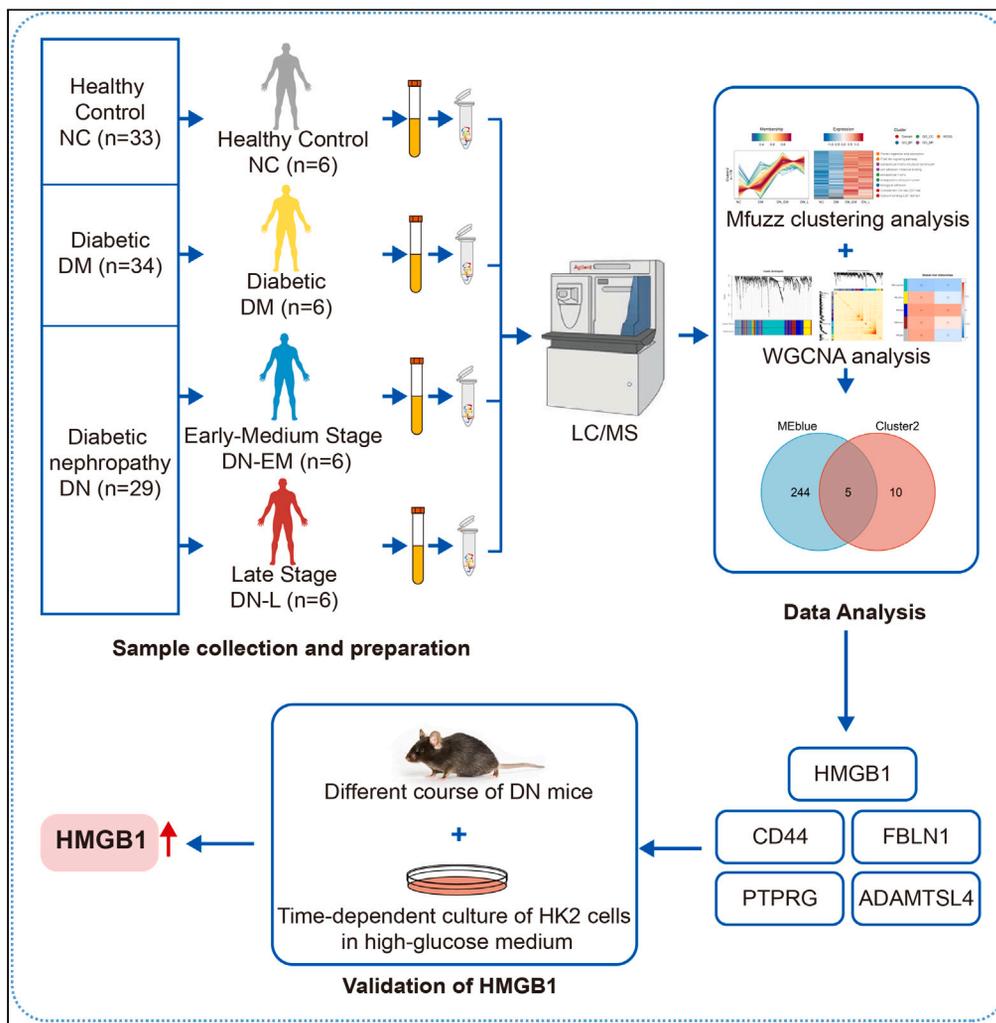**Article**

Investigating HMGB1 as a potential serum biomarker for early diabetic nephropathy monitoring by quantitative proteomics

Rui Peng, Siyang Zuo, Xia Li, ..., Qingqing Zhao, Bing Guo, Lirong Liu

lirongliu@gmc.edu.cn

Highlights

Potential biomarkers increased along with diabetic nephropathy progression

HMGB1 as an acceptable biomarker for the early monitoring of diabetic nephropathy

HMGB1 was elevated under high glucose conditions both in cells and animals

Article

Investigating HMGB1 as a potential serum biomarker for early diabetic nephropathy monitoring by quantitative proteomics

Rui Peng,[1,2,3] Siyang Zuo,[1,2,3] Xia Li,[3,6] Yun Huang,[1] Siyu Chen,[1,2,3] Xue Zou,[6] Hehua Long,[1,2,3] Min Chen,[1,2,3] Yuan Yang,[1,2,3] Huixiong Yuan,[1,2,3] Qingqing Zhao,[6] Bing Guo,[4,5] and Lirong Liu[1,2,3,7,*]

SUMMARY

Current diagnostic methods for diabetic nephropathy (DN) lack precision, especially in early stages and monitoring progression. This study aims to find potential biomarkers for DN progression and evaluate their accuracy. Using serum samples from healthy controls (NC), diabetic patients (DM), early-medium stage DN (DN-EM), and late-stage DN (DN-L), researchers employed quantitative proteomics and Mfuzz clustering analysis revealed 15 proteins showing increased expression during DN progression, hinting at their biomarker potential. Combining Mfuzz clustering with weighted gene co-expression network analysis (WGCNA) highlighted five candidates (HMGB1, CD44, FBLN1, PTPRG, and ADAMTSL4). HMGB1 emerged as a promising biomarker, closely correlated with renal function changes. Experimental validation supported HMGB1's upregulation under high glucose conditions, reinforcing its potential as an early detection biomarker for DN. This research advances DN understanding and identifies five potential biomarkers, notably HMGB1, as a promising early monitoring target. These findings set the stage for future clinical diagnostic applications in DN.

INTRODUCTION

Diabetes mellitus (DM) is a chronic metabolic disease characterized by chronic hyperglycemia caused by impaired insulin secretion or utilization. Globally, over 415 million people suffer from DM, and 693 million are expected to be diagnosed with it by 2045.[1] Diabetes nephropathy (DN) is one of the major microvascular complications of DM, and approximately 30–40% of patients with DM will develop DN.[2] Unfortunately, the majority of DN patients progress without symptoms until they develop renal injury and then irreversible renal failure, which is treated only with kidney transplantation and dialysis. Besides posing a threat to patients' lives, DN is also an enormous economic and medical burden on patients and society.[3] To effectively prevent and treat DN, early diagnosis is therefore crucial.

Renal biopsy is still the golden standard for diagnosing and typing DN. Nonetheless, this invasive approach has inherent limitations, such as the possibility of bleeding complications and the biases in its sampling.[4] Therefore, the diagnosis of DN is increasingly being conducted using noninvasive surrogate techniques. Biomarkers can be used to identify people with diseases and redefine disease classifications.[5,6] Classic markers for assessing the severity of DN include proteinuria, estimated glomerular filtration rate (eGFR), creatinine (Crea), and blood urea nitrogen (BUN).[7,8] These biomarkers accurately quantify the degree of renal injury in patients with DN, but they do not have sufficient accuracy to discern the mild renal insufficiency of early DN. It is now widely accepted that albuminuria, a protein that is filtered through the glomerulus and then reabsorbed by the renal tubules, can monitor the development of DN.[9] However, 20–40% of DM patients already have an eGFR decline before they are detected with albuminuria.[10] Furthermore, the precise role of novel biomarkers such as microRNA (miRNA), long noncoding RNA (lncRNA), and urinary exosomes in DN still remains to be determined.[11] Therefore, DN management requires noninvasive or minimally invasive methods that are more sensitive and selective for the detection of DN as well as monitoring the progression of DN.

Proteomics based on mass spectrometry (MS) is the technology of choice for analyzing proteins to discover potential disease-related biomarkers.[12,13] Serum contains numerous secreted proteins that play crucial roles in physiological and pathological changes.[14] Therefore, comprehensive serum proteomics can be used to discover novel protein biomarkers for DN that may be obtained to learn more about

[1]Center for Clinical Laboratories, the Affiliated Hospital of Guizhou Medical University, Guiyang 550004, China
[2]School of Clinical Laboratory Science, Guizhou Medical University, Guiyang 550004, China
[3]Guizhou Precision Medicine Institute, the Affiliated Hospital of Guizhou Medical University, Guiyang 550004, China
[4]Department of Pathophysiology, Guizhou Medical University, Guiyang 550025, China
[5]Laboratory of Pathogenesis Research, Drug Prevention and Treatment of Major Diseases, Guizhou Medical University, Guiyang 550025, China
[6]Center for Clinical Medical Research, the Affiliated Hospital of Guizhou Medical University, Guiyang 550004, China
[7]Lead contact
*Correspondence: lirongliu@gmc.edu.cn
https://doi.org/10.1016/j.isci.2024.108834

pathophysiology and increase the accuracy of diagnostic stratification. Despite numerous studies exploring serum proteomics for potential DN biomarkers, the detection of only a limited number of proteins has been predominantly attributed to methodological limitations.[15,16] Recent studies involving the proteomics of serum for DN progression had previously been undertaken, but they were unable to distinguish DM from early DN.[17] Moreover, many studies only focus on biomarkers in serum, while very few studies examine molecules identified through *in vitro*, cell-based or animal models. Thus, the biomarkers of DN progression based on serum proteomics need to be further explored.

In this study, we investigated the serum proteome of early-medium stage and late stage DN patients compared with diabetic patients and healthy control subjects. Then, we clustered the associated biomarkers with a similar expressive variation trend, focusing on the cluster containing rise biomarkers along with DN development. With weighted gene co-expression network analyses (WGCNA) of all biomarkers, we explored which of the biomarkers was most relevant for DN progression and whether it was able to discriminate DN or not. We confirmed that the biomarker is elevated in both cell-based and animal models of high glucose. This study identifies a potential candidate biomarker to monitor the progression of DN patients and may be used in clinical practice.

RESULTS

Participant characteristics and assessment of the serum proteome analysis

A total of 96 patients, including healthy control (NC, male/female: 11/22), diabetic (DM, male/female:20/14), and diabetic nephropathy (DN, male/female: 16/13) were recruited. Upon enrollment for serum proteomic analysis, patients' inclusion and exclusion criteria were reset. According to the eGFR categories in chronic kidney disease (CKD) by the Kidney Disease Improving Global Outcomes (KDIGO),[18] patients were further divided into four groups (Figure 1A): healthy control (NC, eGFR $\geq$ 90 mL/min/1.73 m^2), diabetic (DM, eGFR $\geq$ 90 mL/min/1.73 m^2), early medium stage (DN-EM, 30 $\leq$ eGFR <90 mL/min/1.73 m^2), and late stage (DN-L, eGFR $\leq$ 30 mL/min/1.73 m^2), with 6 patients in each group. Demographic and clinical characteristics of the subjects are demonstrated in Table 1. The study included 12 females and 12 males, ages 40–80, and patients in the DN-EM and DN-L groups were older than the DM group; the observed distinction presented a statistically significant difference. No significant discrepancies were found in the gender and levels of uric acid between the four groups of participants. The blood glucose level in the DM, DNEM, and DNL groups was significantly higher than that in NC group. In the DN-L group, serum creatinine and cystatin c levels were significantly elevated, and the blood urea nitrogen to serum creatinine ratio was significantly lower than in the NC and DM groups, while the level of blood urea nitrogen was significantly higher than in the NC, DM and DNEM groups. Meanwhile, the DN-L group had a significantly lower eGFR than the NC, DM and DNEM groups, and the DMEM group had a significantly lower eGFR than DM.

Different protein levels in serum range over large orders of magnitude and are very heterogeneous compared to tissues or cellular samples. To obtain more comprehensive and useful proteins related to the DN pathogenesis, we filtered out the interference of high abundance proteins in serum. Subsequently, 1602 proteins were identified by LC/MS, of which 1402 proteins could be quantified (Figure 1B). It was consistent with the general rule based on enzymatic hydrolysis and mass spectrum fragmentation mode that most peptide segments were distributed in 7–20 amino acids, indicating that the distribution of peptide lengths identified by mass spectrum met quality control standards (Figure 1C). In order to ensure the reliability of the serum samples and assay technique, and to confirm the statistical consistency of the quantitative outcomes, a comprehensive approach was undertaken. This involved the utilization of various methods, including Pearson's correlation coefficient (PCC), principal component analysis (PCA), and assessment of Relative Standard Deviation (RSD). PCC analysis showed that there were high correlations between all specimens, with correlation coefficients generally higher than 0.9 (Figure 1D). Further analysis of these proteins by PCA demonstrated that the four groups could be successfully distinguished, particularly in the NC and DM groups, where reproducibility was high within each group (Figure 1E). It is worth highlighting that the DN-EM and DN-L groups exhibited the most pronounced dissimilarity compared to the other stages. This suggests a substantial alteration in the protein expression profile during the DN stage. Quantitative repeatability of proteins in each group was acceptable as shown by the boxplot of RSD (Figure 1F). Based on quality control data, the samples exhibited high quality, and the accuracy of detection techniques indicates that any alterations observed in serum proteins may be associated with pathological disorders.

Serum proteome profiling of patients with diabetic nephropathy

Comprehensive view of proteomic changes related to DN pathogenesis, quantitative results of serum proteomics were performed differentially expressed proteins analysis (p < 0.05, FC > 1.5, and FC < 1/1.5). After pre-processing and missing value filtering, a total of 484 differential proteins were shown by the differential expressed proteins heatmap (Figure 2A). In the cluster heatmap, most NC and DM groups had the same pattern of differentially expressed proteins, whereas the DN-EM and DN-L groups had the same pattern. In addition, we analyzed the volcano plots of pairwise comparisons between NC, DM, DN-EM, and DN-L groups to visualize differentially expressed proteins (Figures 2B–2G). In the context of pairwise comparisons between groups, proteins exhibiting differential expression were visually represented as those that showed increased and decreased levels (Figure 2H). A closer look at the two DN groups had more differentially expressed proteins from the NC and DM groups, and that the NC group in particular had the most differentially expressed proteins compared to DN-EM and DN-L. Then, we selected all differentially expressed proteins from pairwise comparisons between groups and plotted UpSet plots of them, which hinted at 40 common upregulated proteins and 23 common downregulated proteins in DN-EM versus NC, DN-L versus NC, DN-EM versus DM, and DN-L versus DM (Figures 2I and 2J; black arrow). The list of 40 common upregulated and 23 downregulated differentially expressed proteins was showed in Tables S1 and S2.
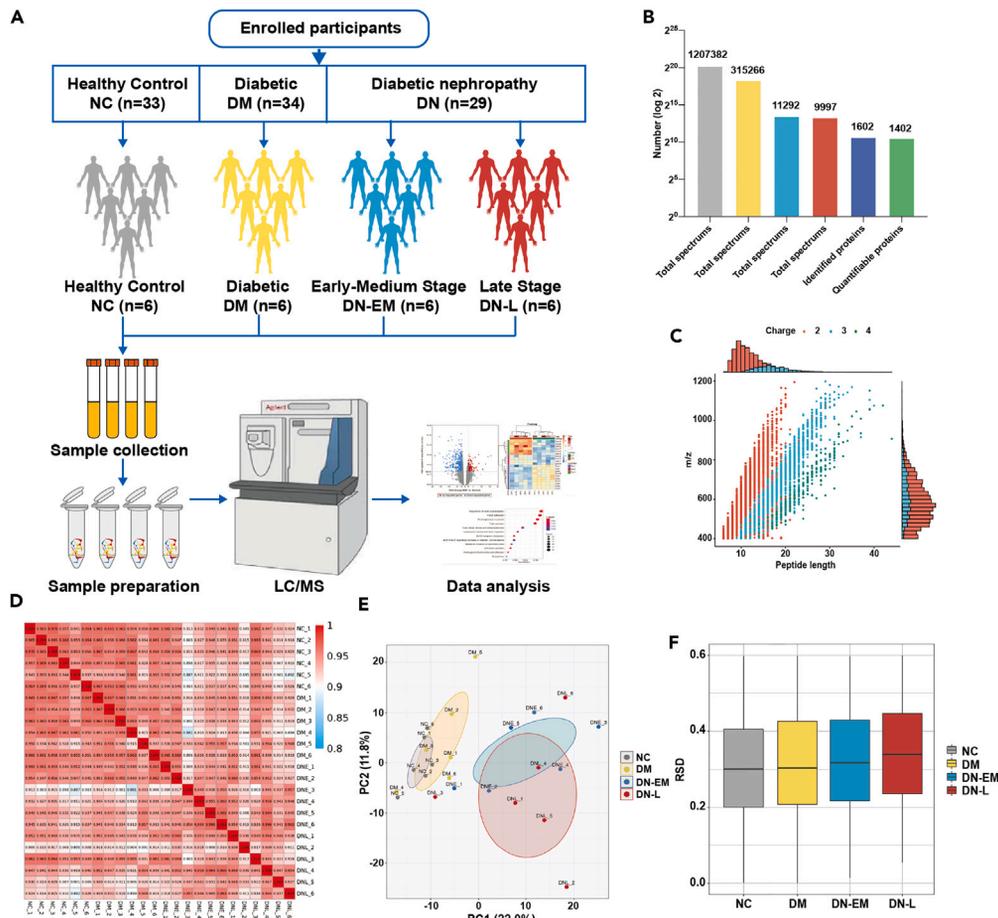
Figure 1. Design and quality control of serum proteomics in DN

(A) Overview of the study structure and cohort particulars: In total, 96 participants from healthy control, diabetic, and diabetic nephropathy were recruited and further divided into four groups, including healthy control (NC), diabetic (DM), early medium stage (DN-EM), and late stage (DN-L), with 6 patients in each group to perform liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

(B) Overview of the number of proteins identified by LC-MS/MS analysis.

(C) Distribution of identifiable peptide lengths from LC-MS/MS analysis.

(D) Pearson's Correlation Coefficient (PCC) analysis from the proteomics data in NC, DM, DN-EM, and DN-L groups. Color saturation in red and blue indicates a degree of correlation present among the samples.

(E) Principal component analysis (PCA) from the proteomics data shows discrimination between NC (gray), DM (yellow), DN-EM (blue), and DN-L (red) groups. Each dot represents a single sample.

(F) Relative Standard Deviation (RSD) from the proteomics data shows intra-group data repeatability in NC (gray), DM (yellow), DN-EM (blue), and DN-L (red) groups.

Serum proteomics revealed the progression of diabetic nephropathy

With serum proteomics, we performed the fuzzy c-means algorithm[19] to cluster the underlying protein determinants of DN progression and onset in circulation. It can cluster the associated protein expression patterns, and proteins in the same cluster display similar expressive variation trends. A total of 5 distinct clusters of temporal patterns representing different regulated proteins were observed (Figure 3). In these clusters, cluster 1 represents downregulated proteins, clusters 2 and 5 represent upregulated proteins, and clusters 3 and 4 represent bimodally expressed proteins. In this study, we focused on proteins that are elevated during DN progression. An analysis of gene ontology (GO) of proteins in each cluster (Figure S1) revealed that the upregulated proteins tend to perform a variety of functions, including extracellular matrix (ECM) structural constituent, cell adhesion molecule binding, ECM, endoplasmic reticulum lumen and biological adhesion (Cluster 2), presumably responsible for cell growth, polarity, shape, migration, and metabolic activity in DN. KEGG pathway enriched proteomics in Cluster 2 are mainly associated with protein digestion and absorption, PI3K-Akt signaling pathway, human papillomavirus infection, ECM-receptor interaction and protein processing in endoplasmic reticulum (Figures 3 and S2).

For reliable DN diagnostic indicators, we set the membership of cluster, the RSD within groups, and the number of unique peptides in cluster 2. A total of 16 peptides were eligible and derived from different 15 proteins, including ADAMDEC1, ADAMTSL4, AMBP, APOA4,

Table 1. Demographic characteristics of the participants

Variable	NC (N = 6)	DM (N = 6)	DN-EM (N = 6)	DN-L (N = 6)	P
Age (Years)	51.50 ± 5.01	48.17 ± 5.04	63.83 ± 12.98[b]	63.50 ± 8.31[b]	0.007~
Gender (Female/Male)	4/2	2/4	4/2	2/4	0.541#
eGFR (ml/min/1.73m2)	99.57 ± 5.68	102.06 ± 7.43	63.67 ± 15.04[ab]	18.89 ± 9.38[abc]	<0.001~
Glu (mmol/L)	4.47 (4.14–4.72)	7.59 (6.02–10.08)[a]	6.68 (5.99–13.66)[a]	8.6 (5.35–11.58)[a]	0.010*
BUN (mmol/L)	5.12 ± 1.45	5.00 ± 1.37	6.29 ± 2.26	10.06 ± 1.76[abc]	<0.001~
SCr (μmol/L)	57.5 (47.25–63.75)	54.0 (48.00–60.75)	67 (62.00–116.25)	180 (177.25–396.25) [ab]	0.001*
BUN/SCr	0.085 (0.075–0.115)	0.09 (0.068–0.118)	0.09 (0.083–0.090)	0.04 (0.028–0.055) [ab]	0.009*
Cys-C (mg/L)	0.85 (0.79–0.89)	0.84 (0.77–0.89)	1.16 (1.05–2.07)	3.27 (2.24–5.42) [ab]	<0.001*
UA (μmol/L)	276.5 (216.75–313.50)	250.0 (228.00–411.75)	335 (235.50–359.50)	414 (289.00–429.75)	0.283*

eGFR: estimated glomerular filtration rate; Glu: glucose; BUN: blood urea nitrogen; SCr: serum creatinine; BUN/Scr; blood urea uitrogen to serum creatinine ratio; Cys-C: cystatin c; UA: uric acid; ~: Analysis of variance (ANOVA); #: Fisher's Exact Test; *: Kruskal-Wallis Test; [a] p < 0.05 vs. NC group; [b] p < 0.05 vs. DM group; [c] p < 0.05 vs. DNEM group.

AZGP1, CD44, COL18A1, COL6A3, EEA1, FBLN1, FBN1, HMGB1, MAN1A1, OAF and PTPRG (Table S3). Each peptide's expression levels across all four groups were compared (Figure 4A), and the panel of FBLN1, AZGP1, CD44, ADAMDEC1, ADAMTSL4 and HMGB1 showed significant differences (p < 0.001). To investigate the mechanism of DN progression, PPI network data were constructed by inputting 15 proteins into the STRING Database, then uploaded into Cytoscape, and selected the top 21 core proteins using the Cytohubba plug-in, based on descending degrees (Figure 4B). Based on the highest scores, CD44, HMGB1 and AMBP may play a crucial role in DN progression. These 15 proteins were also subjected to enrichment analysis for GO and KEGG, whose functions are mainly related to ECM structure and receptor interaction, and ECM deposition plays an important role in DN development (Figure 4C).

High-mobility group protein B1 is a biomarker for monitoring diabetic nephropathy

Analysis of co-expression networks serves as a valuable tool in unraveling the intricate changes characteristic of DN. This is particularly essential because the emergence of DN phenotypes stems from the amalgamation of numerous and gradual alterations in the deregulated expression of multiple proteins, rather than the isolated deregulation of individual proteins.[20] Co-expression network analysis was conducted using all proteins with the weighted gene co-expression network analysis (WGCNA) approach.[21] Soft thresholding power (β = 8; cut-off = 0.85) with increased adjacency was used to create a weighted gene network (Figure S3), resulting in five distinct modules of different colors (Figure 5A). Initial visualization of the Topological Overlap Matrix (TOM) of proteins after DN was performed using a heatmap plot with various module assignments and protein dendrograms (Figure 5B). Based on this, the module-trait relationship between the module eigengene E and the trait DN were then investigated, the DN group contains all of the proteins from the DM, DN-EM, and DN-L groups (Figure 5C). Interestingly, only blue module is most significantly correlated both with the trait NC (cor = 0.51 and p = 0.01) and DN (cor = 0.45 and p = 0.03). Finally, we explored the overlap of proteins in blue module and cluster 2 with a Venn diagram (Figure 5D). The overlapping proteins (HMGB1, CD44, FBLN1, PTPRG and ADAMTSL4) were display in Table 2 with their relevance scores of DN, which are obtained from the GeneCards database.

In our analysis, our research indicates a significant connection between the serum protein High-mobility group protein B1 (HMGB1) and the development of DN. We tested HMGB1's ability to discriminate between patients who developed DN and those who did not. Based on ROC curve analysis, we found that HMGB1 was a good marker for distinguishing DN-EM versus NC (AUC of ROC = 0.917), DN-L versus NC (AUC of ROC = 1.000), DN-EM versus DM (AUC of ROC = 0.889), and DN-L versus DM (AUC of ROC = 1.000) (Figure 6A). Furthermore, intensity of HMGB1 was inversely related to eGFR (*Pearson's R* = −0.787 and p < 0.001) and the blood urea nitrogen to serum creatinine ratio (*Pearson's R* = −0.631 and p < 0.001). Meanwhile, intensity of HMGB1 was positively correlated blood urea nitrogen (*Pearson's R* = 0.646 and p < 0.001), serum creatinine (*Pearson's R* = 0.688 and p < 0.001) and cystatin c (*Pearson's R* = 0.661 and p < 0.001) (Figures 6B–6F). These findings suggest that HMGB1, particularly when it comes to individuals in the early and advanced stages of DN, may accurately monitor the status of the disease.

Investigation of high-mobility group protein B1 in diabetic nephropathy model

Although serum proteome profiling method does not require that specific protein epitopes be detected, further research is needed to determine if their biological response can be applied across species and sources. Injection of streptozotocin (STZ) can generate well-established diabetic mouse models, and these can commonly be used to study the pathogenesis of DN. We took advantage of a diabetic mice to determine the role of HMGB1 on improving the development of DN. Diabetic mice at 28 weeks with severe pathophysiologic alterations observed in the kidney compared to diabetic mice at 20 weeks. As diabetes progressed, the renal tissues showed increasing damage, collagen deposition, and irregular thickening of the basement membrane, indicating that the diabetic mouse model was feasible (Figure 7A). Compared to the control group, mice injected with STZ at 20 and 28 weeks had significantly higher blood glucose levels (Figure 7B), and most importantly,
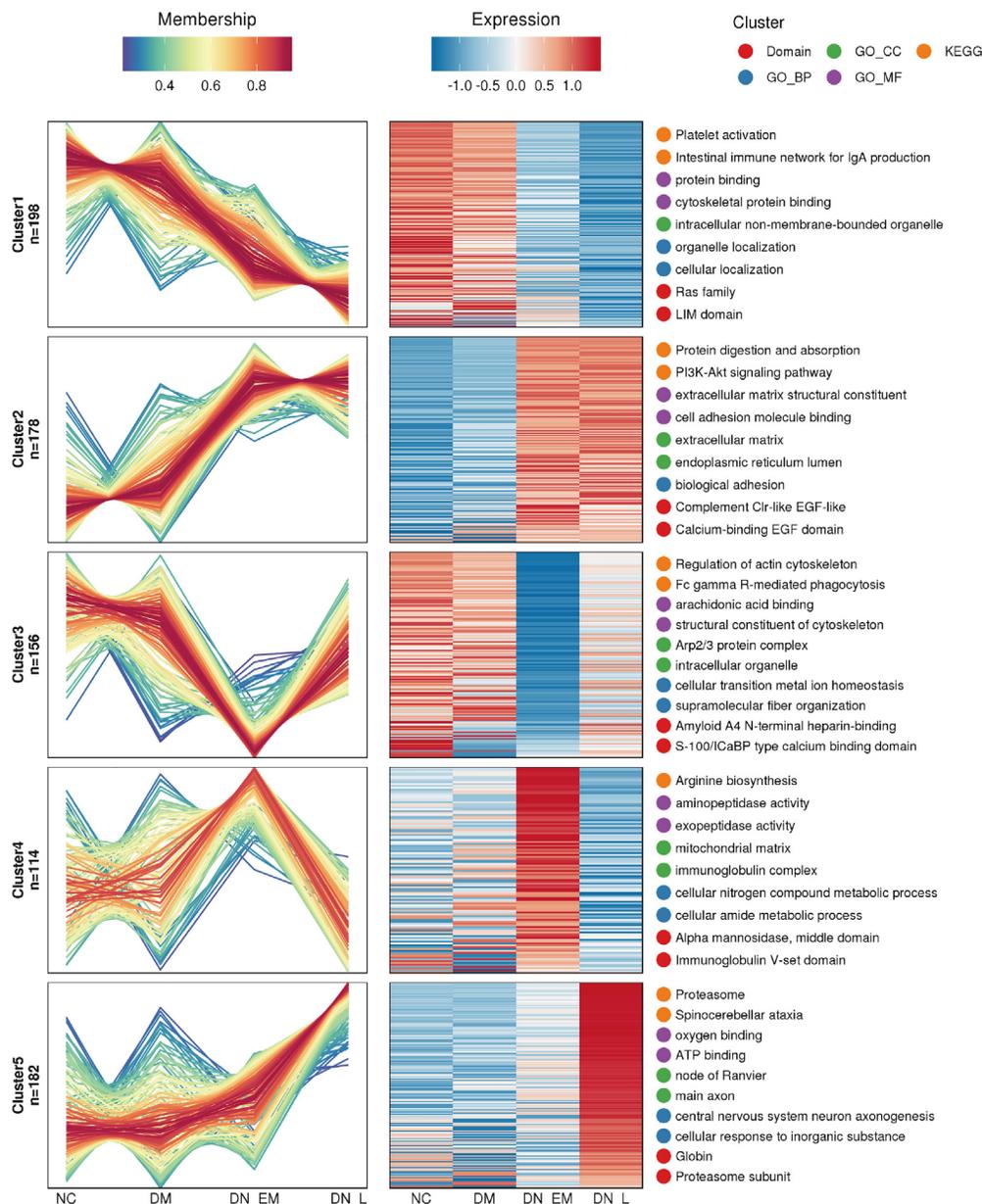
Figure 2. Serum proteome profiling of patients with DN

(A) Heatmap of differential expressed proteins (DEPs) from the proteomics data in NC, DM, DN-EM, and DN-L groups. Red: upregulated proteins; blue: downregulated proteins.

(B–G) Volcano plots of DEPs from the proteomics data for each pairwise comparison (p < 0.05, FC > 1.5 and FC < 1/1.5). Gray: proteins that are not significantly deregulated; red: upregulated proteins; blue: downregulated proteins.

(H) Overview of the increased and decreased DEPs from the proteomics data for each pairwise comparison. Red: upregulated proteins; blue: downregulated proteins.

(I and J) UpSet plots show the intersections between DEPs from the proteomics data for each pairwise comparison (red sets, upregulated proteins; blue sets, downregulated proteins). Circles connected to the intersection indicate which DEPs are included in the intersection, and the size of the intersection is displayed in the main bar (right bars).

Figure 3. Mfuzz analysis reveals different expression patterns of proteins in DN Progression

A protein expression line graph is shown on the left, a heatmap is shown in the middle, and the top 2 enrichment analysis entries are shown on the right. Line graph: the horizontal axis represents the sample, the vertical axis depicts the relative protein expression, a line represents a protein, and the color of the line indicates the affiliation intensity in the cluster. Heatmap: the horizontal axis represents the sample, the vertical axis depicts different proteins, and the heatmap color indicates the relative expression of the protein in the sample. Domain enrichment is red, Gene Ontology Biological Process (GO-BP) enrichment is blue, Gene Ontology Cellular Component (GO-CC) enrichment is green, Gene Ontology Molecular Function (GO-MF) enrichment is purple, and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment is orange.

their serum creatinine (Figure 7C), urea nitrogen (Figure 7D) and urinary microalbumin to creatinine ratio (ACR; Figure 7E) levels, which reflect the change in kidney function, increased gradually with diabetes.

HMGB1, a biomarker we identified as being linked to DN in human serum, also displayed a noteworthy upregulation as diabetic mice progressed through renal tissues. Immunofluorescence (Figure 7F) and Western blot analysis (Figure 7G) revealed that the kidney tissues of diabetic mice with renal injury have higher levels of HMGB1, and that this level increases with the duration of diabetes. Interestingly, the expression of HMGB1 in the kidney tissues of control mice increased with age, although it was not statistically significant (Figure 7G). Additionally, an *in vitro* model of DN renal tubular epithelial cells (HK2 cells) was established. A time-dependent culture of HK2 cells in high-glucose medium
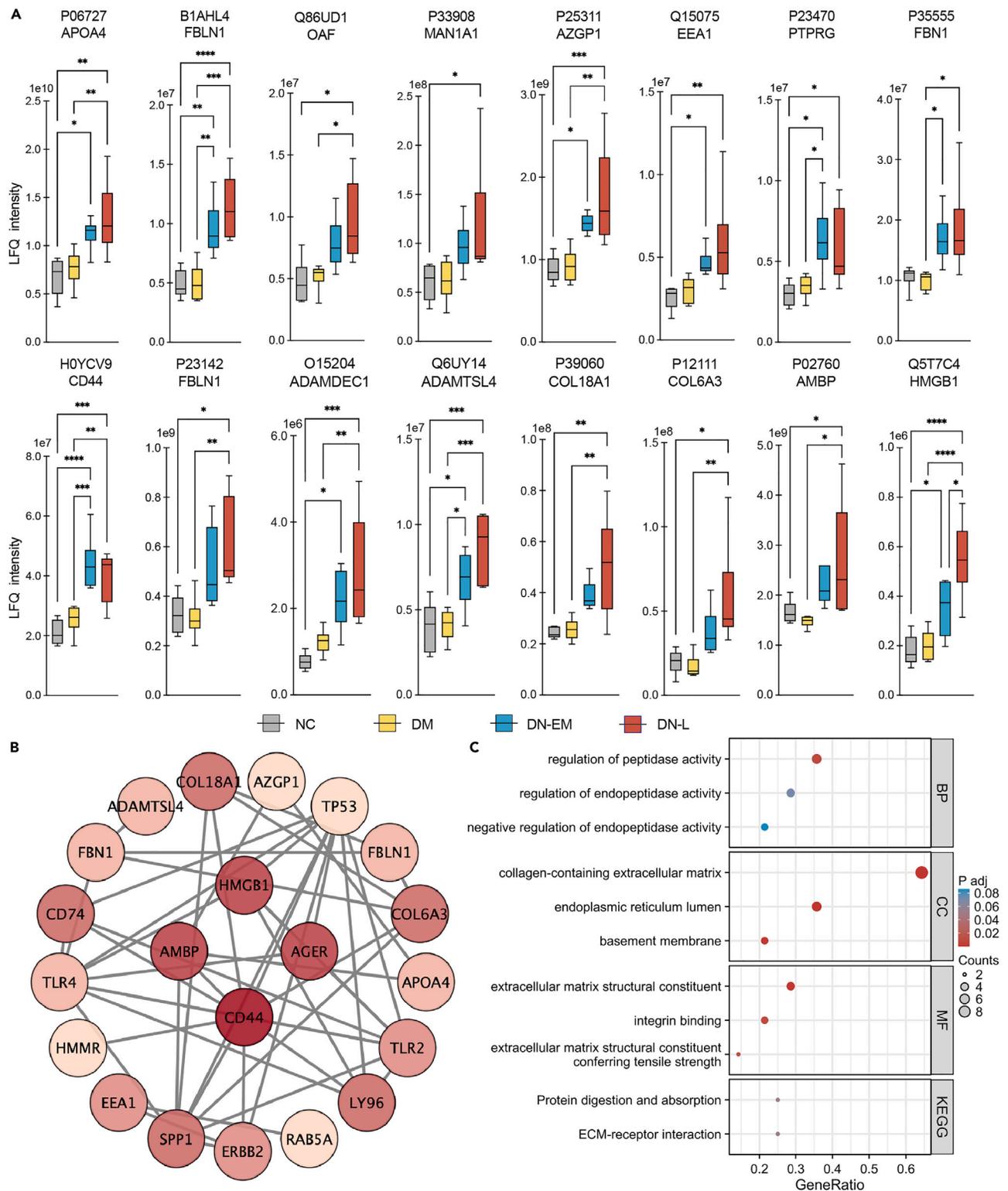
Figure 4. A panel of proteins strongly associated with DN progression

(A) Box-and-whisker plot to visually represent the dispersion of LFQ intensity values for APOA4, FBLN1, OAF, MAN1A1, AZGP1, EEA1, PTPRG, FBN1, CD44, ADAMDEC1, ADAMTSL4, COL18A1, COL6A3, AMBP, and HMGB1 in serum. NC is gray, DM is yellow, DN-EM is blue, DN-L is red. A median line is shown in the middle of the box, the top and bottom represent the upper and lower quartiles, and whiskers indicate the upper and lower limits for outliers. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$ by multiple comparison.

Figure 4. Continued

(B) PPI network showing the interactions of the 15 proteins (shown in Figure 4A) based on STRING. The darker the node, the more core the interaction.

(C) Gene enrichment analysis (BP: GO-BP, CC: GO-CC, MM: GO-MF, KEGG) of 15 proteins (shown in Figure 4A). The horizontal axis represents the ratio of proteins, the vertical axis represents enrichment analysis entries, the color shades represent adjust p-values, and the circle size represents the protein number.

(30 mM) increased HMGB1 protein level at 36 h as epithelial-mesenchymal transition (EMT) changes intensified (Figure 7H). Observing treated HK2 cells for HMGB1 expression by immunofluorescence showed increased expression of HMGB1 in the nucleus and cytoplasm after high-glucose-stimulated HK2 cells (Figure 7I). Our data implicate that HMGB1 is a newly identified marker that we can evaluate as a potential marker to monitor the progress of DN.

DISCUSSION

In this study, comprehensive quantitative proteomics on serum from four independent cohorts was conducted to determine biomarkers of DN progression. First, we quantified thousands of proteins with discovery MS without prior knowledge, which allowed us to identify proteins not previously associated with DN. After protein relative quantification between NC, DM, DN-EM and DN-L patients the elevated expressive variation trend of a total of 15 proteins was further identified in DN progression by Mfuzz clustering analysis. We further identified five proteins from overlapping analysis between rising cluster 2 and WGCNA, which were measured to discriminate between patients who developed DN and those who did not. Based on the results, we explored GeneCard's database of DN scores and found that HMGB1 was the priority biomarker in the development of DN and was tested in both cell and animal models of high glucose.

Although proteomics can be utilized in treating many diseases,[22] early diagnosis of DN by proteomics has been a challenge.[23] Several studies have shown that profiling urinary proteomics can identify novel biomarkers for DN.[24,25] Nevertheless, few trials have been conducted to assess the monitoring value of blood for the course of DN, especially the early diagnosis of DN.[16,17] In early DN diagnosis, proteomics may be limited due to vast heterogeneity and widespread protein abundance in blood, as well as strong proteolytic activity, which can muddle interpretation of the blood proteome.[26] In our study, we filtered out the high abundance of interfering signals using field-asymmetric ion mobility spectrometry (FAIMS), which can selectively identify compounds in a complex background, and detect more key proteins in blood.[27] 1602 proteins were identified by filtering out the high abundance of proteins, of which 1402 could be quantified (Figure 1B). Based on the disease course, PCA analysis of 1402 quantified proteins and heatmap analysis of 484 differentially expressed proteins showed significant protein expression changes in the serum of patients in the DN stage (Figures 1E and 2A), indicating the various molecular alterations induced by DN also affect the expression in serum, especially early stages.[28] Although the serum proteome changed much less both in NC versus DM and DN-EM versus DN-L, the cohort with early DN had clearer differences than the DM cohort, which facilitated finding biomarkers for early DN (Figure 2H).

Variations in eGFRs are known to monitor DN progression, but compensatory changes in the remaining nephrons might overestimate or underestimate the true GFR in the condition.[29] In the present work, we group 4 cohorts: healthy, diabetic, early medium stage, and late stage of DN, to gain a better understanding of serum protein alterations during DN progression. On the basis of this category, we cluster the associated proteins with similar expressive variation trends in DN progression (Figure 3). In Cluster 2, proteomics has increased along with DN progression, and proteins with high expression are generally more detectable. In KEGG pathway enriched proteomics of Cluster 2 (Figures 3 and S2), one of the most prominent signaling pathways was the PI3K-AKT pathway, consistent with previous studies demonstrating that PI3K-AKT signaling is implicated in DN.[30,31] Similar to our result, PI3K/Akt signaling contributes to ECM accumulation, which promotes the progression of renal interstitial fibrosis in DM.[32,33] To differentiate early DN from DM, we further refined screening biomarkers. Overall, 15 biomarkers strongly associated with DN progression were screened in Cluster 2 and mainly related to ECM structure and receptor interaction. In DN progression, ECM proteins are frequently deposited in the mesangium and renal tubule interstices of the glomerulus and basement membranes of patients with DN.[34]

Because DN progression involves multiple factors and has a complex proteins alteration,[35] a comprehensive analysis of multiple deregulated proteins is needed to monitor DN progression.[36] Thus, using WGCAN and Mfuzz, we obtain 5 biomarkers (HMGB1, CD44, FBLN1, PTPRG and ADAMTSL4) that were highly correlated with DN progression (Figure 5). Among the 5 candidate biomarkers, HMGB1, the most promising biomarker relevant to DN, has been verified in both cell and animal models of high glucose (Figure 7). A nonhistone protein HMGB1 is mainly located inside the nucleus of a cell.[37] When the cell is stimulated by various kinds of damage, it releases itself into the extracellular space as a damage-associated molecular pattern (DAMP) molecule that participates in inflammatory responses, differentiation and migration of cells.[38,39] HMGB1, as a pathogenic factor in DN, interacts with TLRs and RAGE, which are its receptors on the cytomembrane, activating innate immune responses by promoting nuclear translocation of transcription factors.[40–42] Interestingly, high levels of HMGB1 in serum from patients with DN can induce podocyte autophagy, apoptosis, and EMT.[43] We speculate that after persistent high glucose stimulation in DN, HMGB1 may be released from the nucleus into the extracellular space, triggering positive feedback through TLRs and RAGE receptors that further exacerbates renal fibrotic factors. However, in order to understand the detailed mechanism, further investigation is required. In addition, studies have shown that HMGB1 is elevated both in DN patients' kidney tissues and in those from DN mice, which is consistent with our findings[44] (Figure 7G). In addition to the HMGB1 mentioned above, CD44 is a cell adhesion molecule, and its ligands are various ECM components.[45,46] CD44 regulates the expression of ECM derived from parietal epithelial cells (PEC) and podocytes in DN.[47] Another candidate biomarker, FBLN1, is a secreted glycoprotein that interacts with ECM proteins to maintain ECM integrity.[48] Exosomal
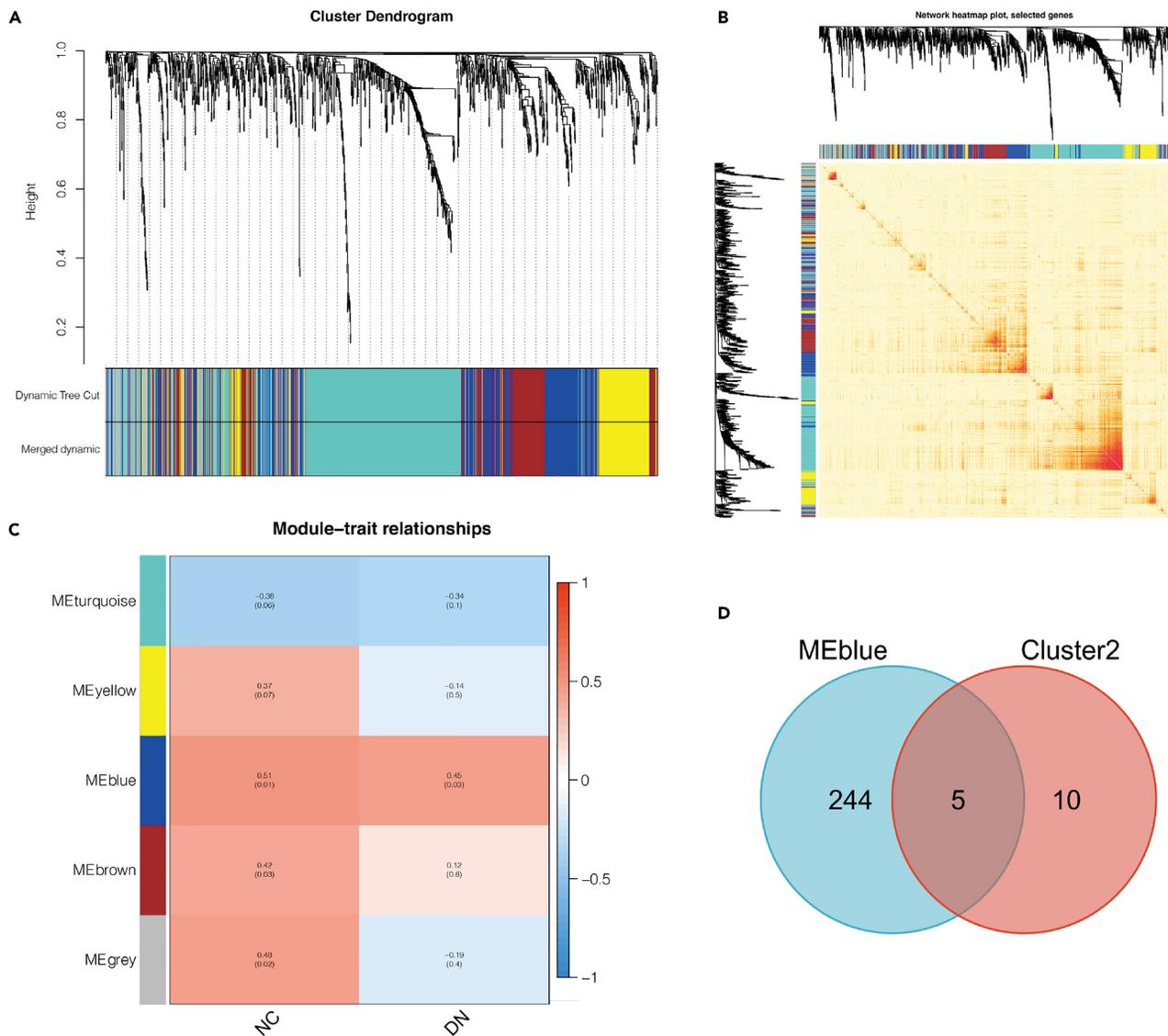
Figure 5. Novel biomarkers associated with DN progression achieved by WGCNA

(A) Cluster dendrogram of the topological overlap between proteins. Highly similar modules are delineated by hierarchical clustering dendrograms and assigned arbitrary colors to each module below.

(B) The dendrogram and network heatmap plot of the topological overlap between proteins. A module membership, color-coded for easy identification, is displayed below and to the right of the dendrograms. The color saturation in yellow and red indicates a greater degree of interconnection between co-expressions.

(C) Heatmap of the module-trait relationship between the module eigengene E and the trait DN. An eigengene corresponds to each row, and a trait to each column.

(D) Venn diagram of DEPs in the blue module and 15 proteins (shown in Figure 4A) from cluster 2.

FBLN1 promotes DN progression through inducing EMT in the proximal renal tubules.[49] Patients with CKD and T2DM showed an increased risk of cardiovascular events when their circulating FBLN1 levels were elevated.[50] The remaining two biomarkers, PTPRG and ADAMTSL4, have not been associated with DN, similar to what we found in our results. Nevertheless, to determine whether these proteins are diagnostic biomarkers for DN progression, additional large-scale investigations are needed.

Our research showed that HMGB1 was highly correlated with several renal function indicators (Figures 6B–6F) and could be a good marker for the early detection of DN, especially distinguishing healthy control and DN patients (Figure 6A). Based on our results, HMGB1 may be identified as a potential novel biomarker for DN progression. Of course, as a single biomarker, HMGB1 also has limitations. As the age confounder factor could not be excluded in DN patients (Table S4; β = 8552.354, p = 0.011), the expression of HMGB1 in the kidney tissues of control mice increased with age, which hinted that the elevation of HMGB1 may be related to age factors (Figure 7G). Furthermore, HMGB1

Table 2. The basic information of shared proteins between MEblue and Cluster2

Protein accession	Protein description	Gene name	Subcellular locations from UniProtKB/Swiss-Prot	Relevance scores from Genecards
Q5T7C4	High-mobility group protein B1	HMGB1	Nucleus. Chromosome. Cytoplasm. Secreted. Cell membrane. Peripheral membrane protein. Extracellular side. Endosome. Endoplasmic reticulum-Golgi intermediate compartment. Endoplasmic reticulum.	10.5875282287598
H0YCV9	CD44 antigen	CD44	Cell membrane. Single-pass type I membrane protein. Cell projection, microvillus.	6.28472328186035
B1AHL2/P23142	Fibulin-1	FBLN1	Secreted, extracellular space, extracellular matrix.	2.0302460193634
P23470	Receptor-type tyrosine-protein phosphatase gamma	PTPRG	Membrane. Single-pass type I membrane protein.	None
Q6UY14	ADAMTS-like protein 4	ADAMTSL4	Secreted, extracellular space, extracellular matrix.	None

proteins could serve as a potential noninvasive biomarker for several inflammation-related diseases or tumors.[51–53] Inflammation is also crucial to DN progression,[54] which seems that HMGB1 might serve as a general biomarker of early inflammation. By combining other biomarkers, the monitoring accuracy of HMGB1 in DN progression might improve. Intriguingly, a candidate biomarker in our study is CD44, which is increased by extracellular HMGB1 in tumor progression.[55] However, it is not clarified whether HMGB1 and CD44 can be combined to monitor DN progression, and whether HMGB1 increases CD44 to regulate it. Thus, further clinical sample prediction models and mechanism studies are urgently needed to clarify the role of HMGB1 as a potential biomarker for early DN progression.

A novel aspect of this study is that, based on the stages of DN patients and clustering analyses of proteins with similar altered trends in progression, we were able to identify biomarkers in serum that can reflect DN progression more accurately. A validated biological response of HMGB1 was found across species and sources, making it a reliable biomarker for DN progression, but clinical translation needs more exploration.

In conclusion, we investigated the proteomes of patients with DM or different stages of DN and healthy control by quantitative proteomics to gain an understanding of serum protein alterations during DN progression. Five promising biomarkers, HMGB1, CD44, FBLN1, PTPRG, and ADAMTSL4, allowed monitoring of the progression of DN, whereas HMGB1 was highly correlated with renal function alterations and could be an appropriate marker for the early detection of DN, especially distinguishing healthy controls and DN patients. Although there is insufficient evidence to conclude that these biomarkers can replace invasive diagnostics for DN, with further research, these proteomics changes may help clinicians identify DN in the early stages.

Limitations of the study

Despite our findings, this study has several limitations. First, as this study only involves a small number of patients, it is necessary to validate the results with additional patients within each cohort. Statistics show significant differences in all results, and with increased sample size, the difference in HMGB1 elevation in serum of DN patients may increase. Second, there were not enough clinical data collected in this retrospective study, so no adjustments for clinical covariates or pathological confounders were made. Third, the HMGB1 is an inflammation-associated molecule, and many proinflammatory factors must be considered before it can be used as a biomarker for DN. Thus, more research cohorts with DN are needed to validate the biomarker. Fourth, the findings from our study were limited to a single hospital, which may limit their generalizability. Finally, nondiabetic chronic kidney disease was not profiled for serum proteomics, so we cannot infer that the serum protein alterations in DN are exclusively due to diabetes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
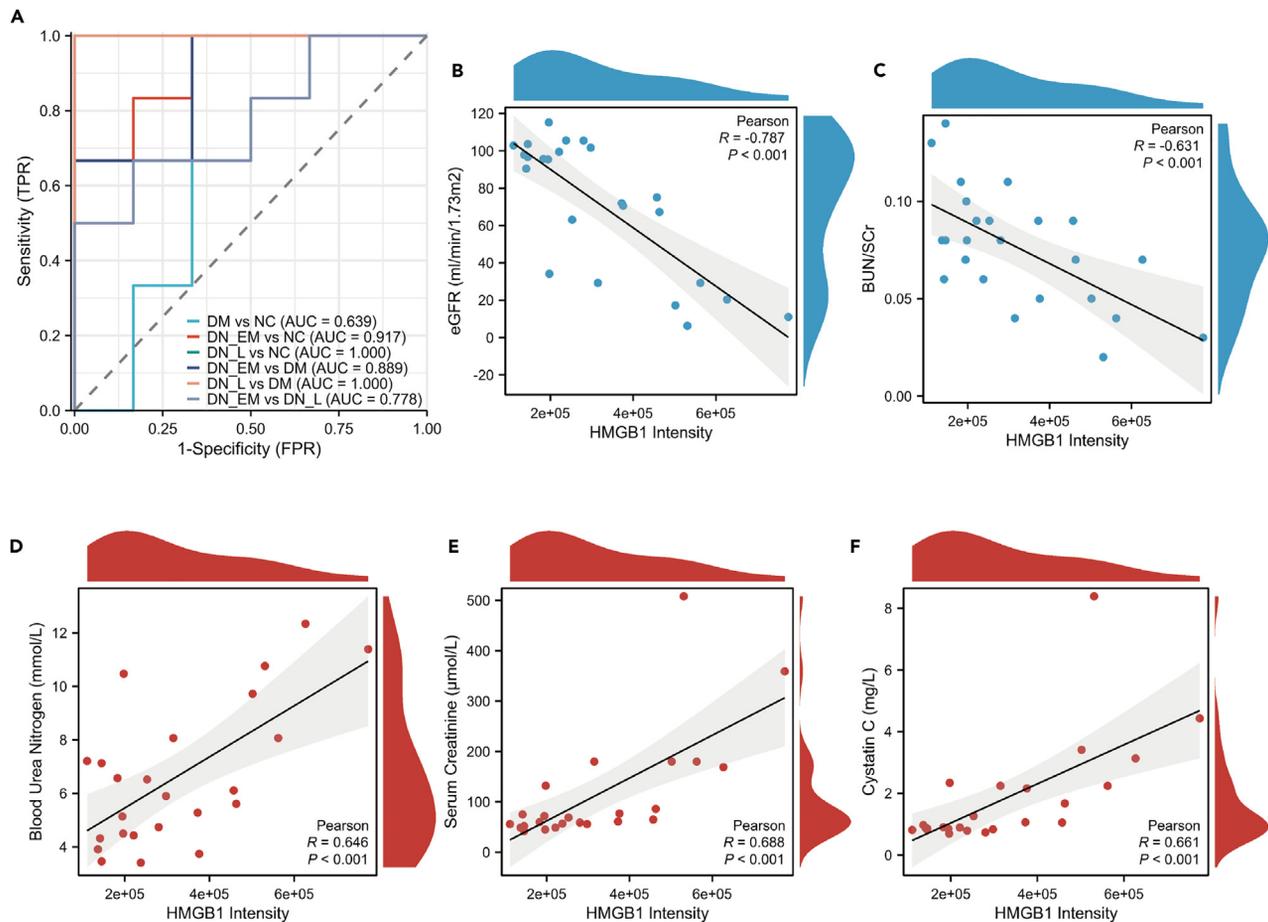 - Clinical samples

Figure 6. The diagnosis values of HMGB1 in DN progression

(A) Receiver-operating characteristic (ROC) curve of HMGB1 for each pairwise comparison. AUC, area under the ROC curve. AUC has low diagnostic accuracy at 0.5 to 0.7, certain diagnostic accuracy at 0.7 to 0.9, and high diagnostic accuracy at 0.9 or more.

(B–F) Pearson correlation between intensity of HMGB1 and eGFR, blood urea nitrogen to serum creatinine ratio (BUN/Scr), blood urea nitrogen, serum creatinine, and cystatin c, respectively. R > 0 represents positively correlated, and R < 0 represents inversely correlated.

- ○ Inclusion and exclusion criteria
- ○ Animals and procedures
- ○ Cell culture
- ● METHOD DETAILS
  - ○ Sample collection and proteomics assays
  - ○ Proteomics data analysis
  - ○ Receiver operating characteristic (ROC) and correlation analysis
  - ○ Immunoblotting
  - ○ Histology analysis
  - ○ Biochemical assays
  - ○ Immunofluorescent staining
- ● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.108834.
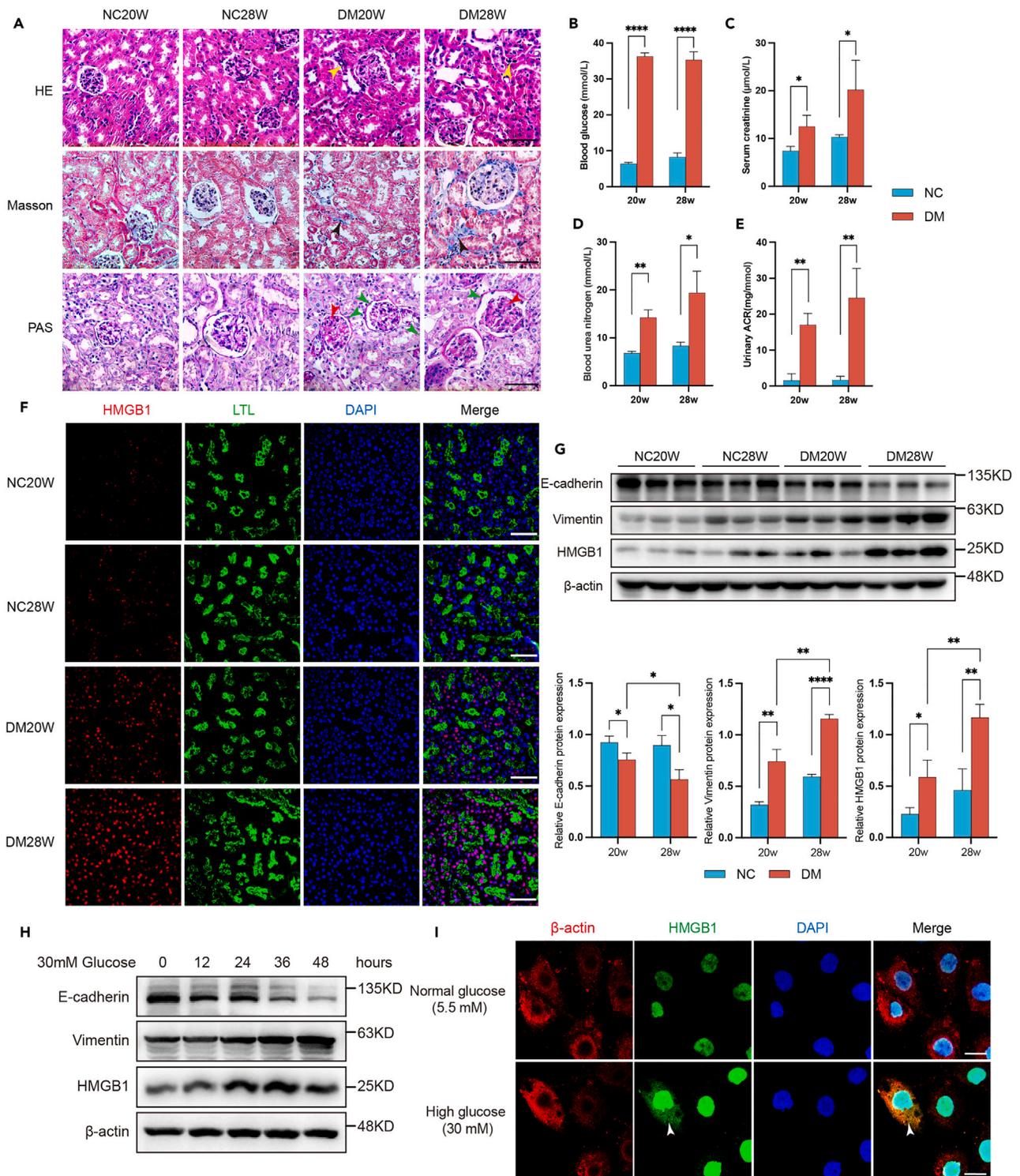
Figure 7. Validation of HMGB1 in DN model

(A) Hematoxylin and eosin staining (HE), Masson trichrome staining, and periodic acid-Schiff staining (PAS) staining show typical DN changes in diabetic mice; scale bars = 50 μm; inflammatory cell infiltration (yellow arrow), collagen deposition (black arrow), and irregular thickening of the glomerular basement membrane (red arrow) and glomerular capsule (green arrow). NC20W and NC28W is normal mice at 20 and 28 weeks, DM20W and DM28W is diabetic mice at 20 and 28 weeks.

(B–E) Blood glucose, serum creatinine, and blood urea nitrogen levels and urine albumin to creatinine ratio of normal mice (blue) and diabetic mice (red).

Figure 7. Continued

(F) Immunofluorescence staining for HMGB1 (red) and lotus tetragonolobus lectin (LTL, green, the marker of renal tubular brush brush) in the kidney tissues of mice; scale bars = 50 μm.

(G) In the analysis of kidney tissue from mice, Western blotting was utilized to ascertain the expression levels of HMGB1 as well as proteins associated with epithelial-mesenchymal transition (EMT); quantification of relative levels of protein expression.

(H) Western blotting was employed to assess the expression of HMGB1 and proteins indicative of EMT in HK2 cells stimulated with high glucose.

(I) Immunofluorescence staining for β-actin (red) and HMGB1 (green) in high-glucose-stimulated HK2 cells; white arrow is cytoplasm; scale bars = 20 μm. All experiments are performed three times and data were expressed as mean ± standard deviation (SD). *p < 0.05, **p < 0.01 and ***p < 0.001 by Student's *t* test.

AUTHOR CONTRIBUTIONS

Conceptualization: L.L. Writing – original draft: R.P. Methodology: R.P. and S.Z. Software: R.P. and S.Z. Formal analysis: R.P. Visualization: R.P. Writing – review and editing: S.Z., X.L., and L.L. Funding acquisition: L.L., X.L., and Q.Z. Sample collection: Y.H. Data collection: X.L. and Y.H. Construct animal model: S.C., X.Z., and H.L. Cell model experiments: M.C., Y.Y., and H.Y. Supervision: L.L. and B.G. Project administration: L.L. and B.G.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Cho, N.H., Shaw, J.E., Karuranga, S., Huang, Y., da Rocha Fernandes, J.D., Ohlrogge, A.W., and Malanda, B. (2018). IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res. Clin. Pract. *138*, 271–281.

2. Afkarian, M., Zelnick, L.R., Hall, Y.N., Heagerty, P.J., Tuttle, K., Weiss, N.S., and de Boer, I.H. (2016). Clinical Manifestations of Kidney Disease Among US Adults With Diabetes. JAMA *316*, 602–610.

3. Fu, H., Liu, S., Bastacky, S.I., Wang, X., Tian, X.J., and Zhou, D. (2019). Diabetic kidney diseases revisited: A new perspective for a new era. Mol. Metab. *30*, 250–263.

4. Gonzalez Suarez, M.L., Thomas, D.B., Barisoni, L., and Fornoni, A. (2013). Diabetic nephropathy: Is it time yet for routine kidney biopsy? World J. Diabetes *4*, 245–255.

5. Califf, R.M. (2018). Biomarker definitions and their applications. Exp. Biol. Med. *243*, 213–221.

6. Group, F.-N.B.W. (2016). In BEST (Biomarkers, EndpointS, and other Tools) Resource (Food and Drug Administration (US) National Institutes of Health (US)).

7. Satirapoj, B., and Adler, S.G. (2014). Comprehensive approach to diabetic nephropathy. Kidney Res. Clin. Pract. *33*, 121–131.

8. Satirapoj, B., and Adler, S.G. (2015). Prevalence and Management of Diabetic Nephropathy in Western Countries. Kidney Dis. *1*, 61–70.

9. Parving, H.H., Oxenbøll, B., Svendsen, P.A., Christiansen, J.S., and Andersen, A.R. (1982). Early detection of patients at risk of developing diabetic nephropathy. A longitudinal study of urinary albumin excretion. Acta Endocrinol. *100*, 550–555.

10. Fioretto, P., Steffes, M.W., and Mauer, M. (1994). Glomerular structure in nonproteinuric IDDM patients with various levels of albuminuria. Diabetes *43*, 1358–1364.

11. Barutta, F., Bellini, S., Canepa, S., Durazzo, M., and Gruden, G. (2021). Novel biomarkers of diabetic kidney disease: current status and potential clinical application. Acta Diabetol. *58*, 819–830.

12. Niu, L., Geyer, P.E., Wewer Albrechtsen, N.J., Gluud, L.L., Santos, A., Doll, S., Treit, P.V., Holst, J.J., Knop, F.K., Vilsbøll, T., et al. (2019). Plasma proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease. Mol. Syst. Biol. *15*, e8793.

13. Liu, Y., Buil, A., Collins, B.C., Gillet, L.C.J., Blum, L.C., Cheng, L.Y., Vitek, O., Mouritsen, J., Lachance, G., Spector, T.D., et al. (2015). Quantitative variability of 342 plasma proteins in a human twin population. Mol. Syst. Biol. *11*, 786.

14. Campion, C.G., Sanchez-Ferras, O., and Batchu, S.N. (2017). Potential Role of Serum and Urinary Biomarkers in Diagnosis and Prognosis of Diabetic Nephropathy. Can. J. Kidney Health Dis. *4*, 2054358117705371.

15. Thippakorn, C., Schaduangrat, N., and Nantasenamat, C. (2018). Proteomic and bioinformatic discovery of biomarkers for diabetic nephropathy. EXCLI J *17*, 312–330.

16. Colombo, M., Valo, E., McGurnaghan, S.J., Sandholm, N., Blackbourn, L.A.K., Dalton, R.N., Dunger, D., Groop, P.H., McKeigue, P.M., Forsblom, C., et al. (2019). Biomarker panels associated with progression of renal disease in type 1 diabetes. Diabetologia *62*, 1616–1627.

17. Liu, S., Gui, Y., Wang, M.S., Zhang, L., Xu, T., Pan, Y., Zhang, K., Yu, Y., Xiao, L., Qiao, Y., et al. (2021). Serum integrative omics reveals the landscape of human diabetic kidney disease. Mol. Metab. *54*, 101367.

18. Kidney Disease: Improving Global Outcomes KDIGO Diabetes Work Group (2020). KDIGO 2020 Clinical Practice Guideline for Diabetes Management in Chronic Kidney Disease. Kidney Int. *98*, S1–S115.

19. Futschik, M.E., and Carlisle, B. (2005). Noise-robust soft clustering of gene expression time-course data. J. Bioinform. Comput. Biol. *3*, 965–988.

20. Gaiteri, C., Ding, Y., French, B., Tseng, G.C., and Sibille, E. (2014). Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. Genes Brain Behav. *13*, 13–24.

21. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559.

22. Dubin, R.F., and Rhee, E.P. (2020). Proteomics and Metabolomics in Kidney Disease, including Insights into Etiology, Treatment, and Prevention. Clin. J. Am. Soc. Nephrol. *15*, 404–411.

23. Tofte, N., Persson, F., and Rossing, P. (2020). Omics research in diabetic kidney disease: new biomarker dimensions and new understandings? J. Nephrol. *33*, 931–948.

24. Tofte, N., Lindhardt, M., Adamova, K., Bakker, S.J.L., Beige, J., Beulens, J.W.J., Birkenfeld, A.L., Currie, G., Delles, C., Dimos, I., et al. (2020). Early detection of diabetic kidney disease by urinary proteomics and subsequent intervention with spironolactone to delay progression (PRIORITY): a prospective observational study and embedded randomised placebo-controlled trial. Lancet Diabetes Endocrinol. *8*, 301–312.

25. Guillén-Gómez, E., Bardají-de-Quixano, B., Ferrer, S., Brotons, C., Knepper, M.A., Carrascal, M., Abian, J., Mas, J.M., Calero, F., Ballarín, J.A., and Fernández-Llama, P. (2018). Urinary Proteome Analysis Identified Neprilysin and VCAM as Proteins Involved in Diabetic Nephropathy. J. Diabetes Res. *2018*, 6165303.

26. Kolch, W., Neusüss, C., Pelzing, M., and Mischak, H. (2005). Capillary electrophoresis-mass spectrometry as a powerful tool in

clinical diagnosis and biomarker discovery. Mass Spectrom. Rev. *24*, 959–977.

27. Covington, J.A., van der Schee, M.P., Edge, A.S.L., Boyle, B., Savage, R.S., and Arasaradnam, R.P. (2015). The application of FAIMS gas analysis in medical diagnostics. Analyst *140*, 6775–6781.

28. Swaminathan, S.M., Rao, I.R., Shenoy, S.V., Prabhu, A.R., Mohan, P.B., Rangaswamy, D., Bhojaraja, M.V., Nagri, S.K., and Nagaraju, S.P. (2023). Novel biomarkers for prognosticating diabetic kidney disease progression. Int. Urol. Nephrol. *55*, 913–928.

29. Colhoun, H.M., and Marcovecchio, M.L. (2018). Biomarkers of diabetic kidney disease. Diabetologia *61*, 996–1011.

30. Dong, R., Zhang, X., Liu, Y., Zhao, T., Sun, Z., Liu, P., Xiang, Q., Xiong, J., Du, X., Yang, X., et al. (2023). Rutin alleviates EndMT by restoring autophagy through inhibiting HDAC1 via PI3K/AKT/mTOR pathway in diabetic kidney disease. Phytomedicine *112*, 154700.

31. Wang, X., Jiang, L., Liu, X.Q., Huang, Y.B., Wang, A.L., Zeng, H.X., Gao, L., Zhu, Q.J., Xia, L.L., and Wu, Y.G. (2022). Paeoniflorin binds to VEGFR2 to restore autophagy and inhibit apoptosis for podocyte protection in diabetic kidney disease through PI3K-AKT signaling pathway. Phytomedicine *106*, 154400.

32. Zhang, Y., Jin, D., Kang, X., Zhou, R., Sun, Y., Lian, F., and Tong, X. (2021). Signaling Pathways Involved in Diabetic Renal Fibrosis. Front. Cell Dev. Biol. *9*, 696542.

33. Li, F., Li, L., Hao, J., Liu, S., and Duan, H. (2017). Src Homology 2 Domain-Containing Inositol 5′-Phosphatase Ameliorates High Glucose-Induced Extracellular Matrix Deposition via the Phosphatidylinositol 3-Kinase/Protein Kinase B Pathway in Renal Tubular Epithelial Cells. J. Cell. Biochem. *118*, 2271–2284.

34. Tuleta, I., and Frangogiannis, N.G. (2021). Diabetic fibrosis. Biochim. Biophys. Acta Mol. Basis Dis. *1867*, 166044.

35. Tuttle, K.R., Agarwal, R., Alpers, C.E., Bakris, G.L., Brosius, F.C., Kolkhof, P., and Uribarri, J. (2022). Molecular mechanisms and therapeutic targets for diabetic kidney disease. Kidney Int. *102*, 248–260.

36. Chen, J., Luo, S.F., Yuan, X., Wang, M., Yu, H.J., Zhang, Z., and Yang, Y.Y. (2022). Diabetic kidney disease-predisposing proinflammatory and profibrotic genes identified by weighted gene co-expression network analysis (WGCNA). J. Cell. Biochem. *123*, 481–492.

37. Amato, J., Cerofolini, L., Brancaccio, D., Giuntini, S., Iaccarino, N., Zizza, P., Iachettini, S., Biroccio, A., Novellino, E., Rosato, A., et al.

(2019). Insights into telomeric G-quadruplex DNA recognition by HMGB1 protein. Nucleic Acids Res. *47*, 9950–9966.

38. Niu, Z., Lin, J., Hao, C., Xu, X., Wang, C., Dai, K., Deng, X., Deng, M., Guo, Y., and Yao, W. (2022). Glycyrrhizic Acid Attenuates Pulmonary Fibrosis of Silicosis by Inhibiting the Interaction between HMGB1 and BRG1 through PI3K/Akt/mTOR Pathway. Int. J. Environ. Res. Public Health *19*, 8743.

39. Czura, C.J., Wang, H., and Tracey, K.J. (2001). Dual roles for HMGB1: DNA binding and cytokine. J. Endotoxin Res. *7*, 315–321.

40. Chen, X., Ma, J., Kwan, T., Stribos, E.G.D., Messchendorp, A.L., Loh, Y.W., Wang, X., Paul, M., Cunningham, E.C., Habib, M., et al. (2018). Blockade of HMGB1 Attenuates Diabetic Nephropathy in Mice. Sci. Rep. *8*, 8319.

41. Zhou, B., Li, Q., Wang, J., Chen, P., and Jiang, S. (2019). Ellagic acid attenuates streptozocin induced diabetic nephropathy via the regulation of oxidative stress and inflammatory signaling. Food Chem. Toxicol. *123*, 16–27.

42. Zhang, J., Yang, X., Zhang, X., Lu, D., and Guo, R. (2021). Electro-Acupuncture Protects Diabetic Nephropathy-Induced Inflammation Through Suppression of NLRP3 Inflammasome in Renal Macrophage Isolation. Endocr., Metab. Immune Disord. Drug Targets *21*, 2075–2083.

43. Jin, J., Gong, J., Zhao, L., Zhang, H., He, Q., and Jiang, X. (2019). Inhibition of high mobility group box 1 (HMGB1) attenuates podocyte apoptosis and epithelial-mesenchymal transition by regulating autophagy flux. J. Diabetes *11*, 826–836.

44. Zhang, Y. (2021). MiR-92d-3p suppresses the progression of diabetic nephropathy renal fibrosis by inhibiting the C3/HMGB1/TGF-β1 pathway. Biosci. Rep. *41*.

45. Aruffo, A., Stamenkovic, I., Melnick, M., Underhill, C.B., and Seed, B. (1990). CD44 is the principal cell surface receptor for hyaluronate. Cell *61*, 1303–1313.

46. Goodison, S., Urquidi, V., and Tarin, D. (1999). CD44 cell adhesion molecules. Mol. Pathol. *52*, 189–196.

47. Chan, G.C., Eng, D.G., Miner, J.H., Alpers, C.E., Hudkins, K., Chang, A., Pippin, J.W., and Shankland, S.J. (2019). Differential expression of parietal epithelial cell and podocyte extracellular matrix proteins in focal segmental glomerulosclerosis and diabetic nephropathy. Am. J. Physiol. Renal Physiol. *317*, F1680–F1694.

48. Segade, F. (2010). Molecular evolution of the fibulins: implications on the functionality of the elastic fibulins. Gene *464*, 17–31.

49. Tsai, Y.C., Hung, W.W., Chang, W.A., Wu, P.H., Wu, L.Y., Lee, S.C., Kuo, M.C., and Hsu, Y.L. (2021). Autocrine Exosomal Fibulin-1 as a Target of MiR-1269b Induces Epithelial-Mesenchymal Transition in Proximal Tubule in Diabetic Nephropathy. Front. Cell Dev. Biol. *9*, 789716.

50. Scholze, A., Bladbjerg, E.M., Sidelmann, J.J., Diederichsen, A.C.P., Mickley, H., Nybo, M., Argraves, W.S., Marckmann, P., and Rasmussen, L.M. (2013). Plasma concentrations of extracellular matrix protein fibulin-1 are related to cardiovascular risk markers in chronic kidney disease and diabetes. Cardiovasc. Diabetol. *12*, 6.

51. Paudel, Y.N., Shaikh, M.F., Chakraborti, A., Kumari, Y., Aledo-Serrano, Á., Aleksovska, K., Alvim, M.K.M., and Othman, I. (2018). HMGB1: A Common Biomarker and Potential Target for TBI, Neuroinflammation, Epilepsy, and Cognitive Dysfunction. Front. Neurosci. *12*, 628.

52. Magna, M., and Pisetsky, D.S. (2014). The role of HMGB1 in the pathogenesis of inflammatory and autoimmune diseases. Mol. Med. *20*, 138–146.

53. Wang, Y., Jiang, Z., Yan, J., and Ying, S. (2019). HMGB1 as a Potential Biomarker and Therapeutic Target for Malignant Mesothelioma. Dis. Markers *2019*, 4183157.

54. Rayego-Mateos, S., Morgado-Pascual, J.L., Opazo-Ríos, L., Guerrero-Hue, M., García-Caballero, C., Vázquez-Carballo, C., Mas, S., Sanz, A.B., Herencia, C., Mezzano, S., et al. (2020). Pathogenic Pathways and Therapeutic Approaches Targeting Inflammation in Diabetic Nephropathy. Int. J. Mol. Sci. *21*, 3798.

55. Li, J., Ren, H., Wang, J., Zhang, P., and Shi, X. (2021). Extracellular HMGB1 promotes CD44 expression in hepatocellular carcinoma via regulating miR-21. Aging (Albany NY) *13*, 8380–8395.

56. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. *50*, D543–D552.

57. Xiao, L., Xu, X., Zhang, F., Wang, M., Xu, Y., Tang, D., Wang, J., Qin, Y., Liu, Y., Tang, C., et al. (2017). The mitochondria-targeted antioxidant MitoQ ameliorated tubular injury mediated by mitophagy in diabetic kidney disease via Nrf2/PINK1. Redox Biol. *11*, 297–311.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-HMGB1	Abcam	ab79823; RRID:AB_1603373
Anti-E-cadherin	Proteintech	20874-1-AP; RRID:AB_10697811
Anti-Vimentin	Proteintech	10366-1-AP; RRID:AB_2273020
Anti-β-actin	Proteintech	66009-1-Ig; RRID:AB_2687938
Anti-rabbit IgG, HRP-linked Antibody	Cell Signaling Technology	7074; RRID:AB_2099233
Anti-mouse IgG, HRP-linked Antibody	Cell Signaling Technology	7076; RRID:AB_330924
Lotus tetragonolobus lectin (LTL)	Vector Laboratories	B-1325; RRID:AB_2336558
FITC Goat Anti-Rabbit IgG (H + L) Antibody	ApexBio Technology	K1203
Cy3 Goat Anti-Mouse IgG (H + L) Antibody	ApexBio Technology	K1207
Cy3 Goat Anti-Rabbit IgG (H + L) Antibody	ApexBio Technology	K1209
Biological samples		
Serum samples from patients	Affiliated Hospital of Guizhou Medical University	This paper
Chemicals, peptides, and recombinant proteins		
Streptozotocin (STZ)	Solarbio	S8050
Dulbecco's modified Eagle's/F12 medium	Gibco	C11330500BT
Fetal bovine serum	Gibco	2497736
D-glucose	Solarbio	G8150
phenylmethanesulfonyl fluoride (PMSF)	Solarbio	P0100
Enhanced chemiluminescence reagents	Absin	abs920
Critical commercial assays		
Pierce™ Top 12 Abundant Protein Depletion Spin Columns Kit	Thermo Fisher scientific	85164
BCA Protein Assay Kit	Beyotime	P0011
Hematoxylin and eosin staining Kit	Solarbio	G1220
Masson trichrome staining Kit	Solarbio	G1346
Periodic acid-schiff staining Kit	Solarbio	G1281
Deposited data		
Mass spectrometry data	ProteomeXchange	PXD047872
Experimental models: Cell lines		
HK-2 cells	Procell Life Science&Technology	Cat# CL-0109; RRID:CVCL_0302
Experimental models: Organisms/strains		
C57BL/6	Charles River	RRID:MGI:2159769
Software and algorithms		
STRING	N/A	http://string-db.org RRID:SCR_005223
Cytoscape (version 3.9.1)	Open source	https://cytoscape.org RRID:SCR_003032
GraphPad Prism	GraphPad Software	RRID:SCR_002798

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ImageJ	National Institutes of Health	RRID:SCR_003070
R (version 4.2.1)	R Project	https://www.r-project.org
SPSS 26.0	IBM	RRID:SCR_002865
Other		
polyvinylidene difluoride (PVDF) membranes	Millipore	ISEQ00010

RESOURCE AVAILABILITY

Lead contact

For additional details and inquiries regarding resources and reagents, please contact the primary point of contact, Lirong Liu (e-mail: lirongliu@gmc.edu.cn).

Materials availability

In this study, no novel reagents were developed.

Data and code availability

- The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[56] partner repository with the dataset identifier PXD047872.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Clinical samples

The study population was all Han Chinese, comprising healthy control (NC, n = 33, 54.67 ± 18.24 years, male: 10, female: 23), diabetic (DM, n = 34, 57.03 ± 16.07 years, male: 20, female: 14), and diabetic nephropathy (DN, n = 29, 62.10 ± 10.69 years, male: 16, female: 13) patients enrolled from July 2021 to January 2022 at the Affiliated Hospital of Guizhou Medical University. In accordance with the Declaration of Helsinki 1975, this study was approved by the Ethics Committees of the Affiliated Hospital of Guizhou Medical University (Ethics Approval No.2023398). Consent forms have been signed by all participants for the extra samples to be used for academic research.

Inclusion and exclusion criteria

Inclusion criteria include (1) all patients except the NC group fulfilling the diagnostic criteria of DM; (2) for NC and DM group, eGFR ≥ 90 mL/min/1.73 m^2 without kidney disease; (3) for DN group, with kidney disease and diabetic retinopathy, eGFR less than 90 mL/min/1.73 m^2.

Exclusion criteria include (1) other urinary diseases as hereditary kidney disease, urinary tract infection, stones or obstruction; (2) systemic diseases, atherosclerosis, severe hypertension, thrombotic microangiopathy, active or chronic infectious diseases; (3) complication of serious diseases as well as a tumor; (4) pregnant woman.

Animals and procedures

We sourced male C57BL/6 mice (6–8 weeks old) bred under specific pathogen-free (SPF) conditions from Charles River Co. Ltd (Beijing, China). All mice were randomly grouped (n = 3), and DM mice were injected with STZ (Solarbio, S8050) through the intraperitoneal at a dose of 55 mg/kg/day, while NC mice received solvent injections of the same volume. Experiments were conducted on mice with random blood glucose ≥ 16.7 mmol/L and positive urine glucose. At 20 and 28 weeks, mice were killed, then kidney tissues, blood and urine were collected. All animal experiments were approved by China's National Health and Medical Research Council's Code for the Care (No. 1800353).

Cell culture

Human renal proximal tubule epithelial cells (HK2) were cultured in Dulbecco's modified Eagle's/F12 medium (DMEM/F12, 1:1) under standard conditions of 37°C and 5% CO2. All mediums were supplemented with 10% fetal bovine serum (Gibco-BRL) and 1% penicillin/streptomycin. HK2 cells were cultured with 30 mM D-glucose (HG; Solarbio, G8150) mediums at different time point.[57]

METHOD DETAILS

Sample collection and proteomics assays

Serum samples from fasting blood were collected by centrifugation at 3500 rpm for 10 min at room temperature and stored at $-80°C$ until assayed. The cell debris was removed by centrifuged at 12000 g for 10 min at 4°C, then depleted of high-abundance proteins using the Pierce Top 12 Abundant Protein Depletion Spin Columns Kit (Thermo Fisher Scientific, 85164) according to the manufacturer's instructions. Protein concentration was determined using BCA kit (Beyotime, P0011).

An equivalent quantity of protein was utilized for enzymatic hydrolysis in every sample. The volume was standardized, and subsequently, dithiothreitol (DTT) was introduced to achieve a final concentration of 5 mM. This mixture was then subjected to reduction at 56°C for a duration of 30 min. Subsequent to that, iodoacetamide (IAA) was incorporated to attain a final concentration of 11 mM. This concoction was left to incubate in darkness for a duration of 15 min at room temperature. Afterward, the samples were carefully transferred into ultrafiltration tubes and subjected to centrifugation at 12000g for a span of 20 min, all at room temperature. Following this centrifugation step, the samples underwent a thorough washing process, involving three successive washes with 8M urea, followed by an additional three rounds of washing with the replacement buffer. In the final step, trypsin was incorporated at a proportion of 1:50 (trypsin to protein, m/m) to facilitate an overnight digestion process. Following centrifugation at 12000g for a duration of 10 min at room temperature to recover the peptides, they were also recovered using ultrapure water once and then mixed together.

For the examination of serum samples, a liquid chromatography-tandem mass spectrometry (LC-MS/MS) approach was employed. This involved utilizing an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) in tandem with an EASY-nLC 1200 Ultra high-performance liquid phase system (UHPLC, Thermo Fisher Scientific). The mass spectrometry was used to analyze the ionized peptides after they had been separated by the UHPLC and introduced into the NSI ion source for ionization. Mass spectrometry was set for primary scan range of 400–1200 m/z, at 60000 resolution; secondary scan range of 100 m/z, at 30000 resolution, without TurboTMT. Data-dependent scanning (DDA) software was utilized in the data collecting mode. Automatic gain control (AGC) was set to 75% to maximize the effective use of the mass spectrum. The signal threshold was defined at 1E4 ions/s, with a maximum injection time of 100ms. Additionally, a dynamic exclusion time of 30s was applied to tandem mass spectrometry scans to ensure avoidance of duplicate scanning for parent ions. The service of proteomics assays was provided by PTM Biolabs, Inc.

Proteomics data analysis

Each protein was subjected to a two-sample t-test for each pairwise comparison across the four groups. Absolute $\log^2$ (fold change) > 1.5 or <1/1.5 and a p-value <0.05 were used to designate differentially expressed (DE) proteins. Following log2 ratio transformation, the relative protein expression was screened by a standard deviation (SD) > 0.3. The 828 proteins remained, and Mfuzz clustering analysis was used to separate proteins in the same cluster with similar expression trends into five clusters. The most relevant peptides from the cluster2 group (n = 16) were selected after setting unique peptides ≥2, membership of cluster2 ≥ 0.4, relative standard deviation (RSD) ≤ 0.4, and a p-value <0.05 among samples. The examination of the protein-protein interaction (PPI) network involved the utilization of the Search Tool for the Retrieval of Interacting Genes (STRING; http://string-db.org). Cytoscape software (version 3.9.1) was used to construct a PPI network from an interaction with a combined score of more than 0.4. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis and Weighted Gene Co-Expression Network Analysis (WGCNA) were performed using R software (version 4.2.1). For WGCNA, soft thresholding power (β = 8; cut-off = 0.85) was chosen to increase the adjacency matrix, which was converted into a topological overlap matrix. The minModuleSize was set to 50 to hierarchically cluster the modules, and then DN related modules were merged. The relevance scores of DN were searched from the Genecards database (https://www.genecards.org/).

Receiver operating characteristic (ROC) and correlation analysis

The diagnostic sensitivity and specificity of HMGB1, along with the correlation between HMGB1 and clinical diagnostic indicators, were examined using R software (version 4.2.1). With R packages "pROC", receiver operator characteristic curves were plotted for HMGB1 in DN and the area under the curve (AUC) estimated. AUC ranging from 0.7 to 0.9 signified a well-fitted model, while an AUC exceeding 0.9 indicated a highly fitting model.

Immunoblotting

Renal tissues were homogenised and sonicated using RIPA lysis buffer, supplemented with 1mM phenylmethanesulfonyl fluoride (PMSF, Solarbio, P0100). Additionally, cells were lysed using RIPA lysis buffer (1mM PMSF). After protein extraction, the separation process was carried out on SDS-PAGE gels. Subsequently, these proteins were transferred onto polyvinylidene difluoride (PVDF) membranes (Millipore, ISEQ00010). After a 1 h blocking step at room temperature using 5% (w/v) nonfat milk, the membrane underwent an overnight incubation at 4°C with antibodies against E-cadherin (Proteintech, 20874-1-AP), Vimentin (Proteintech, 10366-1-AP), β-actin (Proteintech, 66009-1-Ig), and HMGB1 (Abcam, ab79823). This was followed by a subsequent 1 h incubation at room temperature with secondary antibodies (CST, 7074 and 7076). Enhanced chemiluminescence reagents (Absin, abs920) and ImageJ (National Institutes of Health, V1.8.0) were used to visualize and analyze the protein bands.

Histology analysis

For histological observation of renal morphology, the renal tissues were fixed in 4% buffered paraformaldehyde, embedded in paraffin, and sliced at 3 mm. In accordance with the protocol, hematoxylin and eosin staining (HE; Solarbio, G1220), Masson trichrome staining (Solarbio, G1346), and periodic acid-schiff staining (PAS; Solarbio, G1281) were performed, and the stained sections were analyzed using a computational color image analysis system (Leica Microscope, Germany DM2500). Utilizing ImageJ software, the collagen tissue area was calculated to assess renal fibrosis quantitatively.

Biochemical assays

Blood and urine were measured by automatic clinical chemistry analyzer (ARCHITECT, c16000).

Immunofluorescent staining

The renal tissues section for immunofluorescent staining was treated in the same manner as for histology analysis, and heat-mediated antigen retrieval with Tris/EDTA buffer pH 9.0 was carried out before staining was conducted. After fixing with 4% buffered paraformaldehyde, 0.3% Triton X-100 was used to permeabilize the HK2 cells. Rabbit anti-HMGB1 antibody (Abcam, ab79823); Mouse anti-β-actin antibody (Proteintech, 66009-1-Ig) and Lotus Tetragonolobus Lectin (LTL, Biotinylated; Vector Laboratories, B-1325-2) were used for immunofluorescence staining, and images were obtained under a positive fluorescence microscope (ZEISS, Axio Imager) and a confocal microscopy (ZEISS, LMS 710).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses of clinical parameters were exclusively conducted using SPSS 26.0 software. With GraphPad Prism 9.4, 16 candidate peptides and all verification experiments were analyzed statistically. R software (version 4.2.1) was used to analyze of all proteomics data analysis. A Shapiro-Wilk test was performed on the numerical variables of this study for the purpose of testing normality. To compare normal variables (mean $\pm$ SD), we used an analysis of variance (ANOVA), while to compare variables with non-normality (median, interquartile range), we used the Kruskal-Wallis test. Using Fisher's exact test, this study tested classification variables. The factors that influenced the rise of HMGB1 were determined using multiple stepwise regressions. All statistical tests of verification experiments were performed by the two-tailed Student's *t* test. p-value ≤ 0.05 was defined statistically significant.