



OPEN Construction of a diagnostic model utilizing m7G regulatory factors for the characterization of diabetic nephropathy and the immune microenvironment

Jingying Zhong^{1,4}, Pengli Xu^{1,4}, Xuanyi Li¹, Meng Wang¹, Xuejun Chen¹, Huiyu Liang¹, Zedong Chen¹, Jing Yuan^{2,3} & Ya Xiao¹✉

Diabetic nephropathy (DN), a prevalent and severe complication of diabetes, is associated with poor prognosis and limited treatment options. N7-Methylguanosine (m7G) modification plays a crucial role in regulating RNA structure and function, linking it closely to metabolic disorders. However, despite its biological significance, the interplay between m7G methylation and immune status in DN remains largely unexplored. Leveraging data from the GEO database, we conducted consensus clustering of m7G regulators in DN patients to identify distinct molecular subtypes. To construct and validate m7G-related prognostic features and risk scores, we integrated multiple machine learning approaches, including Support Vector Machine-Recursive Feature Elimination, Random Forest, LASSO, Cox regression, and ROC curves analysis. In addition, we employed GSVA, ssGSEA, CIBERSORT, and Gene Set Enrichment Analysis to investigate the associated biological pathways and the immune landscape, providing deeper insights into the role of m7G methylation in DN. Based on the expression levels of 18 m7G-related regulatory factors, we identified nine key regulators. Through machine learning techniques, we identified four significant regulators (METTL1, CYFIP2, EIF3D, and NUDT4). Consensus clustering classified these genes into two distinct m7G-related clusters. To characterize these subtypes, we conducted immune infiltration analysis, differential expression analysis, and enrichment analysis, uncovering significant biological differences between the clusters. Additionally, we developed an m7G-related risk scoring model using the PCA algorithm. The differential expression of the four key regulators was further validated through *in vivo* experiments, reinforcing their potential role in disease progression. The m7G-related genes METTL1, CYFIP2, EIF3D, and NUDT4 may serve as potential diagnostic biomarkers for DN, providing new insights into its molecular mechanisms and immune landscape.

Keywords Diabetic nephropathy, N7-Methylguanosine (m7G) modification, Biomarkers, Scoring model

Diabetic nephropathy (DN), also known as diabetic kidney disease, is a severe chronic microvascular complication of diabetes. It affects approximately 20–40% of diabetic patients¹, with some studies suggesting that up to 50% may develop kidney disease over time². Early manifestations of DN include an increased glomerular filtration rate (GFR) and microalbuminuria, which, if left unchecked, progress to persistent albuminuria, nodular glomerular lesions, and ultimately end-stage renal disease (ESRD)³. From 1990 to 2019, the incidence of DN rose by 156.5%, accompanied by a significant increase in mortality rates⁴. With the global prevalence of diabetes continuing to escalate, the incidence of DN is expected to rise correspondingly. By 2045, the global diabetic population is projected to reach 700 million, with nearly 40% at risk of developing DN⁵. DN is a major clinical and public health concern, imposing significant burdens on global economies and healthcare systems⁶.

¹School of Traditional Chinese Medicine, Jinan University, 601 West Huangpu Avenue, Guangzhou 510632, China.

²School of Traditional Chinese Medicine, Southern Medical University, Guangzhou, China. ³Nanfeng Hospital, Southern Medical University, Guangzhou, China. ⁴These authors contributed equally: Jingying Zhong and Pengli Xu. ✉email: xiaoya0527@126.com

The diagnosis of diabetic nephropathy (DN) primarily relies on clinical presentation and laboratory findings. However, these indicators often emerge only in the advanced stages of the disease. Traditionally, microalbuminuria has been regarded as a primary marker for DN diagnosis. Yet, growing evidence suggests that many DN patients with microalbuminuria may revert to normal urine levels, while only a minority progress to proteinuria. Moreover, approximately one-third of DN patients with normal urinary protein levels experience a gradual decline in renal function, including glomerular filtration rate (GFR). This underscores that relying solely on proteinuria assessment may be inadequate for monitoring DN onset and progression⁷. Furthermore, a decline in GFR without the presence of microalbuminuria has been associated with tubular damage⁸. Several studies suggest that tubular and interstitial damage may occur in the early stages of DN and significantly contribute to disease progression. Hence, investigating biomarkers indicative of tubular damage in diabetes can facilitate the early detection of kidney structural and functional impairments, enhance DN monitoring, and improve prognosis. Timely detection and intervention are essential for slowing the progression of DN^{9,10}.

RNA modification is a critical biochemical mechanism that alters the properties and functions of RNA molecules by introducing or chemically modifying specific functional groups¹¹. Among these modifications, RNA methylation is one of the most widespread, encompassing diverse chemical alterations, with adenosine methylation being particularly prominent. Adenosine methylation occurs in various RNA molecules, including mRNA and non-coding RNAs such as circular RNA (circRNAs), microRNA (miRNAs), and long non-coding RNA (lncRNAs)¹². RNA methylation includes various forms, such as m6A, m5C, and m7G, which collectively regulate gene expression and cellular functions, playing a crucial role in preserving cellular homeostasis and responding to environmental changes^{13,14}. N7-methylguanosine (m7G) is a methylated RNA modification predominantly found in tRNA, specific rRNA subunits, and the 5' cap of mRNA. It plays a pivotal role in RNA stability, transcription, translation, and splicing¹⁵. In recent years, substantial research has focused on the prognostic value of RNA modifications in DN^{16–18}. However, research on m7G modification in DN remains limited. Previous studies have indicated that abnormal m7G RNA modification is associated with the progression of various pathological processes, including tumors and neurodegenerative diseases. However, its role in diabetes mellitus, particularly its influence on immune responses in DN, remains largely unexplored. Notably, m7G cap modification is frequently observed in mRNAs associated with DN and is involved in the regulation of their translation¹⁹. Given its critical role in transcriptional and translational regulation, m7G holds significant potential as a biomarker for DN and may serve as a foundation for developing novel diagnostic models.

With the advancement of DN research to the genetic and epigenetic level, bioinformatics tools have become essential for identifying prognostic and predictive biomarkers²⁰. However, a major challenge in data processing lies in managing high-dimensional features and data redundancy. To address this, machine learning algorithms are increasingly being employed for feature selection and model development, enabling more efficient exploration of DN-related genetic markers and epigenetic modifications. For instance, Ma et al. first applied Best Subset Regression (BSR) analysis to identify feature subsets, then compared the coefficients of LASSO, Ridge Regression, and Elastic Net Regression to select predictive models. They subsequently developed an RF prediction model incorporating all m7G regulatory factors and ultimately identified diagnostic biomarkers for heart failure by integrating results from BSR analysis, Ridge Regression, and the RF algorithm. Building on these methodologies, the integration of machine learning is expected to drive advancements in bioinformatics-based data mining and analysis of DN-related m7G modifications. This approach can effectively identify key genes and epigenetic markers associated with DN, offering novel insights into disease diagnosis and treatment.

Here, utilizing seven publicly available DN microarray datasets—GSE47184²¹, GSE96804^{22,23}, GSE104948²⁴, GSE104954²⁴, GSE30528²⁵, GSE99325²⁶, and GSE99339²⁶—we performed differential expression analysis and algorithmic screening to identify key genes at the intersection of m7G regulatory factors. Through this approach, we identified four key regulatory factors—METTL1, CYFIP2, EIF3D, and NUDT4—and further applied machine learning techniques to analyze their relevance in DN. Based on the expression profiles of m7G regulatory factors, we clustered DN samples into m7G-associated subgroups. Subsequently, we conducted Principal Component Analysis (PCA) to compute the m7G scores for these two subgroups. Through differential expression analysis, enrichment analysis, and immune infiltration profiling, we characterized these two subgroups and identified two distinct subtypes of m7G-modified DN with unique immune profiles and biological functions. Furthermore, we validated these four regulators through *in vivo* experiments.

Materials and methods

Obtainment and preprocessing of DN datasets

Gene expression data from DN patients were retrieved from the GEO database (<http://www.ncbi.nlm.nih.gov/geo>). We selected four datasets (GSE47184²¹, GSE96804^{22,23}, GSE104948²⁴, GSE104954²⁴) for training set, as they have been extensively used in studies on DN and involve transcriptome analysis of renal tissue. GSE30528²⁵, GSE99325²⁶ and GSE99339²⁶ were used as the validation set. Details of the datasets used in our study were presented in Table 1. Each dataset underwent background correction using the ComBat function from the “sva” R package (version 3.42.0) to eliminate batch effects. Subsequently, the data were merged and normalized, resulting in a dataset comprising 66 samples from DN patients and 60 from normal human kidney tissues.

Recent analyses of the GSEA database and previously published literature have identified 30 genes associated with m7G, including gene sets such as GOMF_RNA_CAP_BINDING (SNUPN, EIF4A1, EIF4E, NCBP2, CYFIP1, LARP1, IFIT5, EIF4E1B, GEMIN5, CYFIP2, AGO2, LSM1, DCPs, EIF4E3, NCBP2L, NCBP1, NCBP3, EIF3D, EIF4G3, EIF4E2), GOMF_M7G_5_PPPN_DIPHOSPHATASE_ACTIVITY (NUDT4, NUDT3, NUDT16, NUDT10, NUDT4B, NUDT1, NUDT11), and specific genes like METTL1, WDR4, and NSUN2 documented in current literature²⁷. These are considered commonly associated with m7G modification.

Dataset	Accession	Samples		Tissue
		NM	DN	
Training set	GSE47184	4	11	Kidney
	GSE96804	20	41	Kidney
	GSE104948	18	7	Kidney
	GSE104954	18	7	Kidney
Verification set	GSE30528	13	9	Kidney
	GSE99325	7	18	Kidney
	GSE99339	8	7	Kidney

Table 1. The information of the datasets. *NM* normal, *DN* diabetic nephropathy.

Identification of differentially expressed m7G-related genes

Differential expression genes (DEGs) were identified using the “limma” R package (version 3.52.4)²⁸, with significance set at $P < 0.05$. Among them, only 18 regulators were annotated in our dataset. We explored the association between the expression of these 18 genes and diseases, as well as the self-expression of m7G regulatory factors. The findings were depicted using a heatmap, and important regulatory factors were selected using the Wilcoxon signed-rank test.

Correlation analysis

Spearman’s rank correlation coefficient (method = “spearman”) was used to calculate the expression correlation of 9 m7G-related DEGs across all samples and DN samples, employing the cor.test function from the base R “stats” package. The correlations were visualized using the “corrplot” R package.

Protein–protein interaction (PPI) network analysis

Genes were imported into the Search Tool for the Retrieval of Interacting Genes (STRING, <http://www.string-db.org>) to construct PPI networks, using interactions with a combined score > 0.5 . The network was visualized with Cytoscape (version 3.7.2), and genes were ranked according to their degree centrality values using the cytoHubba plugin.

Functional enrichment analysis

Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed to evaluate the functions of 9 differentially expressed genes across distinct clusters^{29–33}. GO analysis encompassed enriched biological processes (BPs), molecular functions (MFs), and cellular components (CCs). Enrichment analysis results with adjusted $P < 0.05$ were considered statistically significant, and the top 20 findings from each analysis were visualized using the “cluster Profiler (3.18.1)” R package.

Construction of prediction models

To identify m7G regulatory factors associated with DN occurrence, we employed three machine learning techniques: Support Vector Machine Recursive Feature Elimination (SVM-RFE), Random Forest (RF), and Least Absolute Shrinkage and Selection Operator (LASSO) regression. The detailed procedures are described below.

Support vector machine recursive feature elimination (SVM-RFE)

SVM-RFE is a feature selection algorithm that iteratively removes features with minimal contribution to classification performance³⁴. The expression matrix of m7G regulatory factors, along with their classification labels (DN and normal groups), was used to train an SVM model for feature importance ranking. In each iteration, the least important features were removed, followed by model retraining. This process continued until all features were ranked, after which the top-ranked genes were selected for further analysis.

Random forest (RF)

RF is an ensemble learning algorithm that constructs multiple decision trees to evaluate feature importance³⁴. The expression matrix was divided into training and test sets, and the RF model was trained on the training set to classify samples as DN or normal. Gene importance was assessed using metrics such as Mean Decrease in Accuracy (MDA) and the Gini Index. Genes with higher importance scores were selected for further analysis.

Feature intersection

To improve the robustness of feature selection, the top-ranked features from SVM-RFE and RF were analyzed using a Venn diagram. Only the overlapping genes identified by both methods were selected as core candidates for further analysis.

LASSO regression analysis

LASSO regression was applied to the intersecting gene set to identify key predictive features and develop a diagnostic model³⁵. Cross-validation was performed using the “glmnet” R package (version 4.1-4) to determine the optimal regularization parameter (λ), ensuring a balance between model complexity and performance.

Features with non-zero coefficients at the optimal λ were selected, and the diagnostic score was computed using the formula:

$$\text{Diagnostic Score} = \sum (\text{Coefficient} \times \text{Gene Expression Level}).$$

Model performance evaluation and validation

Internal validation and dataset partitioning

For internal validation, the dataset was randomly partitioned into training and validation sets at a 1:1 ratio. The training set was used to train the classification model, while the validation set was employed to evaluate model performance.

Confusion matrices and classification performance metrics

To assess classification performance, a logistic regression model (`glm()`, from the `stats` package) was applied, generating predictions based on a classification threshold of 0.5. Confusion matrices for the training and validation sets were computed using the `confusionMatrix()` function from the `caret` package (version 6.0–94). Heatmaps of the confusion matrices were generated using the `ggplot2` package (version 3.4.0).

Key classification performance metrics were derived from the confusion matrices, including:

- $\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN).$
- $\text{Sensitivity} = TP / (TP + FN).$
- $\text{Specificity} = TN / (TN + FP).$
- $\text{F1-Score} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}).$
- $\text{Precision} = TP / (TP + FP).$

All metrics were calculated using the `confusionMatrix()` function from the `caret` package. A bar plot comparing classification performance between the training and validation sets was created using the `ggplot2` package.

Receiver operating characteristic (ROC) analysis

The model's diagnostic performance was further evaluated through ROC curve analysis using the `pROC` R package (version 1.18.0). The Area Under the Curve (AUC) was computed to quantify classification accuracy.

External validation

To assess the model's generalizability, external validation was performed on three independent datasets (GSE30528²⁵ GSE99325²⁶ GSE99339²⁶) using the same classification model and evaluation metrics.

Risk score calculation

Expression matrices were processed to retain only genes relevant to the model. A logistic regression model (`glm()` from the `stats` package) was then used to compute risk scores, ensuring consistency with internal validation.

Performance metrics

Predictions were made using a classification threshold of 0.5, and confusion matrices were constructed with the `caret` package. Accuracy, sensitivity, specificity, and F1-score were computed for each dataset to evaluate classification performance.

Decision curve analysis (DCA)

The `rmda` package (version 1.6) was employed to assess the model's clinical utility. Decision curves were constructed for risk scores, nomogram predictions, and clinical factors using `decision_curve()`.

Calibration curve and C-index calculation

The `rms` package (version 6.7-0) was employed to generate a calibration curve, evaluating the agreement between predicted and actual probabilities. The concordance index (C-index) was calculated using `rcorrcons()` to quantify predictive accuracy.

Clustering and principal component analysis (PCA)

To investigate gene expression patterns, consensus clustering was conducted using the `ConsensusClusterPlus` package (version 1.62.0), categorizing samples into two distinct groups. Principal Component Analysis (PCA) was performed using the `FactoMineR` (version 2.8) and `factoextra` (version 1.0.7) packages to extract PC1 and PC2 from selected genes.

The m7G score was calculated as: $\text{m7G Score} = \sum (\text{PC1}_i + \text{PC2}_i).$

These analyses offered insights into the biological significance of identified gene signatures and ensured robust model validation across multiple datasets.

m7G-related gene expression in kidney diseases

Public gene expression datasets were obtained from GEO, including GSE73953³⁶ (systemic lupus erythematosus, SLE), GSE93798³⁷ (IgA nephropathy, IgAN), GSE37455³⁸ (hypertensive nephropathy, HN), and GSE99325²⁶ and GSE99339²⁶ (diabetic nephropathy, DN). Data preprocessing involved normalization using the `limma` package in R, with batch effect correction via `ComBat` when necessary. Differential expression analysis of METTL1, CYFIP2, EIF3D, and NUDT4 was conducted using the Wilcoxon rank-sum test or Student's t-test, depending on data distribution. Statistical significance was set at $P < 0.05$, with $P < 0.001$ considered highly significant. Violin

plots were generated using ggplot2 to visualize gene expression across disease groups and controls. Significant differences in gene expression were observed, suggesting a potential role for these genes in disease pathogenesis.

Gene set enrichment analysis (GSEA)

Enrichment analysis was performed using GSEA software (GSEA_4.2.3) on Java version 8.0 platform to investigate differentially activated biological pathways between control and DN groups. The merged dataset included 126 samples categorized into two groups: 60 control samples and 66 DN samples. The gene set used for analysis, c2.cp.kegg.v7.5.symbols.gmt, was obtained from the GSEA official website (<http://www.gsea-msigdb.org/gsea/index.jsp>)³⁹. Enrichment scores (ES) were calculated with 1,000 permutations, and gene sets ranged from 5 to 500. A normalized P-value < 0.05 and false discovery rate (FDR) < 0.25 were considered statistically significant thresholds.

Gene set variation analysis (GSVA)

GSVA analysis, one of the GSEA algorithms, could explore the difference in biological pathways between distinct pattern clusters according to the enrichment score. “GSVA (1.16.0)” R Package was applied to perform functional enrichment analysis of DN disease samples in GSE96804 to obtain the enrichment pathways. We downloaded “c2.cp.kegg.v7.4.symbols.gmt” from MsigDB database to carry out the analysis. Adjusted P value < 0.05 was considered to suggest statistical difference between distinct clusters.

Immune infiltration

Genes associated with 22 types of immune cells were collected from literature sources. Next, we correlated the expression profiles of these immune genes with the distribution of 22 immune cell types using single-sample gene set enrichment analysis⁴⁰. Alongside our m7G regulators, we investigated the connections between the 22 immune cell types and different m7G-related clusters, including immune cells marked by m7G-associated genes. Employing the “IOBR” R package, we used the CIBERSORT algorithm to calculate scores for the 22 immune infiltrating cells in each sample. Violin plots were generated to visually depict the inferred distribution of these 22 cell types, demonstrating their correlations with immune cell infiltration. We conducted Spearman correlation analysis using the “cor.test” function to explore the relationship between m7G-related genes and immune cells.

Construction of diagnostic model of DN

To better explore the characteristics of our two m7G-related clusters, we conducted differential expression analysis, assessing DEGs between the two m7G-related clusters and between NM and DN samples using cutoff criterion $|\log_2FC| > 1$ ($P < 0.05$). Following the intersection of these two sets of DEGs, we built a diagnostic model using the overlapping DEGs identified through LASSO Cox regression.

$$\text{Diagnostic Score} = \sum I \text{ Coefficients} * \text{Expression level of signature}$$

Animals

Male C57BL/6JGpt mice, aged 4–6-weeks, were acquired from GemPharmatech (Nanjing, China) and housed in a specific pathogen-free (SPF) environment. The experimental conditions were maintained at a temperature of 21 ± 3 °C, 50% humidity, and a 12-hour light/dark cycle. The mice were acclimated to these conditions for one week prior to the experiment. All animal experiments were conducted in accordance with the National Institute of Health (NIH) Guidelines for the Care and Use of Laboratory Animals and were approved by the Animal Experiment Ethics Committee of Jinan University (Approval No. 20231102-18). The mice were randomly divided into two groups: normal diet group (control group, CR $n = 6$), the STZ group (model group, MD $n = 6$). Throughout the experiment, the STZ group received daily intraperitoneal injections of streptozotocin (STZ, Sigma Aldrich, St. Louis, USA) at a dosage of 50 mg/kg for five consecutive days. Streptozotocin was dissolved in a 1% citric acid buffer with a pH of 4.5. One week later, diabetes was confirmed by measuring blood glucose levels, with a level exceeding 13.0 mmol/L indicating successful model establishment. At the end of the experiment (week 8), mice were euthanized with an overdose of sodium pentobarbital (100 mg/kg) by intraperitoneal injection. Measurements included body weight, 24-hour urinary protein levels, and fasting blood glucose. Blood glucose was performed using a glucometer (Roche Diabetes Care GmbH, Basel, Switzerland). Mouse tissues were promptly frozen in liquid nitrogen for subsequent analysis. The study was reported in accordance with ARRIVE guidelines.

Histological analysis

Kidney specimens were fixed in a 4% paraformaldehyde solution for 1 to 2 h. After paraffin embedding, the samples were sectioned into 5 µm thick slices. The sections were stained with hematoxylin and eosin (H&E), periodic acid-Schiff (PAS), and Masson’s trichrome. High-resolution images were obtained using an Olympus microscope (Shinjuku, Japan). Image processing was performed using Image-Pro Plus software (version 6.0; Media Cybernetics, Bethesda, MD).

Urine biochemistry assays

Urinary albumin levels were determined using a mouse urinary albumin ELISA kit (Bethyl Laboratories, Montgomery, TX).

Renal function assessment

Renal function was assessed based on blood urea nitrogen (BUN) and serum creatinine (Scr) levels in mice. BUN and Scr levels were detected using corresponding commercial enzyme-linked immunosorbent assay (ELISA) kits (JBS-M0549, SenBeiJia Biological Technology, China; JL20633, JianglaiBio, China), respectively.

Western blot analysis

Radioimmunoprecipitation experiment lysis buffer comprising 50 mmol/L Tris-HCl, pH 7.5, 150 mmol/L NaCl, 0.5% deoxycholate, 1% Nonidet P-40, 0.1% sodium dodecyl sulfate, 1 mmol/L phenylmethylsulfonyl fluoride and 1 µg/mL protease cocktail was used to extract protein from cells and tissues. The protein content was determined using a bicinchoninic acid test kit. Protein samples (80 g/lane) were placed onto gels, separated using polyacrylamide gel electrophoresis with 10% sodium dodecyl sulfate, and then transferred to polyvinylidene difluoride membranes. The membranes were then treated with anti-METTL1, anti-CYFIP2, anti-EIF3D, anti-NUDT4, or anti-GAPDH antibodies at 4°C for an overnight period. A secondary antibody was applied to the membrane after 3 rounds of washing. Using improved chemiluminescence, certain signals were detected.

Immunohistochemical staining

The sections were washed 3 times with phosphate buffer saline and blocked in 10% goat serum for 30 min at room temperature after deparaffinization, rehydration and heat-induced epitope retrieval. Immunohistochemical staining was performed by incubating primary antibodies for CYFIP2 (Ag8749; proteintech), NUDT4 (Ag1346; proteintech), METTL1 (Ag6980; proteintech), and EIF3D (ER63335; HUABIO) overnight at 4°C. Subsequently, an HRP-DAB system (Proteintech, Wuhan, China) was used to detect the immunoactivity, followed by counterstaining with hematoxylin. Goat anti-rabbit immunoglobulin G (IgG; Proteintech, Wuhan, China) was used as the negative control. The sections were imaged under a light microscope. Twenty images under brightfield were randomly taken for per section in a blinded fashion (3 mice per group). Images quantified by Image-Pro Plus 6.0 analysis software.

Results

Identification of significant m7G regulatory factors in diabetic nephropathy

Figure 1 illustrated the study flow diagram. First, Spearman correlation analysis revealed significant correlations among several of these 18 regulatory factors based on their expression levels in the m7G dataset. Subsequently, this dataset was integrated with three GEO datasets for further analysis (Fig. 2A). Furthermore, a protein-protein interaction (PPI) network was constructed, identifying interactions among 16 of these 18 regulatory factors, excluding LSM1 and SNUPN (Fig. 2B). Next, the Wilcoxon signed-rank test identified nine significant regulatory factors ($P < 0.05$) in the training set: DCP2, IFIT5, EIF4E2, NUDT11, NUDT3, LARP1, SNUPN, LSM1, and CYFIP1 (Fig. 2C). A heatmap was generated to visualize the expression levels of these nine significant regulatory factors (Fig. 2D).

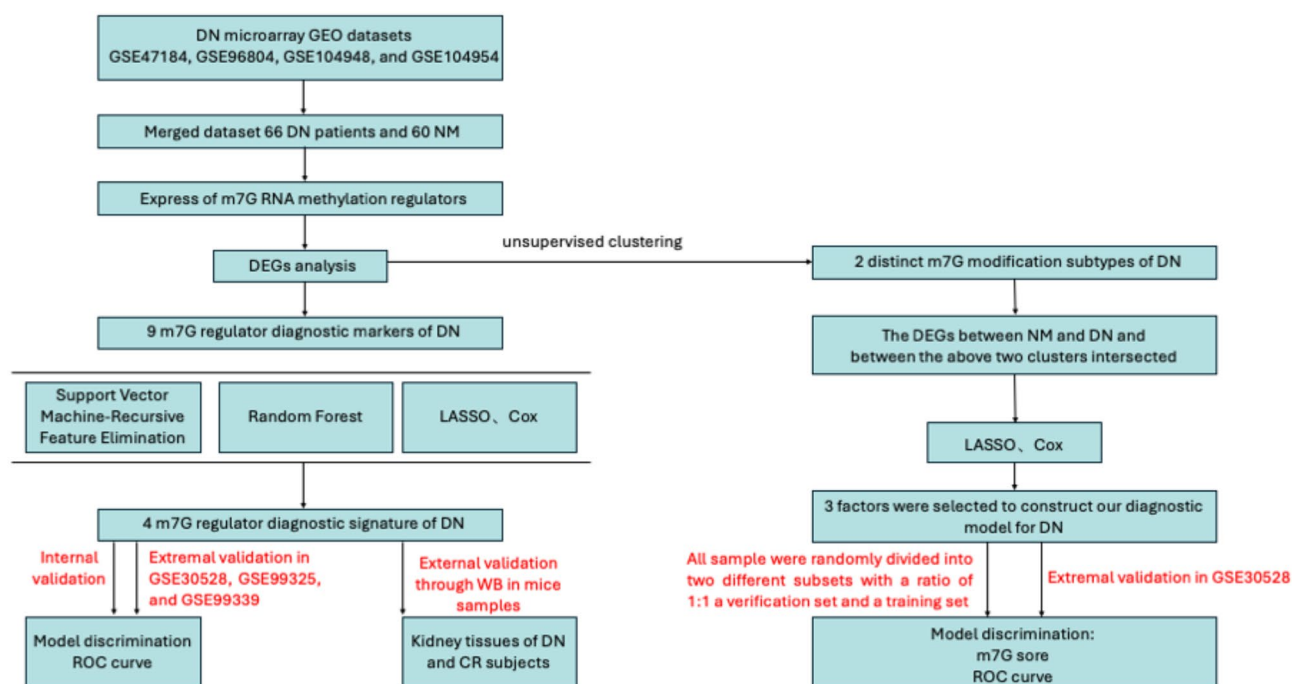


Fig. 1. Study flow diagram. NM normal, DN diabetic nephropathy, GEO Gnen expression omnibus, m7G N7-methylguanosine, DEG differentially expressed gene, CR normal, WB western blot, ROC receiver operating characteristic.

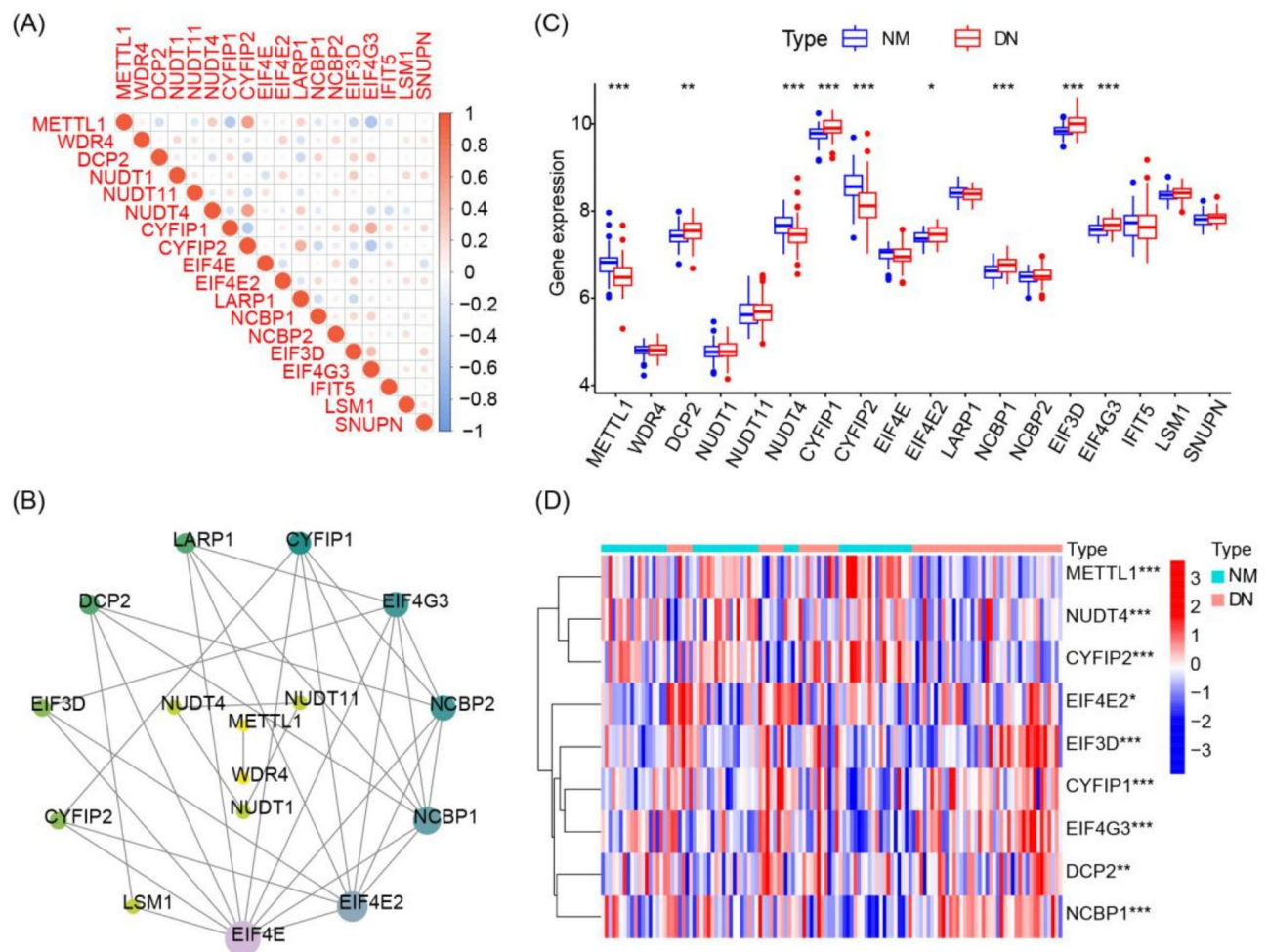


Fig. 2. The basic landscape of m7G-regulators in DN. (A) Bar plot of expression levels of 18 m7G regulatory factors. (B) PPI network of 18 m7G regulatory factors. (C) Expression differences of 18 m7G regulatory factors between NM and DN groups. (D) Heatmap of 9 key regulatory factors between NM and DN. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. NM, Normal; DN, diabetic nephropathy.

Enrichment analysis of significant m7G regulatory factors

GO and KEGG analyses were conducted to assess the reliability of the study findings ($P < 0.05$) (Fig. 3A, B). Enrichment analysis results from multiple approaches indicated that these regulatory factors are predominantly involved in RNA metabolism. Moreover, several immune-related pathways were identified (Fig. 3B), including viral myocarditis, actin cytoskeleton regulation, and infections caused by pathogenic *E. coli* and *Salmonella*, suggesting that these m7G regulatory factors play pivotal roles in RNA modification and immune regulation.

Selection of critical m7G regulatory factors through machine learning

We employed various machine learning approaches to identify key regulatory factors in diabetic nephropathy (DN) and investigate their functional significance. Initially, SVM-RFE was applied to assess the diagnostic performance of these factors, identifying eight regulatory factors (Fig. 4A, B). Meanwhile, RF was utilized to evaluate the significance of these regulatory factors, selecting four with scores exceeding six (Fig. 4C, D). The intersection of SVM-RFE and RF results identified four regulatory factors (METTL1, CYFIP2, EIF3D, and NUDT4) as the core determinants (Fig. 4E). Furthermore, LASSO Cox regression was performed to validate the diagnostic utility of these four regulatory factors (Fig. 4F, G). Thus, all four regulatory factors were confirmed as significant diagnostic markers. Additionally, based on the expression matrix of overlapping regulatory factors, bar plots were generated to illustrate the ability of these four m7G regulatory factors to distinguish early- and late-stage DN. Calibration curves were plotted to assess the model's predictive accuracy (Fig. 4H, I). Machine learning analysis demonstrated that METTL1, CYFIP2, EIF3D, and NUDT4 exhibit high sensitivity and specificity in distinguishing DN (Fig. 4J). ROC analysis further evaluated the specificity and sensitivity of the selected disease-related m7G genes in distinguishing DN. The AUC values for METTL1, CYFIP2, EIF3D, and NUDT4 all exceed 0.698 (Fig. 4K, L, M, N).

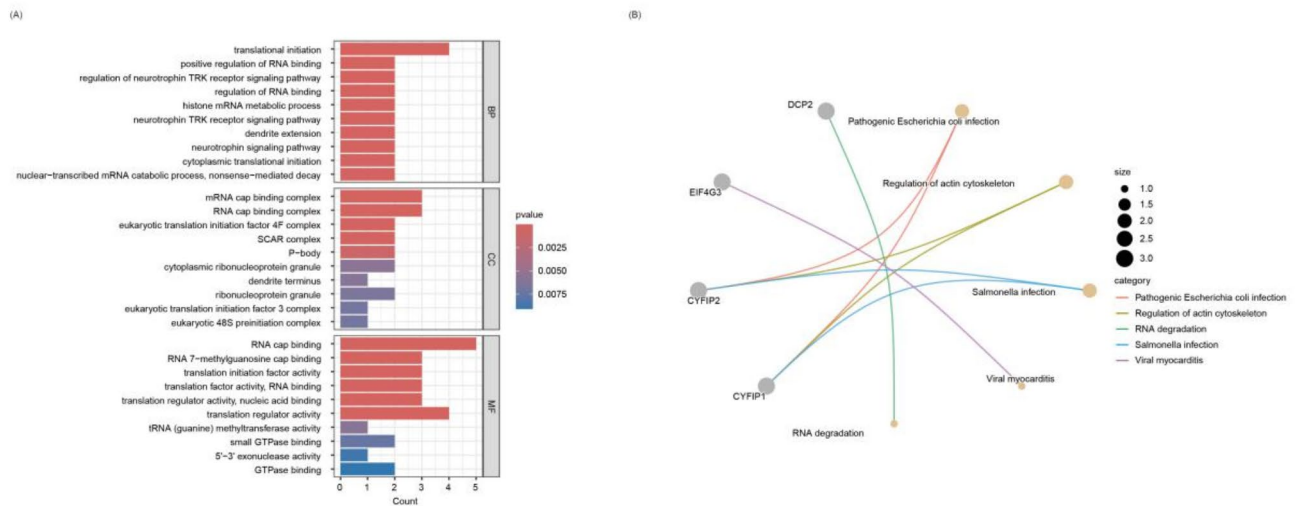


Fig. 3. Enrichment analysis results for 9 m7G-regulated factors. GO analysis is depicted in histogram form (A), while KEGG analysis is shown as a bubble plot^{31–33} (B). Both figures highlight the enrichment of significantly regulated factors associated with differential m7G methylation modifications.

Identification of two distinct m7G-related clusters

Consensus clustering analysis, based on the expression levels of four key regulatory factors identified through machine learning, stratified DN samples in the training set into two m7G-related clusters, with $K = 2$ determined as the optimal solution (Fig. 5A–D). Furthermore, the PCA plot illustrated significant differences between clusters A and B (Fig. 5E). Moreover, all four regulatory factors exhibited significant statistical differences between m7G-related clusters A and B (Fig. 5F, G). According to the PCA algorithm, the m7G scoring module effectively differentiated these clusters ($P < 0.05$), with higher scores observed in m7G-related cluster B and lower scores in cluster A (Fig. 5H).

GSEA, immune infiltration, and immune checkpoint features of m7G-related clusters

GSEA was conducted to elucidate the characteristics and functions of the m7G-related clusters. It identified three pathways ($P < 0.05$), highlighting the association between m7G scoring and both metabolism and the immune system (Fig. 6A). Subsequently, immune infiltration analysis was conducted to illustrate the association between the four key regulatory factors and the infiltration scores of 22 distinct immune cell types. The results indicated stronger associations in cluster B, suggesting a significant increase in immune cell infiltration among DN patients with higher m7G scores (Fig. 6B, C). Furthermore, Pearson correlation coefficients were calculated to evaluate the relationship between the expression of immune checkpoint-related genes and m7G scoring in these clusters, providing a comprehensive assessment of immune characteristics.

Validation of an m7G-related signature model

To validate the m7G regulator-based signature model for diabetic nephropathy (DN), which consists of four key regulatory genes (METTL1, CYFIP2, EIF3D, and NUDT4), we employed multiple evaluation methods to assess its robustness and reliability across independent transcriptomic datasets. The heatmap (Fig. 7A) illustrated the distinct expression profiles of the four key genes in the training dataset, effectively distinguishing the control group from the treatment group. These significant expression differences underscored the potential of these genes as reliable biomarkers for diabetic nephropathy (DN). The calibration curve (Fig. 7B) demonstrated a strong concordance between predicted and actual probabilities, achieving a C-index of 0.892 and confirming the model's high predictive accuracy. The nomogram (Fig. 7C) served as a practical tool for individualized risk assessment, integrating the m7G score and key gene expression levels to quantitatively estimate the probability of DN. The ROC curves (Fig. 7D) validated the model's superior diagnostic performance, with the integrated nomogram achieving an AUC of 0.892, significantly exceeding the AUCs of individual genes (0.698–0.758). The decision curve analysis (Fig. 7E) highlighted the model's clinical utility, demonstrating that the nomogram provided the highest net benefit across a broad range of threshold probabilities. Collectively, these findings confirm the reliability, predictive accuracy, and clinical relevance of the m7G regulator-based signature model for diabetic nephropathy (DN).

To evaluate the disease specificity of the key genes, we analyzed their expression levels in diabetic nephropathy (DN) compared to systemic lupus erythematosus (SLE), IgA nephropathy (IgA), hypertensive nephropathy (HN), and their respective controls (Fig. 7F). METTL1 and CYFIP2 showed significantly higher expression levels in DN ($P < 0.001$), whereas EIF3D and NUDT4 exhibited moderate but statistically significant increases ($P < 0.05$). Although NUDT4 displayed slight variation in hypertensive nephropathy (HN) ($P < 0.05$), no significant changes were detected in SLE or IgA nephropathy, supporting the predominant specificity of these genes for DN (Fig. 7F).

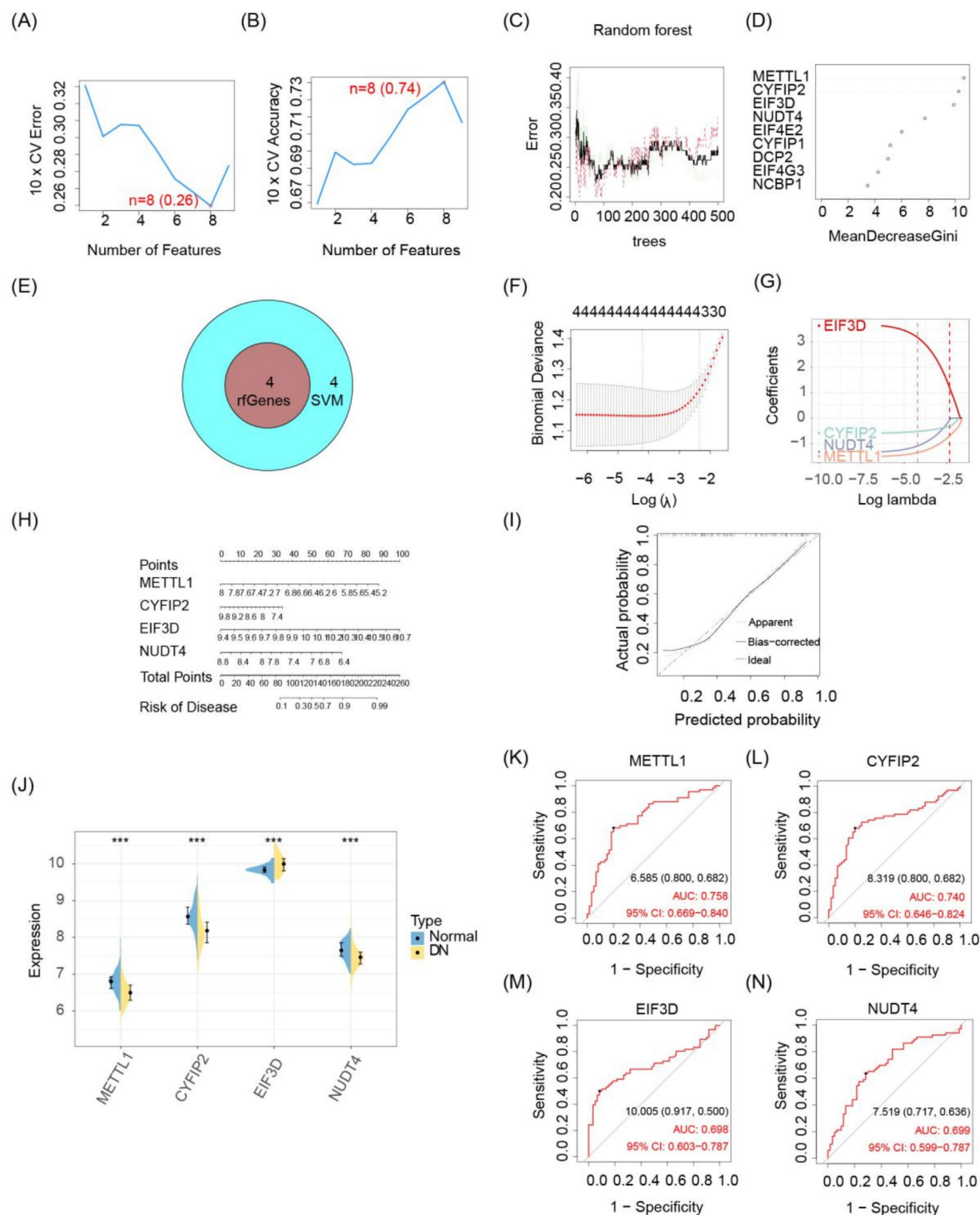


Fig. 4. Further screening of key m7G regulatory factors for DN through machine learning. **(A,B)** SVM-RFE identified 8 significant features. **(C,D)** Random forest identified 4 significant features. **(E)** Venn diagram of overlapping features from RF and SVM-RFE. **(F,G)** Results of lasso-cox regression analysis. **(H)** Nomogram of four key regulatory factors distinguishing early and late DN confirmed by machine learning. **(I)** Calibration curve of the nomogram. **(J)** Differential expression of four key regulatory factors between NM and DN in our training set. **(K–N)** Validation ROC curves of the discriminative power of 4 m7G genes. *** $P < 0.001$.

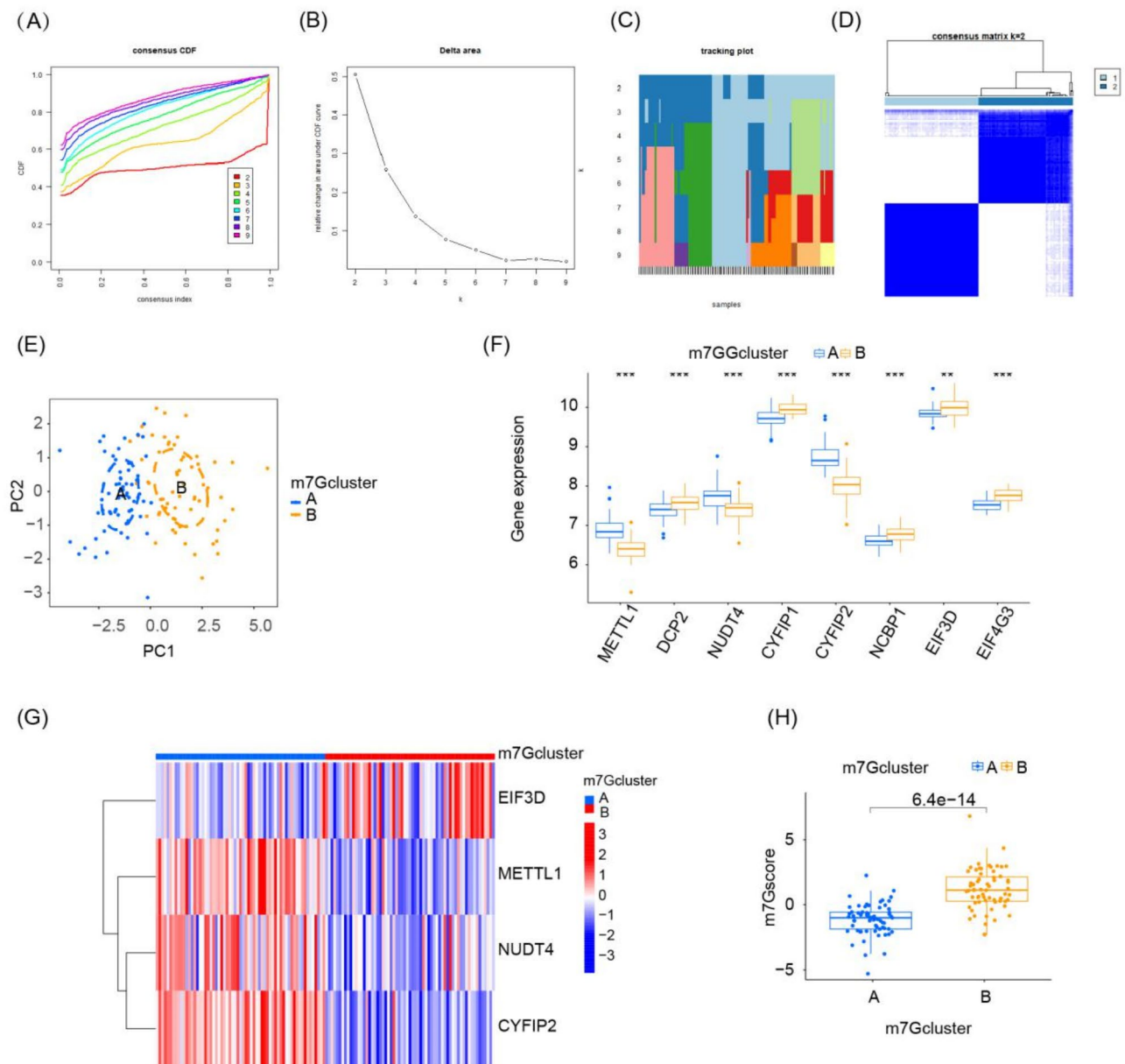


Fig. 5. Identification of m7G-related clusters and construction of the m7G scoring model. (A) Cumulative distribution curve of consensus clustering analysis results. (B) Area under the curve of consensus clustering analysis results. (C) The consensus cluster of items (in column) at $k=2-9$ (in row). (D) Heatmap of two different m7G-related clusters. (E) PCA plot of the above two different clusters. (F) Expression levels of 4 key regulatory factors between clusters A and B. (G) Heatmap of 4 key regulatory factors revealing expression levels between clusters A and B. (H) Difference in PCA scores of 4 key m7G-related regulatory factors between the two clusters. ** $P < 0.01$; *** $P < 0.001$.

External validation was conducted using the GSE30528, GSE99325, and GSE99339 datasets to evaluate the m7G regulator-based signature model. Heatmaps (Fig. 7G, K, O) depicted the consistent dysregulation of key genes, including METTL1, CYFIP2, EIF3D, and NUDT4, in DN samples relative to controls, highlighting their potential as DN-specific biomarkers. Calibration curves (Fig. 7H, L, P) exhibited a strong concordance between predicted and actual probabilities, confirming the robustness of the model. ROC curves (Fig. 7I, M, Q) yielded AUC values ranging from 0.732 to 0.897, validating the model's strong predictive performance in distinguishing DN from controls. Finally, decision curve analyses (Fig. 7J, N, R) highlighted the model's clinical utility, demonstrating significant net benefits across a broad range of threshold probabilities. Collectively, these findings confirm the reliability, diagnostic precision, and clinical applicability of the proposed model for DN.

To further assess the classification performance of our model, we constructed confusion matrices for both the training and test sets (Supplementary Figure S1). The confusion matrix for the training set indicated that the model correctly classified 29 DN samples and 29 normal samples, with 3 false positives and 2 false negatives,

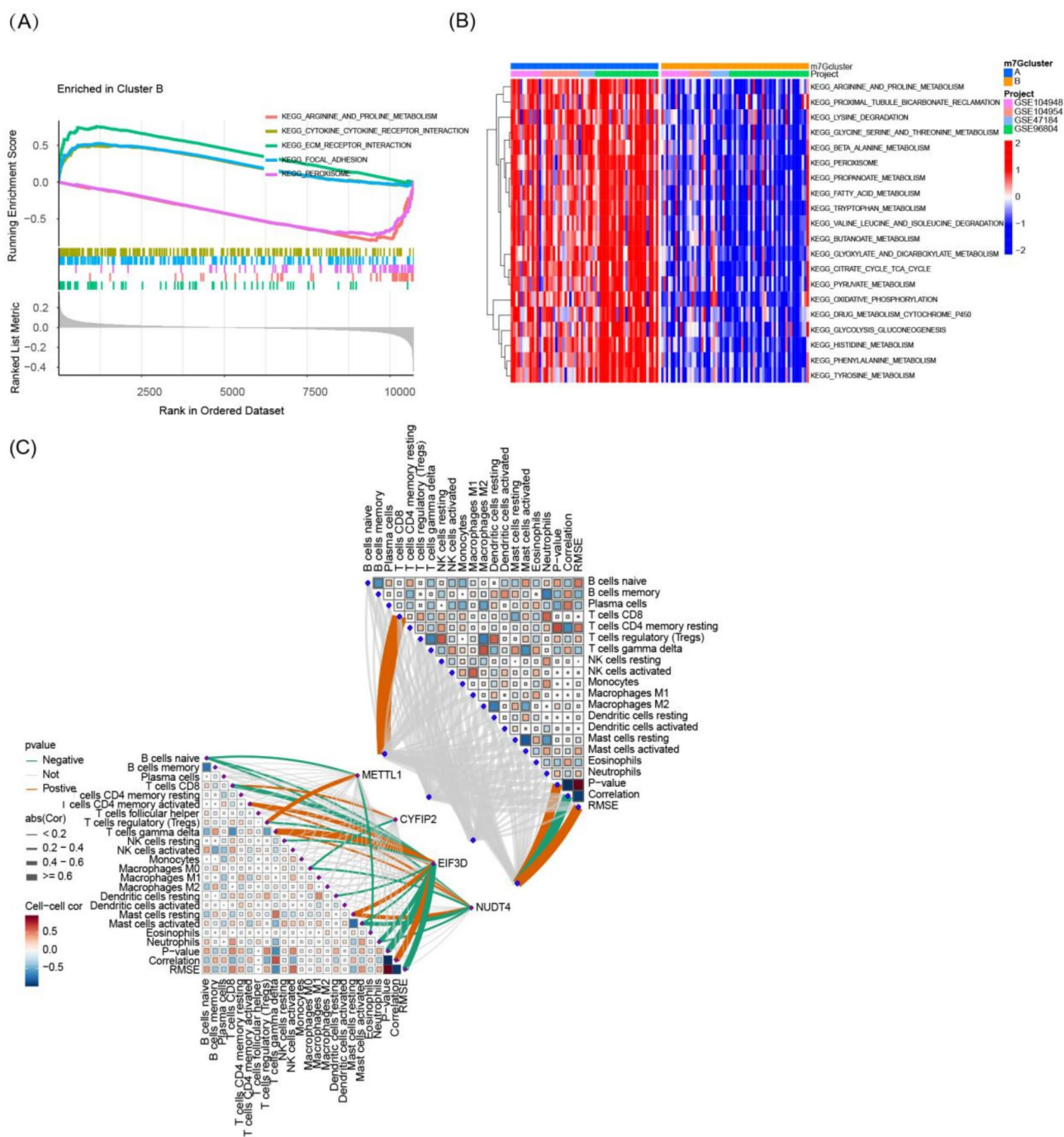


Fig. 6. Immune signature and pathways of two distinct m7G subtypes. **(A)** The results of the GSEA enrichment analysis in two clusters. **(B)** Gene-set variation analysis (GSVA) of biological pathways enrichment between two m7G subtypes. **(C)** The mantel correlation heat map between the expression level of four significant regulators and 28 different kinds of immune cells in two different cluster.

achieving an accuracy of 92.06%, sensitivity of 90.63%, and specificity of 93.55%. Likewise, in the test set, the model correctly classified 30 DN samples and 23 normal samples, with 5 false positives and 5 false negatives, yielding an accuracy of 84.13%, sensitivity of 82.14%, and specificity of 85.71%. These findings confirm the model's strong generalizability and reliability in accurately distinguishing DN from control samples across different datasets.

Explore the differences between the above two clusters and build a diagnostic model

To further emphasize the significance of m7G-related clusters, differential expression analysis was conducted using $|\log_2FC| > 1$ ($P < 0.05$) as the cutoff, identifying 92 differentially expressed genes (DEGs), including 18

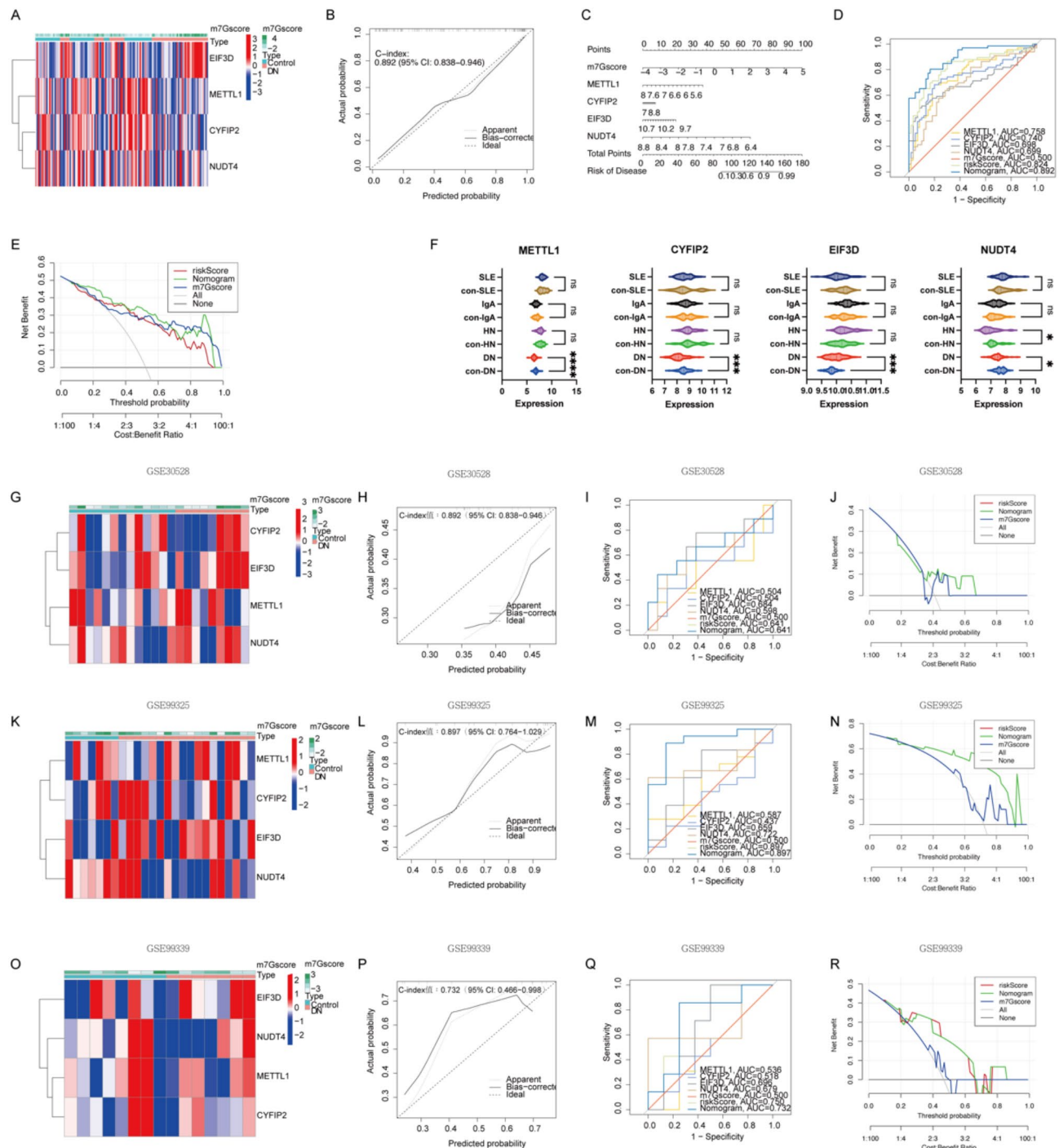


Fig. 7. Validation of an m7G-related signature model for diabetic nephropathy using independent transcriptomic datasets. (A–E) Heatmap (A), calibration curve (B), nomogram (C), ROC curves (D), and decision curve analysis (E) collectively illustrate the expression patterns, predictive accuracy, and clinical utility of key regulatory factors, m7Gscore, risk score, and the nomogram, derived from a training set across multiple datasets (GSE47184, GSE96804, GSE104948, GSE104954). (F) Violin plots showing the expression levels of METTL1, CYFIP2, EIF3D, and NUDT4 across SLE, IgA, HN, DN, and their respective control groups. (G–R) Heatmaps (G,K,O), calibration curves (H,L,P), ROC curves (I,M,Q), and decision curve analyses (J,N,R) collectively present the expression patterns, predictive accuracy, and clinical utility of key regulatory factors, m7Gscore, risk score, and the nomogram, derived from GSE30528, GSE99325, and GSE99339 datasets as validation sets. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

upregulated and 26 downregulated genes, which were visualized using volcano plots and heatmaps (Fig. 8A). Moreover, an overlap of DEGs was observed between NM and DN, as well as between the clusters. In total, 12 DEGs were identified: LYZ, MMP7, VCAN, LUM, FN1, G6PC, FOS, ALB, HPD, GSTA1, EGF, and PVALB (Fig. 8B, C). These DEGs served as potential candidates for LASSO Cox regression.

Using 10-fold cross-validation, $\lambda_{\min} = 0.01490143$ was determined as the optimal regularization parameter, identifying five key genes associated with m7G classification (Fig. 8D, E). Additionally, a Cox logistic regression model was applied to predict and extract three key genes significantly associated with DN from these five feature genes. Hence, LUM, G6PC, and FOS were selected to construct the DN diagnostic model (Fig. 8F, G, H). Moreover, all samples were randomly divided into two subsets (validation and training sets) at a 1:1 ratio. The Wilcoxon signed-rank test was performed using the aforementioned diagnostic scoring formula to evaluate statistical differences between NM and DN samples, revealing significant differences in diagnostic scores between the groups ($P < 0.05$) (Fig. 8I, J). Additionally, the receiver operating characteristic (ROC) curves of the

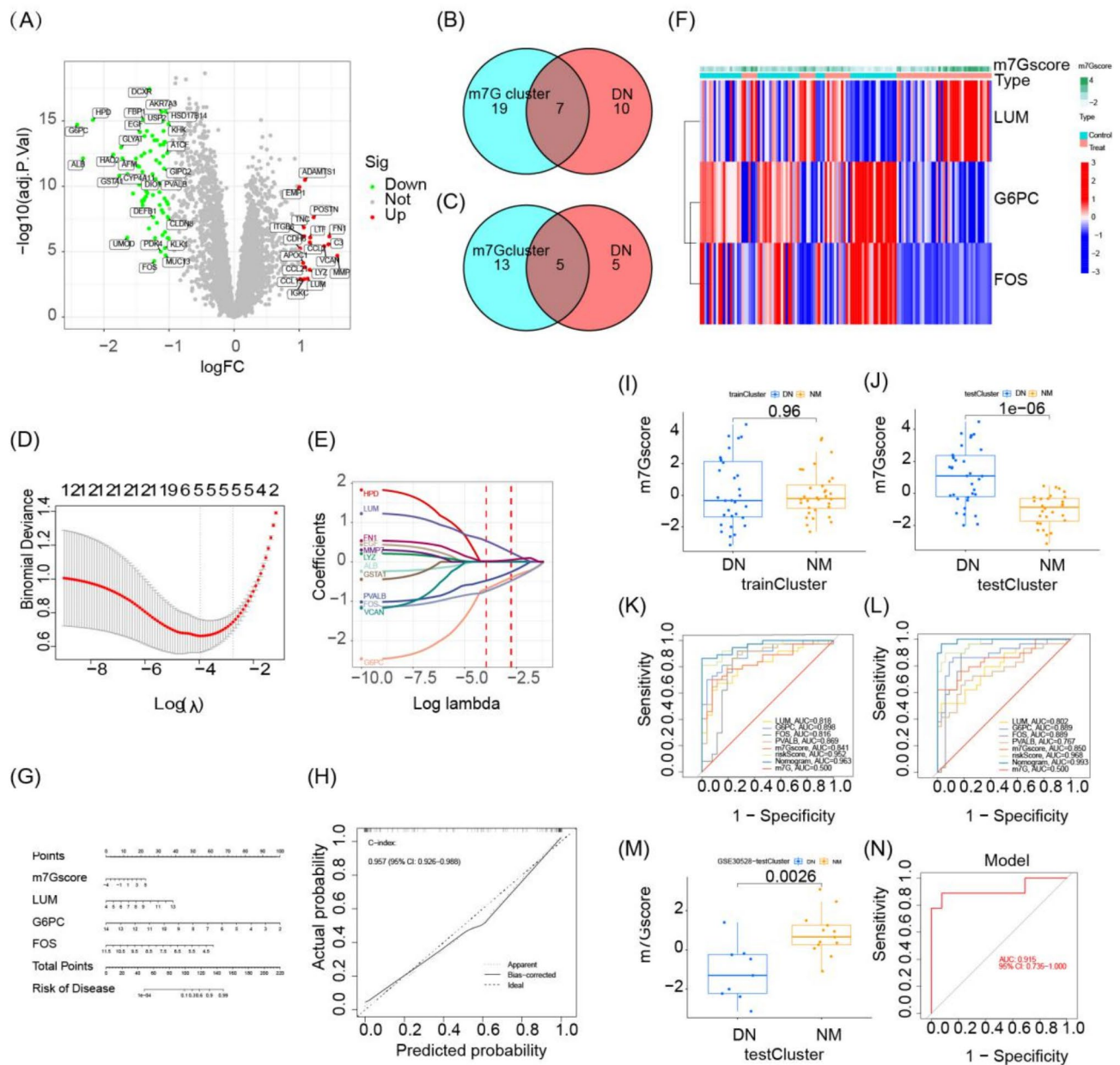


Fig. 8. Establishment of m7G-related diagnostic model. (A) Volcano plot of DEGs between two different m7G clusters. (B,C) Venn diagram of overlapping DEGs; (B) upregulated DEGs between two different clusters. (C) downregulated DEGs between NM and DN groups. (D,E) Results of lasso regression analysis. (F–H) Results of Cox logistic regression analysis. (I,J,M) Difference in diagnostic scores between NM and DN in our (I) training set, (J) internal validation set, and (M) external validation set. (K–N) ROC curves of diagnostic scores: (K) training set, (L) internal validation set, (N) external validation set.

diagnostic model yielded AUC values of 0.963 in the validation set and 0.993 in the training set, confirming its outstanding performance (Fig. 8K, L). Furthermore, to further validate the diagnostic model's accuracy, another external dataset, GSE30528, was utilized. The diagnostic values between NM and DN samples showed significant differences ($P < 0.05$) (Fig. 8M). Additionally, the AUC of the ROC curve for the diagnostic model in GSE30528 was 0.915, confirming its robust accuracy (Fig. 8N).

Validation of screened core genes

To investigate the role of four key biomarkers in diabetic nephropathy (DN), we established a DN mouse model. Immunohistochemical analysis was conducted to assess protein expression levels in biopsy tissues from six normal kidneys and seven DN model mice. The DN model (MD) group exhibited significantly higher body weight, fasting blood glucose, 24-hour urine volume, serum creatinine (Scr), blood urea nitrogen (BUN), albumin (ALB), and urine creatinine (UCR) compared to the control (CR) group (Fig. 9A–G). Renal pathology across the groups was evaluated using Hematoxylin and Eosin (H&E), Periodic Acid-Schiff (PAS), and Masson's trichrome staining. Pathological alterations observed in DN mouse kidneys included mesangial expansion, basement membrane thickening, and increased renal fibrosis. Interstitial fibrosis was also noted in the tubular region (Fig. 9H). The expression levels of EIF3D, CYFIP2, METTL1, and NUDT4 were subsequently analyzed (Fig. 9I). EIF3D expression was significantly upregulated in the MD group, whereas CYFIP2, METTL1, and NUDT4 expression levels were higher in the CR group than in the MD group. Immunohistochemistry revealed that EIF3D expression was upregulated, whereas CYFIP2, METTL1, and NUDT4 expression levels were downregulated in the DN group compared to the normal group (Fig. 9J). Therefore, EIF3D, CYFIP2, METTL1, and NUDT4 exhibit high specificity and sensitivity for DN, particularly when combined for diagnostic purposes.

Discussion

Diabetic nephropathy (DN) is a serious microvascular complication of diabetes mellitus (DM), primarily driven by inflammation, oxidative stress, and epigenetic modifications^{41–44}. Hyperglycemia-induced oxidative stress damages glomerular cells, resulting in proteinuria and tubulointerstitial fibrosis⁴³. Several inflammatory mediators contribute to DN progression, including intrarenal angiotensin (Ang), which enhances glomerular capillary permeability and stimulates mesangial cell proliferation⁴². Persistent renal inflammation exacerbates renal injury⁴³, as inflammatory cells secrete pro-inflammatory cytokines, inducing epithelial-to-mesenchymal transition (EMT).

Epigenetics plays a pivotal role in DN pathogenesis, with epigenetic modifications, including DNA methylation, histone modifications, and non-coding RNA methylation, regulating disease progression⁴⁴. Notably, m7G (7-methylguanosine) methylation, a key tRNA modification, plays an essential role in RNA metabolism and cellular function⁴³.

This study aimed to identify and validate an m7G-related diagnostic signature for DN. Through transcriptomic analysis, machine learning, and experimental validation, four key m7G regulatory genes—METTL1, CYFIP2, EIF3D, and NUDT4—were identified as crucial contributors to DN pathogenesis, exhibiting strong diagnostic potential. This study explores their clinical relevance, the rationale for utilizing a DN mouse model, and their broader implications in disease research. These findings offer novel insights into DN pathogenesis and identify potential biomarkers for clinical application.

Clinical significance of the identified m7G regulators

The four key genes identified in this study play distinct yet interconnected roles in diabetic nephropathy (DN) pathogenesis, highlighting their clinical significance as potential biomarkers.

EIF3D, a subunit of the EIF3 complex, possesses unique cap-binding activity, enabling the translation of specific mRNAs, such as cJUN mRNA, via an EIF4E-independent mechanism. It is essential for adaptive translation under nutrient deprivation. Its cap-binding function is precisely regulated by two phosphorylation sites located at the C-terminal tail, near the cap-binding domain⁴⁵. EIF3D, structurally similar to the DXO family of cap-specific exonucleases, specifically recognizes the 5' cap structure of mRNA. Its activity is regulated by the EIF3 complex, which recruits target mRNAs for translation. Its specificity for N-7 methylated guanosine, as evidenced by competitive inhibition with m7GDP, underscores its selective role in translation regulation⁴⁶. Aberrant EIF3D expression has been linked to tumor progression, particularly in bladder cancer, where its overexpression drives tumor cell proliferation and metastasis. These findings suggest that EIF3D's role in selective mRNA translation may extend beyond cancer, potentially functioning as a key regulator in diseases characterized by dysregulated translation, including inflammatory and metabolic conditions.

The well-established role of EIF3D in modulating mRNA translation under stress conditions, such as nutrient deprivation, suggests its potential involvement in chronic inflammatory diseases, including DN. Although its role in DN remains unexplored, EIF3D is likely involved in modulating the translation of pro-inflammatory cytokines and stress-response genes, thereby contributing to disease progression. The functional similarities between EIF3D's activity in cancer and its potential role in inflammation or fibrosis suggest that EIF3D may serve as a key regulator of pathological translation processes across various diseases.

METTL1, an m7G methyltransferase, regulates RNA metabolism by stabilizing tRNA through m7G methylation, a process essential for translational efficiency^{47,48}. Studies confirm that enhancing m7G modification of tRNA via the METTL1/WDR4 complex stabilizes mRNA decay and improves translation efficiency, which is associated with poor survival rates in human cancers. These findings underscore METTL1's pivotal role in translational regulation and suggest its potential as a key regulator in diseases associated with aberrant RNA metabolism. Mutations in yeast m7G methyltransferase lead to rapid decay of under-methylated tRNA, resulting in growth defects⁴⁹. METTL1 deficiency results in the loss of m7G tRNA modification, reduced stability of under-methylated tRNA, and global translation defects^{50,51}. Given its established role in RNA methylation and

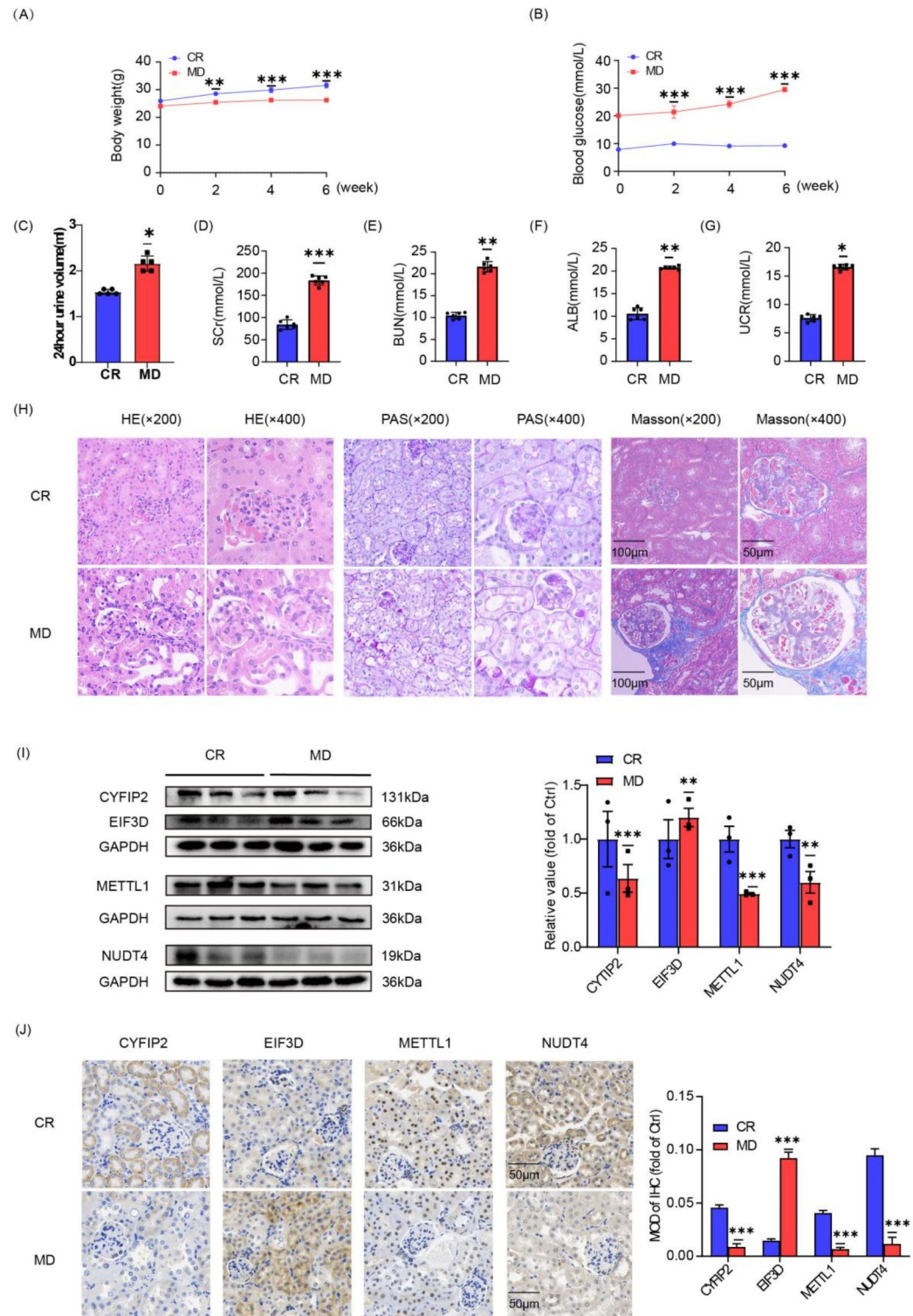


Fig. 9. External validation of screened biomarkers. **(A)** The body weight of all mice was measured weekly ($n = 5$ or 6). **(B)** Fasting blood glucose levels of all mice were measured weekly ($n = 5$ or 6). **(C)** Twenty-four-hour urine samples were analyzed to detect albumin ($n = 5$). **(D–G)** Serum creatinine (S-Cr) **(D)**, blood urea nitrogen (BUN) **(E)**, albumin (ALB) **(F)**, and urine creatinine (UCR) **(G)** were assessed in mice from each group ($n = 5$ or 6). **(H)** Pathological changes in the kidneys of mice from different groups were evaluated using hematoxylin-eosin (H&E), periodic acid-Schiff (PAS), and Masson's trichrome staining. **(I)** Western blotting was employed to detect protein markers; representative blots and density measurements are shown ($n = 3$). **(J)** Immunohistochemical staining was utilized to detect the expression of four core biomarkers in normal kidney tissues (left) and DN kidney tissues (right) (scale bar, $50 \mu\text{m}$). Positive brownish-yellow staining was quantified ($n = 3$). Data are presented as mean \pm SEM. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. CR, normal; MD, diabetic nephropathy.

translational control, METTL1 may influence DN progression by modulating inflammation or oxidative stress, similar to its role in cancer and metabolic disorders.

CYFIP2 is a p53-induced gene that suppresses colon cancer proliferation, activates caspases, and induces apoptosis⁵². In gastric cancer, circRNA CYFIP2 interacts with miRNA-1205 to regulate the E2E1 gene, which is directly targeted by miRNA-1205, thereby promoting tumor proliferation and metastasis⁵³. These findings underscore CYFIP2's pivotal role in apoptosis regulation, immune responses, and tumor progression, suggesting its broader functional relevance in diseases characterized by immune dysregulation and apoptosis, including DN. The ability of CYFIP2 to regulate apoptosis and immune responses in cancer models implies its broader role in modulating immune and apoptotic pathways across other pathological conditions, including DN.

NUDT4, a hydrolase that catalyzes the hydrolysis of inositol pyrophosphates, plays a role in signal transduction and cellular stress responses⁵⁴. Given the pivotal role of cellular stress in DN pathogenesis, including oxidative stress, inflammation, autophagy, and apoptosis, NUDT4's involvement in cellular stress regulation highlights its potential as a key modulator in other stress-related diseases.

Immune cell infiltration serves as a key mechanism in DN pathogenesis, particularly involving T lymphocytes and macrophages. For example, Treg cells are significantly elevated in affected tissues, where interleukin-10 derived from Treg cells regulates macrophage chemokine expression and alleviates DN in humans. The observed correlation between CYFIP2 expression and Treg cell infiltration in DN suggests that CYFIP2 may play a broader role in modulating immune cell behavior in other diseases characterized by immune dysfunction. Reduced CYFIP2 expression significantly contributes to Treg cell infiltration in DN, suggesting that its regulatory role in immune cells may be relevant in other immune-mediated conditions.

Recent studies indicate that NK cell-mediated cytotoxic pathways are activated in human DN, contributing to DN progression¹⁹. The observed positive correlation between NUDT4 expression and resting NK cell levels in DN underscores its broader potential as a modulator of NK cell activity in diseases characterized by immune cell activation and cytotoxicity. This association suggests that NUDT4 may play a protective role by inhibiting NK cell activation, not only in DN but also in other conditions where NK cells contribute to disease progression. These findings emphasize the connection between m7G regulation and immune dysfunction, suggesting that CYFIP2 and NUDT4 may act as critical regulators in diseases involving immune dysfunction and stress-related pathways.

To further validate our findings, we compared the m7G-related gene signature to machine learning-based DN biomarker studies. The identified m7G-related gene signature aligns with key pathways involved in DN pathogenesis, further highlighting its translational value.

Validation of four key m7G regulatory genes

The biological functions of METTL1, CYFIP2, EIF3D, and NUDT4 are highly conserved across species and are expressed in both humans and mice. This evolutionary conservation enhances translational relevance, establishing mouse models as essential tools for validating disease-associated biomarkers. In line with previous studies that integrate machine learning to identify and validate key genes in animal models^{55,56}, our study follows a rigorous approach to confirm the pathological significance of m7G-related biomarkers in diabetic nephropathy (DN).

To assess these biomarkers, we used a well-established streptozotocin (STZ)-induced DN mouse model that closely replicates key pathological features of human DN, such as hyperglycemia, albuminuria, and renal dysfunction⁵⁷. This model is widely used in DN research and serves as a robust platform for biomarker validation^{58–60}. Its capability to replicate human disease characteristics makes it a valuable model for studying molecular mechanisms and potential therapeutic targets.

Experimental validation demonstrated the pathological relevance of METTL1, CYFIP2, EIF3D, and NUDT4 in DN. DN mice showed marked changes in fasting blood glucose, urinary albumin excretion, and renal function markers, such as increased serum creatinine and blood urea nitrogen levels, closely mimicking the human DN phenotype. Furthermore, histopathological analyses revealed glomerular and tubulointerstitial damage, while Western blot and immunohistochemical staining confirmed significant upregulation of these biomarkers in DN kidney tissues, consistent with computational predictions.

To validate these findings, we used independent human transcriptomic datasets, confirming the differential expression and clinical relevance of the identified biomarkers. By integrating bioinformatics, machine learning, in vivo validation, and external dataset analysis, our study provides a comprehensive framework for developing a clinically relevant diagnostic signature for DN, bridging computational predictions with experimental validation.

Limitations and future directions

Despite the promising findings, several limitations warrant further discussion. The reliance on a DN mouse model for experimental validation, while pragmatic, may limit the direct applicability of the findings to human patients. Additionally, the diagnostic model, developed based on three key genes (LUM, G6PC, and FOS) from m7G-related clusters, was validated using transcriptomic datasets but lacked experimental validation, which could impact its clinical accuracy. Finally, while the m7G scoring model effectively differentiated DN patients and linked them to specific immune microenvironment states, its clinical application requires further investigation. Future research should focus on validating the diagnostic signature in larger, multi-center cohorts and investigating its potential to guide therapeutic interventions. Integrating advanced experimental systems, such as human organoids or single-cell sequencing, could provide deeper insights into the roles of m7G regulatory factors in DN.

Conclusion

This study highlights the diagnostic potential of m7G regulatory factors in DN, identifying METTL1, CYFIP2, EIF3D, and NUDT4 as key biomarkers. By integrating machine learning techniques with experimental validation, we developed an m7G scoring model that effectively differentiates DN patients and offers novel insights into disease pathogenesis. While limitations remain, our findings lay a foundation for future clinical applications and highlight the need for further research into m7G regulation in DN.

Data availability

The data used for bioinformatics analysis are openly available in the GEO databases. Below are the repositories and corresponding accession numbers: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE47184>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE96804>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104948>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104954>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30528>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99325>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99339>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73953>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93798>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37455>. The datasets can be accessed and downloaded via the provided links. All data generated or analyzed during this study are included in the manuscript or supplementary information files.

Received: 9 November 2024; Accepted: 10 March 2025

Published online: 17 March 2025

References

- Dai, Y. et al. Probiotics improve renal function, glucose, lipids, inflammation and oxidative stress in diabetic kidney disease: A systematic review and meta-analysis. *Ren. Fail.* **44**, 862–880. <https://doi.org/10.1080/0886022X.2022.2079522> (2022).
- Selby, N. M. & Taal, M. W. An updated overview of diabetic nephropathy: diagnosis, prognosis, treatment goals and latest guidelines. *Diabetes Obes. Metab.* **22**, 3–15. <https://doi.org/10.1111/dom.14007> (2020).
- Deng, Y. et al. Global, regional, and National burden of diabetes-related chronic kidney disease from 1990 to 2019. *Front. Endocrinol.* **12**, 672350. <https://doi.org/10.3389/fendo.2021.672350> (2021).
- Sun, H. et al. Erratum to 'IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045'. *Diabetes Res. Clin. Pract.* **204**, 110945. <https://doi.org/10.1016/j.diabres.2023.110945> (2023).
- Bhatt, D. L. et al. Sotagliflozin in patients with diabetes and chronic kidney disease. *N Engl. J. Med.* **384**, 129–139. <https://doi.org/10.1056/NEJMoa2030186> (2021).
- Bikbov, B. et al. Global, regional, and National burden of chronic kidney disease, 1990–2017: A systematic analysis for the global burden of disease study 2017. *Lancet* **395**, 709–733. [https://doi.org/10.1016/S0140-6736\(20\)30045-3](https://doi.org/10.1016/S0140-6736(20)30045-3) (2020).
- Zhang, J. et al. Advances in early biomarkers of diabetic nephropathy. *Rev. Assoc. Med. Bras.* **64**, 85–92. <https://doi.org/10.1590/1806-9282.64.01.85> (2018).
- Zeni, L. et al. A more tubulocentric view of diabetic kidney disease. *J. Nephrol.* **30**, 701–717. <https://doi.org/10.1007/s40620-017-0423-9> (2017).
- Zhou, H. et al. Integrated analysis of multiple microarray studies to identify core gene-expression signatures involved in tubulointerstitial injury in diabetic nephropathy. *Biomed. Res. Int.* **2022**, 1–20. <https://doi.org/10.1155/2022/9554658> (2022).
- Zhan, M. et al. Perturbations in mitochondrial dynamics by p66Shc lead to renal tubular oxidative injury in human diabetic nephropathy. *Clin. Sci.* **132**, 1297–1314. <https://doi.org/10.1042/CS20180005> (2018).
- Jonkhout, N. et al. The RNA modification landscape in human disease. *RNA* **23**, 1754–1769. <https://doi.org/10.1261/rna.063503.117> (2017).
- Yang, D. et al. N6-Methyladenosine modification of lincRNA 1281 is critically required for mESC differentiation potential. *Nucleic Acids Res.* **46**, 3906–3920. <https://doi.org/10.1093/nar/gky130> (2018).
- Zhang, H. et al. Dynamic landscape and evolution of m6A methylation in human. *Nucleic Acids Res.* **48**, 6251–6264. <https://doi.org/10.1093/nar/gkaa347> (2020).
- Yang, X. et al. 5-Methylcytosine promotes mRNA export — NSUN2 as the methyltransferase and ALYREF as an m5C reader. *Cell.* **177**, 606–625. <https://doi.org/10.1038/cr.2017.55> (2017).
- Luo, Y. et al. The potential role of N7-methylguanosine (m7G) in cancer. *J. Hematol. Oncol.* **15**, 63. <https://doi.org/10.1186/s13045-022-01285-5> (2022).
- Huang, J. et al. N6-Methyladenosine RNA methylation in diabetic kidney disease. *Biomed. Pharmacother.* **171**, 116185. <https://doi.org/10.1016/j.biopha.2024.116185> (2024).
- Wang, Z. et al. METTL1/WD4-mediated tRNA m7G modification and mRNA translation control promote oncogenesis and doxorubicin resistance. *Oncogene* **42**, 1900–1912. <https://doi.org/10.1038/s41388-023-02695-6> (2023).
- Srinivasan, K. et al. Alzheimer's patient microglia exhibit enhanced aging and unique transcriptional activation. *Cell. Rep.* **31**, 107843. <https://doi.org/10.1016/j.celrep.2020.107843> (2020).
- Malbec, L. et al. Dynamic methylome of internal mRNA N7-methylguanosine and its regulatory role in translation. *Cell. Res.* **29**, 927–941. <https://doi.org/10.1038/s41422-019-0230-z> (2019).
- Chan, L. et al. Derivation and validation of a machine learning risk score using biomarker and electronic patient data to predict progression of diabetic kidney disease. *Diabetologia* **64**, 1504–1515. <https://doi.org/10.1007/s00125-021-05444-0> (2021).
- Ju, W. et al. Defining cell-type specificity at the transcriptional level in human disease. *Genome Res.* **23**, 1862–1873. <https://doi.org/10.1101/gr.155697.113> (2013).
- Shi, J.-S. et al. Identification of transcription regulatory relationships in diabetic nephropathy. *Chin. Med. J.* **135**, 1008–1016. <https://doi.org/10.1097/CM9.0000000000001967> (2022).
- Pan, Y. et al. Dissection of glomerular transcriptional profile in patients with diabetic nephropathy: SRGAP2a protects podocyte structure and function. *Diabetes* **67**, 717–730 (2018).
- Grayson, P. C. et al. Metabolic pathways and immunometabolism in rare kidney diseases. *Ann. Rheum. Dis.* <https://doi.org/10.1136/annrheumdis-2017-212935> (2018).
- Woroniecka, K. I. et al. Transcriptome analysis of human diabetic kidney disease. *Diabetes* **60**, 2354–2369. <https://doi.org/10.2337/db10-1181> (2011).
- Shved, N. et al. Transcriptome-based network analysis reveals renal cell type-specific dysregulation of hypoxia-associated transcripts. *Sci. Rep.* **7**, 8576. <https://doi.org/10.1038/s41598-017-09097-1> (2017).
- Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47. <https://doi.org/10.1093/nar/gkv007> (2015).

28. Wilkerson, M. D. et al. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573. <https://doi.org/10.1093/bioinformatics/btq170> (2010).
29. Altermann, E. et al. PathwayVoyager: pathway mapping using the Kyoto encyclopedia of genes and genomes (KEGG) database. *BMC Genom.* **6**, 60. <https://doi.org/10.1186/1471-2164-6-60> (2005).
30. Zhao, C. et al. Special issue: cilia and ciliopathies. *J. Cell. Physiol.* **237**, 2611–2612. <https://doi.org/10.1002/jcp.30816> (2022).
31. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677. <https://doi.org/10.1093/nar/gkac909> (2025).
32. Kanehisa, M. Toward Understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951. <https://doi.org/10.1002/pro.3715> (2019).
33. Kanehisa, M. et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
34. Zhang, G. Z. et al. Development of a machine learning-based autophagy-related LncRNA signature to improve prognosis prediction in osteosarcoma patients. *Front. Mol. Biosci.* **8**, 615084 (2021).
35. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA.* **102**, 15545–15550. <https://doi.org/10.1073/pnas.0506580102> (2005).
36. Nagasawa, Y. et al. IFI27 is a useful genetic marker for diagnosis of Immunoglobulin A nephropathy and membranous nephropathy using peripheral blood. *PLoS One.* **11**, e0153252. <https://doi.org/10.1371/journal.pone.0153252> (2016).
37. Liu, P. et al. Transcriptomic and proteomic profiling provides insight into mesangial cell function in IgA nephropathy. *J. Am. Soc. Nephrol.* **28**, 2961–2972. <https://doi.org/10.1681/ASN.2016121355> (2017).
38. Guo, F. et al. Prediction of drug positioning for Quan-Du-Zhong capsules against hypertensive nephropathy based on the robustness of disease network. *Front. Pharmacol.* **10**, 49. <https://doi.org/10.3389/fphar.2019.00049> (2019).
39. Barbie, D. A. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112. <https://doi.org/10.1038/nature08460> (2009).
40. Sun, T. et al. The role of m6A RNA methylation in cancer. *Biomed. Pharmacother.* **112**, 108613. <https://doi.org/10.1016/j.biopha.2019.108613> (2019).
41. Duni, A. et al. Oxidative stress in the pathogenesis and evolution of chronic kidney disease: untangling Ariadne's thread. *Int. J. Mol. Sci.* **20** (15), 3711. <https://doi.org/10.3390/ijms20153711> (2019).
42. Du, Y. H. et al. Efficacy and safety of angiotensin converting enzyme inhibitors and angiotensin II receptor blockers in diabetic nephropathy: A systematic review and meta-analysis. *Ann. Palliat. Med.* **11** (3), 1093–1101. <https://doi.org/10.21037/apm-22-212> (2022).
43. Rayego-Mateos, S. et al. Pathogenic pathways and therapeutic approaches targeting inflammation in diabetic nephropathy. *Int. J. Mol. Sci.* **21** (11), 3798. <https://doi.org/10.3390/ijms21113798> (2020).
44. Sun, J. et al. Role of epigenetic histone modifications in diabetic kidney disease involving renal fibrosis. *J. Diabetes Res.* **2017**, 1–11. <https://doi.org/10.1155/2017/7242384> (2017).
45. Lee, A. S. Y. et al. eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. *Nature* **536**, 96–99. <https://doi.org/10.1038/nature18954> (2016).
46. Mukhopadhyay, S. et al. eIF3d controls the persistent integrated stress response. *Mol. Cell.* **83**, 3303–3313. e6 (2023).
47. Zhang, L. S. Transcriptome-wide mapping of internal N7-methylguanosine methylome in mammalian mRNA. *Mol. Cell. Doi* : <https://doi.org/10.1016/j.molcel.2019.03.036>
48. Leulliot, N. et al. Structure of the yeast tRNA m7G methylation complex. *Structure* **16**, 52–61. <https://doi.org/10.1016/j.str.2007.10.025> (2008).
49. Alexandrov, A. et al. Rapid tRNA decay can result from lack of nonessential modifications. *Mol. Cell.* **21**, 87–96. <https://doi.org/10.1016/j.molcel.2005.10.036> (2006).
50. Orellana, E. A. et al. METTL1-mediated m7G modification of Arg-TCT tRNA drives oncogenic transformation. *Mol. Cell.* **81**, 3323–3338.e14. <https://doi.org/10.1016/j.molcel.2021.06.031> (2021).
51. Chen, J. et al. Aberrant translation regulated by METTL1/WDR4-mediated tRNA N7-methylguanosine modification drives head and neck squamous cell carcinoma progression. *Cancer Commun.* **42**, 223–244. <https://doi.org/10.1002/cac2.12273> (2022).
52. Mongroo, P. S. et al. IMP-1 displays cross-talk with K-Ras and modulates colon cancer cell survival through the novel proapoptotic protein CYFIP2. *Cancer Res.* **71**, 2172–2182. <https://doi.org/10.1158/0008-5472.CAN-10-3295> (2011).
53. Lin, J. et al. circCYFIP2 acts as a sponge of miR-1205 and affects the expression of its target gene E2F1 to regulate gastric cancer metastasis. *Mol. Therapy - Nucleic Acids.* **21**, 121–132. <https://doi.org/10.1016/j.omtn.2020.05.007> (2020).
54. Hua, L. Molecular cloning of a novel isoform of diphosphoinositol polyphosphate phosphohydrolase: A potential target of lithium therapy. *Neuropsychopharmacology* **24**, 640–651. [https://doi.org/10.1016/S0893-133X\(00\)00233-5](https://doi.org/10.1016/S0893-133X(00)00233-5) (2001).
55. Han, H. et al. Identification and verification of diagnostic biomarkers for glomerular injury in diabetic nephropathy based on machine learning algorithms. *Front. Endocrinol.* **13**, 876960. <https://doi.org/10.3389/fendo.2022.876960> (2022).
56. Wang, H. et al. Identification and functional analysis of the hub ferroptosis-related gene EZH2 in diabetic kidney disease. *Int. Immunopharmacol.* **133**, 112138. <https://doi.org/10.1016/j.intimp.2024.112138> (2024).
57. Giralt-López, A. et al. Revisiting experimental models of diabetic nephropathy. *Int. J. Mol. Sci.* **21** (10), 3587. <https://doi.org/10.3390/ijms21103587> (2020).
58. Syed, R. U. et al. Effect of cubebin against streptozotocin-induced diabetic nephropathy rats via Inhibition TNF- α /NF- κ B/TGF- β : in vivo and in Silico study. *Sci. Rep.* **15** (1), 4369. <https://doi.org/10.1038/s41598-025-87319-7> (2025).
59. Ma, Y. et al. Angiotensin-like protein 3 deficiency combined with Valsartan administration protects better against podocyte damage in streptozotocin-induced diabetic nephropathy mice. *Int. Immunopharmacol.* **115**, 109715. <https://doi.org/10.1016/j.intimp.2023.109715> (2023). Epub 2023 Jan 25.
60. Sunilkumar, S. et al. REDD1 expression in podocytes facilitates renal inflammation and pyroptosis in streptozotocin-induced diabetic nephropathy. *Cell. Death Dis.* **16** (1), 79. <https://doi.org/10.1038/s41419-025-07396-4> (2025).

Author contributions

Y.X. conceived and designed the study. PL.X. conducted the experiments and obtained the data. HY.L. and ZD.C. and J.Y. analyzed and collated the data. JY.Z., XY.L., M.W. and XJ.C. drafted and wrote the final version of the manuscript. All authors reviewed the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (nos.82074307, 82374231, 82104910 and T2341019), Guangdong Basic and Applied Basic Research Foundation (2023A1515011949, 2023A1515220054, 2019A1515010816, and 2019A1515110159).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-93811-x>.

Correspondence and requests for materials should be addressed to Y.X.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025