11* is involved in the pathophysiological immune response regulation and oxidative stress pathway of DN. However, further research should unravel the complex mechanism by which glycolysis controls the pro-coagulant activity of bone marrow cells. In this study, a high coagulation state and low fibrinolytic activity were detected in adipose tissue macrophages isolated from HFD-fed mice, further supporting the key role of enhanced glycolytic activity in driving immune thrombotic activity in vivo. Several studies have extensively elucidated the specific involvement of *F11* in the pathophysiology of DN [54]. It should be noted that we used the entire GSE30122 dataset, which includes glomerular and tubule samples. DN is a systemic kidney disease that affects the entire nephron. While the glomeruli are usually the initial site of injury, pathological processes, such as metabolic disorders and altered signaling pathways, can extend to other compartments, including the tubules. Therefore, by including glomerular and tubule data, we aim to fully capture the transcriptional landscape of DN and its systemic effects, ensuring a more complete understanding of the disease. To address potential biases caused by differences in tissue origin, we performed rigorous batch effect correction using the sva software package. We further evaluated the effectiveness of this correction using PCA analysis (Additional file 3: Fig. S1A-B) to compare the data distribution before and after correction. At the same time, we observed significant differences in the proportion of PT cells in single-cell sequencing, with the five-gene diagnostic model having the highest PT activity score across cell types in the AUCell algorithm. Therefore, we propose the following hypothesis: The core genes in the diagnostic model are mainly expressed in PT cells. Because of the kidney mitochondrial dysfunction associated with diabetes, Increased transport through glucose transporters may result in PT cells requiring increased glycolytic flux to maintain nutrient flux in diabetic capacity [55].

Based on these findings, we selected the HK-2 cell line for experimental verification. This selection was not intended to show that DN was limited to tubules, but rather to focus on elucidating the mechanistic role of GRGs in specific and related cell types.

However, this study has some limitations. The potential associations between immune cells and the identified gene expression in the scRNA-seq data could be explored and additional datasets are needed to validate the robustness of the diagnostic model further. Integrating in vivo data is essential to fully understand the mechanism of action of hub genes in the pathophysiology of DN. Besides, Validation of cell line experiments should also add multiple experimental time conditions to better simulate chronic processes. Integrating various data sets and experimental data may improve the future development of DN research. In addition, the analysis should be stratified according to sex, early or late stages of DN.

Conclusion

Various glycolytic-related clusters were detected within diseased samples through consensus clustering, each characterized by distinct immune signatures. A diagnostic model was developed for DN utilizing the XGB algorithm, and five specific genes were identified. The model showed strong performance, accurately classifying samples from both qRT-PCR and independently validated datasets. In conclusion, these findings explain the understanding of DN heterogeneity and immune microenvironment and may provide a new diagnostic method for DN.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13062-025-00601-6.

Additional file 1. Supplementary table 1: Primer sequences used in qRT-PCR.

Additional file 2. Supplementary table 2: The role of 69 GRGs in the glycolysis pathway.

Additional file 3. Figure S1: Batch effect correction eliminates data bias from different datasets.

Additional file 4. Figure S2: Correlation analysis of immune cell relative abundance and GRGs between DN group and normal group.

Additional file 5. Figure S3: Analysis of relative abundance of immune cells in two glycolytic clusters.

Additional file 6. Figure S4: Construction and module analysis of glycolytic related WGCNA in DN group.

Additional file 7. Figure S5: The intersectional analysis of the two best modules obtained from WGCNA and their different forms of enrichment analysis.

Additional file 8. Figure S6: Comprehensive Evaluation of 5-gene diagnostic model: ROC Analysis, Calibration, and Decision Curve Analysis.

Fan *et al. Biology Direct*   (2025) 20:10

Page 15 of 16

Declarations

Ethics approval and consent to participate

In this study, we confirm that all experiments and methods were conducted strictly in accordance with relevant guidelines and regulations.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Urology, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, China. [2]Department of Urology, Guangxi Medical University Cancer Hospital, Nanning 530000, Guangxi, China.

References

1. Argyropoulos C, et al. Urinary MicroRNA profiling predicts the development of microalbuminuria in patients with type 1 diabetes. JCM. 2015;4(7):1498–517. https://doi.org/10.3390/jcm4071498.
2. Samsu N. Diabetic nephropathy: challenges in pathogenesis, diagnosis, and treatment. Biomed Res Int. 2021;2021:1–17. https://doi.org/10.1155/2021/1497449.
3. Zhang L, et al. Trends in chronic kidney disease in China. N Engl J Med. 2016;375(9):905–6. https://doi.org/10.1056/NEJMc1602469.
4. Duman TT, Ozkul FN, Balci B. Could systemic inflammatory index predict diabetic kidney injury in type 2 diabetes mellitus? Diagnostics. 2023;13(12):2063. https://doi.org/10.3390/diagnostics13122063.
5. Sum SLW, Shi Y. The glycolytic process in endothelial cells and its implications. Acta Pharmacol Sin. 2022;43(2):251–9. https://doi.org/10.1038/s41401-021-00647-y.
6. Hato T, et al. Novel application of complementary imaging techniques to examine in vivo glucose metabolism in the kidney. Am J Physiol-Renal Physiol. 2016;310(8):F717–25. https://doi.org/10.1152/ajprenal.00535.2015.
7. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47–e47. https://doi.org/10.1093/nar/gkv007.
8. Gu Z, Gu L, Eils R, Schlesner M, Brors B. *circlize* implements and enhances circular visualization in R. Bioinformatics. 2014;30(19):2811–2. https://doi.org/10.1093/bioinformatics/btu393.
9. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26(12):1572–3. https://doi.org/10.1093/bioinformatics/btq170.
10. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics. 2013;14(1):7. https://doi.org/10.1186/1471-2105-14-7.
11. Chen S, Sun Y, Zhu X, Mo Z. Prediction of survival outcome in lower-grade glioma using a prognostic signature with 33 immune-related gene Pairs. IJGM. 2021;14:8149–60. https://doi.org/10.2147/IJGM.S338135.
12. Newman AM, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453–7. https://doi.org/10.1038/nmeth.3337.
13. Wada J, Makino H. Innate immunity in diabetes and diabetic nephropathy. Nat Rev Nephrol. 2016;12(1):13–26. https://doi.org/10.1038/nrneph.2015.175.
14. Peng Q-Y, An Y, Jiang Z-Z, Xu Y. The role of immune cells in DKD: mechanisms and targeted therapies. JIR. 2024;17:2103–18. https://doi.org/10.2147/JIR.S457526.
15. Elendu C, et al. Comprehensive advancements in the prevention and treatment of diabetic nephropathy: a narrative review. Medicine. 2023;102(40):e35397. https://doi.org/10.1097/MD.0000000000035397.
16. Tuttle KR, et al. Diabetic kidney disease: a report from an ADA consensus conference. Am J Kidney Dis. 2014;64(4):510–33. https://doi.org/10.1053/j.ajkd.2014.08.001.
17. Colhoun HM, Marcovecchio ML. Biomarkers of diabetic kidney disease. Diabetologia. 2018;61(5):996–1011. https://doi.org/10.1007/s00125-018-4567-5.
18. Jha R, Lopez-Trevino S, Kankanamalage HR, Jha JC. Diabetes and renal complications: an overview on pathophysiology, biomarkers and therapeutic interventions. Biomedicines. 2024;12(5):1098. https://doi.org/10.3390/biomedicines12051098.
19. Li R-Y, Guo L. Exercise in diabetic nephropathy: protective effects and molecular mechanism. IJMS. 2024;25(7):3605. https://doi.org/10.3390/ijms25073605.
20. Liu Y, et al. T cells and their products in diabetic kidney disease. Front Immunol. 2023;14:1084448. https://doi.org/10.3389/fimmu.2023.1084448.
21. Lavoz C, et al. Interleukin-17A blockade reduces albuminuria and kidney injury in an accelerated model of diabetic nephropathy. Kidney Int. 2019;95(6):1418–32. https://doi.org/10.1016/j.kint.2018.12.031.
22. Yan J, Li X, Liu N, He JC, Zhong Y. Relationship between macrophages and tissue microenvironments in diabetic kidneys. Biomedicines. 2023;11(7):1889. https://doi.org/10.3390/biomedicines11071889.
23. Donate-Correa J, Martín-Núñez E, Muros-de-Fuentes M, Mora-Fernández C, Navarro-González JF. Inflammatory cytokines in diabetic nephropathy. J Diabetes Res. 2015;2015:1–9. https://doi.org/10.1155/2015/948417.
24. Torres-Arévalo Á, et al. a2bar antagonism decreases the glomerular expression and secretion of chemoattractants for monocytes and the pro-fibrotic M2 macrophages polarization during diabetic nephropathy. IJMS. 2023;24(13):10829. https://doi.org/10.3390/ijms241310829.
25. Liu J, et al. Hyperoside suppresses renal inflammation by regulating macrophage polarization in mice with type 2 diabetes mellitus. Front Immunol. 2021;12:733808. https://doi.org/10.3389/fimmu.2021.733808.
26. Yang H, et al. Tim-3 aggravates podocyte injury in diabetic nephropathy by promoting macrophage activation via the NF-κB/TNF-α pathway. Mol Metabol. 2019;23:24–36. https://doi.org/10.1016/j.molmet.2019.02.007.
27. Ji L, et al. Overexpression of Sirt6 promotes M2-macrophage transformation, alleviating renal injury in diabetic nephropathy. Int J Oncol. 2019. https://doi.org/10.3892/ijo.2019.4800.
28. Zhang X-L, Guo Y-F, Song Z-X, Zhou M. Vitamin D prevents podocyte injury via regulation of macrophage M1/M2 phenotype in diabetic nephropathy rats. Endocrinology. 2014;155(12):4939–50. https://doi.org/10.1210/en.2014-1020.
29. You H, Gao T, Cooper TK, Reeves WB, Awad AS. Macrophages directly mediate diabetic renal injury. Am J Physiol Renal Physiol. 2013;305(12):1719–27. https://doi.org/10.1152/ajprenal.00141.2013.
30. Moon J-Y, Jeong K-H, Lee T-W, Ihm C-G, Lim SJ, Lee S-H. Aberrant recruitment and activation of T cells in diabetic nephropathy. Am J Nephrol. 2012;35(2):164–74. https://doi.org/10.1159/000334928.
31. Delrue C, De Bruyne S, Speeckaert MM. Application of machine learning in chronic kidney disease: current status and future prospects. Biomedicines. 2024;12(3):568. https://doi.org/10.3390/biomedicines12030568.

Fan *et al. Biology Direct*   (2025) 20:10

Page 16 of 16

32. American Diabetes Association, Nephropathy in Diabetes, *Diabetes Care*, vol. 27, no. suppl_1, pp. s79–s83, 2004, https://doi.org/10.2337/diacare.27.2007.S79.

33. MacIsaac RJ, Jerums G. Diabetic kidney disease with and without albuminuria. Curr Opini Nephrol Hypertension. 2011;20(3):246–57. https://doi.org/10.1097/MNH.0b013e3283456546.

34. Spasovski G, Ortiz A, Vanholder R, El Nahas M. Proteomics in chronic kidney disease: the issues clinical nephrologists need an answer for. Proteom Clini Apps. 2011;5(5–6):233–40. https://doi.org/10.1002/prca.201000150.

35. Huang J, Chen J, Wang C, Lai L, Mi H, Chen S. Deciphering the molecular classification of pediatric sepsis: integrating WGCNA and machine learning-based classification with immune signatures for the development of an advanced diagnostic model. Front Genet. 2024;15:1294381. https://doi.org/10.3389/fgene.2024.1294381.

36. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. εD octor: machine learning and the future of medicine. J Intern Med. 2018;284(6):603–19. https://doi.org/10.1111/joim.12822.

37. Bihorac A, et al. MySurgeryRisk: development and validation of a machine-learning risk algorithm for major complications and death after surgery. Ann Surg. 2019;269(4):652–62. https://doi.org/10.1097/SLA.0000000000002706.

38. Huang J, et al. Integrative analysis of gene expression profiles of substantia nigra identifies potential diagnosis biomarkers in Parkinson's disease. Sci Rep. 2024;14(1):2167. https://doi.org/10.1038/s41598-024-52276-0.

39. Tholen LE, et al. Bifunctional protein PCBD2 operates as a co-factor for hepatocyte nuclear factor 1β and modulates gene transcription. The FASEB J. 2021;35(4):63. https://doi.org/10.1096/fj.202002022R.

40. Ferrè S, et al. Mutations in PCBD1 cause hypomagnesemia and renal magnesium wasting. J Am Soc Nephrol. 2014;25(3):574–86. https://doi.org/10.1681/ASN.2013040337.

41. Heidet L, et al. Spectrum of HNF1B mutations in a large cohort of patients who harbor renal diseases. Clin J Am Soc Nephrol. 2010;5(6):1079–90. https://doi.org/10.2215/CJN.06810909.

42. Ulinski T, et al. Renal phenotypes related to hepatocyte nuclear factor-1β (TCF2) mutations in a pediatric cohort. J Am Soc Nephrol. 2006;17(2):497–503. https://doi.org/10.1681/ASN.2005101040.

43. Ferrè S, Igarashi P. New insights into the role of HNF-1β in kidney (patho) physiology. Pediatr Nephrol. 2019;34(8):1325–35. https://doi.org/10.1007/s00467-018-3990-7.

44. Dwivedi OP, et al. Genome-wide mRNA profiling in urinary extracellular vesicles reveals stress gene signature for diabetic kidney disease. iScience. 2023;26(5):106686. https://doi.org/10.1016/j.isci.2023.106686.

45. Miyamoto T, Sengoku K, Hayashi H, Sasaki Y, Jinno Y, Ishikawa M. GATM, the human ortholog of the mouse imprinted Gatm gene, escapes genomic imprinting in placenta. Genet Mol Biol. 2005;28(1):44–5. https://doi.org/10.1590/S1415-47572005000100008.

46. Si S, Liu H, Xu L, Zhan S. Identification of novel therapeutic targets for chronic kidney disease and kidney function by integrating multi-omics proteome with transcriptome. Genome Med. 2024;16(1):84. https://doi.org/10.1186/s13073-024-01356-x.

47. Zhang Y, Han D, Yu P, Huang Q, Ge P. Genome-scale transcriptional analysis reveals key genes associated with the development of type II diabetes in mice. Exp Ther Med. 2017;13(3):1044–150. https://doi.org/10.3892/etm.2017.4042.

48. Van Schaftingen E, Gerin I. The glucose-6-phosphatase system. Biochem J. 2002;362(3):513–32. https://doi.org/10.1042/bj3620513.

49. Van De Werve G, Lange A, Newgard C, Méchin M, Li Y, Berteloot A. New lessons in the regulation of glucose metabolism taught by the glucose 6-phosphatase system. Eur J Biochem. 2000;267(6):1533–49. https://doi.org/10.1046/j.1432-1327.2000.01160.x.

50. Mohammed BM, et al. An update on factor XI structure and function. Thromb Res. 2018;161:94–105. https://doi.org/10.1016/j.thromres.2017.10.008.

51. Wada J, Makino H. Inflammation and the pathogenesis of diabetic nephropathy. Clin Sci. 2013;124(3):139–52. https://doi.org/10.1042/CS20120198.

52. Patrassi GM, Martinelli S, Picchinenna A, Girolami A. Contact phase of coagulation in diabetes mellitus after aspirin administration. Folia Haematol. 1985;112(2):333–8.

53. Sun J, Liu C. Correlation of vascular endothelial function and coagulation factors with renal function and inflammatory factors in patients with diabetic nephropathy. Exp Ther Med. 2018. https://doi.org/10.3892/etm.2018.6718.

54. Rehill AM, et al. Glycolytic reprogramming fuels myeloid cell-driven hypercoagulability. J Thromb Haemost. 2024;22(2):394–409. https://doi.org/10.1016/j.jtha.2023.10.006.

55. Sas KM, et al. Tissue-specific metabolic reprogramming drives nutrient flux in diabetic complications. JCI Insight. 2016;1(15):63. https://doi.org/10.1172/jci.insight.86976.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.