**Title**: Proteomic analysis identifies dysregulated proteins in albuminuria: a South African pilot study

Siyabonga Khoza[1*], Jaya A. George[2], Previn Naicker[3], Stoyan H. Stoychev[4,5], June Fabian[6,7,†], Ireshyn S. Govender[3,4†*]

[1] Department of Chemical Pathology, National Health Laboratory Service and School of Pathology, University of the Witwatersrand, Faculty of Health Sciences, Johannesburg, South Africa
[2] Wits Diagnostic Innovation Hub, University of the Witwatersrand, Johannesburg, South Africa
[3] Future Production Chemicals, Council for Scientific and Industrial Research, Pretoria, South Africa
[4] ReSyn BioSciences, Edenvale, South Africa
[5] Evosep Biosystems, Odense, Denmark
[6] Wits Donald Gordon Medical Centre, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, South Africa
[7] Medical Research Council/Wits University Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

* Correspondence: professor.khoza@wits.ac.za (S.K), IGovender@csir.co.za (I.S.G)
† Joint senior authors

## Abstract

Albuminuria may precede decreases in glomerular filtration rate (GFR) and both tests are insensitive predictors of early stages of kidney disease. Our aim was to characterise the urinary proteome in black African individuals with albuminuria and well-preserved GFR from South Africa. A case-controlled study that compared urinary proteomes of 52 normoalbuminuric (urine albumin: creatinine ratio (uACR) <3 mg/mmol) and 56 albuminuric (uACR ≥3 mg/mmol) adults of Black African ethnicity. Urine proteins were precipitated, reduced, alkylated, digested, and analysed using an Evosep One LC coupled to a Sciex 5600 Triple-TOF in data-independent acquisition mode. Data were searched on Spectronaut™ 15. Differentially abundant proteins (DAPs) were filtered ≥ 2.25-fold change and false discovery rate ≤ 1%. Receiver operating characteristic curves were used to assess the discriminating ability of proteins of interest. Pathway analysis was performed using Enrichr software. The albuminuric group had a higher uACR (7.9 vs 0.55 mg/mmol, $p$ <0.001). The median eGFR (mL/min/1.73m$^2$) showed no difference between the groups (111 vs 114, $p$=0.707). We identified 80 DAPs in the albuminuria group compared to normoalbuminuria, of which 59 proteins increased while 21 proteins decreased in abundance. We found 12 urinary proteins with AUC > 0.8, and $p$-value <0.001 in the multivariate analysis. Furthermore, an 80-protein model was developed that showed a high AUC >0.907 and a predictive accuracy of 91.3% between the two groups. Pathway analysis associated with DAPs were involved in insulin growth factor (IGF) functions, innate immunity, platelet degranulation, and extracellular matrix organization. In albuminuric individuals with well-preserved eGFR, pathways involved in preventing the release and uptake of IGF by insulin growth factor binding protein were significantly enriched. These proteins are indicative of a homeostatic imbalance in a variety of cellular processes underlying renal dysfunction and are implicated in chronic kidney disease.

**Keywords:** urinary proteomics; albuminuria; chronic kidney disease; biomarker

## 1. Introduction

Chronic kidney disease (CKD) affects between 9.1-13.4% of people and is one of the main causes of death worldwide (1–3). CKD often progresses silently and laboratory tests are used to diagnose, stage, and monitor improvement or progression (2). According to recently published Kidney Disease: Improving Global Outcomes (KDIGO) clinical practice guidelines, CKD is defined as the presence of a reduced glomerular filtration rate (GFR) < 60

ml/min/1.73m$^2$ or albuminuria defined as urine albumin:creatinine ratio (uACR) ≥ 3 mg/mmol) for more than 3 months (4). However, these markers have well-known limitations. Approximately 50% of kidney function is compromised when eGFR falls below 60ml/min/1.73m$^2$, thereby making patients vulnerable to a risk of developing drug toxicity and metabolic and endocrine abnormalities linked with declining kidney function (5). As the GFR falls below this threshold, there is a marked increase in the risk of kidney failure, cardiovascular complications, and all-cause mortality (6,7). Damage to the glomerular filtration barrier leads to impaired size and charge selectivity, causing albumin leakage (8). Regardless of the cause of albumin leakage, data suggest that albuminuria is not only a marker of kidney damage but also exerts a direct toxic effect on renal tubules, resulting in progressive loss of function (9). Some studies show a moderate diagnostic accuracy of albuminuria for CKD, with a lack of sensitivity at low levels (10–12). Unfortunately, albuminuria often manifests after substantial kidney damage has already occurred (13,14). The identification of proteomic markers that are dysregulated in albuminuria could allow non-invasive evaluation of individuals at risk of CKD (15). Early diagnosis and understanding of the pathophysiological mechanism behind CKD progression remains a critical area of research due to the limitations of traditional markers. Proteomic-based biomarkers have the potential to overcome the limitations of current markers for the diagnosis of CKD. Analysis of the urinary proteome is attractive as urine can be collected non-invasively, and in large quantities, and may reflect a biological process specific to the kidney (16). Proteomic-based analysis showed a higher diagnostic precision for CKD and a better ability to identify patients with a rapid kidney function decline beyond the detection of albuminuria in a Norwegian population (17). In a cross-sectional study in the Chinese population, urinary proteome profiling was able to differentiate uncomplicated diabetes from diabetic nephropathy at different stages (18). The discovery of proteins related to CKD has the potential not only to improve the understanding of the pathophysiology and progression, but also to establish new markers for the diagnosis of CKD (19). Before such markers can be applied in clinical settings, multi-ethnic studies should be conducted, especially in African populations, which are largely unrepresented by the current literature. Limited or no studies have been published in South Africa (SA) and Africa. A recent study from SA demonstrated the ability of urinary peptidomics to discriminate normotensive from hypertensive individuals using a classifier that consists of 20 urine peptides (20). Another study in SA, using urinary proteomics coupled with machine learning tools, showed that proteomics-based markers could classify participants with hypertensive-associated albuminuria from healthy controls

(21). Identifying early markers could significantly streamline efforts to delay progression towards end-stage renal disease, especially in low-and middle-income countries experiencing a high burden of CKD with limited resources. Therefore, the detection of proteins associated with albuminuria can serve as early-stage diagnostic markers for CKD. In this pilot study, the aim was to characterise the urinary proteome in individuals with albuminuria with well-preserved eGFR in a South African cohort.

## 2. Materials and Methods

### 2.1 Ethics statement

This study received approval (certificate number: M210128) from the Faculty of Health Sciences Human Research Ethics Committee (Medical) at the University of the Witwatersrand, Johannesburg, South Africa.

### 2.2. The Study Population

Our study was a sub-study of the African Research on Kidney Disease (ARK) study, a population-based sample size of 2021 adults aged 20–80 years from Agincourt in the rural village of Mpumalanga province, SA. The study aimed to determine the prevalence of CKD and associated risk factors. Detailed methods for the ARK study have already been published (22). All demographic and clinical information collected from this study were captured in REDCap (V.22.3.4, Vanderbilt University) (23). For the present study, we included urine samples from 108 participants from the ARK study. We included participants with confirmed albuminuria (cases) on repeated testing and normoalbuminuria (controls) groups, irrespective of their estimated GFR and regardless of CKD risk factors. Cases and controls were age and sex-matched, with a 5-year range applied, if the exact match was not achieved. Cases were defined as having albuminuria (uACR ≥ 3 mg/mmol), while controls were defined as normoalbuminuria (uACR< 3 mg/mmol), in line with the KDIGO guidelines (5). In the original study, participants were classified as being hypertensive ((SBP ≥ 140 mm Hg and/or diastolic blood pressure [DBP] ≥ 90 mm Hg) based on the 7$^{th}$ Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (24,25). Diabetes was defined as a non-fasting glucose ≥ 11.1mmol/L and HIV status was recorded as positive if participant knew their HIV status as positive. Voluntary HIV testing was offered to participants with unknown status or previously tested negative (25).

## 2.3 Clinical laboratory tests

Serum creatinine was measured using Jaffe's method traceable to isotope dilution mass spectrometry, while urine albumin was measured using an immunoturbidimetry method on a Roche analyzer (Roche Diagnostics, Mannheim, Germany). The eGFR was calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation 2009 from serum creatinine without race correction, as suggested by previous literature in the African populations (26).

## 2.4 Urine protein extraction

Urinary proteome preparation was performed using an in-house method as previously described (21). Briefly, four times (1600 µL) ice-cold 80% acetone was added to 400 µL of urine in 2 ml protein Lo-Bind tubes (Eppendorf, Hamburg, Germany). After incubating at –20 ºC for 1hr, samples were centrifuged at 12000 g for 30 min. The supernatant was removed, and the pellet dried using a 70 ºC heating block (AccuBlock Digital dry bath, Labnet International, Inc) for 1 min. Pellets were resuspended in 100 µL 2% sodium dodecyl sulfate (SDS) and sonicated for 5 min. Proteins were reduced with 1 µL of 1M dithiothreitol (DTT) at 70 ºC for 15 min, and thereafter transferred to a 40 ºC heating block for an additional 15 min, followed by alkylation with 6 µL of 500 mM iodoacetamide (IAA) for 30 min at room temperature (RT) in the dark. Proteins were on-bead digested using MagResyn™ HILIC microparticles (ReSyn Biosciences, Edenvale, South Africa) using an automated workflow in a KingFisher™ Duo (Thermo Fisher Scientific, Rockford, IL, USA), as previously described (21). The peptides were dried with a CentriVap vacuum concentrator (Labconco, Kansas City, MO, USA) overnight and resuspended in 40 µL of 2% acetonitrile/0.2% formic acid and were stored at –80 $^{o}$C until LC–MS/MS analysis. Peptide quantification was done using the Pierce™ Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific, Massachusetts, USA) as per the manufacturer's instructions. A pooled sample from 10 patient urine samples was prepared and analysed alongside individual samples as system quality control (QC) for this study. In addition, a commercial Hela digest system suitability QC was analysed.

Tryptic peptides were analysed using an Evosep One LC system (Evosep Biosystems, Odense, Denmark) interfaced to a SCIEX TripleTOF 5600 tandem mass spectrometer (Sciex, Massachusetts, USA) in data-independent acquisition (DIA) mode. An Evosep performance

column (EV1112, 15 cm × 75 μm, 1.9 μm) was used for the Whisper 40SPD method. The Nanospray 3 source settings were as follows: CUR – 20, GS1 – 30, ISVF – 2900. Data were acquired using 48 MS/MS scans of overlapping sequential precursor isolation windows (variable m/z isolation width, 1 m/z overlap, high sensitivity mode), with a precursor MS scan for each cycle. The accumulation time was 50 ms for the MS1 scan (from 400 to 1100 m/z) and 30 ms for each product ion scan (200 to 1800 m/z) for a 1.53 sec cycle time.

A spectral library was built in Spectronaut™ 15 software (Biognosys Schlieren, Switzerland). Default settings were used with minor adjustments as follows: segmented regression was used to determine the normalized retention time (iRT) in each run; iRT were calculated as median for all runs; the digestion rule was set as "Trypsin" and modified peptides were allowed; fragment ions between 300 and 1800 *m/z* and peptides more than three amino acids were considered; peptides with a minimum of three and maximum of 6 (most intense) fragment ions were accepted. This study-specific spectral library was combined with an in-house generated urinary proteome spectral library (using in Spectronaut™ "Search Archives" feature). Raw (.wiff) data files were analysed using Spectronaut™ 15 with default settings targeted analysis. These default settings included: dynamic iRT retention time prediction with a correction factor for window 1; mass calibration was set to local; decoy method was set as scrambled; the false discovery rate (FDR), according to the mProphet approach (27), was set at 1% on the precursor, peptide and protein group levels; protein inference was set to "default" which is based on the ID picker algorithm (28), and global cross-run normalization on the median was selected. The final urinary proteome spectral library (peptides—20,616, protein groups—2604) was used as a reference for targeted data extraction.

### 2.5 Retrospective power analysis

The appropriate fold-change cut-off for group comparisons was determined by performing retrospective power analysis using the MSstats package (Northeastern University, MSstat 4.4.1, Bioconductor version: release 3.15, R v4.2.0). A group comparison was performed to compare the protein changes between cases and controls. To achieve desirable statistical power, parameters that were false discovery rate (FDR)=0.01 (1%), and n=minimum sample size for each comparison was 52. At power=80%, proteins that showed a fold-change ≥ 2.25 were statistically significant.

### 2.6 Data analysis and pathway analysis

The demographic and biochemical characteristics were analysed using STATA 17 SE (Stata Corp, Texas, USA) and GraphPad Prism 10 (GraphPad Software, San Diego California USA). Continuous and categorical variables were analysed by chi-square ($\chi^2$) and Mann-Whitney U tests, respectively. The continuous variables are expressed as medians (interquartile ranges), and categorical variables as proportions. A p-value < 0.05 was considered statistically significant. A volcano plot was generated using online data analysis and visualization tool http://www.bioinformatics.com.cn/srplot (accessed on the 15th January 2024) and principal component analysis (PCA) was performed on Metaboanalyst v.5.0 (https://www.metaboanalysts.ca/). Differentially abundant proteins between cases and controls were calculated by a two-sided t-test, with a minimum fold change ≥ 2.25 and *p*-values adjusted for multiple testing by FDR at 1% from Spectronaut™ 15. Receiver operating characteristic (ROC) curves were generated using MetaboAnalyst v.5.0 on all differentially abundant proteins (DAPs) to identify putative marker proteins that differentiate albuminuria patients from those without. Data was log-transformed, interquartile range filtering, and zero imputation strategy was used. Normalization by median was performed. Multivariate ROC exploratory analysis was generated using the Monte-Carlo cross-validation algorithm to identify proteins or models with high sensitivity and specificity. In each validation algorithm, two-thirds of the participant data were used to build a model, while the remaining one-third were used to validate the model. This process was replicated several times to calculate the performance and confidence intervals (29). The results of the ROC curve analysis are reported as the area under the curve (AUC) (30). Functional enrichment analysis on DAPs was performed using a free online tool Enrichr/Enrichr-KG (https://www.maayanlab.cloud/, accessed on 20th January 2024), with the Reactome 2022 human database (31,32). The top enriched pathways ($p$<0.05) were selected.

## 3. Results

### 3.1. Clinical and demographic characteristics of patients

The baseline characteristics and biochemical variables of the study population are summarized in **Table 1**. From the initial 116 participants selected for analysis, eight were excluded post-data acquisition due to poor chromatographic separation or low peptide recovery. The median age of the sample (N= 108) was 42 years, with 57% being female. There were significantly more HIV-positive individuals ($p$=0.033) in the albuminuric group, and uACR was significantly higher in this group (7.9 mg/mmol vs 0.6 mg/mmol, $p$ <0.001).

**Table 1. Baseline characteristics of participants**

| Variable | Total | Cases ( N=56) | Controls (N=52) | P value |
|---|---|---|---|---|
| Age, years, | 42 (30–54) | 42 (30–55) | 42 (31–53) | 0.987 |
| Women | 61/108 (57) | 32/56 (57) | 29/52 (56) | 0.886 |
| BMI, kg/m$^2$ | 25 ( 22–29) | 25 ( 21–28) | 25 (23–33) | 0.428 |
| Serum creatinine, µmol/L | 63 ( 53–74) | 63 ( 53–76) | 63 (52–71) | 0.550 |
| eGFR, mL/min/1.73m$^2$ | 113 ( 95–124) | 111(93–124) | 114 (99–124) | 0.707 |
| uACR, mg/mmol | 3.9 ( 0.6–8.4) | 7.9 (5.5–18.5) | 0.6 ( 0.30–1.1) | **<0.001** |
| HPT status | 12/ 108 (11) | 8 /45 ( 18) | 4/ 46 ( 9) | 0.439 |
| Diabetes status | 3/108 (2.7) | 3/26 (12) | 0/28 (0.0) | 0.064 |
| HIV status | 35/108 (32) | 22/56 (39) | 13/52 (25) | **0.033** |
| Smoking | 17/108 (16) | 8/56 (14) | 9/52 (17) | 0.667 |
| Glucose, mmol/L | 6.3 (5.6–7.7) | 6.3 ( 5.7–7.7) | 6.4 ( 5.6–7.5) | 0.848 |

Abbreviations: BMI: body mass index, SBP: systolic blood pressure, DBP: diastolic blood pressure, HPT: hypertension, HIV: human immunodeficiency virus, eGFR: estimated glomerular filtration (CKD-EPI$_{creatinine}$ equation 2009 without correction for race), ACR: urine albumin: creatinine ratio. For continuous variables, data was reported as median (25$^{th}$ to 75$^{th}$ percentile), for categorical variables, data reported as number (N) of participants and percentages (%).

### 3.2 Performance of study-specific system suitability-quality control

The performance of study-specific and commercial QCs (Hela) is shown in **Supplementary Material Figure S1**. The coefficient of variation (CV) at the protein group level was 15% in the Hela digest system suitability assessment and 14% for the urine pool over 5 days of analysis. The CV at the peptide level was 18.2% and 17.5 % for Hela and urine pool, respectively. Counts for protein, peptide, and precursors remained consistent throughout the data acquisition process (**Supplementary Material Figure S1: A-F**).

### 3.3. Differentially abundant proteins between cases and controls

Differentially abundant proteins (FDR ≤ 1%, fold-change ≥ 2.25) are shown in **Figure. 1**. Eighty urinary proteins (**Supplementary Material, Tables S1 and S2**) were differentially abundant between the groups. Among them, 59 urine proteins (74 %) showed higher abundance, while 21 proteins (26 %) showed lower abundance in cases compared to controls. PCA revealed modest clustering between the groups (**Figure. 2**). The top ten DAPs in higher abundance (based on q values) in cases comprised of alpha-1-antitrypsin (SERPINA1), albumin, afamin, antithrombin III (SERPINC1), vitamin-D binding protein, alpha-1B glycoprotein, beta-ala His dipeptidase, myoglobin, fibrinogen beta chain, and apolipoprotein

A-1 (**Supplementary Material Table S1**). The top ten DAPs found in lower abundance in the cases (higher abundance in the control group) included polymeric immunoglobulin receptor, collagen alpha-1(VI) chain, mucin-1, collagen alpha-1(XV), inducible costimulatory (ICOS) ligand, collagen alpha-2(IV) chain, uromodulin, intercellular adhesion molecule, osteopontin, and platelet endothelial cell adhesion (**Supplementary Table S2**).

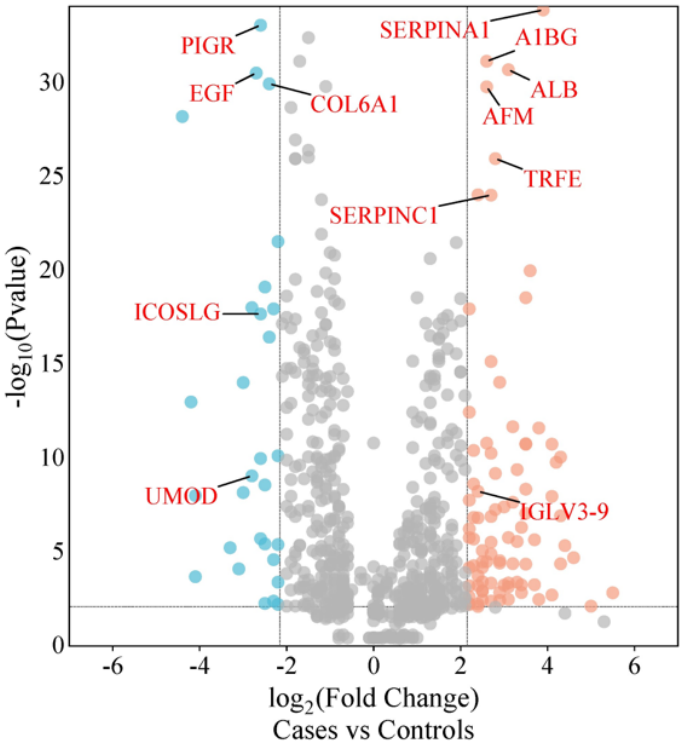



**Figure. 1** Volcano plot showing differentially abundant proteins. Candidate protein markers (orange dots indicate proteins with higher abundance in the cases and blue dots indicate proteins with lower abundance in the cases) with a cut-off of minimum fold change ≥ 2.25 and maximum 0.01% (1% FDR).

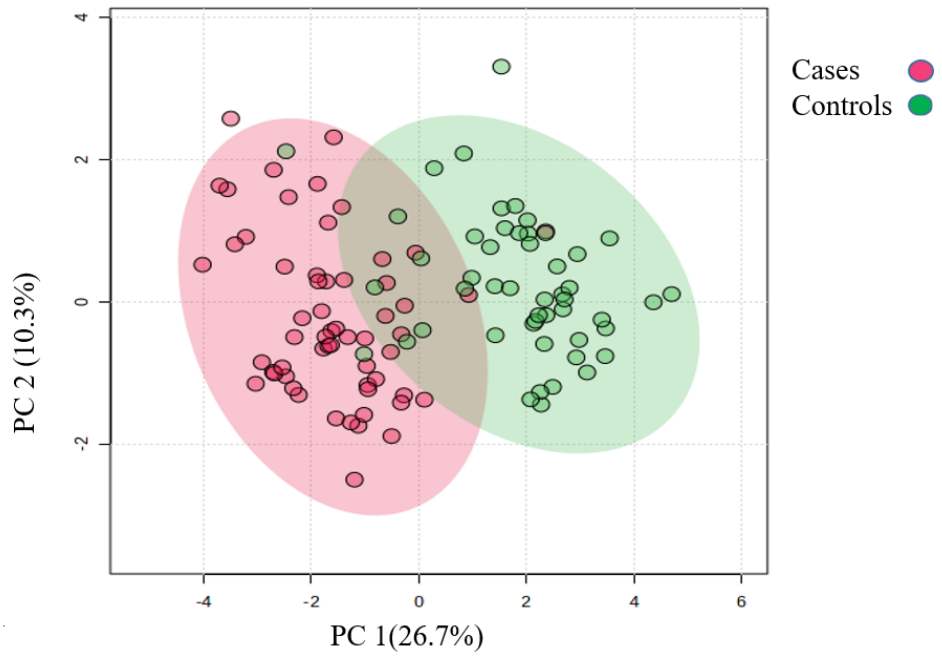
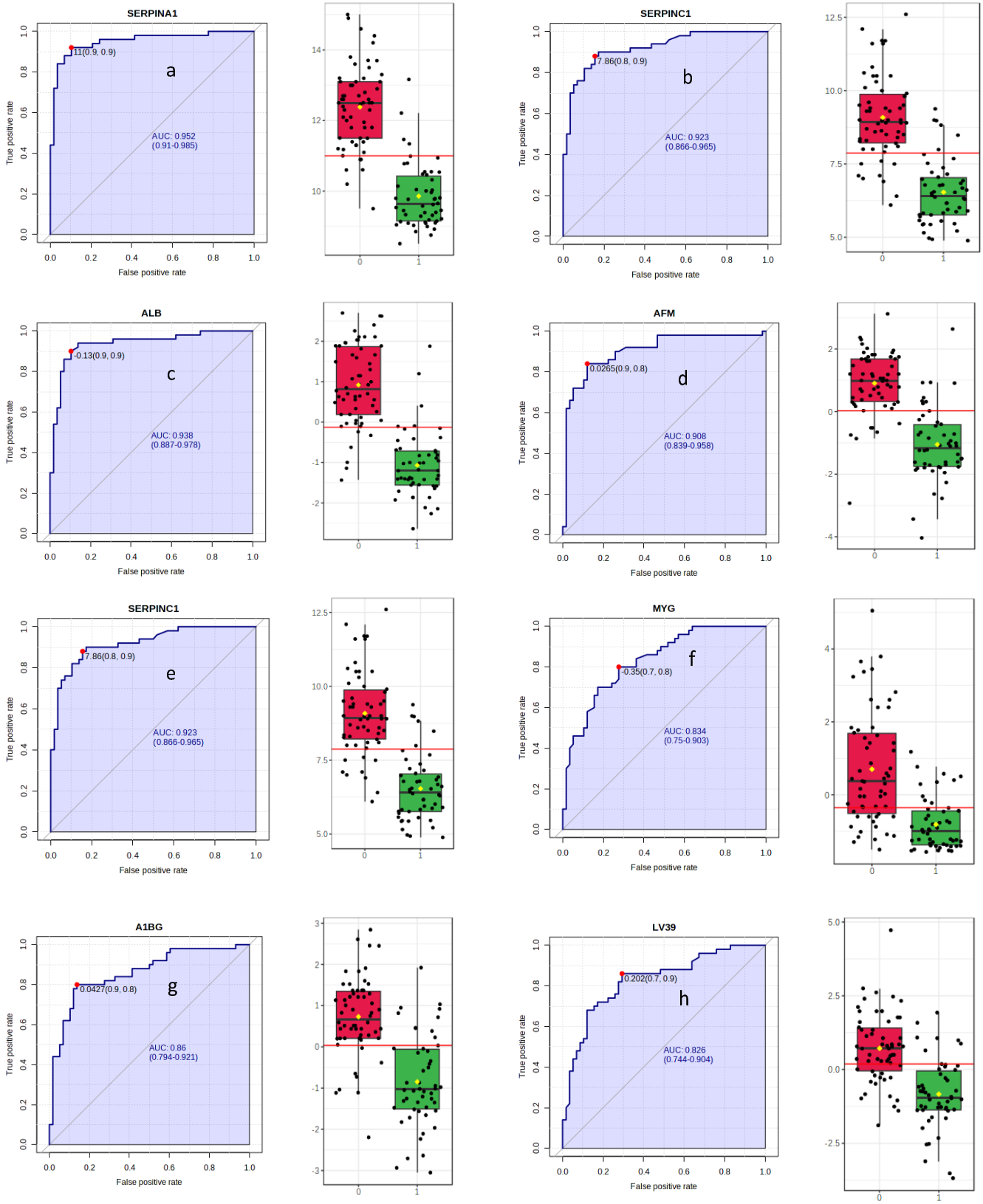



**Figure 2.** Principal component analysis of the urinary proteome of albuminuria (red) and normoalbuminuria (green).

### 3.4 Potential markers for albuminuria and normoalbuminuria classification

ROC curve analysis was performed to evaluate the diagnostic potential of differentially abundant proteins. Using an AUC >0.8 and p-valve <0.05 as criteria for diagnostic potential, 12 urine proteins were identified as potential markers for albuminuria (**Figure 1 and Figure 3, Supplementary Material Table S3).** Among 12 urine proteins, SERPINA1, albumin, SERPINC1, and afamin exhibited AUCs greater than 0.9 (AUC=0.951, 0.936, 0.920, 0.903, respectively). Among the 12 urine proteins with the highest AUC, seven proteins were higher in abundance (SERPINA1, alpha-1B glycoprotein, albumin, afamin, SERPINC1, myoglobin, and immunoglobulin lambda variable 3-9), while five proteins (polymeric immunoglobulin receptor, collagen alpha-1 (VI) chain, mucin-1, uromodulin, and inducible costimulatory ligand) were lower in abundance among cases compared to controls. Finally, a multivariate ROC curve analysis was performed using a combination of proteins selected through logistic regression analysis, as shown in **Figure 4A**. All 6 urine protein models had AUC > 0.8. However, the first three models had wider confidence intervals than models 3 and 4 (**Figure 4A**). The highest AUC of 0.907 (confidence interval (CI): 0.836-0.954) was achieved with the 6$^{th}$ model, comprising 80 urine proteins (**Supplementary Material Table S4**). Among all

models, the 6th model had the highest predictive accuracy of 91.3% to classify cases, even though the increase was modest compared to the 4th and 5th urine protein models (**Figure 4B**).


SERPINA1
11(0.9, 0.9)
a
AUC: 0.952
(0.91-0.985)
SERPINC1
7.86(0.8, 0.9)
b
AUC: 0.923
(0.866-0.965)
ALB
-0.13(0.9, 0.9)
c
AUC: 0.938
(0.887-0.978)
AFM
0.0265(0.9, 0.8)
d
AUC: 0.908
(0.839-0.958)
SERPINC1
7.86(0.8, 0.9)
e
AUC: 0.923
(0.866-0.965)
MYG
-0.35(0.7, 0.8)
f
AUC: 0.834
(0.75-0.903)
A1BG
0.0427(0.9, 0.8)
g
AUC: 0.86
(0.794-0.921)
LV39
0.202(0.7, 0.9)
h
AUC: 0.826
(0.744-0.904)
True positive rate
False positive rate

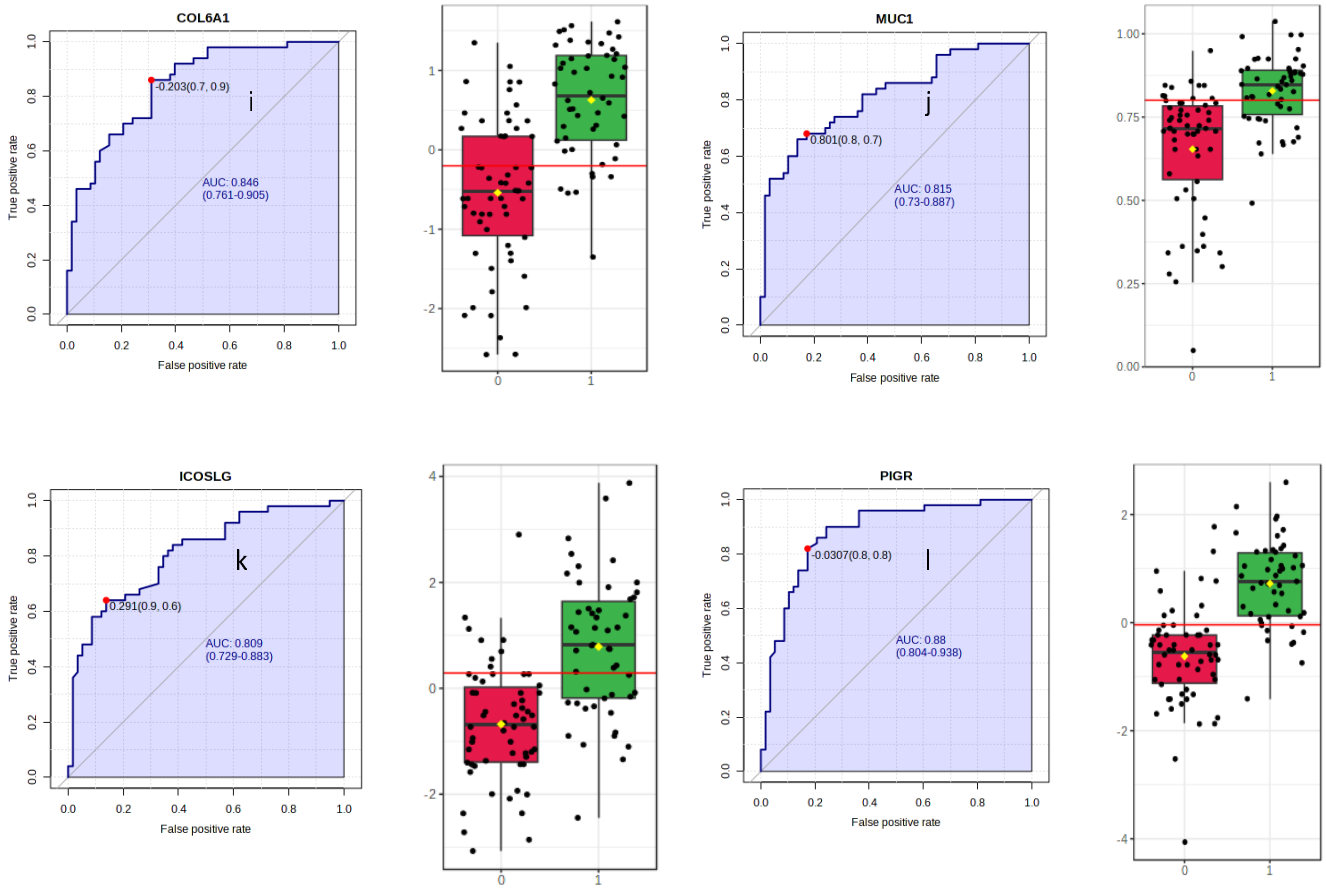
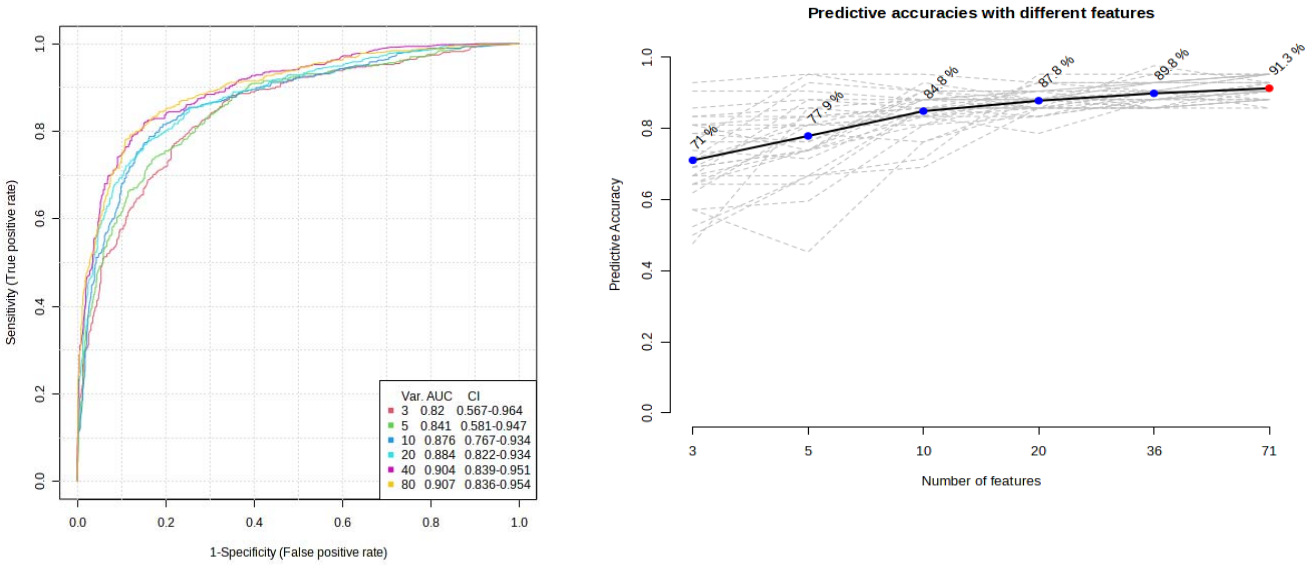



**Figure 3a-l.** The receiver operating characteristic curves (A-L) of the top 12 urine proteins. The box-cum whisker plots show significantly different proteins between albuminuric and normoalbuminuric patients. The box denotes interquartile ranges, and the bottom and top boundaries of boxes are the 25$^{th}$ and 75$^{th}$ percentiles, respectively. Lower and upper whiskers correspond to the 5$^{th}$ and 95$^{th}$ percentiles, respectively. A horizontal line inside the box denotes the median. P value < 0.001 for all proteins. SERPINA1-alpha-1 antitrypsin, ALB-albumin, SERPINC1-antithrombin III, AFM-afamin, PIGR-polymeric immunoglobulin receptor, AIBG-alpha-1B-glycoprotein, COL6A1-collagen alpha-1 (VI) chain, MYG-myoglobin, LV39-immunoglobulin lambda variable 3-9, MUC-1-mucin-1, ICOSLG-inducible costimulatory ligand, UMOD-uromodulin. Red boxes represent cases (albuminuria group), green boxes represent controls (normoalbuminuric group)




(**a**) (**b**)

**Figure 4.** Multivariate ROC of 6 models and their predictive accuracies. (**a**) Performance of 6 models; and (**b**) predictive accuracies of all models for discriminating albuminuria from normoalbuminuria.

3.5 **Functional enrichment analysis of differentially abundant proteins**

Pathway analysis was performed by querying the Reactome library, within the Enrichr free online tool, to identify enriched pathways in which differentially abundant proteins are involved. Significantly enriched functional pathways based on FDR (≤ 1%) and p-value < 0.05 are illustrated in **Supplementary Material Figure S2** and **Table S4**). The top five highly enriched pathways for all DAPs included: the innate immune system, regulation of insulin-like growth factor (IGF) transport and uptake by insulin-like growth factor binding proteins (IGFBPs), platelet degranulation, response to elevated cytosolic $Ca^{++}$, and haemostasis (**Supplementary Material Figure S2**). In the biological function (GO) analysis for urinary proteins that showed higher abundance in cases, the top five highly enriched terms were involved in the following biological processes (**Figure 5**): regulated exocytosis, platelet degranulation, high-density particle remodelling, hydrogen peroxide catabolic process, and hydrogen peroxide metabolic process. While proteins that showed lower abundance among cases were enriched in processes such as extracellular structure organization, extracellular matrix organization, extracellular encapsulating structure organization, and homophilic cell adhesion molecule (**Figure 6, Supplementary Material Figure 3S**).

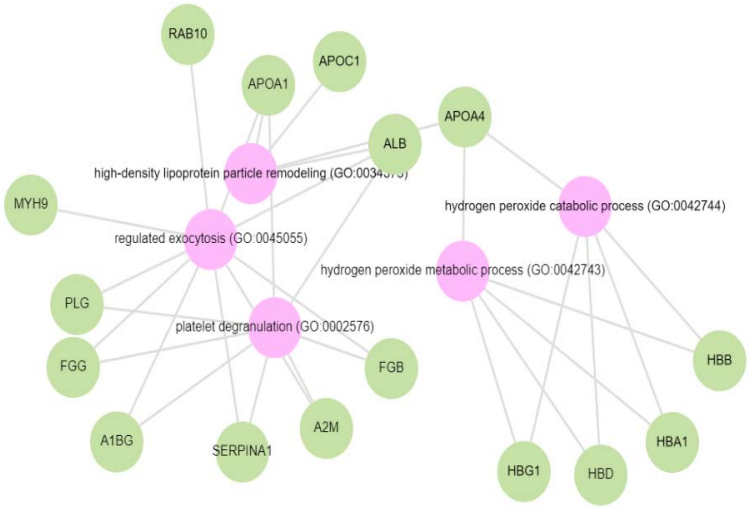



**Figure 5.** GO biological terms for differential abundant proteins increased in the albuminuria group through Enrich-KG (Reactome library). Pink circles indicate highly enriched terms

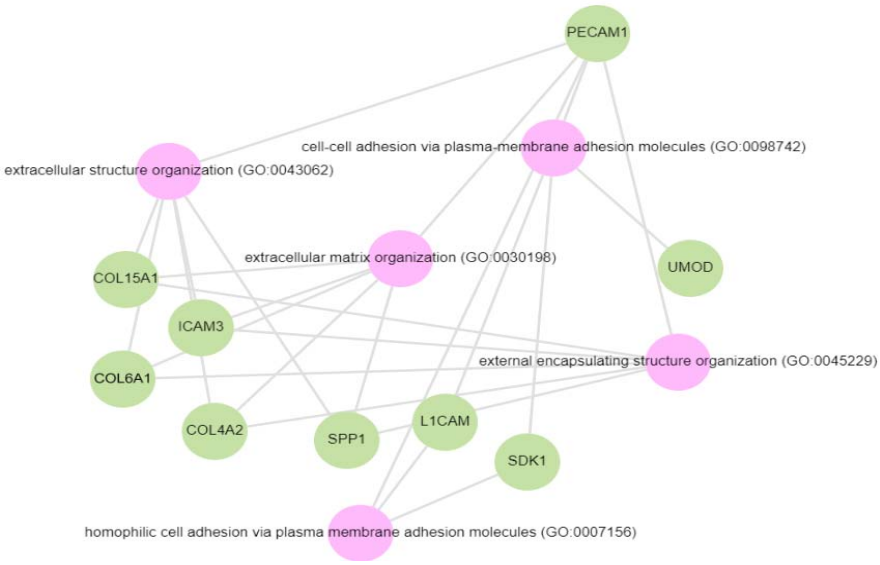



**Figure 6.** GO biological process for proteins decreased in the albuminuria (increased in normoalbuminuria).

## 4. Discussion

In this study which profiled the urinary proteome in albuminuria with well-preserved eGFR, we found 80 differentially expressed proteins when comparing albuminuric cases to non-albuminuric controls. Using multivariable analysis, we identified a combination of 80 proteins (AUC= 0.907) that outperformed other models with a predictive accuracy of 91.3%, suggesting that a combination of markers may offer added clinical utility compared to a single marker (33). In this model, five proteins (ALB, AFM, PIGR, A1AT, ANT3) were found to be differentially expressed similarly to a 20-protein classifier identified in the South African study of hypertension-associated albuminuria (21) and in previous studies (34–36), suggesting generalisability of these markers to different populations.

We observed a higher rate of albuminuria in HIV-infected individuals, compared to those uninfected. HIV-infected individuals are likely to have kidney damage due to direct virus-induced kidney injury or secondary antiretroviral therapy (ART) (37). It has been shown that the prevalence of albuminuria significantly increases, ranging between 34–68% in HIV-infected individuals (38,39).

Most DAPs were involved in pathways involving the innate immune system, regulation of IGF transport and uptake by IGFBPs, platelet degranulation, response to elevated cytosolic $Ca^{++}$, and haemostasis. Dysregulation of the immune system, and inflammatory events have been postulated as key mechanisms in the pathogenesis of CKD (40). Kidney injury from any cause eventually evolves into CKD in the presence of unresolved inflammation. Following kidney injury, damage-associated molecules initiate inflammatory responses such as proinflammatory cytokine activation, and subsequent infiltration of immune cells, such as neutrophils, macrophages, and natural killer cells (41,42). The role of IGFBPs in regulating transcription, cell migration induction, cell-cycle arrest, and preventing tissue renewal, triggering apoptosis, is linked to the development of fibrosis in kidney disease (43). Similarly to the findings of this study, evidence indicates that IGFBPs are upregulated in various kidney diseases (44–46). One mechanism of action of IGFBPs is through binding to insulin-like growth factor (IGF) (47). IGFs play a significant role in enhancing kidney injury repair (47). Upregulation of pathways that prevent the release and actions of IGFs is implicated in the kidney in CKD (48). Administration of sodium-glucose co-transporter 2 (SGLT2) inhibitors has been shown to abolish the effect of IGFBP, especially IGFBP-7 (49). This

effect is believed to be the main cause of the kidney-protective effect of SGLT2 inhibitors, independent of the hypoglycaemic effect in the treatment of type 2 diabetes (50).

We identified nine proteins that were linked to the activation of platelet degranulation pathways. Among these, seven were increased in the albuminuric group. Participants with CKD demonstrated increased platelet activation (51). In addition to their role in thrombosis, platelets appear to modulate inflammatory processes that are integral to the pathophysiology of CKD and its progression (39,47,48). Importantly, other studies have conflicting results of reduced platelet activation or unchanged platelet activation. However, these studies included mainly patients on chronic dialysis and with uraemic syndrome. Both the haemostasis and platelet degranulation pathways often occur together, which could present an opportunity for treatment targets (53). Furthermore, elevated cytosolic calcium leads to the activation of protein kinase C (PKC), a hallmark of the response to platelet degranulation. It is not surprising that pathways to the cytosolic calcium response pathway are also highly enriched together with platelet degranulation and haemostasis (40). Coupled with inflammatory pathways, CKD is a highly oxidative state (54). High levels of oxidative stress have already been detected in the early stages of CKD, and increase with the progression to end-stage renal disease (55,56). These observations are supported by highly enriched biological function processes associated with hydrogen peroxidase catabolic and metabolic processes among the DAPs found in this study.

In keeping with growing evidence, DAPs that were decreased in the albuminuria group were enriched in processes such as extracellular structure organization, extracellular matrix organization, extracellular encapsulating structure organization, and homophilic cell adhesion molecule. It is well accepted that accumulation of extracellular matrix disrupts cellular organization and function of the kidney. An increase in extracellular structure matrix proteins is a key finding in renal fibrosis (57,58). Decreased urinary uromodulin, osteopontin, and collagen fragments all have been implicated in fibrosis-promoting processes in CKD (17,59). There is mounting evidence that collagen fragments as biomarkers reflect structural changes in the renal extracellular matrix (ECM) and provide insight into the progression of fibrosis (60). We observed an increase in urinary collagen alpha-1 (III) in the albuminuria group while other collagen fragments were low (i.e. collagen alpha-1(VI), collagen alpha-1(XV), collagen alpha-2 (IV)). Different types of collagen are involved in maintaining the integrity and function of the renal structure. Upregulation of collagen III is regarded as an early event

in renal fibrosis and is associated with adverse CKD outcomes (61). Elevated levels of collagen alpha-1(III) in the urine could indicate increased turnover or breakdown of collagen in the renal interstitium, possibly as a result of kidney injury or inflammation (62). While collagen type IV (COL VI), a major component of basement membranes is crucial for maintaining the structural integrity of the glomeruli. Studies have shown high urinary COL VI levels, especially in diabetic patients compared to healthy controls (63,64). Reduced urinary COL VI in this study, may partly be explained by a shift in a deposition associated with the severity of the disease in favour of degradation. This may be due to reduced matrix metalloproteases activity and the increased cross-linking of collagen fibers that render them resistant to proteolytic processes (60,65,66).

Similarly, other studies found a lower abundance of urine UMOD in cases (67,68). The reduced levels are believed to be secondary to tubular dysfunction (69). Similarly on the South African study in hypertensive-associated albuminuria, UMOD was found in lower abundance (21). UMOD, a mucoprotein synthesized by the thick ascending loop of Henle, has been shown to be a marker of renal reserve and plays an important role in the innate immunity of the kidney (59,70). Low urine UMOD levels have been associated with acute kidney injury and its progression to CKD (71).

This case-control study characterized urinary proteome in South African participants with and without albuminuria. Most of the urinary proteins identified in this study have also been found in a variety of diseases related to kidney conditions, such as glomerulonephritis, diabetic nephropathy, and hypertension, to name a few. Despite the diverse causes of kidney damage, commonalities in upregulated or downregulated pathways suggest that therapeutic targeting of these pathways may be crucial in preventing kidney progression, regardless of the underlying cause. The main limitation of this study was the analysis of a single cohort. The findings need to be verified in a larger and more diverse cohort of patients to ensure applicability to a wider population.

## 5. Conclusions

Pathway analysis suggested significant enrichment in innate immune pathways, platelet degranulation, IGF, and IGFBPs, as well as haemostasis. Urine proteins that were in lower

abundance were involved in pathways associated with the extracellular matrix. This information can be utilized to discern potential molecular mechanisms underlying kidney damage and its progression to severe stages of chronic kidney disease (CKD). Additionally, urine proteins such as SERPINA1, ALB, SERPINC1, AFM, PIGR, A1BG, COL6A1, MYG, LV39, MUC1, ICOSLG and UMOD had the highest discriminating abilities between cases and controls. Our study identified an 80-protein model that was able to classify albuminuria from normoalbuminuria with well-preserved GFR with the 91.3% predictive accuracy.

## Supplementary Materials

The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: Project suitability quality control; Figure S2: Network Analysis (Reactome library) through Enrichr-KG. Table S1: High abundant dysregulated proteins in albuminuria; Table S2: Low abundant dysregulated proteins in albuminuria group; Table S3: ROC curve analysis of all abundant proteins; Table S4: List of genes in each biological process.

## Author Contributions

Conceptualization, S.K., J.A.G, P.N, S.H.S, I.S.G, J.F; Data acquisition and analysis – S.K., P.N., I.S.G; writing—original draft preparation, S.K.; writing—review and editing, S.K., J.A.G, P.N, S.H.S, I.S.G, J.F. All authors contributed to the manuscript and approved the version submitted. All authors have read and agreed to the published version of the manuscript.

## Funding

This work was funded by the National Research Fund (NRF), Discovery Academic Fellowship, Prof Bongani Mayosi Netcare Clinician Scholarship, and DIPLOMICS, a research infrastructure initiative of the Department of Science and Innovation of South Africa.

## Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki, and received approval (certificate number: M210128) from the Faculty of Health Sciences Human Research Ethics Committee (Medical) at the University of the Witwatersrand, Johannesburg, South Africa.

## Informed Consent Statement

Informed consent was obtained from all subjects involved in the original African Research on Kidney Disease (ARK) study.

## Conflicts of Interest

ISG and SHS are employees of ReSyn Biosciences who are the proprietors of MagReSyn® HILIC technology. SHS is also an employee of Evosep Biosystems who are the proprietors of Evosep technology. The remaining authors have no competing interests associated with this manuscript. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References


1. GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Lond Engl. 2020;395(10225):709–33.

2. Hill NR, Fatoba ST, Oke JL, Hirst JA, O'Callaghan CA, Lasserson DS, et al. Global prevalence of chronic kidney disease – A Systematic Review and Meta-Analysis. PLoS One. 2016;11(7):e0158765.

3. Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet. 2016;388(10053):1459–544.

4. Stevens PE, Ahmed SB, Carrero JJ, Foster B, Francis A, Hall RK, et al. KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Int. 2024;105(4):S117–314.

5. Kidney Diseases Improving Global Outcomes (KDIGO) CKD Work Group. Clinical Practice Guidelines for Evaluation and Management of Chronic Kidney Disease. Kidney international Supplement. 2013;3:1–150.

6. Levey AS, De Jong PE, Coresh J, El Nahas M, Astor BC, Matsushita K, et al. The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. Kidney Int. 2011;80(1):17–28.

7. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. The Lancet. 2010;375(9731):2073–81.

8. Lambers Heerspink HJ, Gansevoort RT. Albuminuria is an appropriate therapeutic target in patients with CKD: The pro view. Clin J Am Soc Nephrol CJASN. 2015;10(6):1079–88.

9. Abbate M, Zoja C, Remuzzi G. How does proteinuria cause progressive renal damage? J Am Soc Nephrol. 2006;17(11):2974–84.

10. Berhane AM, Weil EJ, Knowler WC, Nelson RG, Hanson RL. Albuminuria and estimated glomerular filtration rate as predictors of diabetic end-stage renal disease and death. Clin J Am Soc Nephrol. 2011;6(10):2444–51.

11. Leong A, Ekinci EI, Nguyen C, Milne M, Hachem M, Dobson M, et al. Long-term intra-individual variability of albuminuria in type 2 diabetes mellitus: implications for categorization of albumin excretion rate. BMC Nephrol. 2017;18(1):355.

12. Currie GE, Von Scholten BJ, Mary S, Flores Guerrero JL, Lindhardt M, Reinhard H, et al. Urinary proteomics for prediction of mortality in patients with type 2 diabetes and microalbuminuria. Cardiovasc Diabetol. 2018;17(1):50.

13. Rodríguez-Ortiz ME, Pontillo C, Rodríguez M, Zürbig 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(1):15940.

14. Ahn HS, Kim JH, Jeong H, Yu J, Yeom J, Song SH, et al. Differential urinary proteome analysis for predicting prognosis in type 2 diabetes patients with and without renal dysfunction. Int J Mol Sci. 2020;21(12):4236.

15. Zürbig P, Jerums G, Hovind P, MacIsaac RJ, Mischak H, Nielsen SE, et al. Urinary Proteomics for Early Diagnosis in Diabetic Nephropathy. Diabetes. 2012;61(12):3304–13.

16. Dubin RF, Rhee EP. Proteomics and metabolomics in kidney disease, including Insights into etiology, treatment, and prevention. Clin J Am Soc Nephrol. 2020;15(3):404–11.

17. Øvrehus MA, Zürbig P, Vikse BE, Hallan SI. Urinary proteomics in chronic kidney disease: diagnosis and risk of progression beyond albuminuria. Clin Proteomics. 2015;12(1):21.

18. Fan G, Gong T, Lin Y, Wang J, Sun L, Wei H, et al. Urine proteomics identifies biomarkers for diabetic kidney disease at different stages. Clin Proteomics. 2021;18(1):32.

19. Gajjala PR, Bruck H, Noels H, Heinze G, Ceccarelli F, Kribben A, et al. Novel plasma peptide markers involved in the pathology of CKD identified using mass spectrometric approach. J Mol Med. 2019;97(10):1451–63.

20. De Beer D, Mels CMC, Schutte AE, Delles C, Mary S, Mullen W, et al. Identifying a urinary peptidomics profile for hypertension in young adults: The African PREDICT study: Urinary peptidomics and hypertension. PROTEOMICS. 2023;23(11):2200444.

21. Govender MA, Stoychev SH, Brandenburg JT, Ramsay M, Fabian J, Govender IS. Proteomic insights into the pathophysiology of hypertension-associated albuminuria: Pilot study in a South African cohort. Clin Proteomics. 2024;21(1):15.

22. Kalyesubula R, Fabian J, Nakanga W, Newton R, Ssebunnya B, Prynn J, et al. How to estimate glomerular filtration rate in sub-Saharan Africa: design and methods of the African Research into Kidney Diseases (ARK) study. BMC Nephrol. 2020;21(1):20.

23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377–81.

24. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, et al. Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension. 2003;42(6):1206–52.

25. Fabian J, Gondwe M, Mayindi N, Chipungu S, Khoza B, Gaylard P, et al. Chronic kidney disease (CKD) and associated risk in rural South Africa: a population-based cohort study. Wellcome Open Res. 2022;7:236.

26. George JA, Brandenburg JT, Fabian J, Crowther NJ, Agongo G, Alberts M, et al. Kidney damage and associated risk factors in rural and urban sub-Saharan Africa (AWI-Gen): a cross-sectional population study. Lancet Glob Health. 2019;7(12):e1632–43.

27. Reiter L, Rinner O, Picotti P, Hüttenhain R, Beck M, Brusniak MY, et al. mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nat Methods. 2011;8(5):430–5.

28. Zhang B, Chambers MC, Tabb DL. Proteomic Parsimony through Bipartite Graph Analysis Improves Accuracy and Transparency. J Proteome Res. 2007;6(9):3549–57.

29. Costanzo M, Caterino M, Salvatori I, Manganelli V, Ferri A, Misasi R, et al. Proteome data of neuroblastoma cells overexpressing Neuroglobin. Data Brief. 2022;41:107843.

30. Shan G. Monte Carlo cross-validation for a study with binary outcome and limited sample size. BMC Med Inform Decis Mak. 2022;22(1):270.

31. Evangelista JE, Xie Z, Marino GB, Nguyen N, Clarke DJB, Ma'ayan A. Enrichr-KG: bridging enrichment analysis across multiple libraries. Nucleic Acids Res. 2023;51(W1):W168–79.

32. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14(1):128.

33. Barratt J, Topham P. Urine proteomics: the present and future of measuring urinary protein components in disease. Can Med Assoc J. 2007;177(4):361–8.

34. Chebotareva N, Vinogradov A, McDonnell V, Zakharova NV, Indeykina MI, Moiseev S, et al. Urinary protein and peptide markers in chronic kidney disease. Int J Mol Sci. 2021;22(22):12123.

35. Kaburagi Y, Takahashi E, Kajio H, Yamashita S, Yamamoto-Honda R, Shiga T, et al. Urinary afamin levels are associated with the progression of diabetic nephropathy. Diabetes Res Clin Pract. 2019;147:37–46.

36. Ning J, Xiang Z, Xiong C, Zhou Q, Wang X, Zou H. Alpha1-Antitrypsin in Urinary Extracellular Vesicles: A Potential Biomarker of Diabetic Kidney Disease Prior to Microalbuminuria. Diabetes Metab Syndr Obes Targets Ther. 2020;13:2037–48.

37. Boulle A. Antiretroviral therapy and early mortality in South Africa. Bull World Health Organ. 2008 Sep 1;86(9):678–87.

38. Kim PS, Woods C, Dutcher L, Georgoff P, Rosenberg A, Mican JAM, et al. Increased Prevalence of Albuminuria in HIV-Infected Adults with Diabetes. Pett S, editor. PLoS ONE. 2011;6(9):e24610.

39. Assaram S, Magula NP, Mewa Kinoo S, Mashamba-Thompson TP. Renal manifestations of HIV during the antiretroviral era in South Africa: a systematic scoping review. Syst Rev. 2017;6(1):200.

40. Zoccali C, Vanholder R, Massy ZA, Ortiz A, Sarafidis P, Dekker FW, et al. The systemic nature of CKD. Nat Rev Nephrol. 2017;13(6):344–58.

41. Ikezumi Y, Hurst L, Atkins RC, Nikolic-Paterson DJ. Macrophage-mediated renal injury is dependent on signaling via the JNK pathway. J Am Soc Nephrol. 2004;15(7):1775–84.

42. Zhang ZX, Wang S, Huang X, Min WP, Sun H, Liu W, et al. NK cells induce apoptosis in tubular epithelial cells and contribute to renal ischemia-reperfusion injury. J Immunol. 2008;181(11):7489–98.

43. Wang S, Chi K, Wu D, Hong Q. Insulin-Like Growth Factor Binding Proteins in Kidney Disease. Front Pharmacol. 2021; 12:807119.

44. Worthmann K, Peters I, Kümpers P, Saleem M, Becker JU, Agustian PA, et al. Urinary excretion of IGFBP-1 and -3 correlates with disease activity and differentiates focal segmental glomerulosclerosis and minimal change disease. Growth Factors. 2010;28(2):129–38.

45. Ravassa S, Beaumont J, Cediel G, Lupón J, López B, Querejeta R, et al. Cardiorenal interaction and heart failure outcomes. A role for insulin-like growth factor binding protein 2? Rev Esp Cardiol Engl Ed. 2020;73(10):835–43.

46. Gokulakrishnan K, Velmurugan K, Ganesan S, Mohan V. Circulating levels of insulin-like growth factor binding protein–1 in relation to insulin resistance, type 2 diabetes mellitus, and metabolic syndrome (Chennai Urban Rural Epidemiology Study 118). Metabolism. 2012;61(1):43–6.

47. Bach LA, Hale LJ. Insulin-like Growth Factors and Kidney Disease. Am J Kidney Dis. 2015;65(2):327–36.

48. Gao L, Zhong X, Jin J, Li J, Meng X ming. Potential targeted therapy and diagnosis based on novel insight into growth factors, receptors, and downstream effectors in acute kidney injury and acute kidney injury-chronic kidney disease progression. Signal Transduct Target Ther. 2020;5(1):9.

49. Cai D, Chen Q, Mao L, Xiao T, Wang Y, Gu Q, et al. Association of SGLT2 inhibitor dapagliflozin with risks of acute kidney injury and all-cause mortality in acute myocardial infarction patients. Eur J Clin Pharmacol. 2024;80(4):613–20.

50. Heerspink HJL, Stefánsson BV, Correa-Rotter R, Chertow GM, Greene T, Hou FF, et al. Dapagliflozin in Patients with Chronic Kidney Disease. N Engl J Med. 2020;383(15):1436–46.

51. Gremmel T, Muller M, Steiner S, Seidinger D, Koppensteiner R, Kopp CW, et al. Chronic kidney disease is associated with increased platelet activation and poor response to antiplatelet therapy. Nephrol Dial Transplant. 2013;28(8):2116–22.

52. Huang HS, Chang HH. Platelets in inflammation and immune modulations: Functions beyond hemostasis. Arch Immunol Ther Exp (Warsz). 2012;60(6):443–51.

53. Jain N, Corken AL, Kumar A, Davis CL, Ware J, Arthur JM. Role of platelets in chronic kidney disease. J Am Soc Nephrol JASN. 2021;32(7):1551–8.

54. Podkowińska A, Formanowicz D. Chronic kidney disease as oxidative stress- and inflammatory-mediated cardiovascular disease. Antioxidants. 2020;9(8):752.

55. Annuk M, Zilmer M, Lind L, Linde T, Fellström B. Oxidative stress and endothelial function in chronic renal failure. J Am Soc Nephrol. 2001;12(12):2747–52.

56. Dounousi E, Papavasiliou E, Makedou A, Ioannou K, Katopodis KP, Tselepis A, et al. Oxidative stress is progressively enhanced with advancing stages of CKD. Am J Kidney Dis. 2006;48(5):752–60.

57. Kim KP, Williams CE, Lemmon CA. Cell–matrix interactions in renal fibrosis. Kidney Dial. 2022;2(4):607–24.

58. Pontillo C, Jacobs L, Staessen JA, Schanstra JP, Rossing P, Heerspink HJL, et al. A urinary proteome-based classifier for the early detection of decline in glomerular filtration. Nephrol Dial Transplant. 2016;32(9):1510–6.

59. Pruijm M, Ponte B, Ackermann D, Paccaud F, Guessous I, Ehret G, et al. Associations of urinary uromodulin with clinical characteristics and markers of tubular function in the general Population. Clin J Am Soc Nephrol. 2016;11(1):70–80.

60. Rasmussen DGK, Boesby L, Nielsen SH, Tepel M, Birot S, Karsdal MA, et al. Collagen turnover profiles in chronic kidney disease. Sci Rep. 2019;9(1):16062.

61. Genovese F, Rasmussen DGK, Karsdal MA, Jesky M, Ferro C, Fenton A, et al. Imbalanced turnover of collagen type III is associated with disease progression and mortality in high-risk chronic kidney disease patients. Clin Kidney J. 2021;14(2):593–601.

62. Bülow RD, Boor P. Extracellular Matrix in Kidney Fibrosis: More Than Just a Scaffold. J Histochem Cytochem. 2019;67(9):643–61.

63. Mahendran KB, Bhaskar MV, Santha K, Inmozhi R, Perumal KK. Plasma and Urinary Type IV Collagen Levels for Early Detection of Nephropathy in Type 2 Diabetes Mellitus Patients. Int J Health Sci. 2016;10(4):492–8.

64. Furumatsu Y, Nagasawa Y, Shoji T, Yamamoto R, Iio K, Matsui I, et al. Urinary Type IV Collagen in Nondiabetic Kidney Disease. Nephron Clin Pract. 2010;117(2):c160–6.

65. Catania JM, Chen G, Parrish AR. Role of matrix metalloproteinases in renal pathophysiologies. Am J Physiol-Ren Physiol. 2007;292(3):F905–11.

66. Good DM, Zürbig P, Argilés A, Bauer HW, Behrens G, Coon JJ, et al. Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. Mol Cell Proteomics MCP. 2010;9(11):2424–37.

67. Prikryl P, Vojtova L, Maixnerova D, Vokurka M, Neprasova M, Zima T, et al. Proteomic approach for identification of IgA nephropathy-related biomarkers in urine. Physiol Res. 2017;621–32.

68. Devuyst O, Bochud M. Uromodulin, kidney function, cardiovascular disease, and mortality. Kidney Int. 2015;88(5):944–6.

69. Shao C, Yan W, Gao Y. Application of urinary proteomics in biomarker discovery. Sci China Life Sci. 2011, 54(5);409–17.

70. Puthumana J, Thiessen-Philbrook H, Xu L, Coca SG, Garg AX, Himmelfarb J, et al. Biomarkers of inflammation and repair in kidney disease progression. J Clin Invest. 2021;131(3):e139927, 139927.

71. Wen Y, Xu L, Melchinger I, Thiessen-Philbrook H, Moledina DG, Coca SG, et al. Longitudinal biomarkers and kidney disease progression after acute kidney injury. JCI Insight. 2023;8(9):e167731.