

## **Urine complement proteins are associated with kidney disease progression of type 2 diabetes in Korean and American cohorts**

Donghwan Yun,<sup>1,2\*</sup> Sohyun Bae,<sup>1\*</sup> Yuqian Gao,<sup>3</sup> Lauren Lopez,<sup>4</sup> Dohyun Han,<sup>5</sup> Carrie D. Nicora,<sup>3</sup> Tae Youn Kim,<sup>6</sup> Kyung Chul Moon,<sup>7</sup> Dong Ki Kim,<sup>1</sup> Thomas L. Fillmore,<sup>3</sup> Yon Su Kim,<sup>1,2</sup> Avi Z. Rosenberg,<sup>8</sup> Weijie Wang,<sup>9</sup> Pinaki Sarder,<sup>10</sup> Wei-Jun Qian,<sup>3</sup> Maryam Afkarian<sup>11†</sup> and Seung Seok Han<sup>1†</sup>

<sup>1</sup>Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea

<sup>2</sup>Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea

<sup>3</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA

<sup>4</sup>Division of Nephrology, Department of Medicine, University of California, Davis, CA, USA

<sup>5</sup>Transdisciplinary Department of Medicine and Advanced Technology, Seoul National University Hospital, Seoul, Korea

<sup>6</sup>School of Nursing, University of California, Davis, Sacramento, CA, USA

<sup>7</sup>Department of Pathology, Seoul National University College of Medicine, Seoul, Korea

<sup>8</sup>Department of Pathology, Johns Hopkins University, MD, USA

<sup>9</sup>University of California, Berkeley, Berkeley, CA, USA

<sup>10</sup>Department of Medicine-Quantitative Health, University of Florida College of Medicine, FL, USA

<sup>11</sup>Department of Internal Medicine, University of California, Davis, CA, USA

\*Co-first authors; †co-corresponding authors.

Running title: complements in kidney disease of type 2 diabetes

Correspondence:

Seung Seok Han

Department of Internal Medicine, Seoul National University College of Medicine

103 Daehak-ro, Jongno-gu, Seoul, 03080, Republic of Korea

Tel: +82-2-2072-4785; Fax: +82-2-745-2264; E-mail: [hansway7@snu.ac.kr](mailto:hansway7@snu.ac.kr)

Maryam Afkarian

Room 5317, Genome and Biomedical Sciences Facilities, 451 Health Sciences Drive, University

of California, Davis, CA 95616, USA

Tel: (916) 734-3014; Fax: (916) 734-7920; E-mail address: [mafkarian@ucdavis.edu](mailto:mafkarian@ucdavis.edu)

## Abstract

**Background:** Mechanisms of progression of diabetic kidney disease (DKD) are not completely understood. This study uses untargeted and targeted mass spectrometry-based proteomics in two independent cohorts on two continents to decipher the mechanisms of DKD in patients with type 2 diabetes.

**Methods:** We conducted untargeted mass spectrometry on urine samples collected at the time of kidney biopsy from Korean patients with type 2 diabetes and biopsy-proven diabetic nephropathy at Seoul National University Hospital (SNUH-DN cohort; n = 64). These findings were validated using targeted mass spectrometry in urine samples from a Chronic Renal Insufficiency Cohort subgroup with type 2 diabetes and DKD (CRIC-T2D; n = 282). Urinary biomarkers/pathways associated with kidney disease progression (doubling of serum creatinine,  $\geq 50\%$  decrease in estimated glomerular filtration rates, or the development of end-stage kidney disease) were identified.

**Results:** SNUH-DN patients had an estimated glomerular filtration rate (eGFR) of 55 mL/min/1.73 m<sup>2</sup> (interquartile range [IQR], 44–75) and random urine protein-to-creatinine ratio of 3.1 g/g (IQR, 1.7–7.0). Urine proteins clustered into two groups, with cluster 2 having a 4.6-fold greater hazard (95% confidence interval [CI], 1.9–11.5) of disease progression than cluster 1 in multivariable-adjusted, time-to-event analyses. Proteins in cluster 2 mapped to 10 pathways, four of the top five of which were complement or complement-related. A high complement score, constructed from urine complement protein abundance, was strongly correlated to 4 of 5 histopathologic DN features and was associated with a 2.4-fold greater hazard (95% CI, 1.0–5.4) of disease progression than a low complement score. Targeted mass spectrometry of the CRIC-

T2D participants, who had an eGFR of 42 mL/min/1.73 m<sup>2</sup> (IQR, 37–49) and 24-hr urine protein of 0.48 g (IQR, 0.10–1.87), showed that the complement score similarly segregated them into rapid and slow DKD progression groups. In both cohorts, the complement score had a linear association with disease progression.

**Conclusions:** Urinary proteomic profiling confirms the association between the complement pathway and rapid DKD progression in two independent cohorts. These results suggest a need to further investigate complement pathway inhibition as a novel treatment for DKD.

Keywords: biomarker; complement; diabetic kidney disease; histopathology; proteomics.

## Introduction

Diabetic kidney disease (DKD) is a major complication of diabetes, significantly contributing to morbidity and mortality worldwide <sup>1,2</sup>. As the leading cause of end-stage kidney disease (ESKD) globally, DKD accounts for about half of all ESKD cases, thereby posing a significant burden on healthcare systems worldwide <sup>3</sup>. Despite recent advances in diabetes management, the mechanisms underlying progression (and their respective biomarkers) in human DKD remain incompletely understood. Clinically, this translates into highly variable rates of disease progression, with some patients experiencing rapid progression leading to ESKD <sup>4,5</sup>. This variability presents substantial challenges in clinical practice, as it is currently impossible to identify and intervene in patients who are rapid progressors.

Although extensive research is being conducted, the pathophysiological mechanisms responsible for the DKD progression still need to be completely understood. Traditional risk factors, such as hyperglycemia, hypertension, and diabetic duration, are recognized <sup>6,7</sup>, but they only partially explain the inter-individual differences in disease progression. Prior studies have explored various biomarkers, including cytokines and kidney injury-relevant molecules <sup>8-10</sup>, but a comprehensive understanding that integrates these factors with clinical outcomes remains elusive. This knowledge gap hinders the development of precise predictive models and targeted interventions that are essential for preventing rapid progression in DKD patients. Therapeutic strategies such as renin-angiotensin-system inhibitors, sodium-glucose cotransporter-2 inhibitors, and mineralocorticoid receptor antagonists are employed to mitigate DKD progression <sup>11-13</sup>. However, despite these advances in therapy, it is evident that more targeted therapies are needed to improve DKD outcomes.

In response to these challenges, -omics technologies, including genomics, transcriptomics, metabolomics, and proteomics, emerge as powerful tools for unraveling the complex pathobiology of DKD. Recent advances in urinary proteomics have significantly enhanced the ability to predict prognosis and analyze subtypes in DKD by revealing the molecular dynamics of kidney injury<sup>14,15</sup>. This approach allows for the development of personalized treatment plans based on specific molecular profiles. We used comparative proteomic analyses across two independent cohorts (South Korea and the US) to identify the pathways associated with rapid DKD progression in type 2 diabetes. Utilizing both untargeted and targeted urinary proteomics, we found that differential abundance of complement proteins in urine predicted rapid DKD progression. This data highlights the importance of the complement pathway as a promising source of novel biomarkers for identifying rapid progressors and as a therapeutic target for novel DKD interventions.

## **Methods and Materials**

### **Study patients**

From March 2011 to June 2021, 93 patients with type 2 diabetes and biopsy-proven diabetic nephropathy were recruited at Seoul National University Hospital (SNUH), referred to as the SNUH-DN cohort. Of this group, 29 were excluded due to an estimated glomerular filtration rate (eGFR)  $<30$  mL/min/1.73 m<sup>2</sup> (n = 25), combined type 1 and type 2 diabetes (n = 2), the presence of concomitant nondiabetic glomerular disease (n = 1), or the absence of glomeruli in biopsied kidney tissues (n = 1). The remaining 64 patients were included in the final analyses as the SNUH-DN cohort after conducting untargeted proteomics.

Additionally, this study incorporated a subset of patients with type 2 diabetes from the Chronic Renal Insufficiency Cohort (CRIC) study, termed CRIC-T2D, consisting of American patients with risk factors for cardiovascular disease, progression of chronic kidney disease, and mortality. A total of 3,939 patients aged 21–74 with an eGFR of 20–70 mL/min/1.73 m<sup>2</sup> were enrolled at seven clinical centers throughout the US from June 2003 to December 2008. CRIC over-sampled participants with diabetes, who constitute nearly half of the CRIC cohort. This study excluded participants with monogenic renal disease, liver cirrhosis, class III or IV heart failure, human immunodeficiency virus, cancer, autoimmune disease, polycystic kidney disease, pregnant women, recipients of organ or bone marrow transplants, and those receiving immunotherapy for primary renal disease or systemic vasculitis within the six months preceding recruitment or systemic chemotherapy.

### **Ethical considerations**

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki and was approved by the institutional review boards (IRBs) of all participating institutions. For the SNUH-DN cohort, ethical approval was granted by the IRB at SNUH (no. H-2403-066-1519). The CRIC study was approved by the IRBs at all participating institutions and was conducted in accordance with the Declaration of Helsinki<sup>16, 17</sup>. Urine samples and data for the present study were obtained from the CRIC subset housed at the NIDDK Central Repository (R01 5R01DK104706). The use of de-identified CRIC samples was approved by the IRB at the University of California, Davis. The use of de-identified samples and data from the CRIC-T2D subgroup from the NIDDK Central Repository was approved by the Human Subjects Division, University of California, Davis. All participants provided written informed consent, having been fully informed of the study purpose, the procedures involved, potential risks, and their rights as study participants, including the right to withdraw at any point without consequence. The confidentiality of patient data was rigorously maintained throughout the study. All personal identifiers for SNUH-DN patients were removed, and data was stored securely to prevent unauthorized access. CRIC-T2D data and samples were de-identified at the NIDDK Central Repository.

### **Study outcomes**

The primary outcome was the progression of DKD (named kidney disease progression) based on any of the following events: a doubling of baseline serum creatinine, a decline greater than 50% in the eGFR, or the onset of ESKD. The eGFR was calculated using the Chronic Kidney Disease Epidemiology Collaboration equation<sup>18</sup>. ESKD was specifically defined as the



need for kidney replacement therapy, including hemodialysis, peritoneal dialysis, or kidney transplantation.

### **Study variables**

We collected comprehensive patient data encompassing demographic characteristics, such as age, sex, race, body mass index, diabetic duration, and comorbidities of hypertension, ischemic heart disease, cerebrovascular disease, retinopathy, and neuropathy. Laboratory findings included hemoglobin A1c, eGFR, and random urine protein-to-creatinine ratio (PCR) for the SNUH-DN cohort, while the CRIC-T2D cohort included the 24-hour urine protein amount.

We employed a detailed histopathologic classification system to evaluate kidney biopsies from patients diagnosed with diabetic nephropathy in the SNUH-DN cohort. This classification system follows the criteria described in the reference paper<sup>19</sup>. Each biopsy was required to contain at least ten glomeruli for accurate assessment, excluding any incomplete glomeruli along the biopsy edges. The glomerular classification was segmented into five categories: class I, isolated glomerular basement membrane thickening; class IIA, mild mesangial expansion; class IIB, severe mesangial expansion; class III, nodular sclerosis (Kimmelstiel-Wilson lesions); and class IV, advanced diabetic glomerulosclerosis with more than 50% global glomerulosclerosis. The assessment of interstitial fibrosis and tubular atrophy (IFTA), interstitial inflammation, arterial hyalinosis, and arteriosclerosis was also quantified using a four-point or three-point scale according to the criteria.

### **Urine sample collection**

In the SNUH-DN cohort, urine samples were collected at the time of kidney biopsy. Patients were instructed to provide a midstream clean-catch urine sample, which was immediately processed to minimize protein degradation. Upon collection, urine samples were centrifuged at 3000 rpm for 10 minutes to remove cellular debris and then aliquoted into cryovials. The aliquots were promptly frozen and stored at  $-80^{\circ}\text{C}$  to preserve the integrity of the proteins until preparation for proteomics.

In the CRIC-T2D cohort, 24-hour urine samples were collected at home by study participants during their CRIC visit number 3. Participants were instructed to keep the samples at  $4^{\circ}\text{C}$  during collection. Collections that were less than 500 ml or collected for  $<22$  hours or  $>26$  hours were repeated. Accepted 24-hour collections were thoroughly mixed, and three 10 ml aliquots were obtained. The aliquots were shipped to the CRIC central laboratory on ice packs, where they were further aliquoted and stored at  $-80^{\circ}\text{C}$ . A portion of these samples was shipped to the NIDDK Central Repository when they were sent to a study author (MA). An aliquot of these samples was then shipped to the Pacific Northwest National Laboratory on dry ice and were prepared for targeted proteomics.

## **Untargeted proteomics**

### **Sample preparation**

A volume of 1-2 ml of urine from each participant was concentrated down to 250  $\mu\text{l}$  using a centrifugal filter with a molecular weight cutoff of 3 kDa (Millipore, Billerica, MA). Protein concentrations were then determined using the Bradford method, employing a commercial kit (Bio-Rad, Hercules, CA). For the analysis, 50  $\mu\text{g}$  of protein from each sample was precipitated using ice-cold acetone and added at five times the protein volume. The precipitate was

redissolved in 50  $\mu$ l of SDT buffer (comprising 2% SDS, 0.1 M dithiothreitol, in 0.1 M Tris HCl, pH 8.0), and subjected to heating at 95°C to denature the proteins. Following denaturation, proteins were enzymatically digested utilizing a filter-aided sample preparation approach, adapted from previously established methods with some modifications<sup>20</sup>. Specifically, proteins were transferred to an Amicon 30K filter (Millipore, Billerica, MA) and washed multiple times with UA solution (8 M urea in 0.1M Tris-HCl, pH 8.5) through centrifugation. After three washes, cysteines were alkylated using 0.05 M iodoacetamide in the UA solution for 30 minutes at room temperature in the dark. Subsequent buffer exchanges to 40 mM ammonium bicarbonate were performed twice to prepare for enzymatic digestion. Overnight digestion was carried out at 37°C using trypsin/LysC with a ratio of enzyme to substrate of 1:100. The peptides produced were then collected in new tubes through additional centrifugation, and an extra elution step was performed using 40 mM ammonium bicarbonate mixed with 0.5 M sodium chloride. The eluted peptides were purified and fractionated using custom-made styrene-divinylbenzene reversed-phase sulfonate-StageTips, employing a gradient of acetonitrile (40%, 60%, and 80%) in 1% ammonium hydroxide<sup>21, 22</sup>. Following fractionation, peptides were vacuum-dried and stored at –80°C until further analysis.

### **Construction of a peptide spectral library**

To create a peptide spectral library for using match between runs (MBRs) algorithm of maxquant, a pooling of urine proteins was performed<sup>23</sup>. We combined equal sub-aliqouts of proteins, each containing 5  $\mu$ g, to a total of 100  $\mu$ g. This pooled sample was then subjected to a two-step filter-aided digestion process, consistent with the methods previously outlined<sup>21, 24</sup>. Following digestion, peptides were purified using Oasis HLB solid-phase extraction cartridges to

ensure their readiness for high-performance liquid chromatography. For the library construction, we utilized 100  $\mu\text{g}$  of these clean peptides and processed them through an Agilent 1260 bioinert HPLC system (Agilent, Santa Clara, CA) equipped with a standard analytical column ( $4.6 \times 250$  mm; 5- $\mu\text{m}$  particle size). The peptides were fractionated using a high-pH reversed-phase method, employing a flow rate of 0.8 ml/min across a 60-minute gradient. Solvents included 15 mM ammonium hydroxide in water (solvent A) and 15 mM ammonium hydroxide in 90% acetonitrile (solvent B). We collected 96 separate fractions at 30-second intervals throughout a 48-minute period, which were then concatenated into 24 fractions in step-wise manner. Early-, middle-, and late-eluting peptides were pooled by mixing every 24<sup>th</sup> original fraction for the proteome (e.g., combining fractions 1, 25, 49, and so on)<sup>25</sup>. These concatenated fractions were subsequently dried using a vacuum centrifuge and stored at  $-80^{\circ}\text{C}$ , ready for analysis by liquid chromatography-tandem mass spectrometry (LC-MS/MS).

### **LC-MS/MS analytical procedure**

LC-MS/MS analysis was conducted on a Q-Exactive Plus Quadrupole Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA), which was interfaced with an Ultimate 3000 RSLC nano system (Dionex, Sunnyvale, CA) featuring a nanoelectrospray ion source. The process incorporated slight modifications from previously established protocols<sup>20, 22</sup>. Individual peptide fractions from urine samples underwent separation using a dual-column arrangement consisting of a C18 trap column (300  $\mu\text{m}$  internal diameter  $\times$  0.5 cm, with 3  $\mu\text{m}$  particles and 100  $\text{\AA}$  pore size) and a C18 analytical column (50  $\mu\text{m}$  internal diameter  $\times$  50 cm, with 1.9  $\mu\text{m}$  particles and 100  $\text{\AA}$  pore size). Prior to injection, dried peptide fractions were dissolved in solvent A (2% acetonitrile and 0.1% formic acid). The nano-LC system loaded the samples and

applied a chromatographic gradient over 90 minutes, increasing from 8% to 30% of solvent B (100% acetonitrile with 0.1% formic acid). The setup operated under a spray voltage of 2.0 kV in positive ion mode, and the capillary heater was maintained at 320°C. Mass spectrometric data were collected in a data-dependent acquisition format, prioritizing the top 15 most abundant precursor ions. The Orbitrap detector analyzed precursor ions across a mass range from 300 to 1650 m/z, achieving a resolution of 70,000 at m/z 200. Fragmentation was performed via higher-energy collisional dissociation (HCD) at a resolution of 17,500 for m/z 200, using a normalized collision energy setting of 28. The system's ion injection times were set to a maximum of 20 ms for survey scans and 120 ms for MS/MS scans.

### **Label-free quantitative data analysis**

For the analysis of mass spectrometric data, we employed MaxQuant software (version 1.6.1.0) for label-free quantification<sup>26</sup>. MS/MS spectra were meticulously matched against the Human UniProt protein sequence database (as of December 2014, containing 88,657 entries) via the Andromeda search engine<sup>27</sup>. The analysis parameters were carefully selected, with a precursor ion tolerance set at 6 ppm for total protein analysis and an MS/MS ion tolerance of 20 ppm. We designated carbamido-methylation of cysteine as a fixed modification, while N-acetylation of proteins and methionine oxidation was considered variable modifications. The digestion parameter was configured for full tryptic digestion with allowances for up to two missed cleavages per peptide, which had to be at least six amino acids long. We rigorously maintained a false discovery rate (FDR) of 1% across peptide, protein, and modification levels to ensure high data integrity. Additionally, to enhance quantification accuracy across samples, we utilized a strategy of matching between runs anchored by the peptide library constructed from

pooled urine samples. The Intensity Based Absolute Quantification (iBAQ) algorithm, a component of the MaxQuant suite, facilitated the estimation of protein abundances<sup>28</sup>. In iBAQ, raw intensities were normalized by the count of theoretical peptides, rendering them reflective of the proteins' molar quantities.

## **Targeted proteomics**

### **Selected reaction monitoring (SRM)**

Assay development was conducted as described in detail previously<sup>29</sup>. Briefly, final surrogate peptides were selected based on the uniqueness of their occurrence in proteins of interest and their presence and chromatographic behavior in shotgun MS/MS data in our prior urine proteomics data repository (Supplementary Table 1). Crude synthetic heavy isotope-labeled versions of the selected peptides were used to generate MS/MS data in an Orbitrap HCD. This data was used to identify the best transitions and optimal collision energy for each transition.

### **Sample processing and digestion**

Urine samples were processed and underwent LC-MS/MS, as described previously<sup>29</sup>. Briefly, urine samples were thawed on ice, run through a 10 kDa prewashed (50 mM AMBIC, i.e.,  $\text{NH}_4\text{HCO}_3$ , pH 8.0) Amicon ultrafiltration column, and underwent buffer exchange with AMBIC twice. The final urine protein concentration was determined using the BCA protein assay. Urea was added to the urine sample to a final concentration of 8 M, followed by reduction with 500mM DTT to a final concentration of 10 mM, brief sonication, and incubation at 37°C for 1 hour while shaking. Alkylation was performed by adding IAA to a final concentration of 40 mM and incubation for one hour at 37°C in the dark while shaking. The sample was then diluted

ten-fold with digestion buffer, and trypsin was added to a protein/trypsin ratio of 50:1 (w/w), followed by incubation at 37°C for 3 hours. Digestion was stopped by adding TFA to a final concentration of 0.1%. The tryptic digest was separated using SPE C18 columns, preconditioned using methanol and then SPE conditioning buffer, and washed with SPE washing buffer. Peptides were eluted from the SPE C18 columns with 1ml SPE elution buffer and, dried under a reduced vacuum using a speed vac and redissolved in MS-grade water. The peptide concentration was determined using the BCA protein assay, and peptide samples were stored at -80°C until further use. Heavy isotope-labeled peptide standards were mixed in a single 1 ml stock solution, having been reconstituted at 1000 fmol/ml per peptide in 0.1% TFA in water. The digested peptide mix and the heavy peptide stock were combined while shaking and advanced to LC-SRM analysis.

### **LC-SRM analysis and quantification**

The urine sample/heavy peptide mixes were analyzed using the final transition list, derived from the orbitrap HCD MS/MS, and optimized in one typical urine sample. LC-SRM was performed using a nanoACQUITY UPLC system coupled online to a TSQ Vantage triple quadrupole mass spectrometer. The peak area ratios (PAR) of endogenous transitions and heavy internal standard transitions represent the molar ratios between the amounts of endogenous and heavy internal standard peptides. The PARs were calculated using Skyline software<sup>30</sup>. The data was imported into a Skyline file, peak boundaries for each peptide were manually verified for each dataset, and detection and optimal transition (for best peak area ratio, PAR) were confirmed for each peptide. Protein concentration was calculated using the equation below.

Target protein concentration  $\left(\frac{\text{ng}}{\text{mL}}\right)$

$$= \frac{L}{H} \times C_{\text{heavy, LC-SRM}} \times C_{\text{digest, LC-SRM}} \times C_{\text{protein}} \times 400 \mu\text{L} \times \text{MW} \times \frac{10^{-6} \text{ ng}}{\text{fg}} \div 10 \text{ mL}$$

where L/H is PAR of endogenous (L) to heavy isotope-labeled internal standard peptide (H),  $C_{\text{heavy, LC-SRM}}$  is the molar concentration of heavy internal standard peptides (fmol/ $\mu\text{L}$ ) in the final LC-SRM solutions; MW is the molecular weight of targeted protein (Da or g/mol);  $C_{\text{digest, LC-SRM}}$  is the mass concentration of total protein digest ( $\mu\text{g}/\mu\text{L}$ ) in the final LC-SRM solutions; and  $C_{\text{protein}}$  is mass concentration of total protein ( $\mu\text{g}/\mu\text{L}$ ) in the 400  $\mu\text{L}$  concentrated retentate from 10 ml of original urine.

### **Proteomics data processing and analysis**

The proteomic data obtained using label-free quantification of urine proteins were initially processed using the *Seurat* package (version 5.0.3)<sup>31</sup>. These values were normalized for urine dilution by indexing (division) by urine creatinine, and then log2 transformed for further normalization using the *NormalizeData* function from the *Seurat* package. Finally, for each identified protein, the normalized data was scaled using the *ScaleData* function from the *Seurat*. Unsupervised clustering techniques were utilized to group proteins based on their expression patterns, and the *FindNeighbor* function and *RunUmap* function were used to create cluster groups and to draw UMAP, respectively. Gene names related to each protein were identified using the *UniProt* accession numbers<sup>32</sup>. Differentially expressed proteins between rapid and slow progressors were identified (*FindMarkers* function). Using the *UniProt* database, proteins with well-documented gene associations were selected for pathway analysis, and differentially represented pathways were identified using *WikiPathways*<sup>33</sup>.

### **Complement scores**



The complement score was devised to evaluate the prevalence of complement pathway proteins in urinary proteomics data. We identified the proteins involved in the complement pathway by referencing the Kyoto Encyclopedia of Genes and Genomes database<sup>34</sup>. The list of complement proteins used for scoring is presented in Supplementary Table 2. The average protein expression of denoted genes was calculated based on the assumption of previous studies<sup>35-37</sup>. Additionally, we formalized the distinction between activators and inhibitors. For each patient, this score was calculated by taking the mean of the normalized abundance of identified complement proteins, adjusted for their functional roles: inhibitory proteins were assigned a modifier of  $-1$ , and activating proteins a modifier of  $1$ . The formula used was:

$$\text{Complement score}_i = \sum_{j=1}^n (\text{Normalized expression}_{ij} \times \text{Modifier}_j) / n$$

where  $i$  denotes the sample number, and  $j$  is the individual complement protein. The continuous variable complement score was binarized into low vs. high complement score bins by the median value.

## Statistical analysis

Descriptive statistical analyses were conducted on patient information. For continuous variables with a normal distribution, data are presented as the mean  $\pm$  standard deviation, and for those without a normal distribution, as medians with interquartile ranges (IQR). The Kolmogorov-Smirnov test was used to assess the normality of the data distribution. Categorical variables are shown as percentages. Comparison of categorical variables was performed using either the chi-square test or Fisher's exact test, while continuous variables were compared using

either Student's t-test or the Mann-Whitney U test, depending on their distribution. When comparing multiple groups for continuous variables, analysis of variance (ANOVA) was used, with the groups treated as ordinal variables. Pairwise comparisons were conducted with Bonferroni correction.

Survival analysis was employed to investigate the association between variables and the risk of kidney disease progression. The hazard ratio (HR) and corresponding 95% confidence intervals were calculated using the Cox proportional hazards regression model. Variables included in the model were selected based on clinical relevance, such as age, sex, body mass index, diabetic duration, and baseline kidney function. Adjusted survival curves were then generated to visualize the differences in adjusted cumulative survival between the two groups using the *adjustedCurves* package<sup>38</sup>. Additionally, penalized spline regression was utilized to further elucidate the relationship between complement scores and kidney disease progression. All statistical analyses were conducted using the R program (version 4.2.2). A two-sided *P* value of less than 0.05 was considered statistically significant.

## Results

### Baseline characteristics in the SNUH-DN cohort

This cohort had a mean age of  $53.3 \pm 11.1$  years, with 71.9% male subjects. Hypertension was present in 92.2% of the subjects, and the median duration of diabetes was 10 years (IQR, 4–14 years). Diabetic retinopathy and neuropathy were observed in 54.7% and 29.7% of all patients, respectively. The median hemoglobin A1c was 7.2% (IQR, 6.6–8.4), while the median PCR and eGFR values were 3.1 g/g (IQR, 1.7–7.0) and 55 mL/min/1.73 m<sup>2</sup> (IQR 44–75), respectively. Details of histopathologic features are provided in Table 1.

### Proteomic clustering in the SNUH-DN cohort

Unsupervised clustering of the iBAQ-normalized urine protein abundances from the SNUH-DN cohort identified two distinct clusters, designated as clusters 1 and 2 (Figure 1A). Heatmap of iBAQ-normalized urine proteins that were abundant and scarce in clusters 0 and 1 are illustrated in Supplementary Figures 1A and B, respectively. These data showed that of the 1,877 identified proteins, 71 and 85 were significantly abundant and scarce in cluster 2, compared to cluster 1, after the Bonferroni adjustment, respectively. Many of the abundant proteins mapped to the complement pathway, either complement activators such as complement factor D (CFD), component 2 (C2), component 8 beta chain (C8B), component 6 (C6), and component 5 (C5)<sup>39</sup>, or complement inhibitors such as complement factor H (CFH) and serpin family G member 1 (SERPING1)<sup>40, 41</sup>. Conversely, among the scarce proteins, CD55 and CD59 act as complex inhibitors<sup>42</sup>, while the protein C receptor (PROCR) functions indirectly as a complement inhibitor<sup>43</sup>.

Kidney disease progression within one year after biopsy occurred in 6 out of 24 subjects (25%) with urine proteins in cluster 2, and 2 out of 40 subjects (5%) in cluster 1. Compared to subjects in cluster 1, those in cluster 2 had a 4.62-fold (95% CI, 1.86–11.50) greater hazard of DKD progression in multivariable time-to-event (Cox proportional hazard) regression, adjusted for age, sex, body mass index, hypertension, diabetic duration, PCR, and eGFR (Table 2). The adjusted cumulative incidence of DKD progression displayed more rapid kidney disease progression in cluster 2 than in cluster 1 (Figure 1B). These results indicate that the unsupervised clustering of urine proteins was able to segregate subjects into two clusters, displaying slower and more rapid DKD progression. Notably, the accelerated DKD progression in cluster 2, compared to cluster 1, was associated with high glomerulosclerosis scores in the paired kidney biopsies from these subjects (Figure 1C).

### **Different enrichment in complement proteins**

Pathway analysis identified a notable enrichment of complement and related pathways in cluster 2 (Figure 2A and Supplementary Table 3). The volcano plot of differentially abundant proteins highlighted that complement activators, including CFD, component 3 (C3), C5, C6, and C8B, and some complement inhibitors, including CFH, complement factor I (CFI), and clusterin (CLU), were prominently ranked in cluster 2. Conversely, other complement inhibitors (CD55 and CD59) were scarce in cluster 2 (Figure 2B). Detailed information regarding the proteins related to the complement pathway is included in Supplementary Table 2. The heatmap with the normalized abundance of complement proteins depicted that complement pathway-related proteins were more prominently altered in cluster 2 than in cluster 1 (Figure 2C).

## **Complement score and kidney disease progression in the SNUH-DN cohort**

To comprehensively assess the effect of complement proteins, we calculated a complement score that considers their expression levels and modifies their contribution by their roles in pathway activation (activators or inhibitors). The overall median complement score for all samples was 0.28 (IQR, 0.08–0.62), with a range from –0.25 to 0.92. Cluster 2 had a median complement score of 0.69 (IQR, 0.51–0.76), while cluster 1 had a median score of 0.14 (IQR, 0.01–0.28) ( $P < 0.001$ ) (Figure 3A). Based on the median score of 0.28 from all samples, we divided the patients into high ( $>0.28$ ) and low ( $\leq 0.28$ ) complement score groups (Supplementary Figure 2).

Patients with high complement scores had a 2.41-fold (1.07–5.40) higher risk of kidney disease progression than those with low scores, after adjustment for demographic and clinical variables (Table 3). Multivariable-adjusted cumulative incidence of DKD progression also demonstrated a significant difference between the high and low complement groups (Figure 3B). Notably, in multivariable-adjusted spline regression, the association between the complement score and the hazard of DKD progression was linear (Figure 3C).

To eliminate the potential contribution of the urine creatinine denominator to the outcome, a sensitivity analysis was conducted using the urine abundance of proteins without normalization for urine creatinine. These analyses yielded the same results, with cluster 2 and the high complement score group both significantly associated with rapid DKD progression, compared to cluster 1 and the low complement score group, respectively (Supplementary Figure 3).

## **Histopathologic correlation with complement scores in SNUH-DN cohort**

In the SNUH-DN cohort, which included paired urine and kidney tissues from the same individuals, urine complement scores were significantly correlated with three of five features used to classify DKD: glomerular lesions, interstitial fibrosis and tubular atrophy (IFTA), and arteriolar hyalinosis. The correlation between complement score and interstitial inflammation did not reach statistical significance. Urine complement score was not correlated with arteriosclerosis in this sample set (Supplementary Figure 4).

### **Baseline characteristics in the CRIC-T2D cohort**

The association between urinary complement proteins and kidney disease progression was examined in an independent cohort: the CRIC-T2D. The CRIC-T2D cohort (Table 4), comprising 282 patients, was a majority (62%) male group with a mean age of  $60.8 \pm 7.5$  years. Of these patients, 51.8% were non-Hispanic black and 30.1% were non-Hispanic white. Hypertension and cardiovascular disease were identified in 93.6% and 46.5% of the participants, respectively. The median 24-hour urine protein and eGFR were 0.48 g (IQR, 0.10–1.87) and 42 mL/min/1.73 m<sup>2</sup> (IQR, 37–49), respectively. The median follow-up duration was 9.7 years (IQR, 6.3–12.0).

### **Complement Score in the CRIC-T2D cohort**

In the CRIC-T2D cohort, targeted urinary proteomics was used to quantify 12 complement proteins, including complement activators such as factor B (CFB), C3, component 4 (C4), C5, C6, component 7 (C7), component 8, alpha chain (C8A), and component 9 (C9), and inhibitors such as CD59, CFI, clusterin, and CFH-related protein 2 (Supplementary Table 2 and Supplementary Figure 5). The complement score was calculated for each participant as described

in the Complement scores section. The median complement score for all participants was 1.07 (IQR, 0.42–2.10), ranging from –0.76 to 3.04. Based on this median value, the participants were divided into groups with high ( $>1.07$ ) and low ( $\leq 1.07$ ) complement scores (Figure 4A). The median complement scores for the high and low groups were 2.16 (IQR, 1.63–2.45) and 0.41 (IQR, 0.17–0.68), respectively. The group with high scores had a 2.51-fold (95% CI, 1.76–3.57) higher risk of kidney disease progression than the low score group, after adjustment for age, sex, body mass index, hypertension, ethnicity and race, 24-hour urine protein, and eGFR (Figure 4B and Table 5). Multivariable-adjusted spline regression revealed a linear association between the complement score and the hazard of DKD progression (Figure 4C).

## Discussion

We used untargeted and targeted mass spectrometry to investigate the association between complement proteins and DKD progression in the SNUH-DN and CRIC-T2D cohorts. Unsupervised clustering of proteins identified in untargeted urinary proteomics from the SNUH-DN cohort revealed two groups with different kidney outcomes. Pathway analysis revealed a marked prominence of complement proteins in the group with rapidly progressive DKD. Complement scores, constructed based on the abundance and function of complement proteins, were associated with histopathologic severity and pace of DKD progression. These results were replicated in the CRIC-T2D cohort, where high complement scores similarly correlated with rapid progression. Together, these findings underscore the potential of urinary complement proteins as predictive biomarkers for kidney disease progression and the pathway as a target for therapeutic intervention in DKD.

While agents, including renin-angiotensin-system inhibitors, sodium-glucose cotransporter-2 inhibitors, and mineralocorticoid receptor antagonists, have improved kidney outcomes in DKD patients, a significant number still experience rapid disease progression<sup>44, 45</sup>. The underlying pathophysiology contributing to this residual risk may involve the complement pathway, characterized by an imbalance between activators and inhibitors<sup>46, 47</sup>. High urine levels of complement proteins in DKD, comparable to those in autoimmune glomerulonephritis, have been associated with worse kidney outcomes<sup>48, 49</sup>. Even in patients with biopsy-confirmed DN, the association of complement proteins with worse outcomes has been observed<sup>50, 51</sup>. Our results support and extend these findings and highlight the importance of further exploring the involvement of the complement pathway in DKD progression.



We add strengths to prior studies in several ways. Firstly, we combined the use of untargeted and targeted proteomics. Employing untargeted proteomics enabled an unbiased examination of the urinary proteome, leading to the initial identification of an increase in complement pathway proteins. Targeted proteomics was then utilized to validate these observations. Secondly, to our knowledge, this is the first study to replicate the association between urine complement components and DKD progression in two independent cohorts. Thirdly, our approach goes beyond examining individual complement proteins and measures a complement score, which accounts for both activator and inhibitor functions, similar to methodologies in previous studies that used expression levels of molecules involved in specific pathways for scoring<sup>36,37</sup>. This approach allowed both a qualitative and quantitative analysis by categorizing patients into high and low complement score groups and using spline regression to analyze the shape of the association between the complement score and the hazard of disease progression.

We identified several complement activators, such as CFD, C3, C4, C5, C6, C8, and mannan-binding lectin serine protease 1 and 2, that were elevated in patients with rapid kidney disease progression. Upregulation of several complement proteins (e.g., C3, component 5 alpha chain (C5A), mannan-binding lectin serine protease 1 and 2) has been noted in rats receiving a high-fat diet combined with low-dose streptozotocin<sup>52</sup>. Previous human data on DKD identified C4 and C8 as significant risk factors for ESKD and all-cause mortality<sup>49</sup>. Inhibiting these activators may serve as a therapeutic intervention, as one previous study showed that an inhibitor targeting the C5A receptor attenuated kidney fibrosis in streptozotocin-treated mice<sup>53</sup>. The monoclonal antibody targeting C5, eculizumab, is an established clinical intervention in atypical hemolytic uremic syndrome and also tried in glomerulonephritis such as immunoglobulin A

nephropathy and anti-neutrophil cytoplasmic antibody-mediated vasculitis<sup>54,55</sup>. As a C3 inhibitor, pegcetacoplan, originally effective for paroxysmal nocturnal hematuria, is currently undergoing clinical trials as a treatment for C3 glomerulopathy and immune complex-mediated membranoproliferative glomerulonephritis<sup>56,57</sup>. The time is approaching to examine complement inhibition as a novel treatment in DKD.

We observed an increase in some complement inhibitors (e.g., CFH, CLU, and CFI) and a drop in others (CD55, CD59) in rapid progressors. These findings are consistent with previous studies, which have noted similar patterns of increase in other inhibitors alongside decreased CD55 during DKD progression<sup>51</sup>. The increased complement inhibitors may suggest a compensatory response to an activated complement pathway, perhaps displaying an inadequate regulatory mechanism that is insufficient to counterbalance the activation of the complement pathway<sup>49</sup>. CD55 depletion exacerbates DKD in an animal model.<sup>58</sup> Low urine CD59 was associated with an increased risk of ESKD, cardiovascular complications, and all-cause mortality in patients with DKD<sup>49,59</sup>, and overexpression of CD55 and CD59 attenuated kidney damage in ischemia-reperfusion injury mouse models<sup>60</sup>. Accordingly, CD55 and CD59 may exhibit a different pattern and thus play different roles in complement regulation compared to other complement inhibitors, which showed a compensatory increase in DKD.

Although the study provides valuable insights, there are certain limitations that warrant further examination. As with other human studies, except randomized controlled trials, it is challenging to confirm causality and elucidate the underlying mechanisms solely based on our findings. We did not examine the trend of changes in complement proteins over time, which limits our understanding of their dynamics. Additionally, we did not identify the source of the complement proteins. While the liver is considered a major source of complement proteins, the

kidney may also contribute significantly, especially in the context of injury and inflammation<sup>36, 61</sup>. Furthermore, the present cohorts primarily consisted of patients with reduced eGFR or advanced stage of DKD, hindering the application of the results to those with earlier stages of DKD.

We find a significant association between the complement pathway and rapid DKD progression in two independent populations. Our findings highlight urine complement proteins as potential biomarkers of pathway status and the complement pathway itself as a potential therapeutic target. Comparable findings in two ethnically dissimilar cohorts suggest that current observations are likely generalizable. By integrating various complement proteins and validating the results in an independent cohort, we provide robust evidence for the involvement of the complement pathway in rapidly progressive DKD. These findings lay the foundation for its clinical application and advocate for continued research into complement-targeting interventions to improve outcomes in the complex landscape of DKD.

## Figure legends

### **Figure 1. Unsupervised clustering based on untargeted proteomics in the SNUH-DN cohort.**

**A.** Uniform manifold approximation and projection (UMAP) plot with two distinct clusters (clusters 1 and 2). **B.** Multivariable-adjusted cumulative incidence of kidney disease progression according to clusters, after adjustment for age, sex, body mass index, hypertension, diabetic duration, urine protein-to-creatinine ratio, and eGFR. **C.** Sankey plot illustrating the relationships among glomerular classes, clusters, and kidney disease progression within one year.

### **Figure 2. Complement pathway and kidney disease progression in the SNUH-DN cohort. A.**

Pathway analysis of the differentially abundant proteins in cluster 2 compared to cluster 1. **B.** Volcano plot of the differentially abundant proteins in cluster 2 compared to cluster 1. The dashed line denotes the threshold of significance. Red and blue dots represent proteins increased and decreased in cluster 2, respectively, compared to cluster 1. **C.** Heatmap of protein abundance for urinary complement components according to clustering and complement function.

### **Figure 3. Complement score and kidney disease progression in the SNUH-DN cohort. A.**

Bar plot of complement score in clusters 1 and 2.  $***P < 0.001$ . **B.** Multivariable adjusted cumulative incidence of kidney disease progression according to complement score groups after adjustment for age, sex, body mass index, hypertension, diabetic duration, urine protein-to-creatinine ratio, and eGFR. **C.** Penalized spline regression according to complement score after adjusting for age, sex, body mass index, hypertension, diabetic duration, urine protein-to-creatinine ratio, and eGFR.

**Figure 4. Complement score and kidney disease progression in the CRIC-T2D cohort. A.**

Bar plot of complement score in high and low score groups. \*\*\* $P < 0.001$ . **B.** Multivariable-adjusted cumulative incidence of kidney disease progression according to complement score groups after adjustment for age, sex, body mass index, race, hypertension, 24-hour urine protein, and eGFR. **C.** Penalized spline regression according to complement score after adjusting for age, sex, body mass index, race, hypertension, 24-hour urine protein, and eGFR.

## **Acknowledgment**

Not applicable.

## **Authors' contributions**

SSH and MA designed the study. DY, SB analyzed the collected data. YG, CDN, TLF, WJQ performed targeted proteomics. LL analyzed the targeted proteomic data. DH performed untargeted proteomics. KCM reviewed kidney biopsy section and classified histopathologic parameters. SSH, DKK and YSK assisted the development of the SNUH-DN cohort and sample collections. MA was the recipient of an NIH award with access to CRIC-T2D samples. MA and WW supervised and funded the targeted proteomics study in CRIC-T2D. AZR and PS assisted the interpretation of untargeted proteomics and assisted the evaluation of histopathologic parameters. DY, SSH, and MA wrote the manuscript. SSH and MA critically reviewed the manuscript. All authors approved the final manuscript.

## **Funding**

The work is supported by the grant from the SNUH Research Fund (26-2018-0040 to SSH; 26-2022-0040 to SSH), the Richard A. and Nora Eccles Foundation award (A200111 to MA), Richard A. and Nora Eccles Harrison Endowed Chair in Diabetes Research and the Depner Endowed Chair in Nephrology Research.

## **Availability of data and materials**

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request. The untargeted proteomic data have been

deposited in the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (accession no. PXD037505).

### **Competing interests**

The authors declare that they have no competing interests.

## References

1. Afkarian M, Sachs MC, Kestenbaum B, Hirsch IB, Tuttle KR, Himmelfarb J, et al.: Kidney disease and increased mortality risk in type 2 diabetes. *Journal of the American Society of Nephrology*, 24: 302-308, 2013
2. Hahr AJ, Molitch ME: Management of diabetes mellitus in patients with CKD: core curriculum 2022. *American Journal of Kidney Diseases*, 79: 728-736, 2022
3. Tuttle KR, Bakris GL, Bilous RW, Chiang JL, De Boer IH, Goldstein-Fuchs J, et al.: Diabetic kidney disease: a report from an ADA Consensus Conference. *Diabetes care*, 37: 2864-2883, 2014
4. Furuichi K, Shimizu M, Yamanouchi M, Hoshino J, Sakai N, Iwata Y, et al.: Clinicopathological features of fast eGFR decliners among patients with diabetic nephropathy. *BMJ Open Diabetes Research and Care*, 8: e001157, 2020
5. Oshima M, Shimizu M, Yamanouchi M, Toyama T, Hara A, Furuichi K, et al.: Trajectories of kidney function in diabetes: a clinicopathological update. *Nature Reviews Nephrology*, 17: 740-750, 2021
6. Maclsaac RJ, Ekinici EI, Jerums G: Markers of and risk factors for the development and progression of diabetic kidney disease. *American journal of kidney diseases*, 63: S39-S62, 2014
7. Siddiqui K, George TP, Joy SS, Alfadda AA: Risk factors of chronic kidney disease among type 2 diabetic patients with longer duration of diabetes. *Frontiers in Endocrinology*, 13: 1079725, 2022
8. Anderson AH, Xie D, Wang X, Baudier RL, Orlandi P, Appel LJ, et al.: Novel risk factors for progression of diabetic and nondiabetic CKD: findings from the chronic renal insufficiency cohort (CRIC) study. *American Journal of Kidney Diseases*, 77: 56-73. e51, 2021
9. Colombo M, Looker HC, Farran B, Hess S, Groop L, Palmer CN, et al.: Serum kidney injury molecule 1 and  $\beta$  2-microglobulin perform as well as larger biomarker panels for prediction of rapid decline in renal function in type 2 diabetes. *Diabetologia*, 62: 156-168, 2019
10. Looker HC, Colombo M, Hess S, Brosnan MJ, Farran B, Dalton RN, et al.: Biomarkers of rapid chronic kidney disease progression in type 2 diabetes. *Kidney international*, 88: 888-896, 2015
11. Rossing P, Caramori ML, Chan JC, Heerspink HJ, Hurst C, Khunti K, et al.: KDIGO 2022 clinical practice guideline for diabetes management in chronic kidney disease. *Kidney international*, 102: S1-S127, 2022
12. Barrera-Chimal J, Lima-Posada I, Bakris GL, Jaisser F: Mineralocorticoid receptor antagonists in diabetic kidney disease—Mechanistic and therapeutic effects. *Nature Reviews Nephrology*, 18: 56-70, 2022
13. Tong L-L, Adler SG: Diabetic kidney disease treatment: new perspectives. *Kidney research and clinical practice*, 41: S63, 2022
14. Tofte N, Lindhardt M, Adamova K, Bakker SJ, Beige J, Beulens JW, et al.: 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. *The lancet Diabetes & endocrinology*, 8: 301-312, 2020
15. Zaghlool SB, Halama A, Stephan N, Gudmundsdottir V, Gudnason V, Jennings LL, et al.: Metabolic and proteomic signatures of type 2 diabetes subtypes in an Arab population. *Nature communications*, 13: 7121, 2022
16. Gupta J, Mitra N, Kanetsky PA, Devaney J, Wing MR, Reilly M, et al.: Association between albuminuria, kidney function, and inflammatory biomarker profile in CKD in CRIC. *Clinical journal of the American Society of Nephrology : CJASN*, 7: 1938-1946, 2012 10.2215/CJN.03500412



17. Mehta NN, Matthews GJ, Krishnamoorthy P, Shah R, McLaughlin C, Patel P, et al.: Higher plasma CXCL12 levels predict incident myocardial infarction and death in chronic kidney disease: findings from the Chronic Renal Insufficiency Cohort study. *European heart journal*, 2013 10.1093/eurheartj/eh481
18. Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro III AF, Feldman HI, et al.: A new equation to estimate glomerular filtration rate. *Annals of internal medicine*, 150: 604-612, 2009
19. Tervaert TWC, Mooyaart AL, Amann K, Cohen AH, Cook HT, Drachenberg CB, et al.: Pathologic classification of diabetic nephropathy. *Journal of the American Society of Nephrology*, 21: 556-563, 2010
20. Grzegorski SJ, Hu Z, Liu Y, Yu X, Ferguson AC, Madarati H, et al.: Disruption of the kringle 1 domain of prothrombin leads to late onset mortality in zebrafish. *Scientific reports*, 10: 4049, 2020
21. Han D, Jin J, Woo J, Min H, Kim Y: Proteomic analysis of mouse astrocytes and their secretome by a combination of FASP and StageTip-based, high pH, reversed-phase fractionation. *Proteomics*, 14: 1604-1609, 2014
22. Rhee SJ, Han D, Lee Y, Kim H, Lee J, Lee K, et al.: Comparison of serum protein profiles between major depressive disorder and bipolar disorder. *BMC psychiatry*, 20: 1-11, 2020
23. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M: Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics*, 13: 2513-2526, 2014
24. Han D, Jin J, Yu J, Kim K, Kim Y: Integrated approach using multistep enzyme digestion, TiO<sub>2</sub> enrichment, and database search for in-depth phosphoproteomic profiling. *Proteomics*, 15: 618-623, 2015
25. Mertins P, Tang LC, Krug K, Clark DJ, Gritsenko MA, Chen L, et al.: Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography–mass spectrometry. *Nature protocols*, 13: 1632-1661, 2018
26. Tyanova S, Temu T, Cox J: The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature protocols*, 11: 2301-2319, 2016
27. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M: Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research*, 10: 1794-1805, 2011
28. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al.: Global quantification of mammalian gene expression control. *Nature*, 473: 337-342, 2011
29. Gao Y, Wang H, Nicora CD, Shi T, Smith RD, Sigdel TK, et al.: LC-SRM-Based targeted quantification of urinary protein biomarkers, Springer, 2018
30. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al.: Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, 26: 966-968, 2010
31. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al.: Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature biotechnology*, 42: 293-304, 2024
32. Consortium U: UniProt: a worldwide hub of protein knowledge. *Nucleic acids research*, 47: D506-D515, 2019
33. Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, et al.: WikiPathways: connecting communities. *Nucleic acids research*, 49: D613-D621, 2021
34. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M: KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic acids research*, 51: D587-D592, 2023
35. Broquet A, Gourain V, Goronflot T, Le Mabecque V, Sinha D, Ashayeripanah M, et al.: Sepsis-trained macrophages promote antitumoral tissue-resident T cells. *Nature Immunology*: 1-18, 2024

36. Kim M-G, Yun D, Kang CL, Hong M, Hwang J, Moon KC, et al.: Kidney VISTA prevents IFN- $\gamma$ /IL-9 axis-mediated tubulointerstitial fibrosis after acute glomerular injury. *The Journal of clinical investigation*, 132, 2022
37. Kuppe C, Ibrahim MM, Kranz J, Zhang X, Ziegler S, Perales-Patón J, et al.: Decoding myofibroblast origins in human kidney fibrosis. *Nature*, 589: 281-286, 2021
38. Denz R, Klaaßen-Mielke R, Timmesfeld N: A comparison of different methods to adjust survival curves for confounders. *Statistics in Medicine*, 42: 1461-1479, 2023
39. Barratt J, Weitz I: Complement factor D as a strategic target for regulating the alternative complement pathway. *Frontiers in immunology*, 12: 712572, 2021
40. Haslund D, Ryø LB, Majidi SS, Rose I, Skipper KA, Fryland T, et al.: Dominant-negative SERPING1 variants cause intracellular retention of C1 inhibitor in hereditary angioedema. *The Journal of Clinical Investigation*, 129: 388-405, 2019
41. Meri S, Haapasalo K: Function and dysfunction of complement factor H during formation of lipid-rich deposits. *Frontiers in Immunology*, 11: 611830, 2020
42. De Boer EC, Van Mourik AG, Jongerius I: Therapeutic lessons to be learned from the role of complement regulators as double-edged sword in health and disease. *Frontiers in immunology*, 11: 578069, 2020
43. Mohan Rao LV, Esmon CT, Pendurthi UR: Endothelial cell protein C receptor: a multiliganded and multifunctional receptor. *Blood, The Journal of the American Society of Hematology*, 124: 1553-1562, 2014
44. Naaman SC, Bakris GL: Slowing diabetic kidney disease progression: where do we stand today?, 2021
45. Zoccali C, Floege J: Moderator's view: after SGLT2i and MRA antagonists, where do we go? *Clinical Kidney Journal*, 17: sfae013, 2024
46. Flyvbjerg A: The role of the complement system in diabetic nephropathy. *Nature Reviews Nephrology*, 13: 311-318, 2017
47. Tan SM, Snelson M, Østergaard JA, Coughlan MT: The complement pathway: New insights into immunometabolic signaling in diabetic kidney disease. *Antioxidants & redox signaling*, 37: 781-801, 2022
48. Pelletier K, Bonnefoy A, Chapdelaine H, Pichette V, Lejars M, Madore F, et al.: Clinical value of complement activation biomarkers in overt diabetic nephropathy. *Kidney International Reports*, 4: 797-805, 2019
49. Vaisar T, Durbin-Johnson B, Whitlock K, Babenko I, Mehrotra R, Rocke DM, et al.: Urine complement proteins and the risk of kidney disease progression and mortality in type 2 diabetes. *Diabetes Care*, 41: 2361-2369, 2018
50. Li X-Q, Chang D-Y, Chen M, Zhao M-H: Complement activation in patients with diabetic nephropathy. *Diabetes & Metabolism*, 45: 248-253, 2019
51. Zhao L, Zhang Y, Liu F, Yang H, Zhong Y, Wang Y, et al.: Urinary complement proteins and risk of end-stage renal disease: quantitative urinary proteomics in patients with type 2 diabetes and biopsy-proven diabetic nephropathy. *Journal of Endocrinological Investigation*, 44: 2709-2723, 2021
52. Huang Y, Xu J, Wu X, Chen X, Bai X, Zhuang Y, et al.: High expression of complement components in the kidneys of type 2 diabetic rats with diabetic nephropathy. *Frontiers in Endocrinology*, 10: 459, 2019
53. Tan SM, Ziemann M, Thallas-Bonke V, Snelson M, Kumar V, Laskowski A, et al.: Complement C5a induces renal injury in diabetic kidney disease by disrupting mitochondrial metabolic agility. *Diabetes*, 69: 83-98, 2020
54. Rizk DV, Maillard N, Julian BA, Knoppova B, Green TJ, Novak J, et al.: The emerging role of complement proteins as a target for therapy of IgA nephropathy. *Frontiers in immunology*, 10: 442135, 2019

55. Zipfel PF, Wiech T, Rudnick R, Afonso S, Person F, Skerka C: Complement inhibitors in clinical trials for glomerular diseases. *Frontiers in immunology*, 10: 464149, 2019
56. Hillmen P, Szer J, Weitz I, Röth A, Höchsmann B, Panse J, et al.: Pegcetacoplan versus eculizumab in paroxysmal nocturnal hemoglobinuria. *New England Journal of Medicine*, 384: 1028-1037, 2021
57. Wooden B, Tarragon B, Navarro-Torres M, Bomback AS: Complement inhibitors for kidney disease. *Nephrology Dialysis Transplantation*, 38: ii29-ii39, 2023
58. Angeletti A, Cantarelli C, Petrosyan A, Andrighetto S, Budge K, D'Agati VD, et al.: Loss of decay-accelerating factor triggers podocyte injury and glomerulosclerosis. *Journal of Experimental Medicine*, 217: e20191699, 2020
59. Ghosh P, Sahoo R, Vaidya A, Chorev M, Halperin JA: Role of complement and complement regulatory proteins in the complications of diabetes. *Endocrine reviews*, 36: 272-288, 2015
60. Bongoni AK, Lu B, Salvaris EJ, Roberts V, Fang D, Mcrae JL, et al.: Overexpression of human CD55 and CD59 or treatment with human CD55 protects against renal ischemia-reperfusion injury in mice. *The Journal of Immunology*, 198: 4837-4845, 2017
61. Kragesteen BK, Giladi A, David E, Halevi S, Geirsdóttir L, Lempke OM, et al.: The transcriptional and regulatory identity of erythropoietin producing cells. *Nature medicine*, 29: 1191-1200, 2023

Table 1. Baseline characteristics of the SNUH-DN cohort

Variables	Total (n = 64)	Cluster 1 (n = 40)	Cluster 2 (n = 24)	<i>P</i>
Age (years)	53.3 ± 11.1	52.1 ± 10.9	55.2 ± 11.5	0.282
Male (%)	71.9	75.0	66.7	0.667
Body mass index (kg/m <sup>2</sup> )	25.0 ± 3.3	25.2 ± 3.1	24.7 ± 3.5	0.594
Hypertension (%)	92.2	95.0	87.5	0.355
Diabetic duration (years)	10 (4–14)	8 (2–12)	11 (4–20)	0.278
Diabetic retinopathy (%)	54.7	52.5	58.3	0.846
Diabetic neuropathy (%)	29.7	30.0	29.2	1.000
Ischemic heart disease (%)	7.8	12.5	0	0.148
Cerebrovascular disease (%)	4.7	2.5	8.3	0.551
Hemoglobin A1c (%)	7.2 (6.6–8.4)	7.2 (6.6–7.6)	7.4 (6.8–8.9)	0.178
PCR (g/g)	3.1 (1.7–7.0)	2.2 (1.2–3.6)	7.1 (4.9–9.5)	<0.001
eGFR (mL/min/1.73 m <sup>2</sup> )	56 (44–75)	61.2 (48–85)	43 (33–60)	<0.001
Glomerular classification (%)				
I	14.1	17.5	8.3	0.003
IIA	12.5	17.5	4.2	
IIB	32.8	42.5	16.7	
III	12.5	10.0	16.7	
IV	28.1	12.5	54.2	
Interstitial fibrosis and Tubular atrophy (%)				
0	4.7	7.5	0.0	<0.001
1	35.9	47.5	16.7	
2	46.9	45.0	50.0	
3	12.5	0.0	33.3	
Interstitial Inflammation (%)				
0	10.9	17.5	0.0	0.044
1	84.4	80.0	91.7	
2	4.7	2.5	8.3	
Arteriolar hyalinosis (%)				
0	25.0	35.0	8.3	0.052

1	28.1	22.5	37.5	
2	46.9	42.5	54.2	
Arteriosclerosis (%)				
0	25.0	27.5	20.8	0.674
1	35.9	37.5	33.3	
2	39.1	35.0	45.8	
Follow-up time (years)	6.2 (2.4–7.8)	7.1 (4.9–8.4)	2.7 (1.1–5.1)	<0.001

---

PCR, random urine protein-to-creatinine ratio; eGFR, estimated glomerular filtration rate.

Table 2. Association between cluster and kidney disease progression in the SNUH-DN cohort

Variables	Univariable		Multivariable*	
	HR (95% CI)	<i>P</i>	HR (95% CI)	<i>P</i>
Cluster 2 (vs. 1)	4.63 (2.16–9.93)	<0.001	4.62 (1.86–11.50)	0.001
Age (per 1 year)	1.01 (0.98–1.03)	0.759	0.94 (0.90–0.98)	0.004
Female (vs. male)	0.47 (0.20–1.08)	0.074	0.27 (0.10–0.70)	0.007
Body mass index (per 1 kg/m <sup>2</sup> )	0.93 (0.83–1.04)	0.224	0.88 (0.78–0.99)	0.039
Hypertension (vs. none)	1.14 (0.35–3.74)	0.827	1.03 (0.29–3.68)	0.964
Diabetic duration (per 1 year)	1.02 (0.98–1.07)	0.380	1.00 (0.95–1.05)	0.969
PCR (per 1 g/g)	1.16 (1.08–1.25)	<0.001	1.16 (1.06–1.28)	0.002
eGFR (per 1 mL/min/1.73 m <sup>2</sup> )	0.98 (0.96–0.99)	0.006	0.99 (0.97–1.00)	0.142

\*Adjusted for age, sex, body mass index, hypertension, diabetic duration, proteinuria, and estimated glomerular filtration rate.

HR, hazard ratio; CI, confidence interval; PCR, random urine protein-to-creatinine ratio; eGFR, estimated glomerular filtration rate.

Table 3. Association between complement score and kidney disease progression in the SNUH-DN cohort

Variables	Univariable		Multivariable*	
	HR (95% CI)	<i>P</i>	HR (95% CI)	<i>P</i>
High complement score (vs. low)	2.67 (1.37–5.18)	0.004	2.41 (1.07–5.40)	0.033
Age (per 1 year)	1.00 (0.98–1.03)	0.759	0.95 (0.91–0.99)	0.011
Female (vs. male)	0.47 (0.20–1.08)	0.074	0.30 (0.11–0.79)	0.015
Body mass index (per 1 kg/m <sup>2</sup> )	0.93 (0.83–1.04)	0.224	0.88 (0.78–1.00)	0.050
Hypertension (vs. none)	1.14 (0.35–3.74)	0.827	0.89 (0.25–3.11)	0.852
Diabetic duration (per 1 year)	1.02 (0.98–1.07)	0.380	1.03 (0.98–1.08)	0.311
PCR (per 1 g/g)	1.16 (1.08–1.25)	<0.001	1.14 (1.03–1.26)	0.010
eGFR (per 1 mL/min/1.73 m <sup>2</sup> )	0.98 (0.96–0.99)	0.006	0.99 (0.97–1.00)	0.142

\*Adjusted for age, sex, body mass index, hypertension, diabetic duration, proteinuria, and estimated glomerular filtration rate.

HR, hazard ratio; CI, confidence interval; PCR, random urine protein-to-creatinine ratio; eGFR, estimated glomerular filtration rate.

Table 4. Baseline characteristics of the CRIC-T2D cohort

Variables	Total (n = 282)	No progression (n = 101)	Progression (n = 181)	<i>P</i>
Age (years)	60.8 ± 7.5	61.8 ± 7.3	60.3 ± 7.5	0.103
Male (%)	62.4	58.4	64.6	0.365
Body mass index (kg/m <sup>2</sup> )	34.7 ± 7.6	34.7 ± 8.2	34.7 ± 7.2	0.997
Race and ethnicity				
Other	18.1	11.9	21.5	0.105
White non-Hispanic	30.1	34.7	27.6	
Black non-Hispanic	51.8	53.5	50.8	
Hypertension (%)	93.6	91.1	95.0	0.297
Cardiovascular disease (%)	46.5	44.6	47.5	0.724
Hemoglobin A1c (%)	7.1 (6.3–8.1)	7.0 (6.2–7.8)	7.1 (6.4–8.2)	0.128
24-hour urine protein (g)	0.48 (0.10–1.87)	0.10 (0.05–0.38)	1.03 (0.30–3.30)	<0.001
eGFR (mL/min/1.73 m <sup>2</sup> )	42 (37–49)	43.7 (38–48)	41 (35–49)	0.031
Follow-up time (years)	9.7 (6.3–12.0)	9.7 (5.2–12.0)	9.7 (6.5–12.0)	0.668

eGFR, estimated glomerular filtration rate.

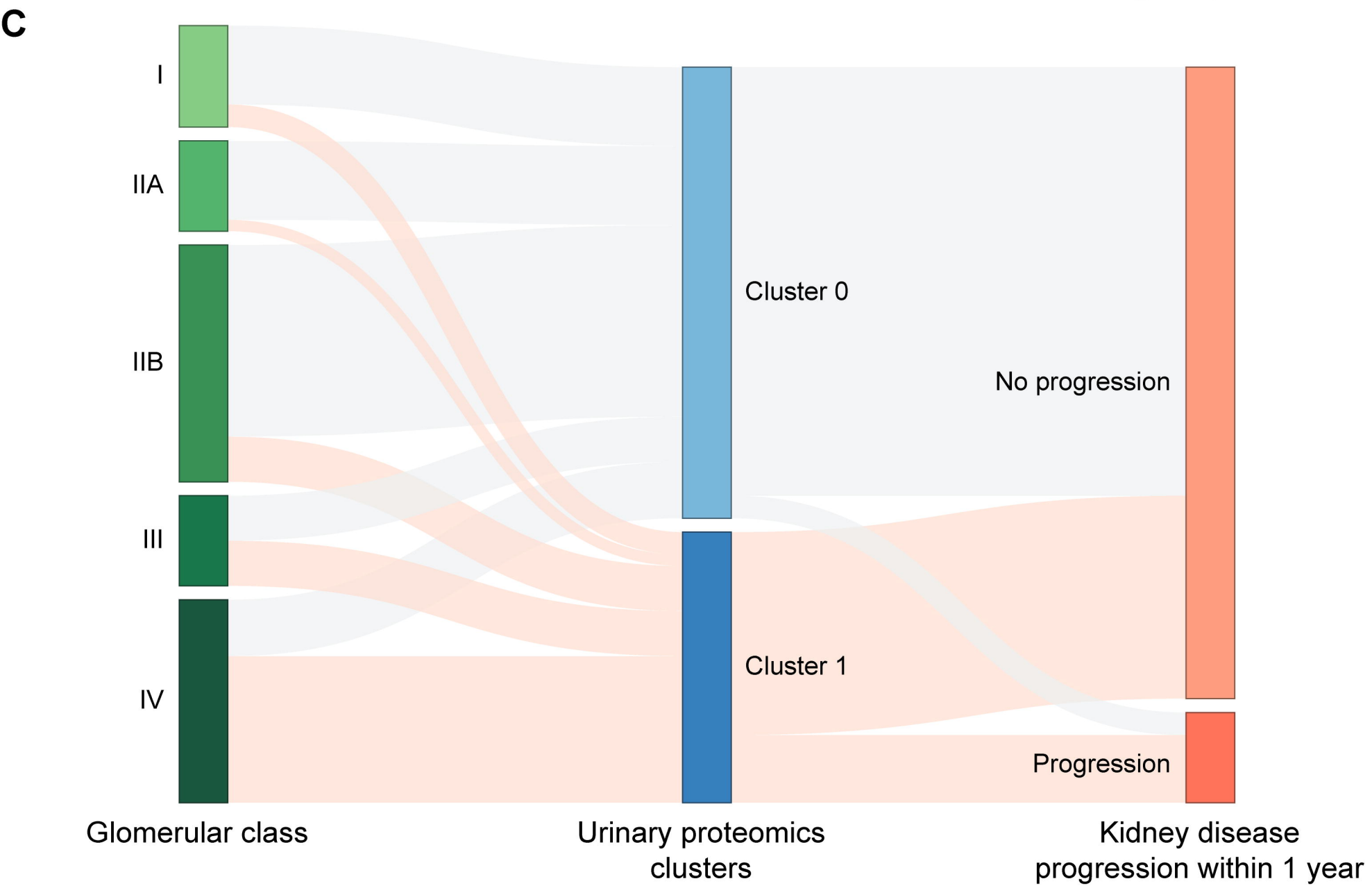
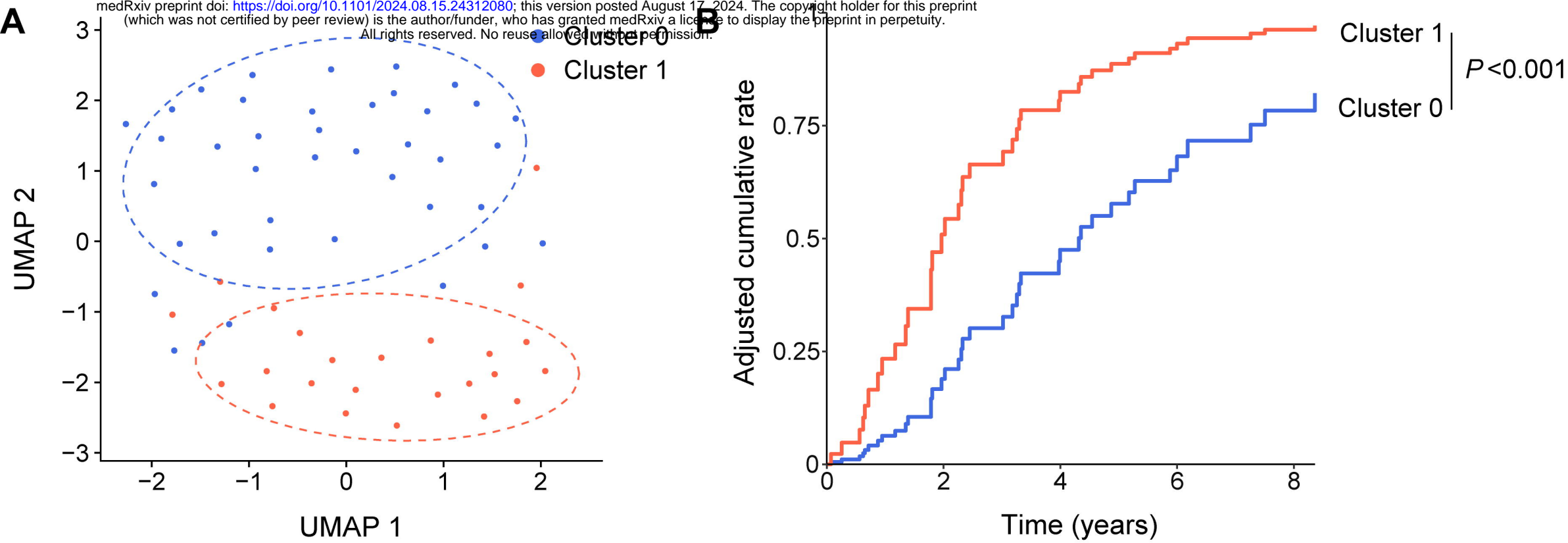


Table 5. Association between complement score and kidney disease progression in the CRIC-T2D cohort

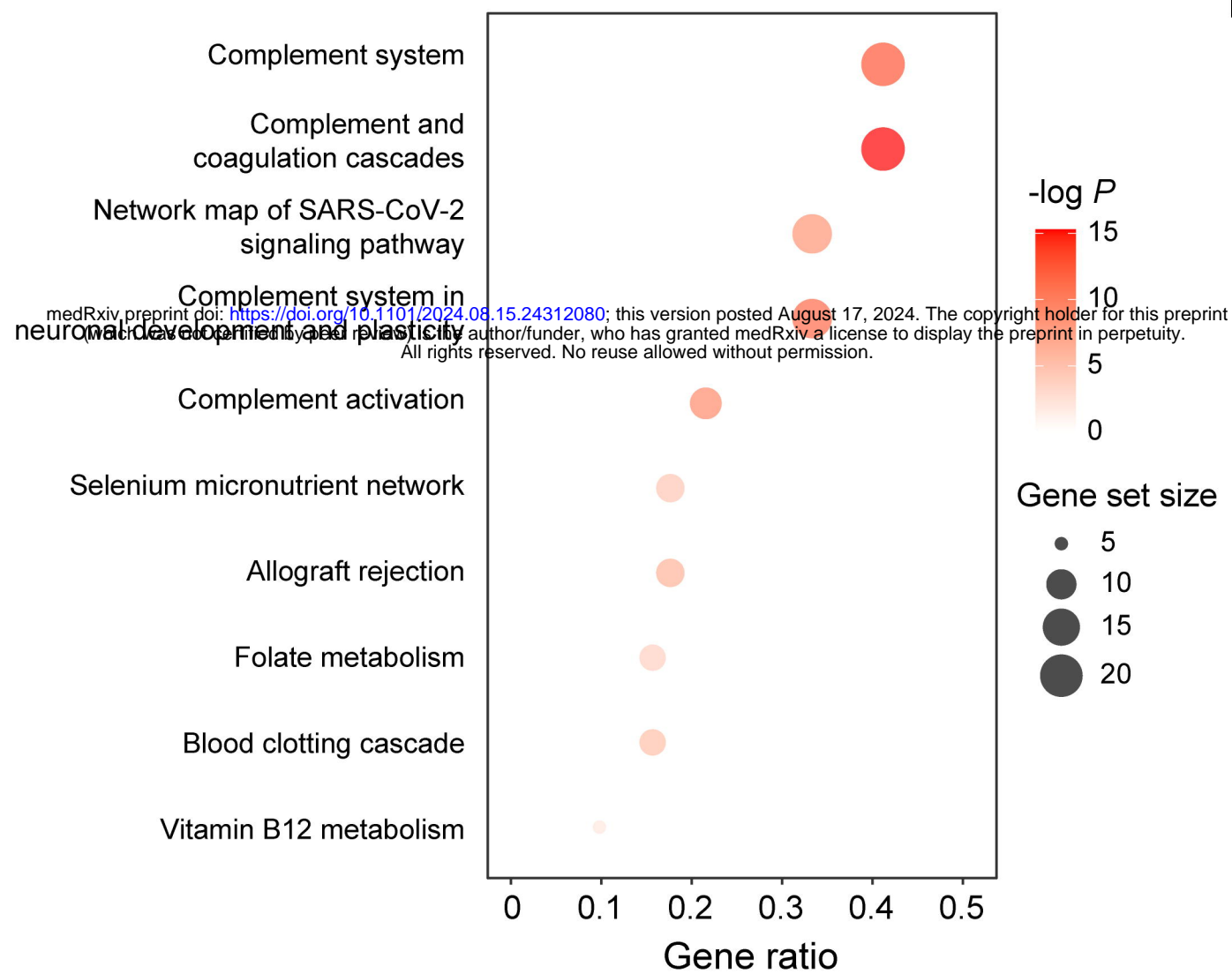
Variables	Univariable		Multivariable*	
	HR (95% CI)	<i>P</i>	HR (95% CI)	<i>P</i>
High score (vs. low)	3.76 (2.74–5.16)	<0.001	2.51 (1.76–3.57)	<0.001
Age (per 1 year)	0.98 (0.96–1.00)	0.120	1.00 (0.97–1.02)	0.720
Female (vs. male)	1.02 (0.75–1.38)	0.903	1.21 (0.87–1.68)	0.263
Body mass index (per 1 kg/m <sup>2</sup> )	1.00 (0.98–1.02)	0.682	1.00 (0.97–1.02)	0.822
Race and ethnicity (vs. other)				
White non-Hispanic	0.55 (0.36–0.49)	0.707	0.79 (0.51–1.23)	0.296
Black non-Hispanic	0.01 (0.84–1.03)	0.070	0.96 (0.65–1.42)	0.839
Hypertension (vs. none)	1.60 (0.82–3.13)	0.171	1.71 (0.87–3.37)	0.121
24-hour urine protein (per 1 g)	1.31 (1.25–1.37)	<0.001	1.23 (1.17–1.30)	<0.001
eGFR (per 1 mL/min/1.73 m <sup>2</sup> )	0.97 (0.96–0.99)	0.003	0.99 (0.97–1.01)	0.263

\*Adjusted for age, sex, body mass index, race and ethnicity, hypertension, 24-hour proteinuria, and estimated glomerular filtration rate.

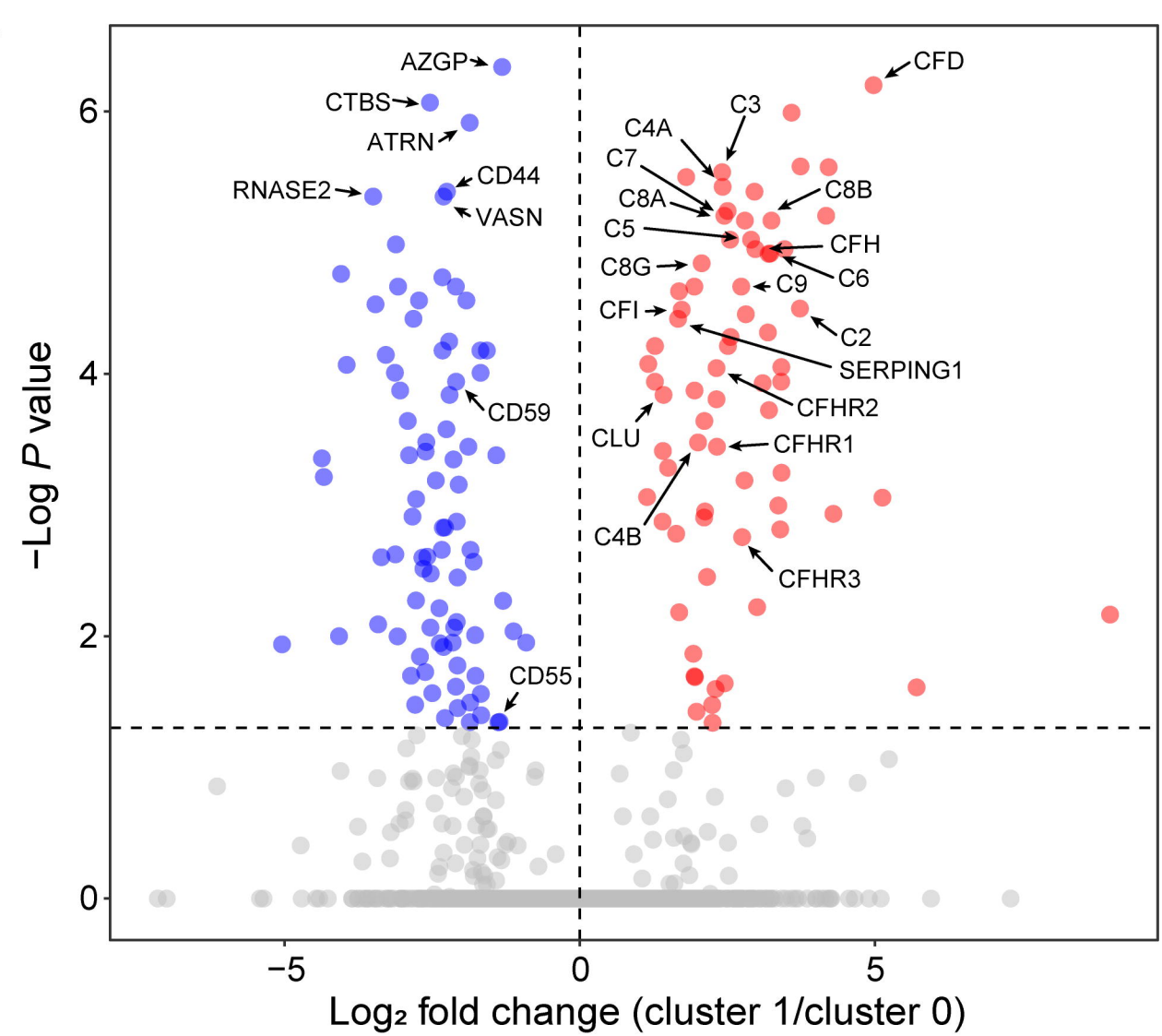
HR, hazard ratio; CI, confidence interval; eGFR, estimated glomerular filtration rate.



A



B



C

