International Journal of
Molecular Sciences

MDPI

Article

Differential Urinary Proteome Analysis for Predicting Prognosis in Type 2 Diabetes Patients with and without Renal Dysfunction

Hee-Sung Ahn [1,†], Jong Ho Kim [2,†], Hwangkyo Jeong [3], Jiyoung Yu [1], Jeonghun Yeom [4], Sang Heon Song [2], Sang Soo Kim [2], In Joo Kim [2,*] and Kyunggon Kim [1,3,5,6,*]

1 Asan Institute for Life Sciences, Asan Medical Center, Seoul 05505, Korea; zaulim3@gmail.com (H.-S.A.); yujiyoung202@gmail.com (J.Y.)
2 Department of Internal Medicine and Biomedical Research Institute, Pusan National University Hospital, Busan 49241, Korea; bedaya@hanmail.net (J.H.K.); shsong0209@gmail.com (S.H.S.); drsskim7@gmail.com (S.S.K.)
3 Department of Biomedical Sciences, University of Ulsan College of Medicine, Seoul 05505, Korea; hkyo723@naver.com
4 Convergence Medicine Research Center, Asan Institute for Life Sciences, Seoul 05505, Korea; nature8309@gmail.com
5 Clinical Proteomics Core Laboratory, Convergence Medicine Research Center, Asan Medical Center, Seoul 05505, Korea
6 Bio-Medical Institute of Technology, Asan Medical Center, Seoul 05505, Korea
* Correspondence: injkim@pusan.ac.kr (I.J.K.); kkkon1@amc.seoul.kr (K.K.); Tel.: +82-51-240-7224 (I.J.K.); +82-2-1688-7575 (K.K.)
† These authors contributed equally to this work.

Received: 3 April 2020; Accepted: 12 June 2020; Published: 14 June 2020



Abstract: Renal dysfunction, a major complication of type 2 diabetes, can be predicted from estimated glomerular filtration rate (eGFR) and protein markers such as albumin concentration. Urinary protein biomarkers may be used to monitor or predict patient status. Urine samples were selected from patients enrolled in the retrospective diabetic kidney disease (DKD) study, including 35 with good and 19 with poor prognosis. After removal of albumin and immunoglobulin, the remaining proteins were reduced, alkylated, digested, and analyzed qualitatively and quantitatively with a nano LC-MS platform. Each protein was identified, and its concentration normalized to that of creatinine. A prognostic model of DKD was formulated based on the adjusted quantities of each protein in the two groups. Of 1296 proteins identified in the 54 urine samples, 66 were differentially abundant in the two groups (area under the curve (AUC): p-value < 0.05), but none showed significantly better performance than albumin. To improve the predictive power by multivariate analysis, five proteins (ACP2, CTSA, GM2A, MUC1, and SPARCL1) were selected as significant by an AUC-based random forest method. The application of two classifiers—support vector machine and random forest—showed that the multivariate model performed better than univariate analysis of mucin-1 (AUC: 0.935 vs. 0.791) and albumin (AUC: 1.0 vs. 0.722). The urinary proteome can reflect kidney function directly and can predict the prognosis of patients with chronic kidney dysfunction. Classification based on five urinary proteins may better predict the prognosis of DKD patients than urinary albumin concentration or eGFR.

Keywords: urine; diabetic kidney disease; kidney function; proteomics; mass spectrometry; statistical clinical model; machine learning

1. Introduction

About 30% of people with diabetes develop diabetic kidney disease (DKD), and the spread of diabetes is increasing worldwide [1,2]. Complications of type 2 diabetes (T2D) mainly cause end-stage renal disease, which is related to high heart disease incidence and mortality [2,3]. Early detection and screening of patients at risk for DKD is important, which may reduce the global burden of T2D.

Because the kidneys filter waste from blood and discharge it as urine, urine can directly reflect kidney function. Unlike plasma, urine can be easily collected non-invasively, with proteins in urine being stable and not vulnerable to sudden degradation [4]. Albuminuria and estimated glomerular filtration rate (eGFR), have been generally used to assess kidney function [2,5]. However, albuminuria is only evaluated after glomerular damage has occurred, and sometimes kidney disease develops before the outbreak of albuminuria [6,7]. Better markers are required to help delay progression to DKD.

Multiple-biomarker approaches based on proteomics, including urinary proteomics, may overcome the limitations of markers diagnostic for DKD [4,8]. This study was designed to identify a urinary multi-protein panel that could predict progression to DKD in patients with T2D.

2. Results

2.1. Baseline Characteristics of Clinical Samples Used in the Study

Table 1 summarizes the baseline characteristics of patients in the poor and good prognosis groups. All factors did not differ significantly in the two groups, including sex, age, body mass index (BMI), duration of follow-up, systolic blood pressure (SBP), glycated hemoglobin concentration (HbA1c), lipid profile, percent with diabetic retinopathy, and the percentages treated with RAS inhibitors, anti-hypertensive agents, and lipid-lowering agents (Bonferroni corrected p-value > 0.05/17 = 0.0029; Table S1).

Table 1. Baseline characteristics of the patients with type 2 diabetes (T2D) with and without renal outcomes.

Variable	With Renal Outcome	Without Renal Outcome
Sex, *n* (%)		
Male	8 (42.1)	11 (31.4)
Female	11 (57.9)	24 (68.6)
Age at diagnosis of diabetic kidney disease, mean ± SD (years)	54.58 ± 11.66	58.66 ± 9.19
BMI, mean ± SD (kg/m^2)	22.64 ± 3.46	23.81 ± 3.00
Duration of follow-up, mean ± SD (years)	4.80 ± 1.96	4.73 ± 1.94
SBP, mean ± SD (mmHg)	126.58 ± 15.70	125.97 ± 12.07
LDL cholesterol, mean ± SD (mg/dL)	104.89 ± 41.00	99.83 ± 32.32
HDL cholesterol, mean ± SD (mg/dL)	48.42 ± 7.50	52.51 ± 11.51
Triglycerides, mean ± SD (mg/dL)	145.74 ± 99.80	154.57 ± 128.88
eGFR after 1 years, mean ± SD (mL/min/1.73 m^2)	91.52 ± 17.57	88.33 ± 15.46
HbA1c, mean ± SD (%)	8.34 ± 2.09	7.16 ± 1.36
ACR, mean ± SD (mg/g)	213.66 ± 446.75	126.11 ± 419.70
NAPCR, mean ± SD (mg/g)	178.18 ± 209.30	154.76 ± 299.68
PCR, mean ± SD (mg/g)	391.84 ± 652.79	280.87 ± 711.04
Diabetic retinopathy, *n* (%)	9 (47.37)	11 (31.43)
RAS inhibitor, *n* (%)	6 (31.58)	15 (42.86)
Anti-hypertensive agent, *n* (%)	5 (26.32)	12 (34.29)
Lipid-lowering agent, *n* (%)	10 (52.63)	20 (57.14)

Abbreviations: BMI, body mass index; SBP, systolic blood pressure; LDL, low-density lipoprotein; HDL, high-density lipoprotein; eGFR, estimated glomerular filtration rate; HbA1c, glycated hemoglobin; ACR, urine albumin-to-creatinine ratio; NAPCR, urine nonalbumin protein-to-creatinine ratio; PCR, urine protein-to-creatinine ratio.

2.2. Urinary Proteome Analysis for Identification and Label-Free Quantitation

The workflow of data processes contained identified and quantified urinary proteins indicative of disease status, as well as significant proteins to build the clinical models (Figure 1A). Liquid chromatography–mass spectrometry (LC-MS) analysis of the 54 urine samples identified 1296 proteins (Table S2). Of these proteins, 1244 were quantified, and their quantities were adjusted relative to the concentration of creatinine [9,10]. Sample-to-sample variation was subsequently fixed by the amount of proteins, leading to the selection of six endogenous normalization proteins that showed stable abundance in all LC-MS analyses. A boxplot of protein abundances in the 54 samples, composed of 35 patients in good-prognostic group (GPG) and 19 patients in poor-prognostic group (PPG), is depicted in Figure 1B. The normalized abundance of 68 proteins significantly correlated with their immunoassay [11] determined concentrations in urine, with a Pearson's coefficient of 0.502 (permutation p-value < 0.001; Figure 1C).

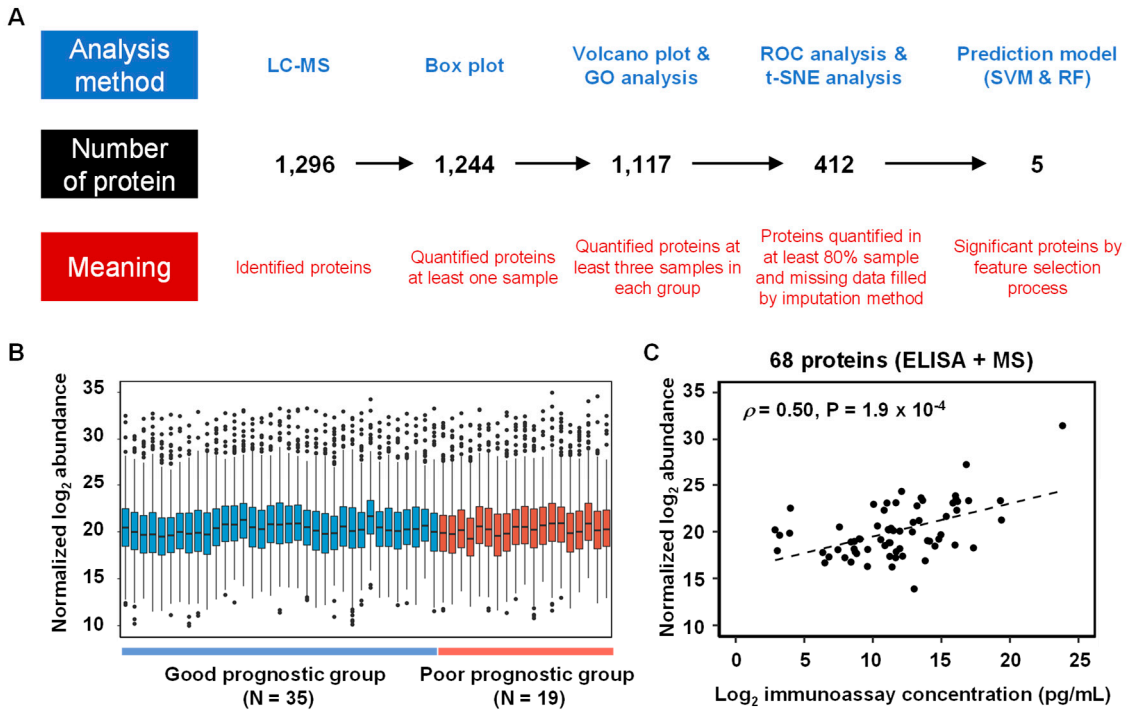


Figure 1. (**A**) Analysis workflow of urinary proteins in the 54 diabetic kidney disease (DKD) patients. The analysis method is written in the upper part, the number of proteins in the middle, and the meaning of the protein in the bottom part. (**B**) Boxplots of normalized urinary protein abundances in the 54 samples (35 patients in good-prognostic group and 19 patients in poor-prognostic group) measured by LC-MS analysis. (**C**) Scatter plot of 68 urine proteins between normalized log2 abundance and log2 immunoassays concentration (Pearson correlation coefficient (ρ): 0.5 and p-value: 1.9×10^{-4}).

2.3. Functional Annotation of Differential Protein Expression in the PPG and GPG Groups

To find the differential abundant proteins (DAPs) from among the 1117 proteins, fold-changes and p-values were calculated by Mann–Whitney U tests of the two groups. A volcano plot showing log2-fold-changes against minus $\log_{10}$ p-values identified 46 proteins as being upregulated in the PPG and 54 proteins in the GPG ($|\log_2$ fold-change$| > 0.5$; p-value < 0.05; Figure 2 and Table S3). These differentially expressed proteins included the six previously described candidate urinary biomarkers (APOE, CO3, COF1, NID1, OSTP-5, and PODXL) of glomerular or tubular injury [11].

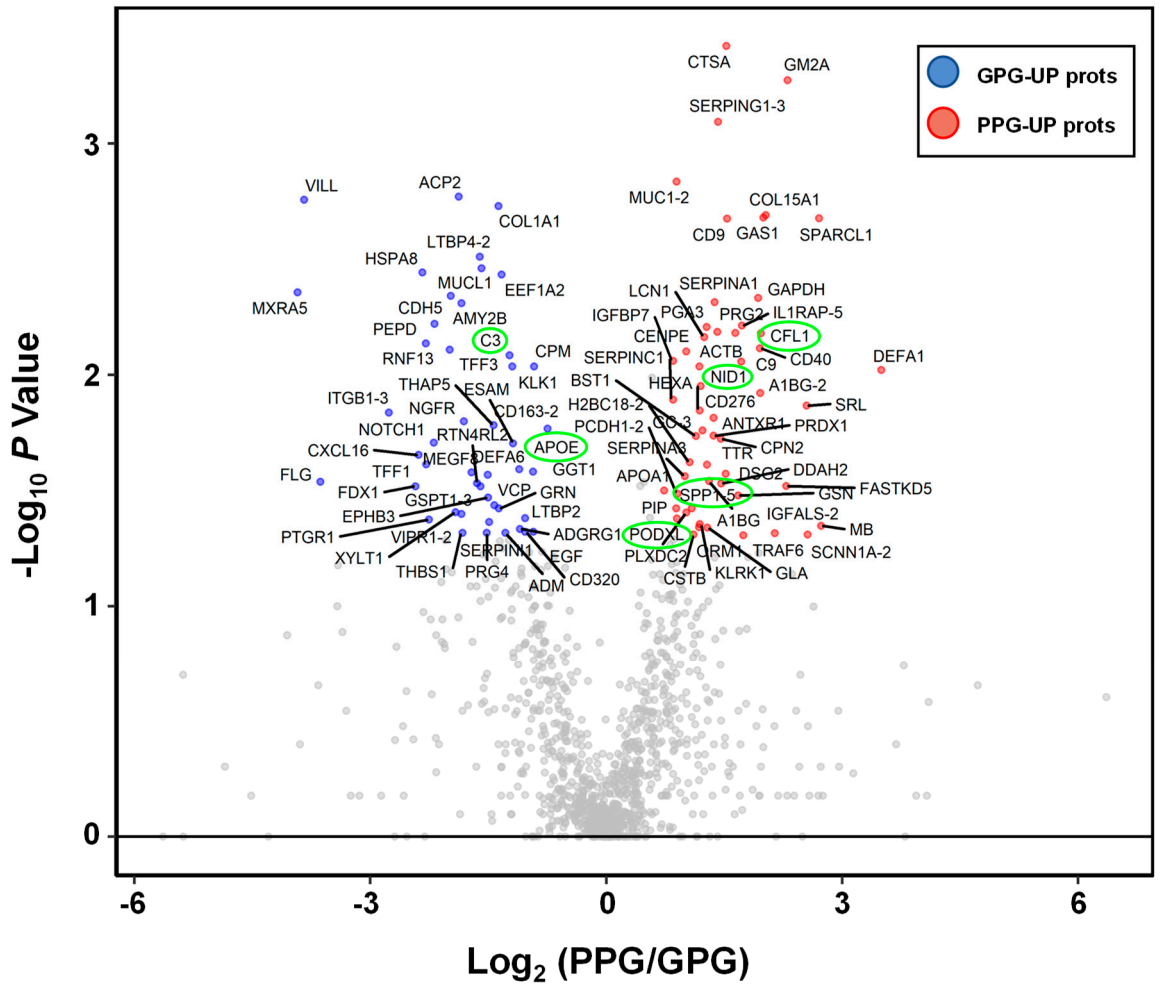


Figure 2. Volcano plot of urinary proteomic data. Volcano plots are depicted with the fold change of each protein abundance and the *p* value was calculated by performing a Mann–Whitney U-test. The averages of the urinary proteomic abundance data of good prognostic group (N = 35) were compared with the averages of the data for poor prognostic group (N = 19). Red circles show 54 urinary proteins that have significant increases in PPG. Blue circles show 46 urinary proteins which have significant decreases in PPG. Gray circles are urinary proteins without statistical meaning. Green circles are previously released as urinary protein markers for glomerular injury or tubular injury.

To determine whether urinary DAPs were associated with specific biological processes, up- and downregulated proteins in the PPG were subjected to gene ontology (GO) enrichment analysis. To integrate the three domains of GO and easily visualize the relationship between terms, ClueGO tools were applied with default settings (kappa score 0.4 and group merger of 50% of genes) to functionally organize the GO term networks [12].

Downregulated proteins in PPG were significantly enriched with an FDR < 0.01 (Figure 3A,B). Biological processes associated with these proteins included negative regulation of lipid localization, collagen catabolic process, positive regulation of neural precursor cell proliferation, and neuron projection regeneration. Resulting analysis of molecular function indicated that transforming growth factor beta binding and cargo receptor activity were annotated. The networks between proteins and functional GO terms showed that three proteins were negative regulators of lipid transport, as well as being associated with another GO term (Figure 3C). These included APOE, which is involved in neuron projection regeneration; THBS1, which is involved in transforming growth factor beta binding; and EGF, which is involved in positive regulation of neural precursor cell proliferation.

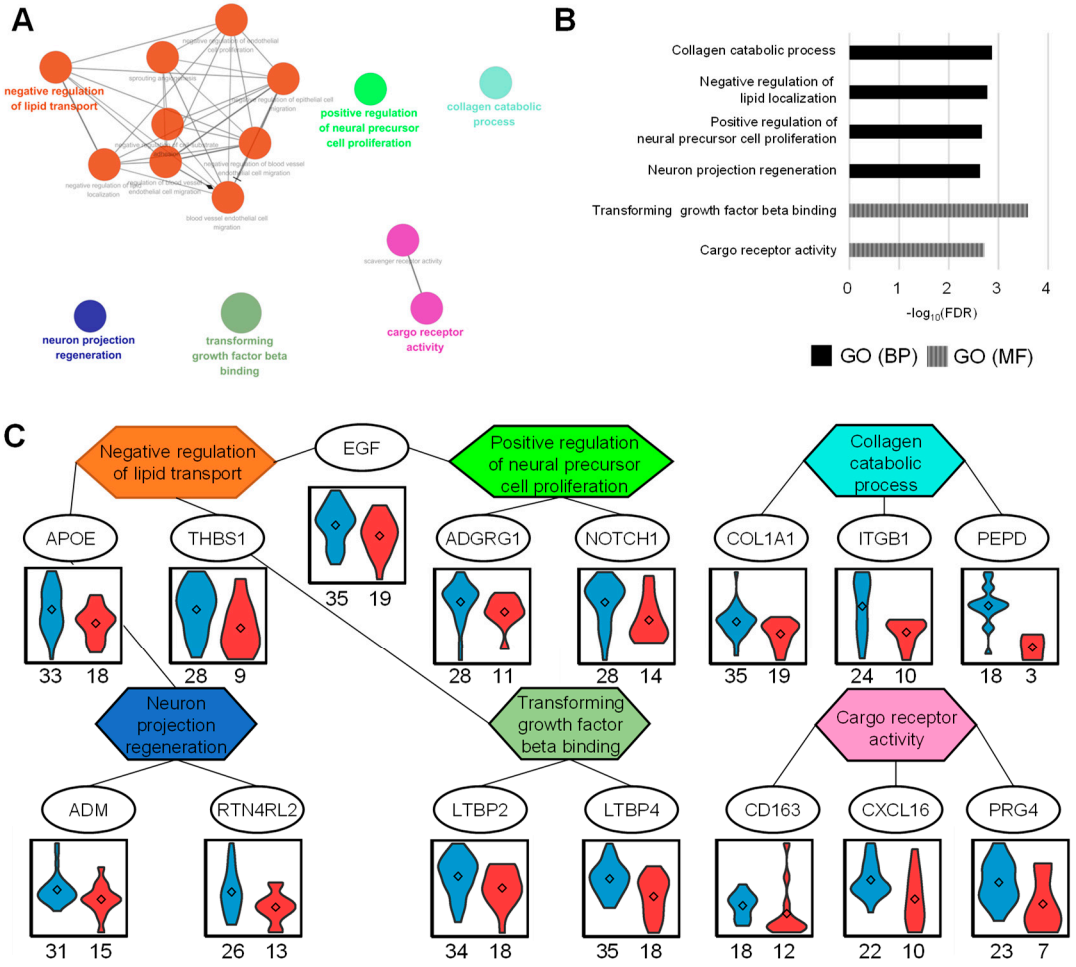


Figure 3. Up-regulated proteome in good prognosis group (GPG) and gene ontology (GO) analysis. (**A**) Functional GO network displaying grouping of GO terms enriched in GPG up-regulated proteins. (**B**) Enriched GO terms in biological process and molecular function. (**C**) The network between GO terms and corresponding proteins represents the relationship between GO terms via the proteins. The abundances of each protein represent the violin plots in two groups. The numbers listed below represent the measured numbers for each group.

Upregulated proteins in PPG were identified in enriched functional GO groups with an FDR < 0.01 (Figure 4A,B). Biological processes associated with these proteins included platelet degranulation, retina homeostasis, and heterotypic cell–cell adhesion. The molecular functional processes related with these proteins contained collagen binding. These urinary proteins were located in the lysosomal lumen and blood microparticles. The networks between proteins and functional GO terms indicated that the proteins in blood microparticles were functionally involved in platelet degranulation (Figure 4C). Platelets in patients with CKD are deficient in reactivity [13]. Leukocytes adhere to and destroy damaged kidney cell walls in patients with CKD [14], accompanied by bone marrow-derived kidney fibrosis, which is highly associated with cell–cell adhesion [15]. CKD is also associated with retinal abnormalities [16] and the possible destruction of retinal homeostasis, as confirmed in this study.

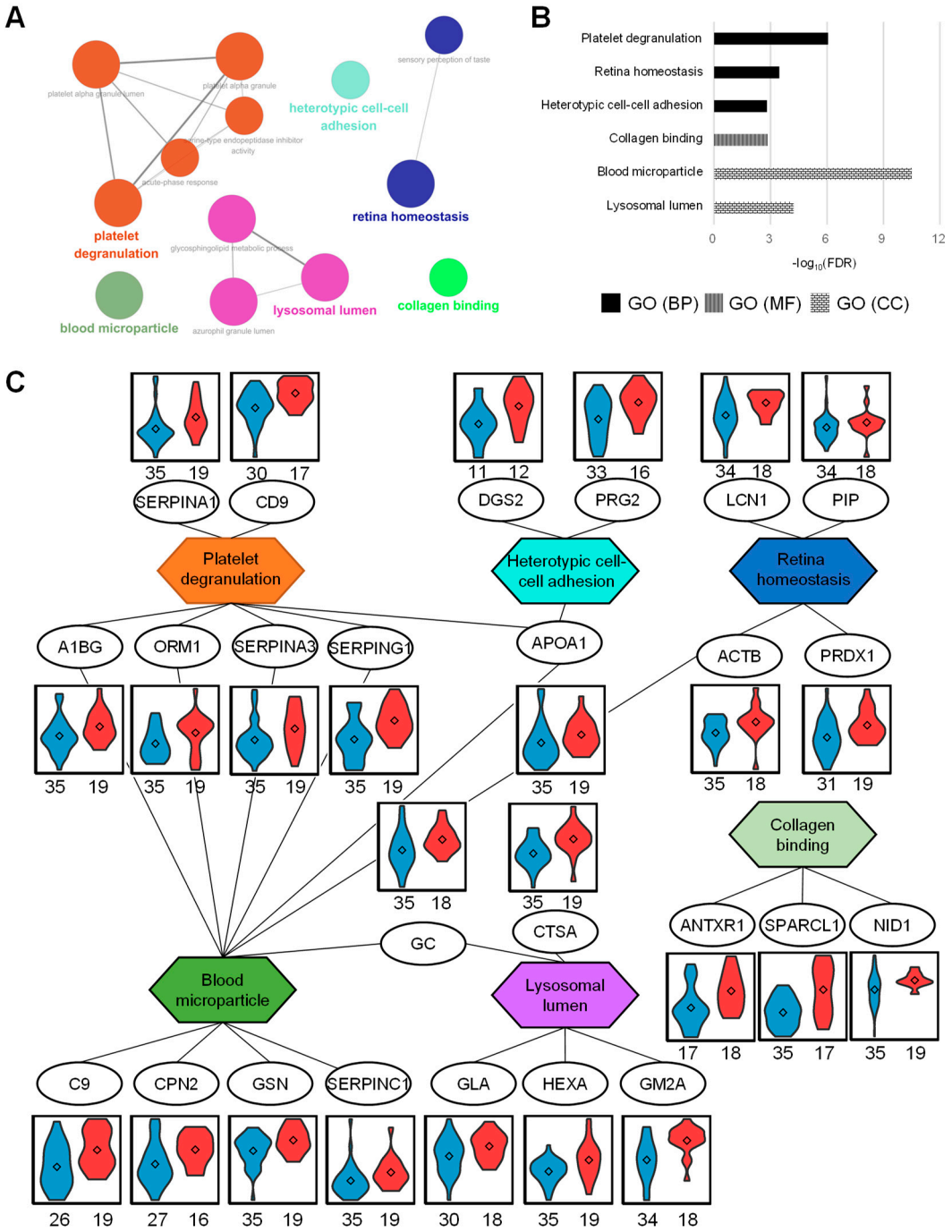


Figure 4. Up-regulated proteome in poor-prognostic group (PPG) and GO analysis. (**A**) Functional GO network displaying grouping of GO terms enriched in PPG up-regulated proteins. (**B**) Enriched GO terms in biological process, molecular function, and cellular component. (**C**) The network between GO terms and their contained proteins represents the relationship between GO terms via the proteins. The abundance of proteins represents the violin plots in two sample groups. The numbers listed below represent the measured numbers for each group.

2.4. Univariate ROC Analysis for Predicting Renal Outcome

To ensure statistical reliability, this study focused on 412 proteins quantified in more than 80% of urine samples [17], with missing values filled by local least squares imputation [18] (Table S4). To confirm that quantified urinary proteins could act as individual biomarkers, univariate receiver operating characteristic (ROC) analysis was performed in samples from the PPG and GPG, with the

resulting histogram of AUC values shown in Figure 5. The AUC values of MUC1, CTSA, ACP2, SERPING1, AMY2B, GM2A, and COL1A1 were 0.791, 0.786, 0.773, 0.771, 0.768, 0.759, and 0.753, respectively. ACP2, AMY2B, and COL1A1 were significantly more abundant, whereas MUC1, CTSA, SERPING1, and GM2A were significantly less abundant, in the GPG than in the PPG (p-value < 0.05 each). The 66 urinary proteins showed significance with AUCs of 0.5 (p-value < 0.05; Table S5). Clinically, urinary albumin is a common marker of DKD [2,5]. The AUCs of 18 proteins were higher than that of albumin (0.722), but the differences were not statistically significant based on likelihood ratio tests.

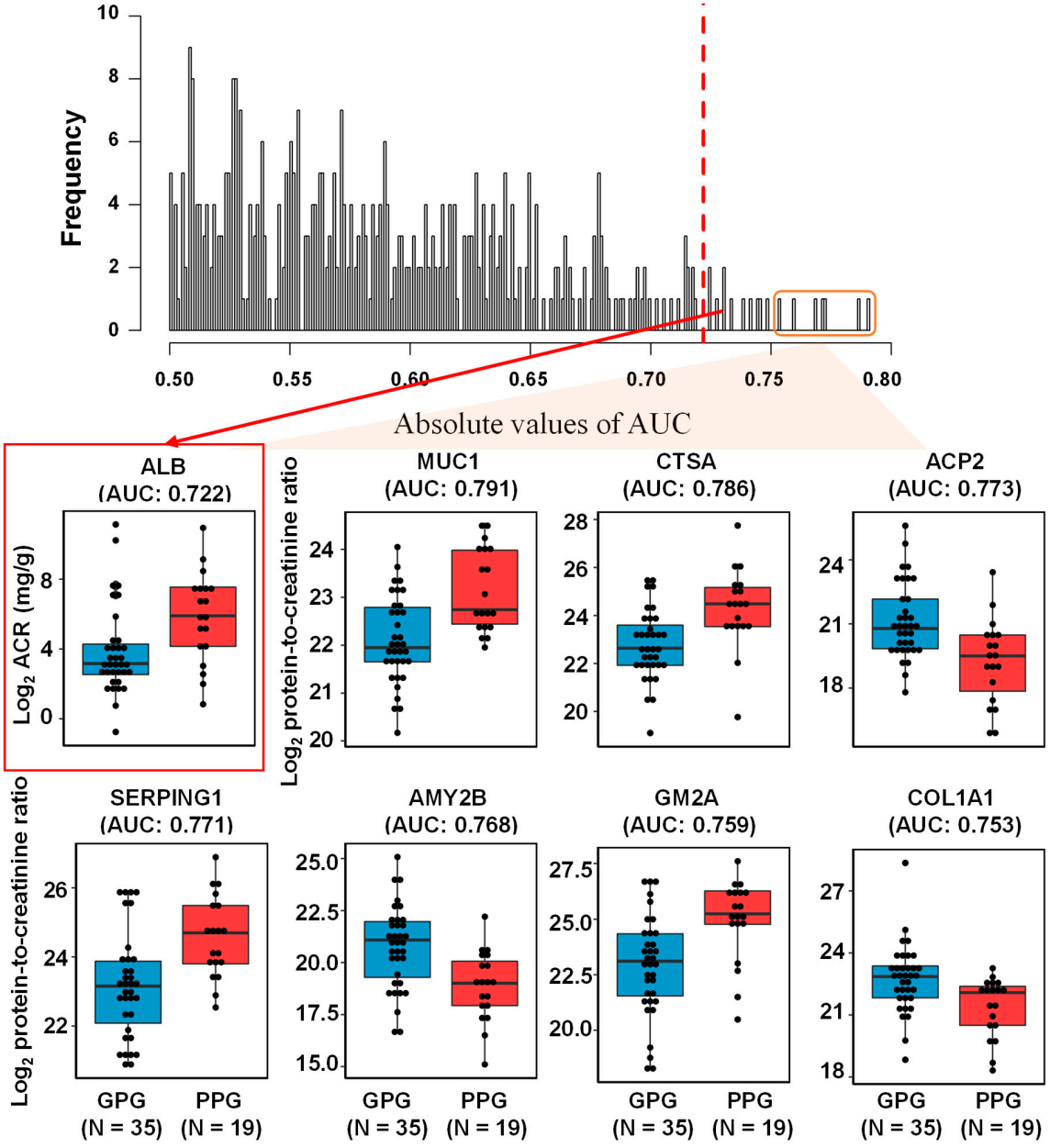


Figure 5. Histogram of area under the ROC curves (AUC) of 412 urinary proteins and ACR. Top seven proteins (MUC1, CTSA, ACP2, SERPING1, AMY2B, GM2A, and COL1A1) and ACR are represented with box plots.

2.5. Multivariate Analysis for Predicting Renal Outcome

To improve predictive performance and find a meaningful combination of proteins that could distinguish patients who were and were not at risk of disease progression, two classifiers of the 412 proteins were generated, one based on random forest (RF) [19] and the other on support vector machine (SVM) [20]. Both the RF and SVM methods selected five proteins (ACP2, CTSA, GM2A, MUC1, and SPARCL1) by an AUC-based RF backward-elimination process [21], according to a >0.3 importance of selection (Table 2). These variables were used to establish a RF model by generating 20,000 decision trees, and a linear SVM model by three repeated iterations of 10-fold cross-validation. Evaluation of the performance of these classifiers showed that the AUC values for RF and SVM were 1.000 and 0.935, respectively (Figure 6A). The nominal binary results of RF and SVM models were transformed in disease prediction scores, which ranged from 0 to 1 (Figure 6B and Table S6). The two classifiers differed significantly from albumin-to-creatinine ratio (likelihood ratio test: p-value < 0.05). These five proteins were located in extracellular exosomes, vesicles, or organelles, with three (ACP2, CTSA, and GM2A) located in the lysosomal lumen, MUC1 placed in plasma membrane, and SPARCL1 interacted with collagen in extracellular matrix.

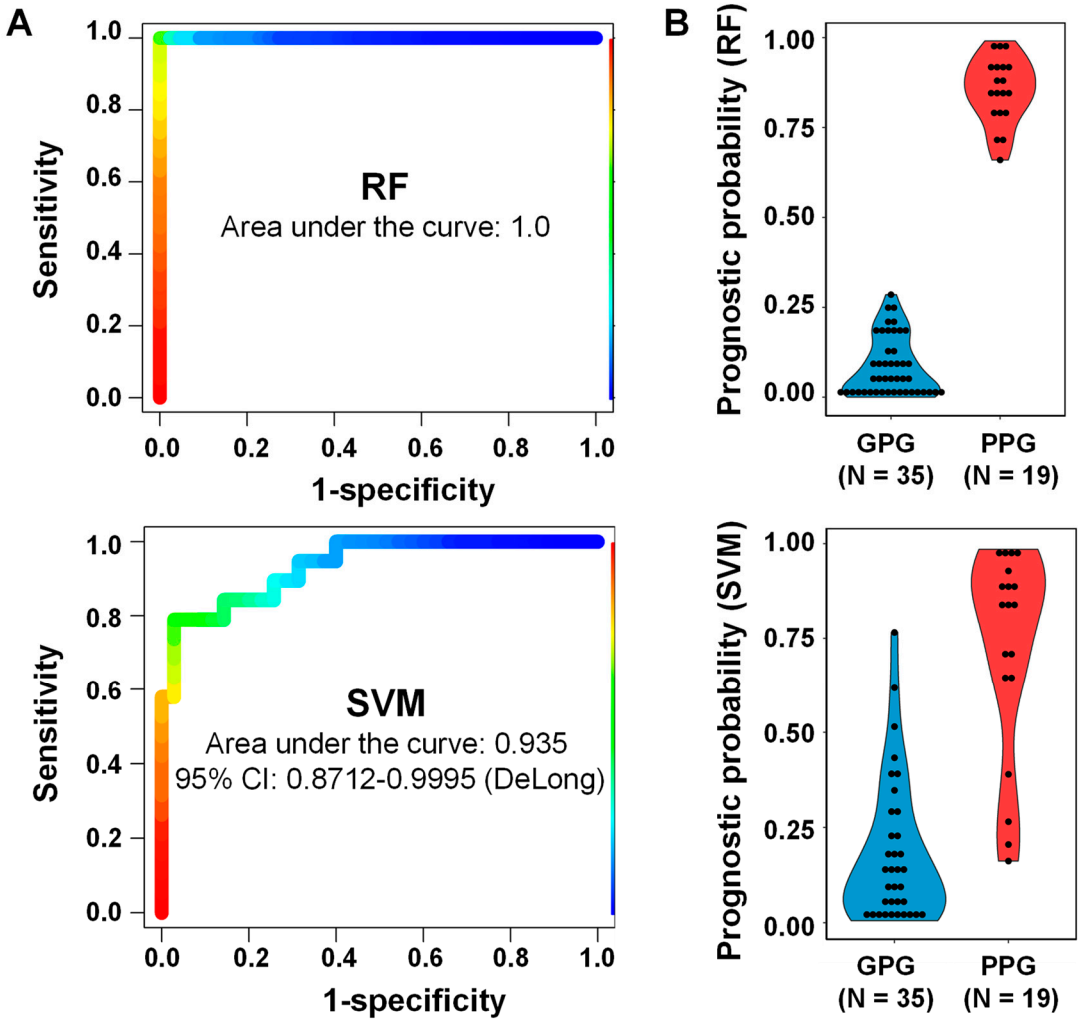


Figure 6. ROC curves of RF and SVM classifiers for five selected proteins (ACP2, CTSA, GM2A, MUC1 and SPARCL1). Performance of the two classifiers in the set of 54 samples, 35 from patients with good prognosis and 19 from patients with poor prognosis. (**A**) Areas under the curve (AUC) for the RF (1.0) and SVM (0.935) classifiers. (**B**) Clinical indices (0–1) of the two classifiers.

Table 2. AUC-based RF backward-elimination process-based selected feature proteins.

Uniprot Accession No.	Gene Name	Importance	Prob. Select	Selection	Univariate AUC
P10619	*CTSA*	0.422	0.700	Y	0.737
Q14515	*SPARCL1*	0.378	0.583	Y	0.659
P17900	*GM2A*	0.373	0.613	Y	0.726
P15941-2	*MUC1*	0.332	0.543	Y	0.791
P11117	*ACP2*	0.312	0.563	Y	0.718
P19961	*AMY2B*	0.299	0.510	N	0.779
P00734	*F2*	0.296	0.483	N	0.694
P06865	*HEXA*	0.274	0.466	N	0.651
P05155-3	*SERPING1*	0.275	0.377	N	0.771
P11142	*HSPA8*	0.238	0.330	N	0.734
P10451	*SPP1*	0.228	0.323	N	0.680

Probability of selection for each variable.

2.6. External Validation of Clinical Models in Public Studies

Since we were unable to find a benchmarking study in the discovery of urine protein biomarkers that could validate our statistical model, we validated the models with mRNA expression in the kidney, an organ that undoubtedly affects urine samples. The SVM and RF models consisting of five urine proteins were applied to four publicly available GEO datasets (GSE99339 [22], GSE47185 [23], GSE30122 [24], and GSE96804 [25,26]) without model adjustment. In the first GSE99339 dataset, mRNA expression in the renal glomerulus of 187 patients was studied, and the 11 disease groups are diabetic nephropathy (DN), rapidly progressive glomerulonephritis (RPGN), tumor nephrectomies (TN), hypertensive nephropathy (HT), IgA nephropathy, membranous glomerulonephritis (MGN), systemic lupus erythematosus (SLE), thin membrane disease (TMD), focal and segmental glomerulosclerosis (FSGS), focal and segmental glomerulosclerosis and minimal change disease (FSGS&MCD), and minimal change disease (MCD). The two classifiers' prognostic probabilities were highly correlated with each other in 187 samples (ρ = 0.817, Pearson correlation coefficient). In both models, the highest value in the DN group was higher than the other ten disease groups (Figure 7A). RF prediction values in the DN group were significantly higher than other eight groups except for RPGN and HT group (Mann-Whitney U Test: p-value < 0.05). SVM prediction values in the DN group were significantly higher than the other nine groups excluding the RPGN group (p-value < 0.05). In the second GSE99339 data set, there are a total of 223 kidney glomerulus (N = 122) and tubulointerstitia (N = 101) mRNA expression levels. The eight disease groups include DN, RPGN, TN, MGN, TMD, FSGS, FSGS&MCD, and MCD. The two classifiers' prognostic probabilities were also highly correlated in 223 samples (r = 0.637). The tendency of the predicted values was different depending on the cell type of the kidney (Figure 7B). In the glomeruli, the two model predictions are the highest in the DN group and are statistically significant with other seven groups. However, in the tubulointerstitium, the SVM model prediction values in DN were significant with four other groups except RPGN, TN, FSGS&MCD, and RF model prediction values in DN is only significant with MCD. It indicated that the five urine proteins are more closely related to the glomeruli than the kidney tubulointerstitium.

Meanwhile, we tried to verify whether the prognostic models could predict DKD. In the third GSE30122 data set, of the total of 69 samples, 26 of the 35 kidney glomerulus were normal obtained from living allograft donors, 9 of which were DKD, 34 of which were renal tubulus, of which 24 were normal and 10 were DKD. The results of the RF model in the glomeruli statistically were divided the normal and disease groups (p-value < 0.05), but the SVM model were not (p-value > 0.05; Figure 7C). The results of the both models in the tubulus statistically were not divided the normal and disease groups (p-value > 0.05). In the fourth GSE30122 dataset, 20 kidney glomerulus out of a total of 62 samples were glomerulus from the non-neoplastic part of tumor nephrectomies and 41 of them were

from DN. The results of the both models statistically were not divided the normal and disease groups (p-value > 0.05; Figure 7D). It indicated that models for predicting kidney prognosis with urine protein markers in diabetics are difficult to distinguish DKD from normal groups by mRNA expression level in kidney.

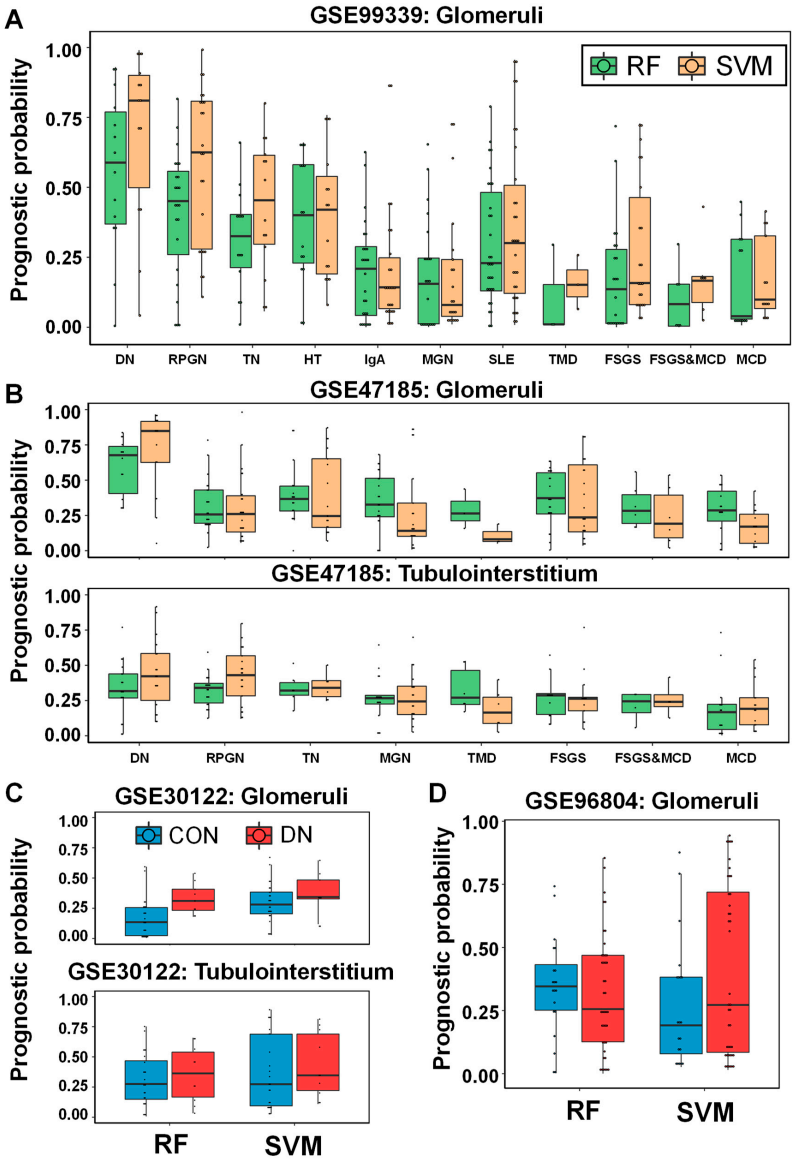


Figure 7. External validation of RF and SVM clinical models in public four GEO datasets (GSE99339, GSE47185, GSE30122 and GSE96804). (**A**) In the GSE99339 dataset, boxplot of the prognostic probabilities of the two classifiers in 11 disease groups including DN (N = 14), RPGN (N = 23), TN (N = 14), HT (N = 15), IgA nephropathy (N = 26), MGN (N = 21), SLE (N = 30), TMD (N = 3), FSGS (N = 22), FSGS&MCD (N = 6), and MCD (N = 13). (**B**) In the GSE30122 data set, the prognostic indexes of the two classifiers in the eight disease groups in the renal glomeruli with DN (N = 14), RPGN, (N = 23), TN (N = 17), MGN (N = 21), TMD (N = 3), FSGS (N = 23), FSGS&MCD (N = 6), and MCD (N = 15) and in the renal tubulointerstitia with DN (N = 18), RPGN (N = 21), TN (N = 6), MGN (N = 18), TMD (N = 6), FSGS (N = 13), FSGS&MCD (N = 4), and MCD (N = 15). (**C**) In the GSE30122 data set, the prediction values of the two classifiers in the control and disease groups in renal glomerulus (N = 26; control and N = 9; disease) and in renal tubulus (N = 24; control and N = 10; disease). (**D**) In the GSE30122 data set, the prediction probabilities of the two classifiers in the control (N = 20) and disease (N = 41) groups in renal glomeruli.

3. Discussion

Urine-based approaches for measuring internal biomolecules can be normalized. Ideally, urine should be collected for 24 h and urinary biomolecules measured. Because this method is practically difficult, urinary proteins in random spot samples were calibrated relative to creatinine concentrations [9]. Prolonged storage of urine samples for studying proteins is important because of the activity of urinary proteases depending on the temperature and pH [27]. In this retrospective study, urine samples were stored at −80 °C for 7–8 years before LC-MS/MS measurements. In general, it is known that it is stable without urine preservatives stored at −70 or −80 °C, and urine samples stored for more than 2.3 years have no significant change in not only most proteins including albumin but also metabolites including creatinine [28–31].

Proteins were extracted from urine samples using an equal volume-based approach similar to ELISA [32]. This procedure for protein standardization was suitable for downstream analysis. Urinary proteins normalized by this method showed lower sample-to-sample variation and higher correlation with immunoassay results.

Albuminuria is primarily used to detect DKD in clinical practice [2,5]. Because glomeruli filter blood, albumin is a good biomarker of chronic kidney disease (CKD) caused by glomerular abnormalities but is insufficient to determine subsequent prognosis [5]. Rather than this, it was determined that finding and measuring specific protein markers that affect pathological function is more clinically meaningful [33,34]. Although causality between albuminuria and prognostic values from the five-protein panel-based clinical models (RF and SVM) cannot be clarified in this retrospective study, it can be inferred by correlation analysis. Correlation analysis between two classifiers and ACR in the 54 enrolled patients reveals a little of bit correlation but no significance (r = 0.086; p-value > 0.05; SVM and r = 0.094; p-value > 0.05; RF). Therefore, it was confirmed that there was no causal relationship as well as a correlation. To consider closely at the relationship between them, we divided the three classes based on the ACR value (normal; <30 mg/g, microalbuminuria; 30–300 mg/g and macroalbuminuria; >300 mg/g) and plotted the predicted values of the SVM model according to the two prognostic groups (Figure S1). In T2D patients with normal range and microalbuminuria, SVM results were almost separated between two groups. Rather, it seems to have problems with predictive power in patients with macroalbuminuria. It means that SVM results did not depend on the development of albuminuria in T2D patients and showed the possibility to predict the earlier disease stage before the development of albuminuria. Moreover, the predicted value of RF results accurately separates two prognostic groups regardless of the ACR value.

As a rule, diabetics are persistently exposed to miscellaneous metabolic and hemodynamic risks [35], with DKD resulting from multiple pathophysiological processes. Multiple-biomarker approaches using proteomics and metabolomics may better reveal the complicated disease status thought to be associated with the onset of DKD [4,8]. CKD273, a panel consisting of 273 urinary peptides currently undergoing Phase 3 testing, was a high performance urine peptidomic classifier for CKD diagnosis [36]. Moreover, this classifier was recently validated as a predictor of the development of microalbuminuria in normoalbuminuric with diabetic patients [37]. These 273 intact peptides were derived from 30 independent proteins, 24 of which were quantified in this study. CKD273, which includes cleaved collagenase peptides and SERPINA1 peptides, is a good prognostic marker, showing that the concentrations of cleaved collagenase peptides decrease and those of SERPINA1 peptides increase in the urine of patients with CKD [38,39]. The present study showed a similar pattern of abundance in the urine of PPG patients despite artificial digestion. Our approach, based on protein concentrations in urine samples, could better explain the pathological pathway associated with DKD than the peptidome approach. Indicators of kidney dysfunction include increased blood particles in urine; lysosomal dysfunction in glomerular cells [40], which is related to the autophagy-lysosome pathway [41] abnormal heterotypic cell–cell adhesion among glomerular, tubular, and immune cell compartments, collagenase, and binding proteins (driven by rapid changes in glycolipids) [42] and platelet activation [43].

Our clinical models consist of five selected proteins, four proteins (CTSA, MUC1, GM2A, and SPARCL1) are high in PPG and other one protein (ACP2) is low in PPG. The family of cathepsin proteins has a variety of roles in kidney disease [44] and is known as a new drug target [45], including cathepsin A. In addition, cathepsin L is known to be important for the early development of diabetic nephropathy [46]. Muc1 is a multifaceted tumor protein, and its relationship with the kidney has recently been highlighted [47] and has been identified as a mutant that causes mendelian disorder medullary cystic kidney disease type 1 [48]. In a meta-analysis study of rat glomerular transcriptome profiling, it was confirmed that GM2A was highly expressed in various diabetic kidney disease rat [49]. Through the mouse kidney injury model experiment, SPARCL1 showed that mRNA expression was not changed in the acute phase, but the expression level was high in the fibrosis of the kidney [50], and it inhibited the movement and invasion of renal cell carcinoma [51]. Lastly, ACP 2, one of the lysosomal enzymes, is a protein used in peptiduria [52] or lysosomal enzymuria [53] that measures kidney disease in diabetic patients. Through the external kidney mRNA published studies, these urine biomarkers we found confirmed differential expression in kidney tissue with DKD.

This study had several limitations. First, the patient population in this study was homogeneous and of small sample size. These results require further validation in a multiethnic cohort including larger numbers of patients to assess applicability to a wider population with T2D, a study currently in progress. Second, DKD was clinically diagnosed in the absence of renal biopsies. Third, it is unclear which organ is derived from the urinary protein signatures. More research is needed to determine whether urinary protein signatures are biomarkers of tubular damage in pathological conditions with a glomerular protein load.

4. Materials and Methods

4.1. Patients and Urine Samples

Urine samples were collected from 54 outpatients with T2D and eGFR ≥ 60 mL/min/1.72 m^2 who were enrolled in the DKD study at Pusan National University Hospital, South Korea, from February 2010 to January 2011 and who met previously described inclusion and exclusion criteria [54]. After one year, patients were followed-up with until September 2017. Patients were managed according to standard guidelines, including treatment with RAS inhibitors, and eGFR was measured at least twice during a follow-up period ≥12 months. Renal function decline was defined as an eGFR < 60 mL/min/1.72 m^2, annual eGFR reduction > 3 mL/min/1.72 m^2, or CKD progression, defined as a reduction in GFR category, accompanied by a ≥ 25% deterioration in eGFR from baseline. The patients were divided into two groups—19 with renal outcomes (poor prognosis group (PPG)) and 35 without renal outcomes (good prognosis group (GPG)). The protocols and consent procedures were approved by the Institutional Review Board of Pusan National University Hospital (approval No. 2013033). Total proteinuria and albuminuria, as well as creatinine concentrations, were measured in random spot urine samples [55].

4.2. Measurements of Nephrology Parameters

eGFR was calculated using the equation eGFR = 141 × min (serum creatinine/kappa, 1) alpha × max (serum creatinine/kappa, 1) − 1.209 × 0.993 × age × sex × race. For females, sex = 1.018; alpha = −0.329; and kappa = 0.7; for males, sex = 1; alpha = −0.411; and kappa = 0.9. Renal outcomes were chronic kidney disease (CKD) progression based on guidelines of the International Society of Nephrology; accelerated eGFR decline, defined as an annual eGFR reduction > 3 mL/min/1.72 m^2; or the development of CKD stage ≥ 3. CKD stages 1, 2, 3a, 3b, 4, and 5 were defined as eGFRs of ≥ 90, 60–89, 45–59, 30–44, 15–29, and < 15 mL/min/1.73 m^2, respectively, and CKD progression was defined as a decline in eGFR category accompanied by a ≥ 25% deterioration in eGFR from baseline [56].

4.3. Urinary Protein Sample Preparation

Urine samples were centrifuged at 13,000 rpm for 30 min to remove debris, and 300 µL of each supernatant was mixed with 100 µL High Select™ HSA/Immunoglobulin Depletion Resin (Cat. No: A36368, Thermo Fisher Scientific, Waltham, MA, USA) and incubated for 1 h at 4 °C to remove albumin and immunoglobulin. Following centrifugation at 13,000 rpm for 10 min, the supernatant was dried using a speed vac with a cold trap (CentriVap Cold Traps, Labconco, Kansas City, MO, USA).

4.4. Enzymatic Digestion in-Solution

Each dried urine sample was dissolved in 100 µL of 8 M urea, reduced with 20 mM dithiothreitol in 50 mM NH_4HCO_3 for 60 min at 25 °C, and alkylated with 40 mM iodoacetamide in 50 mM NH4HCO3 for 60 min in the dark. Urea concentration was diluted to less than 1.0 M. Each urine sample was incubated overnight at 37 °C with 12.5 µg sequencing grade modified trypsin/LysC (Promega, Madison, WI, USA) in 50 mM NH4HCO3 buffer (pH 7.8), followed by quenching with 10uL of 5% formic acid and lyophilization with a cold trap. The samples were re-suspended in 0.1% formic acid, desalted using C18 ZipTips (Millipore, Burlington, MA, USA), and dried for LC-MS analysis.

4.5. Nano-LC-ESI-MS/MS Analysis

Digested peptides were separated using a Dionex UltiMate 3000 RSLCnano system (Thermo Fisher Scientific, Waltham, MA, USA). Tryptic peptides from the bead column were reconstituted in 100 µL of 0.1% formic acid and separated on an Acclaim™ Pepmap 100 C18 column (500 mm × 75 µm i.d., 3 µm, 100 Å) equipped with a C18 Pepmap trap column (20 mm × 100 µm i.d., 5 µm, 100 Å; Thermo Fisher Scientific, Waltham, MA, USA) over 200 min (250 nL/min) using a 0–48% acetonitrile gradient in 0.1% formic acid and 5% DMSO for 150 min at 50 °C. The LC was coupled to a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ mass spectrometer with a nano-ESI source. Mass spectra were acquired in a data-dependent mode with an automatic switch between a full scan and 10 data-dependent MS/MS scans. The target value for the full scan MS spectra, selected from a 350 to 1800 m/z, was 3,000,000 with a maximum injection time of 50 ms and a resolution of 70,000 at m/z 400. The selected ions were fragmented by higher-energy collisional dissociation in the following parameters: 2 Da precursor ion isolation window and 27% normalized collision energy. The ion target value for MS/MS was set to 1,000,000 with a maximum injection time of 100 ms and a resolution of 17,500 at m/z 400. Repeated peptides were dynamically excluded for 20 s. All MS data were measured once per sample and have been deposited in the PRIDE archive (www.ebi.ac.uk/pride/archive/projects/PXD016571) [57].

4.6. Database Searching and Label-Free Quantitation

The SwissProt human database (May 2017) was searched for acquired MS/MS spectra using SequestHT on Proteome discoverer (version 2.2, Thermo Fisher Scientific, USA) [58]. The search parameters were set as default including cysteine carbamidomethylation as a fixed modification, and N-terminal acetylation and methionine oxidation as variable modifications with two miscleavages. Peptides were identified based on a search with an initial mass deviation of the precursor ion of up to 10 ppm, with the allowed fragment mass deviation set to 20 ppm. When assigning proteins to peptides, both unique and razor peptides were used. Label-free quantitation (LFQ) was performed using peak intensity for unique peptides of each protein.

4.7. Normalization of Protein Abundance

To correct for sampling variations resulting from random spot urine collection, the raw LFQ values for each protein were divided by the amounts of total protein and creatinine in each sample, followed by normalization of the corrected LFQ values by endogenous proteins without spike-in standards [59]. To identify endogenous urinary proteins for normalization, the 112 initial completely quantified proteins were considered, with six selected based on the following criteria: (1) quantified in

all 54 samples; (2) corrected LFQ values did not differ significantly in the poor and good prognosis groups by the Mann–Whitney U Test (p-value > 0.05); and (3) had nearly persistent urine concentrations throughout the sample as top-ranked by NormFinder stability value [60].

The corrected LFQ values of the six selected normalization proteins in each sample were divided by their median value in all samples. The median of these six ratios was defined as the normalization scaling factor (NSF) for that sample. For example, NSF for sample s can be determined using the equation:

$$NSF_s = median\left(\frac{N_{1,s}}{\hat{N}_1}, \frac{N_{2,s}}{\hat{N}_2}, \cdots, \frac{N_{6,s}}{\hat{N}_6}\right) \quad (1)$$

where $N_{i,\,s}$ is the corrected LFQ value of normalization protein i in sample s and $\hat{N}_i$ is the median corrected LFQ value of normalization protein i in all samples. Except for the six normalization proteins in a sample, the normalized LFQ value of each protein was calculated by dividing its corrected LFQ value by NSF.

$$\breve{LFQ}_{j,s} = \frac{LFQ_{j,\,s}}{NSF_s} \quad (2)$$

where $\breve{LFQ}_{j,s}$ is the normalized LFQ of urinary protein j in sample s and $LFQ_{j,s}$ is the corrected LFQ of the corresponding protein [61].

4.8. Differential Data Analysis by Filling Missing Data

For clinical utility, the LFQ data were filtered to <20% of quantified proteins in each sample group to analyze the differential urinary proteins in these groups, with the missing data filled by the local least squared imputation method at the normalized abundance [18].

4.9. GO Analysis

Differential abundant proteins (DAPs) in the poor and good prognosis groups were analyzed using the ClueGO (version 2.5.1) [12] plugin for Cytoscape (version 3.6.1) [62]. To group GO terms, the kappa score was set at 0.4 and the number of overlapping genes to combine groups was set at 50%.

4.10. Statistical Clinical Model Generation Based on Feature Selection

The process of feature selection was to find the best subset for classifying two disease progression groups out of 412 proteins. There are two steps. In the first step, 50,000 decision trees containing eight variables were randomly generated 50,000 trees and had AUC values. Based on the AUCs values, the optimal number of proteins were determined by out-of-bag error estimation and the value is 11. Second, through the 100 iterations with three-fold cross-validation for from the selected 11 optimal variables, the probability and importance that each variable was included in the model was calculated. We selected five proteins (>0.3 importance). Prior to model building, centering and scaling were performed as preprocessing on the data. In the clinical models, SVM model with linear kernel was generated by a 10 repeated three-fold cross validation method (parameter C = 0.1052) and The RF model was made by a three-fold cross validation method repeated 100 times with 1000 trees, mtry = 5 and nodesize = 5.

4.11. Mining Public Microarray Data

We downloaded the mRNA expression data (series accession number: GSE99339, GSE47185, GSE30122, and GSE96804) in the Gene Expression Omnibus database [63]. Then, using GEO2R interactive web tool, five identifiers matching the five selected genes according to the platform record and their expression values were extracted.

4.12. Statistical Analysis

Data were analyzed using RStudio (version 1.1.456) including R (version 3.6.0). Statistical R software packages included ggplot2 for drawing box, scattering, volcano and violin plots, permcor for calculating permutation-based *p*-values for Pearson correlation [64], pcaMethods for missing value estimation [65], pROC for univariate ROC analysis [66], ROCR for multivariate ROC analysis, AUCRF for feature selection [21], caret for building statistical classifiers [67], randomForest for building a RF classifier, and e1071 for building a SVM classifier.

5. Conclusions

These results suggest that measurement of urinary proteome was more promising than albuminuria alone for predicting renal outcomes in patients with type 2 diabetes. A panel of five proteins had the potential for use as a biomarker in clinical practice.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/21/12/4236/s1. Figure S1. Prognostic probability of 54 patients by RF and SVM classifier divided into three groups by albumin-to-creatinine ratio.; Table S1. Demographic and clinical characteristics of the 54 patients with type 2 diabetes.; Table S2. Search result of MS/MS spectra of protein sequences obtained from the 54 urine samples in the Human Swissprot proteome database using the SequestHT search engine.; Table S3. Results of volcano plot analysis.; Table S4. Normalized abundance of 412 urinary proteins in the 54 patients with type 2 diabetes.; Table S5. Univariate receiver operating curve analysis of the areas under the curves of 412 urinary proteins.; Table S6. Prognostic probability by RF and SVM classifier of 54 patients.

Author Contributions: Conceptualization, J.H.K., S.H.S., S.S.K., I.J.K., and K.K.; methodology, H.J.; formal analysis, H.-S.A.; investigation, J.Y. (Jiyoung Yu); data curation, H.-S.A., S.S.K., and I.J.K.; writing—original draft preparation, H.-S.A. and K.K.; writing—review and editing, J.H.K., J.Y. (Jiyoung Yu), J.Y. (Jeonghun Yeom), S.H.S., S.S.K., and I.J.K; visualization, H.-S.A.; supervision, I.J.K and K.K.; funding acquisition, J.H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a Biomedical Research Institute Grant (2017-20) from Pusan National University Hospital.

Acknowledgments: The authors thank the Department of Biostatistics, Clinical Trial Center, Biomedical Research Institute, Pusan National University Hospital.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

T2D	Type 2 diabetes
eGFR	Estimated glomerular filtration rate
DKD	Diabetic kidney disease
BMI	Body mass index
SBP	Systolic blood pressure
HbA1c	Glycated hemoglobin concentration
LDL	Low-density lipoprotein
HDL	High-density lipoprotein
ACR	Albumin-to-creatinine ratio
NAPCR	Nonalbumin protein-to-creatinine ratio
PCR	Urine protein-to-creatinine ratio
GPG	Good-prognostic group
PPG	Poor-prognostic group
LC-MS	Liquid chromatography–mass spectrometry
DAP	Differential abundant protein
GO	Gene ontology
FDR	False discovery rate
ROC	Receiver operating characteristic
AUC	Area under the receiver operating characteristic curve
RF	Random forest
SVM	Support vector machine

ELISA	Enzyme-linked immunosorbent assay
CKD	Chronic kidney disease
MS/MS	Tandem mass spectrometry
LFQ	Label-free quantitation
NSF	Normalization scaling factor
DN	Diabetic nephropathy
FSGS	Focal and segmental glomerulosclerosis
FSGS&MCD	Focal and segmental glomerulosclerosis and minimal change disease
HT	Hypertensive nephropathy
MCD	Minimal change disease
MGN	Membranous glomerulonephritis
RPGN	Rapidly progressive glomerulonephritis
SLE	Systemic lupus erythematosus
TMD	Thin membrane disease
TN	Tumor nephrectomies

References

1. Ahn, J.H.; Yu, J.H.; Ko, S.H.; Kwon, H.S.; Kim, D.J.; Kim, J.H.; Kim, C.S.; Song, K.H.; Won, J.C.; Lim, S.; et al. Prevalence and determinants of diabetic nephropathy in Korea: Korea national health and nutrition examination survey. *Diabetes Metab. J.* **2014**, *38*, 109–119. [CrossRef]
2. Tuttle, K.R.; Bakris, G.L.; Bilous, R.W.; Chiang, J.L.; de Boer, I.H.; Goldstein-Fuchs, J.; Hirsch, I.B.; Kalantar-Zadeh, K.; Narva, A.S.; Navaneethan, S.D.; et al. Diabetic kidney disease: A report from an ADA Consensus Conference. *Diabetes Care* **2014**, *37*, 2864–2883. [CrossRef] [PubMed]
3. Collins, A.J.; Foley, R.N.; Chavers, B.; Gilbertson, D.; Herzog, C.; Johansen, K.; Kasiske, B.; Kutner, N.; Liu, J.; St Peter, W.; et al. United States Renal Data System 2011 Annual Data Report: Atlas of chronic kidney disease & end-stage renal disease in the United States. *Am. J. Kidney Dis.* **2012**, *59*, A7. [CrossRef] [PubMed]
4. Currie, G.; Delles, C. Urinary Proteomics for Diagnosis and Monitoring of Diabetic Nephropathy. *Curr. Diabetes Rep.* **2016**, *16*, 104. [CrossRef] [PubMed]
5. KDIGO Working Group. KDIGO clinical practice guideline for the evaluation and management of chronic kidney disease. Chapter 2: Definition, identification, and prediction of CKD progression. *Kidney Int. Suppl.* **2013**, *3*, 63–72. [CrossRef]
6. Barratt, J.; Topham, P. Urine proteomics: The present and future of measuring urinary protein components in disease. *CMAJ* **2007**, *177*, 361–368. [CrossRef] [PubMed]
7. Retnakaran, R.; Cull, C.A.; Thorne, K.I.; Adler, A.I.; Holman, R.R.; Group, U.S. Risk factors for renal dysfunction in type 2 diabetes: U.K. Prospective Diabetes Study 74. *Diabetes* **2006**, *55*, 1832–1839. [CrossRef] [PubMed]
8. Kamijo-Ikemori, A.; Sugaya, T.; Kimura, K. Novel urinary biomarkers in early diabetic kidney disease. *Curr. Diabetes Rep.* **2014**, *14*, 513. [CrossRef]
9. Abitbol, C.; Zilleruelo, G.; Freundlich, M.; Strauss, J. Quantitation of proteinuria with urinary protein/creatinine ratios and random testing with dipsticks in nephrotic children. *J. Pediatr.* **1990**, *116*, 243–247. [CrossRef]
10. Lemann, J., Jr.; Doumas, B.T. Proteinuria in health and disease assessed by measuring the urinary protein/creatinine ratio. *Clin. Chem.* **1987**, *33*, 297–299. [CrossRef]
11. Zhao, M.; Li, M.; Yang, Y.; Guo, Z.; Sun, Y.; Shao, C.; Li, M.; Sun, W.; Gao, Y. A comprehensive analysis and annotation of human normal urinary proteome. *Sci. Rep.* **2017**, *7*, 3024. [CrossRef] [PubMed]
12. Bindea, G.; Mlecnik, B.; Hackl, H.; Charoentong, P.; Tosolini, M.; Kirilovsky, A.; Fridman, W.H.; Pages, F.; Trajanoski, Z.; Galon, J. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **2009**, *25*, 1091–1093. [CrossRef] [PubMed]
13. Van Bladel, E.R.; de Jager, R.L.; Walter, D.; Cornelissen, L.; Gaillard, C.A.; Boven, L.A.; Roest, M.; Fijnheer, R. Platelets of patients with chronic kidney disease demonstrate deficient platelet reactivity in vitro. *BMC Nephrol.* **2012**, *13*, 127. [CrossRef] [PubMed]
14. Prozialeck, W.C.; Edwards, J.R. Cell adhesion molecules in chemically-induced renal injury. *Pharmacol. Ther.* **2007**, *114*, 74–93. [CrossRef]

15. Yan, J.; Zhang, Z.; Jia, L.; Wang, Y. Role of Bone Marrow-Derived Fibroblasts in Renal Fibrosis. *Front. Physiol.* **2016**, *7*, 61. [CrossRef]
16. Deva, R.; Alias, M.A.; Colville, D.; Tow, F.K.; Ooi, Q.L.; Chew, S.; Mohamad, N.; Hutchinson, A.; Koukouras, I.; Power, D.A.; et al. Vision-threatening retinal abnormalities in chronic kidney disease stages 3 to 5. *Clin. J. Am. Soc. Nephrol.* **2011**, *6*, 1866–1871. [CrossRef]
17. Dziura, J.D.; Post, L.A.; Zhao, Q.; Fu, Z.; Peduzzi, P. Strategies for dealing with missing data in clinical trials: From design to analysis. *Yale J. Biol. Med.* **2013**, *86*, 343–358.
18. Karpievitch, Y.V.; Dabney, A.R.; Smith, R.D. Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinform.* **2012**, *13* (Suppl. 16), S5. [CrossRef]
19. Breiman, L. Random forests. *Mach Learn* **2001**, *45*, 5–32. [CrossRef]
20. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
21. Calle, M.L.; Urrea, V.; Boulesteix, A.L.; Malats, N. AUC-RF: A new strategy for genomic profiling with random forest. *Hum. Hered.* **2011**, *72*, 121–132. [CrossRef] [PubMed]
22. Shved, N.; Warsow, G.; Eichinger, F.; Hoogewijs, D.; Brandt, S.; Wild, P.; Kretzler, M.; Cohen, C.D.; Lindenmeyer, M.T. Transcriptome-based network analysis reveals renal cell type-specific dysregulation of hypoxia-associated transcripts. *Sci. Rep.* **2017**, *7*, 8576. [CrossRef] [PubMed]
23. Ju, W.; Greene, C.S.; Eichinger, F.; Nair, V.; Hodgin, J.B.; Bitzer, M.; Lee, Y.S.; Zhu, Q.; Kehata, M.; Li, M.; et al. Defining cell-type specificity at the transcriptional level in human disease. *Genome Res.* **2013**, *23*, 1862–1873. [CrossRef] [PubMed]
24. Woroniecka, K.I.; Park, A.S.; Mohtat, D.; Thomas, D.B.; Pullman, J.M.; Susztak, K. Transcriptome analysis of human diabetic kidney disease. *Diabetes* **2011**, *60*, 2354–2369. [CrossRef] [PubMed]
25. Shi, J.S.; Qiu, D.D.; Le, W.B.; Wang, H.; Li, S.; Lu, Y.H.; Jiang, S. Identification of Transcription Regulatory Relationships in Diabetic Nephropathy. *Chin. Med. J.* **2018**, *131*, 2886–2890. [CrossRef] [PubMed]
26. Pan, Y.; Jiang, S.; Hou, Q.; Qiu, D.; Shi, J.; Wang, L.; Chen, Z.; Zhang, M.; Duan, A.; Qin, W.; et al. Dissection of Glomerular Transcriptional Profile in Patients With Diabetic Nephropathy: SRGAP2a Protects Podocyte Structure and Function. *Diabetes* **2018**, *67*, 717–730. [CrossRef]
27. Kania, K.; Byrnes, E.A.; Beilby, J.P.; Webb, S.A.; Strong, K.J. Urinary proteases degrade albumin: Implications for measurement of albuminuria in stored samples. *Ann. Clin. Biochem.* **2010**, *47*, 151–157. [CrossRef]
28. Parekh, R.S.; Kao, W.H.; Meoni, L.A.; Ipp, E.; Kimmel, P.L.; La Page, J.; Fondran, C.; Knowler, W.C.; Klag, M.J.; Family Investigation of Nephropathy and Diabetes; et al. Reliability of urinary albumin, total protein, and creatinine assays after prolonged storage: The Family Investigation of Nephropathy and Diabetes. *Clin. J. Am. Soc. Nephrol.* **2007**, *2*, 1156–1162. [CrossRef]
29. Chapman, D.P.; Gooding, K.M.; McDonald, T.J.; Shore, A.C. Stability of urinary albumin and creatinine after 12 months storage at −20 degrees C and −80 degrees C. *Pract. Lab. Med.* **2019**, *15*, e00120. [CrossRef]
30. Herrington, W.; Illingworth, N.; Staplin, N.; Kumar, A.; Storey, B.; Hrusecka, R.; Judge, P.; Mahmood, M.; Parish, S.; Landray, M.; et al. Effect of Processing Delay and Storage Conditions on Urine Albumin-to-Creatinine Ratio. *Clin. J. Am. Soc. Nephrol.* **2016**, *11*, 1794–1801. [CrossRef]
31. Klasen, I.S.; Reichert, L.J.; de Kat Angelino, C.M.; Wetzels, J.F. Quantitative determination of low and high molecular weight proteins in human urine: Influence of temperature and storage time. *Clin. Chem.* **1999**, *45*, 430–432. [CrossRef] [PubMed]
32. Voller, A.; Bartlett, A.; Bidwell, D.E. Enzyme immunoassays with special reference to ELISA techniques. *J. Clin. Pathol.* **1978**, *31*, 507–520. [CrossRef] [PubMed]
33. Frantzi, M.; Bhat, A.; Latosinska, A. Clinical proteomic biomarkers: Relevant issues on study design & technical considerations in biomarker development. *Clin. Transl. Med.* **2014**, *3*, 7. [CrossRef] [PubMed]
34. Borrebaeck, C.A. Precision diagnostics: Moving towards protein biomarker signatures of clinical utility in cancer. *Nat. Rev. Cancer* **2017**, *17*, 199–204. [CrossRef]
35. Thomas, M.C.; Burns, W.C.; Cooper, M.E. Tubular changes in early diabetic nephropathy. *Adv. Chronic Kidney Dis.* **2005**, *12*, 177–186. [CrossRef]
36. Good, D.M.; Zurbig, P.; Argiles, A.; Bauer, H.W.; Behrens, G.; Coon, J.J.; Dakna, M.; Decramer, S.; Delles, C.; Dominiczak, A.F.; et al. Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. *Mol. Cell. Proteomics* **2010**, *9*, 2424–2437. [CrossRef]

37. Lindhardt, M.; Persson, F.; Zurbig, P.; Stalmach, A.; Mischak, H.; de Zeeuw, D.; Lambers Heerspink, H.; Klein, R.; Orchard, T.; Porta, M.; et al. Urinary proteomics predict onset of microalbuminuria in normoalbuminuric type 2 diabetic patients, a sub-study of the DIRECT-Protect 2 study. *Nephrol. Dial. Transplant.* **2017**, *32*, 1866–1873. [CrossRef]
38. Rodriguez-Ortiz, M.E.; Pontillo, C.; Rodriguez, M.; Zurbig, P.; Mischak, H.; Ortiz, A. Novel Urinary Biomarkers For Improved Prediction Of Progressive Egfr Loss In Early Chronic Kidney Disease Stages And In High Risk Individuals Without Chronic Kidney Disease. *Sci. Rep.* **2018**, *8*, 15940. [CrossRef]
39. Pontillo, C.; Mischak, H. Urinary peptide-based classifier CKD273: Towards clinical application in chronic kidney disease. *Clin. Kidney J.* **2017**, *10*, 192–201. [CrossRef] [PubMed]
40. Surendran, K.; Vitiello, S.P.; Pearce, D.A. Lysosome dysfunction in the pathogenesis of kidney diseases. *Pediatr. Nephrol.* **2014**, *29*, 2253–2261. [CrossRef] [PubMed]
41. Liu, W.J.; Shen, T.T.; Chen, R.H.; Wu, H.L.; Wang, Y.J.; Deng, J.K.; Chen, Q.H.; Pan, Q.; Huang Fu, C.M.; Tao, J.L.; et al. Autophagy-Lysosome Pathway in Renal Tubular Epithelial Cells Is Disrupted by Advanced Glycation End Products in Diabetic Nephropathy. *J. Biol. Chem.* **2015**, *290*, 20499–20510. [CrossRef]
42. Rops, A.L.; van der Vlag, J.; Lensen, J.F.; Wijnhoven, T.J.; van den Heuvel, L.P.; van Kuppevelt, T.H.; Berden, J.H. Heparan sulfate proteoglycans in glomerular inflammation. *Kidney Int.* **2004**, *65*, 768–785. [CrossRef] [PubMed]
43. Landray, M.J.; Wheeler, D.C.; Lip, G.Y.; Newman, D.J.; Blann, A.D.; McGlynn, F.J.; Ball, S.; Townend, J.N.; Baigent, C. Inflammation, endothelial dysfunction, and platelet activation in patients with chronic kidney disease: The chronic renal impairment in Birmingham (CRIB) study. *Am. J. Kidney Dis.* **2004**, *43*, 244–253. [CrossRef] [PubMed]
44. Shlipak, M.G.; Matsushita, K.; Arnlov, J.; Inker, L.A.; Katz, R.; Polkinghorne, K.R.; Rothenbacher, D.; Sarnak, M.J.; Astor, B.C.; Coresh, J.; et al. Cystatin C versus creatinine in determining risk based on kidney function. *N. Engl. J. Med.* **2013**, *369*, 932–943. [CrossRef]
45. Inker, L.A.; Schmid, C.H.; Tighiouart, H.; Eckfeldt, J.H.; Feldman, H.I.; Greene, T.; Kusek, J.W.; Manzi, J.; Van Lente, F.; Zhang, Y.L.; et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N. Engl. J. Med.* **2012**, *367*, 20–29. [CrossRef] [PubMed]
46. Ingelfinger, J.R.; Marsden, P.A. Estimated GFR and risk of death–is cystatin C useful? *N. Engl. J. Med.* **2013**, *369*, 974–975. [CrossRef] [PubMed]
47. Al-Bataineh, M.M.; Sutton, T.A.; Hughey, R.P. Novel roles for mucin 1 in the kidney. *Curr. Opin. Nephrol. Hypertens.* **2017**, *26*, 384–391. [CrossRef]
48. Kirby, A.; Gnirke, A.; Jaffe, D.B.; Baresova, V.; Pochet, N.; Blumenstiel, B.; Ye, C.; Aird, D.; Stevens, C.; Robinson, J.T.; et al. Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. *Nat. Genet.* **2013**, *45*, 299–303. [CrossRef]
49. Tryggvason, S.H.; Guo, J.; Nukui, M.; Norlin, J.; Haraldsson, B.; Jornvall, H.; Tryggvason, K.; He, L. A meta-analysis of expression signatures in glomerular disease. *Kidney Int.* **2013**, *84*, 591–599. [CrossRef]
50. Feng, D.; Ngov, C.; Henley, N.; Boufaied, N.; Gerarduzzi, C. Characterization of Matricellular Protein Expression Signatures in Mechanistically Diverse Mouse Models of Kidney Injury. *Sci. Rep.* **2019**, *9*, 16736. [CrossRef]
51. Ye, H.; Wang, W.G.; Cao, J.; Hu, X.C. SPARCL1 suppresses cell migration and invasion in renal cell carcinoma. *Mol. Med. Rep.* **2017**, *16*, 7784–7790. [CrossRef] [PubMed]
52. Gudehithlu, K.P.; Hart, P.D.; Vernik, J.; Sethupathi, P.; Dunea, G.; Arruda, J.A.L.; Singh, A.K. Peptiduria: A potential early predictor of diabetic kidney disease. *Clin. Exp. Nephrol.* **2019**, *23*, 56–64. [CrossRef]
53. Gatsing, D.; Garba, I.H.; Adoga, G.I. The use of lysosomal enzymuria in the early detection and monitoring of the progression of diabetic nephropathy. *Indian J. Clin. Biochem.* **2006**, *21*, 42–48. [CrossRef]
54. Kim, S.S.; Song, S.H.; Kim, I.J.; Yang, J.Y.; Lee, J.G.; Kwak, I.S.; Kim, Y.K. Clinical implication of urinary tubular markers in the early stage of nephropathy with type 2 diabetic patients. *Diabetes Res. Clin. Pract.* **2012**, *97*, 251–257. [CrossRef] [PubMed]
55. Lane, C.; Brown, M.; Dunsmuir, W.; Kelly, J.; Mangos, G. Can spot urine protein/creatinine ratio replace 24 h urine protein in usual clinical nephrology? *Nephrology* **2006**, *11*, 245–249. [CrossRef] [PubMed]
56. Summary of Recommendation Statements. *Kidney Int* **2013**, *3*, 5–14. [CrossRef] [PubMed]

57. Deutsch, E.W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J.J.; Kundu, D.J.; Garcia-Seisdedos, D.; Jarnuczak, A.F.; Hewapathirana, S.; Pullman, B.S.; et al. The ProteomeXchange consortium in 2020: Enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* **2019**. [CrossRef] [PubMed]
58. The UniProt Consortium. UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158–D169. [CrossRef]
59. Wisniewski, J.R.; Hein, M.Y.; Cox, J.; Mann, M. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* **2014**, *13*, 3497–3506. [CrossRef]
60. Andersen, C.L.; Jensen, J.L.; Orntoft, T.F. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **2004**, *64*, 5245–5250. [CrossRef] [PubMed]
61. Ahn, H.S.; Sohn, T.S.; Kim, M.J.; Cho, B.K.; Kim, S.M.; Kim, S.T.; Yi, E.C.; Lee, C. SEPROGADIC—Serum protein-based gastric cancer prediction model for prognosis and selection of proper adjuvant therapy. *Sci. Rep.* **2018**, *8*, 16892. [CrossRef] [PubMed]
62. Saito, R.; Smoot, M.E.; Ono, K.; Ruscheinski, J.; Wang, P.L.; Lotia, S.; Pico, A.R.; Bader, G.D.; Ideker, T. A travel guide to Cytoscape plugins. *Nat. Methods* **2012**, *9*, 1069–1076. [CrossRef] [PubMed]
63. Barrett, T.; Edgar, R. Gene expression omnibus: Microarray data storage, submission, retrieval, and analysis. *Methods Enzymol.* **2006**, *411*, 352–369. [CrossRef] [PubMed]
64. Legendre, P. Comparison of permutation methods for the partial correlation and partial Mantel tests. *J Stat Comput. Sim.* **2000**, *67*, 37–73. [CrossRef]
65. Stacklies, W.; Redestig, H.; Scholz, M.; Walther, D.; Selbig, J. pcaMethods–a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **2007**, *23*, 1164–1167. [CrossRef]
66. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.C.; Muller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **2011**, *12*, 77. [CrossRef]
67. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw* **2008**, *28*, 1–26. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).