KI REPORTS
KIReports.org
CLINICAL RESEARCH

A Urinary Fragment of Mucin-1 Subunit α Is a Novel Biomarker Associated With Renal Dysfunction in the General Population

Zhen-Yu Zhang[1], Susana Ravassa[2,3], Martin Pejchinovski[4], Wen-Yi Yang[1], Petra Zürbig[4], Begoña López[2,3], Fang-Fei Wei[1], Lutgarde Thijs[1], Lotte Jacobs[1], Arantxa González[2,3], Jens-Uwe Voigt[5], Peter Verhamme[6], Tatiana Kuznetsova[1], Javier Díez[2,3,7], Harald Mischak[4,8] and Jan A. Staessen[1,9]

[1]Studies Coordinating Centre, Research Unit Hypertension and Cardiovascular Epidemiology, KU Leuven Department of Cardiovascular Diseases, University of Leuven, Leuven, Belgium; [2]Program of Cardiovascular Diseases, Centre for Applied Medical Research, University of Navarra, Navarra Institute for Health Research, Pamplona, Spain; [3]CIBERCV, Carlos III Institute of Health, Madrid, Spain; [4]Mosaiques Diagnostic and Therapeutics AG, Hannover, Germany; [5]Research Unit Cardiology, KU Leuven Department of Cardiovascular Diseases, University of Leuven, Leuven, Belgium; [6]Centre for Molecular and Vascular Biology, KU Leuven Department of Cardiovascular Diseases, University of Leuven, Leuven, Belgium; [7]Department of Cardiology and Cardiac Surgery, University of Navarra Clinic, Pamplona, Spain; [8]BHF Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, United Kingdom; and [9]R&D Group VitaK, Maastricht University, Maastricht, The Netherlands

Introduction: Sequencing peptides included in the urinary proteome identifies the parent proteins and may reveal mechanisms underlying the pathophysiology of chronic kidney disease.

Methods: In 805 randomly recruited Flemish individuals (50.8% women; mean age, 51.1 years), we determined the estimated glomerular filtration rate (eGFR) from serum creatinine using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation. We categorized eGFR according to the National Kidney Foundation Kidney Disease Outcomes Quality Initiative guideline. We analyzed 74 sequenced urinary peptides with a detectable signal in more than 95% of participants. Follow-up measurements of eGFR were available in 597 participants.

Results: In multivariable analyses, baseline eGFR decreased ($P \leq 0.022$) with urinary fragments of mucin-1 (standardized association size expressed in ml/min/1.73 m^2, −4.48), collagen III (−2.84), and fibrinogen (−1.70) and was bi-directionally associated ($P \leq 0.0006$) with 2 urinary collagen I fragments (+2.28 and −3.20). The eGFR changes over 5 years (follow-up minus baseline) resulted in consistent estimates ($P \leq 0.025$) for mucin-1 (−1.85), collagen (−1.37 to 1.43) and fibrinogen (−1.45) fragments. Relative risk of having or progressing to eGFR <60 ml/min/1.73 m^2 was associated with mucin-1. Partial least-squares analysis confirmed mucin-1 as the strongest urinary marker associated with decreased eGFR, with a score of 2.47 compared with 1.80 for a collagen I fragment as the next contender. Mucin-1 predicted eGFR decline to <60 ml/min/1.73 m^2 over and above microalbuminuria ($P = 0.011$) and retained borderline significance ($P = 0.05$) when baseline eGFR was accounted for.

Discussion: In the general population, mucin-1 subunit α, an extracellular protein that is shed from renal tubular epithelium, is a novel biomarker associated with renal dysfunction.

Kidney Int Rep (2017) **2,** 811–820; http://dx.doi.org/10.1016/j.ekir.2017.03.012
KEYWORDS: collagen; fibrosis; glomerular filtration rate; mucin-1; population science; proteomics

Chronic kidney disease (CKD) is a major health problem affecting the quality of life of millions of people. The Global Burden of Disease Study 2010 estimated that worldwide 0.40 million of nearly 50 million deaths occurring annually were attributable to CKD in 1990 and 0.74 million in 2010, representing an increase by 82.3%.[1] In the United States, the prevalence of CKD, defined as an estimated glomerular filtration rate (eGFR) < 60 ml/min/1.73 m^2, rose from 10.0% in 1988 to 1994 to 13.1% in 1999 to 2004.[2] The International Society of Nephrology appealed to address knowledge gaps related to kidney injury.[3]

Correspondence: Jan A. Staessen, Studies Coordinating Centre, Research Unit Hypertension and Cardiovascular Epidemiology, KU Leuven Department of Cardiovascular Sciences, University of Leuven, Campus Sint Rafaël, Kapucijnenvoer 35, Box 7001, BE-3000 Leuven, Belgium. E-mail: jan.staessen@med.kuleuven.be

Received 17 December 2016; revised 4 March 2017; accepted 31 March 2017; published online 7 April 2017

As recently reviewed,[4] proteins mainly appear in urine when the glomerular barrier is failing. In contrast, peptides are freely filtered into urine and, if incompletely reabsorbed by the tubules, the urinary excretion of peptides captures the early stages of renal dysfunction.[5] Capillary electrophoresis coupled with high-resolution mass spectrometry enables detection of more than 5000 peptide fragments in urine samples.[6,7] We recently identified a unique urinary proteomic signature, which, in patient cohorts,[8–10] reproducibly predicted progression of CKD, independent of other risk factors. To deepen insight in the pathophysiological pathways leading to renal injury and to extend previous findings in patients[8–10] to the general population, we analyzed the database of the Flemish Study on Environment, Genes and Health Outcomes (FLEMENGHO)[11] and searched for association of eGFR with individual urinary peptide fragments, of which the amino acid sequence revealed the parent protein.

MATERIALS AND METHODS

Recruitment of Participants

FLEMENGHO complies with the Declaration of Helsinki[12] for research in human subjects. The Ethics Committee of the University of Leuven approved the study.[11] All participants gave informed written consent. Recruitment started in 1985 and continued until 2004. The initial participation rate was 78.0%. The participants were repeatedly followed up at the field center in the catchment area (North Limburg, Belgium). From May 2005 until May 2010, we mailed an invitation letter to 1208 former participants for a follow-up examination. However, 153 were unavailable, because they had died (n = 26), had been institutionalized or were too ill (n = 27), or had moved out of the area (n = 100). Of the remaining 1055 former participants, 828 (78.5%) renewed informed consent. We excluded 23 participants from analysis because urine samples were unavailable (n = 23). Thus, the number of participants statistically analyzed totaled 805.

Assessment of Renal Function

Venous blood samples were drawn after at least 8 hours of fasting. We measured the concentration of creatinine in serum, using the Jaffe method[13] with modifications described elsewhere,[14,15] in a single certified laboratory that applied isotope-dilution mass spectrometry for calibration of the serum creatinine measurements. We derived eGFR from serum creatinine by the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI)[16] equation. The CKD stages, defined according to the National Kidney Foundation Kidney Disease Outcomes Quality Initiative (KDOQI) guideline, were eGFR $\geq$ 90, 60 to 89, 30 to 59, 15 to 29, and $<$ 15 ml/min/1.73 m^2 for stages 1, 2, 3, 4, and 5, respectively. eGFR was assessed at baseline in all participants, of whom 597 had a follow-up assessment (Supplementary Figure S1). Participants also collected a 24-hour urine sample within 1 week of the baseline clinical examination at the field center for measurement of microalbuminuria and creatinine. Microalbuminuria was defined as an albumin-to-creatinine ratio of $\geq$ 3.5 mg/mmol in women and $\geq$ 2.5 mg/mmol in men, and macroalbuminuria as an albumin-to-creatinine ratio of $\geq$ 30 mg/mmol.[17] Guideline-based[18] staging of CKD requires repeated measurement of eGFR or albuminuria or additional evidence for renal disease. However, as this is impracticable in the context of population studies, because multiple visits decrease the participation rate, staging of CKD at baseline and follow-up relied in our current study, as often done in landmark epidemiological research,[19–22] on a single serum sample.

At baseline, we also measured plasma glucose, serum total and high-density lipoprotein (HDL) cholesterol, serum γ-glutamyltransferase as index of alcohol intake, and 24-hour microalbuminuria. Diabetes mellitus was a self-reported diagnosis, a fasting plasma glucose of $\geq$126 mg/dl, or use of antidiabetic agents.[23]

Measurements of Biomarkers

Aliquots (0.7 ml) were stored at −80 °C and thawed immediately before analysis. Capillary electrophoresis coupled with high-resolution mass spectrometry analysis, sequencing of peptides, mass spectrometry data evaluation, and calibration were performed, as described in detail in the Supplementary Material and in previous publications.[6,7] A total of 2129 urinary peptides were sequenced. We selected 74 peptides (Supplementary Table S5), which had a detectable signal in more than 95% of participants (Supplementary Figure S2). For the current analysis, the values of peptides undetectable in less than 5% of randomly varying study participants were set to the minimum of the distribution of each sequenced peptide. This strategy has been validated by other research groups.[24,25]

Other Measurements

Blood pressure was the average of 5 consecutive auscultatory readings obtained according to European guidelines[26] with a standard mercury sphygmomanometer after participants had rested in the sitting position for at least 10 minutes. Hypertension was a blood pressure of at least 140 mm Hg systolic or 90 mm Hg diastolic or use of antihypertensive drugs. Body mass index was calculated as weight in kilograms divided by height in meters squared. Study nurses administered a standardized questionnaire inquiring into each participant's medical history, smoking and drinking habits, and intake of medications.

Statistical Analyses

For database management and statistical analysis, we used SAS version 9.4 (SAS Institute Inc., Cary, NC). Supplementary Table S1 summarizes the statistical analysis work flow. Means were compared using the large-sample z test, proportions by the Fisher exact test, and survival function estimates by the log-rank test. We normalized the distributions of γ-glutamyltransferase and 24-hour albuminuria by a logarithmic transformation. We rank normalized the distributions of the urinary peptides by sorting measurements from the smallest to the largest and then applying the inverse cumulative normal function.[27]

We identified covariables to be retained in the analyses by a best-subset approach, with the number of covariables allowed to be retained in the model set at 11. In continuous analyses, we standardized eGFR to the average in the whole study population (mean or ratio) of the covariables identified by stepwise regression. While accounting for covariables, we regressed the indices of renal function on the urinary peptide markers and constructed –log10 probability plots. Based on the number of parent proteins, we used a Bonferroni-corrected P-value threshold of 0.005 (0.05/10). Renal function and changes in renal function were analyzed as continuous or categorical variables, using multivariable-adjusted linear regression, logistic regression, and Cox modeling, as appropriate. In Cox regression, the start date to event and censoring date coincided with the baseline and follow-up visit.

In the next step of our analyses, we applied partial least-squares (PLS) analysis, which is a statistical technique that constructs models for continuous outcomes in relation to correlated high-dimensional explanatory variables.[28] In our study, PLS allowed identification of a set of independent latent factors that are linear combinations of the urinary peptides and that maximize the covariance between the urinary peptides and the variables describing renal function. We studied the multivariable-adjusted eGFR in relation to the latent factors. We retained the smallest number of latent factors for which the predicted residual sums of squares (calculated using leave-one-out cross-validation) did not differ significantly ($P > 0.10$) from the model with the minimum predicted residual sums of squares value as assessed by the van der Voet T^2 statistic. The importance of each urinary peptide in the construction of the PLS factors was assessed from the Variable Importance in Projection (VIP) scores of Wold, with the threshold set at 1.5. Finally, we evaluated the capability to discriminate between participants with or without renal impairment by constructing receiver operating characteristic curves and by calculating the area under the receiver operating characteristic curve (AUC). The 95% confidence interval (CI) of the AUC was calculated by the DeLong method.

RESULTS

Baseline Characteristics of Participants

Of 805 participants, 409 (50.8%) were female. All were Europeans of white ethnicity. Mean values (± SD) in the 805 participants were 51.1 ± 15.7 years for age, 26.5 ± 4.3 kg/m^2 for body mass index, 129.5 ± 17.7 mm Hg for systolic blood pressure, 79.7 ± 9.6 mm Hg for diastolic blood pressure, and 203 ± 38 and 55 ± 14 mg/dl for total and HDL cholesterol. Among all participants, 344 (42.7%) had hypertension, of whom 212 (61.6%) were on antihypertensive drug treatment, and 20 (2.5%) had diabetes.

At baseline, the prevalence of renal dysfunction, defined as an eGFR of < 60 ml/min/1.73 m^2, amounted to 74 (9.2%). Table 1 shows that age, body mass index, central obesity, systolic and mean arterial pressure, total cholesterol, plasma glucose, urinary albumin-to-creatinine ratio, 24-hour albuminuria, proportion of women, prevalence of hypertension, and probability of being treated for hypertension were all higher ($P \leq 0.028$) in participants with renal dysfunction. The opposite was the case for the prevalence of reported alcohol intake and serum HDL cholesterol levels ($P \leq 0.049$). Reduced eGFR was associated with a higher prevalence of microalbuminuria (18 participants [2.5%] vs. 11 participants [14.9%]; $P < 0.0001$), with no difference in the prevalence of macroalbuminuria (1 [0.1%] vs. 1 [1.4%] participant; $P = 0.18$). Compared with participants with normal renal function, patients with renal dysfunction more frequently ($P < 0.0001$) used diuretics (8.3% vs. 33.8%), β-blockers (13.1% vs. 40.5%), angiotensin-converting enzyme inhibitors, and angiotensin-I type-1 receptor blockers (7.4% vs. 21.6%), whereas this was not the case for vasodilators, including calcium-channel blockers and α-blockers (4.7% vs. 8.1%; $P = 0.19$). The baseline characteristics of the 597 participants who had a follow-up measurement of eGFR appear in Supplementary Table S2 and mirror those presented in Table 1.

Change in eGFR and eGFR Category

Median follow-up of eGFR in 597 participants was 4.7 years (5th to 95th percentile interval, 3.7–5.4). In these participants, eGFR decreased by 1.5 ml/min/1.73 m^2 per year. Of 172 participants with baseline eGFR ≥ 90 ml/min/1.73 m^2, 75 (43.6%) maintained this category, and 97 (56.4%) progressed to eGFR 60–89 ml/min/1.73 m^2. Of the 386 participants with baseline eGFR 60–89 ml/min/1.73 m^2, 21 participants (5.4%) regressed to eGFR ≥ 90 ml/min/1.73 m^2, 309 (80.1%) remained within this category, and 55 (14.2%) and 1 (0.3%)

Table 1. Baseline characteristics of participants by renal function

Characteristic	All	eGFR < 60	eGFR ≥ 60	*P* value
No. of participants (%)				
All participants in category	805	731	74	
Women	409 (50.8)	359 (49.1)	50 (67.6)	0.003
Smokers	162 (20.1)	153 (20.9)	9 (12.2)	0.073
Drinking ≥ 5 g/d alcohol	328 (40.7)	314 (43.0)	14 (9.2)	<0.0001
Hypertension	344 (42.7)	285 (39.0)	59 (79.7)	<0.0001
Antihypertensive treatment	212 (61.6)	165 (22.6)	47 (63.5)	<0.0001
Diabetes mellitus	20 (2.5)	16 (2.2)	4 (5.4)	0.090
Mean (SD) of characteristic				
Age (yr)	51.1 ± 15.7	49.1 ± 14.8	70.6 ± 10.4	<0.0001
Body mass index (kg/m^2)	26.5 ± 4.3	26.4 ± 4.4	27.8 ± 3.8	0.008
Waist-to-hip ratio	0.87 ± 0.08	0.86 ± 0.08	0.89 ± 0.08	0.013
Systolic pressure (mm Hg)	129.5 ± 17.7	127.8 ± 16.3	145.8 ± 21.9	<0.0001
Diastolic pressure (mm Hg)	79.7 ± 9.6	79.6 ± 9.7	80.2 ± 8.1	0.55
Mean arterial pressure (mm Hg)	96.3 ± 10.7	95.7 ± 10.6	102.1 ± 10.0	<0.0001
Heart rate (beats/min)	63.5 ± 9.8	63.5 ± 9.7	63.7 ± 10.7	0.82
Serum total cholesterol (mg/dl)	203 ± 38	202 ± 37	215 ± 39	0.006
Serum HDL cholesterol (mg/dl)	55 ± 14	55 ± 14	52 ± 13	0.049
Plasma glucose (mg/dl)	87 ± 14	88 ± 14	92 ± 13	0.028
Serum creatinine (mg/dl)	0.95 ± 0.18	1.00 ± 0.14	1.19 ± 0.29	<0.0001
eGFR (ml/min/1.73 m^2)	81.8 ± 17.3	77.9 ± 14.0	51.3 ± 7.2	<0.0001
Geometric mean (IQR) of characteristic				
UACR (mg/mmol)	0.52 (0.32–0.76)	0.49 (0.30–0.72)	1.06 (0.48–1.28)	<0.0001
24-h Albuminuria (mg)	6.0 (4.1–7.4)	5.8 (4.1–7.1)	9.3 (4.6–9.5)	<0.0001
γ-Glutamyltransferase (units/l)	23 (15–32)	23 (15–32)	25 (16–33)	0.42

eGFR, estimated glomerular filtration rate, derived from serum creatinine by the Chronic Kidney Disease Epidemiology Collaboration equation; HDL, high-density lipoprotein; IQR, interquartile range; UACR, urinary albumin-to-creatinine ratio. Values are arithmetic mean (SD) or geometric mean (interquartile range). Office blood pressure was the average of 5 consecutive readings. Hypertension was defined as an office blood pressure ≥ 140 mm Hg systolic or ≥ 90 mm Hg diastolic or use of antihypertensive drugs. Diabetes mellitus was a self-reported diagnosis, a fasting glucose level ≥ 126 mg/dl, or use of antidiabetic agents.

progressed to eGFR 30–59 or 15–29 ml/min/1.73 m^2, respectively. Of the 39 participants with baseline eGFR 30 to 59 ml/min/1.73 m^2, 9 participants (23.1%) regressed to eGFR 60 to 89 ml/min/1.73 m^2, 26 (66.7%) remained at the same stage, and 4 (10.3%) progressed to eGFR 15 to 29 ml/min/1.73 m^2. No participant proceeded to the eGFR category < 15 ml/min/1.73 m^2 or to renal replacement therapy.

Analysis of eGFR as Continuous Variable

Based on the best-subset regression procedure, we adjusted the cross-sectional associations between eGFR and the urinary peptides in 805 participants for mean arterial pressure, waist-to-hip ratio, smoking, plasma glucose, γ-glutamyltransferase, total-to-HDL cholesterol ratio, 24-hour albuminuria and use of diuretics, inhibitors of the renin-angiotensin system (β-blockers, angiotensin-converting enzyme inhibitors, and angiotensin type-1 receptor blockers) and vasodilators (calcium-channel blockers and α-blockers). These covariables explained 21.9% of the variance in eGFR. Analyses with change in eGFR in 597 participants were additionally adjusted for baseline eGFR and follow-up duration.

Supplementary Table S3 lists the multivariable-adjusted associations of eGFR and change in eGFR with the urinary peptides measured at baseline. Figure 1 identifies the peptides that retained significance with Bonferroni correction applied. Table 2 lists the urinary peptides that, with adjustment for multiple testing, were associated with both baseline eGFR and change in eGFR over follow-up. Per 1-SD increment in the urinary peptides (Table 2), the association sizes, expressed in ml/min/1.73 m^2, amounted to −4.48 ($P < 0.0001$) for p8342 (mucin-1), −3.20 ($P < 0.0001$) for p77763 (collagen I), −2.84 ($P < 0.0001$) for p105352 (collagen III), −1.70 ($P = 0.022$) for p61573 (fibrinogen), and +2.28 ($P = 0.0006$) for p57531 (collagen I). Consistent with the cross-sectional analysis, over time (Table 2), eGFR increased with p57531 (+1.43 ml/min/1.73 m^2 per 5 years; $P = 0.005$) but decreased with the other urinary peptides, with effect sizes ranging from −1.23 to −1.85 ml/min/1.73 m^2 per 5 years; $P \leq 0.025$), among which p8342 (mucin-1) was the strongest predictor.

The PLS procedure yielded 3 latent factors that accounted for 20.1% of the overall variance in the urinary peptides and 28.9% of the variance in multivariable-adjusted eGFR. Figure 2 depicts the PLS-derived VIP scores versus the centered and rescaled correlation coefficients. The dependent variable in this analysis was baseline eGFR standardized for the

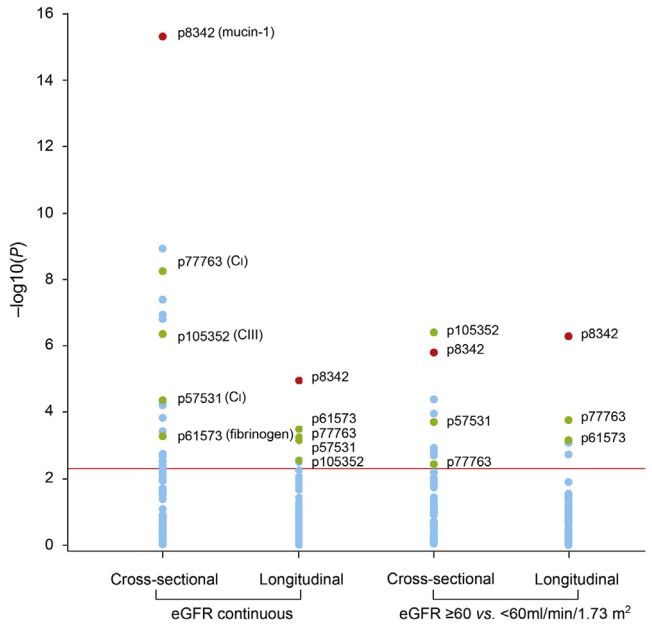

Figure 1. −Log10(p) probability plot of the multivariable-adjusted associations of renal function measures with the urinary peptides. All analyses were adjusted for mean arterial pressure, waist-to-hip ratio, smoking, plasma glucose, γ-glutamyltransferase, total-to–high-density-lipoprotein cholesterol ratio, 24-hour albuminuria, and use of diuretics, inhibitors of the renin-angiotensin system (β-blockers, angiotensin-converting enzyme inhibitors, and angiotensin type-1 receptor blockers), and vasodilators (calcium-channel blockers and α-blockers). The longitudinal analysis of change in estimated glomerular filtration rate (eGFR) as continuous variable was additionally adjusted for baseline eGFR and follow-up duration. The horizontal line denotes the significance level with Bonferroni correction applied. Red dots represent mucin-1 and green dots the other peptides listed in Tables 2 and 3.

aforementioned covariables. The urinary peptides associated with lower eGFR (left side of the V plot in Figure 2) included, among others, mucin-1 subunit α

Table 2. Multivariable-adjusted associations of eGFR with urinary proteomic biomarkers

Biomarker (baseline)	Parent protein	eGFR baseline (n = 805)	eGFR change over 5 yr (n = 597)
		Estimate (95% CI)	Estimate (95% CI)
p8342	mucin-1	−4.48 (−6.00 to −2.96)[a]	−1.85 (−3.03 to −0.66)[b]
p57531	collagen I	2.28 (0.69 to 3.87)[b]	1.43 (0.29 to 2.57)[d]
p77763	collagen I	−3.20 (−4.73 to −1.68)[a]	−1.37 (−2.49 to −0.26)[d]
p105352	collagen III	−2.84 (−4.40 to −1.27)[a]	−1.23 (−2.38 to −0.10)[c]
p61573	fibrinogen	−1.70 (−3.26 to −0.15)[c]	−1.45 (−2.56 to −0.33)[d]

eGFR, estimated glomerular filtration rate, derived from serum creatinine by the Chronic Kidney Disease Epidemiology Collaboration equation. Change in eGFR was the follow-up minus the baseline value. Estimates given with 95% confidence interval, express the change in the dependent variable associated with a 1-SD increase in the normalized urinary peptides measured at baseline. The cross-sectional analyses were adjusted for mean arterial pressure, waist-to-hip ratio, smoking, plasma glucose, γ-glutamyltransferase, total-to–high-density lipoprotein cholesterol ratio, 24-hour albuminuria, and use of diuretics, inhibitors of the renin-angiotensin system (β-blockers, angiotensin-converting enzyme inhibitors, and angiotensin type-1 receptor blockers) and vasodilators (calcium-channel blockers and α-blockers). Longitudinal analyses were additionally adjusted for baseline eGFR and follow-up duration. *P* values reflect Bonferroni-corrected significance of the associations.
[a] $P \leq 0.0001$.
[b] $P \leq 0.001$.
[c] $P \leq 0.05$.
[d] $P \leq 0.01$.

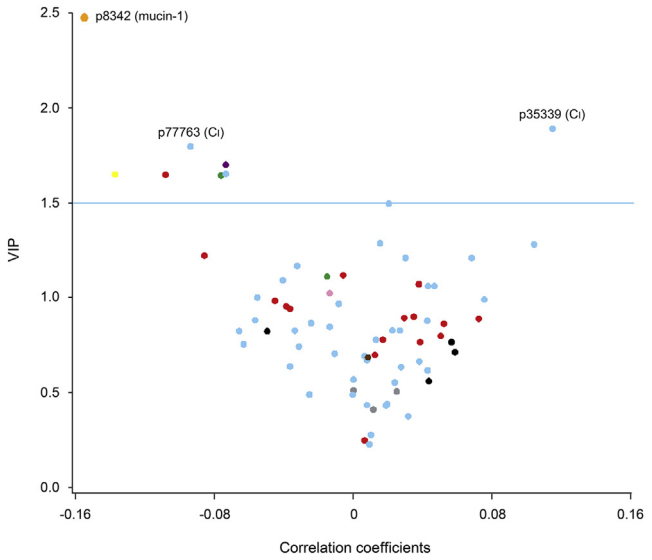

Figure 2. V-plots generated by partial least-squares analysis. Variable Importance in Projection (VIP) scores indicate the importance of each urinary fragment in the construction of the partial least-squares factors and are plotted against the centered and rescaled correlation coefficients. The correlation coefficients reflect the associations of the multivariable-adjusted estimated glomerular filtration rate (eGFR) with the urinary fragments. Fragments associated with reduced eGFR (left side of the V-plot) include, among others, p8342 and p77763 (Table 2). p35339 was associated with a higher eGFR (right side of the V-plot). Colors identify fragments derived from collagens I (blue), II (gray), III (red), IV (brown), the mucin-1 subunit α (orange), fibrinogen (green), protocadherin-12 (purple), retinol-binding protein 4 (pink), stabilin-2 (yellow), and uromodulin (black).

(p8342) and collagen I (p77763). The urinary peptide associated with higher eGFR (right side of the V plot in Figure 2) was collagen I fragment (p35339).

Analysis of eGFR Categories

Supplementary Table S4 lists the multivariable-adjusted associations of eGFR categories (≥ 60 vs. < 60 ml/min/1.73 m^2) or longitudinal change in eGFR categories (from ≥ 60 to < 60 ml/min/1.73 m^2) with the urinary peptides measured at baseline. Figure 1 shows the peptides that retained significance with Bonferroni correction applied. In the cross-sectional analysis (Table 3), the odds ratios associated with a 1-SD increment in the urinary peptides were 2.03 ($P < 0.0001$) for p8342 (mucin-1), 0.58 ($P = 0.0002$) for p57531 (collagen I), 1.45 ($P = 0.043$) for p77763 (collagen I), 2.18 ($P < 0.0001$) for p105352 (collagen III), and 1.18 ($P > 0.99$) for p61573 (fibrinogen). The risk of eGFR decline from ≥ 60 to < 60 ml/min/1.73 m^2 increased ($P < 0.0001$) across thirds of the mucin-1 distribution (Figure 3). With multivariable adjustments applied, the hazard ratios associated with a 1-SD increase were 2.10 ($P < 0.0001$) for p8342, 0.78 ($P = 0.76$) for p57531, 1.78 ($P = 0.0017$) for p77763, 1.16 ($P > 0.99$) for p105352, and 1.66 ($P = 0.007$) for p61573 (Table 3).

Table 3. Multivariable-adjusted associations of eGFR category with urinary proteomic biomarkers

Biomarker (baseline)	Parent protein	eGFR ≥ 60 versus < 60 (731 versus 74) Odds ratio (95% CI)	eGFR ≥ 60 → < 60 (502 versus 56) Hazard ratio (95% CI)
p8342	mucin-1	2.03 (1.34–3.08)[a]	2.10 (1.39–3.17)[a]
p57531	collagen I	0.58 (0.39–0.87)[b]	0.78 (0.52–1.16)
p77763	collagen I	1.45 (1.01–2.10)[c]	1.78 (1.16–2.73)[d]
p105352	collagen III	2.18 (1.42–3.34)[a]	1.16 (0.74–1.80)
p61573	fibrinogen	1.18 (0.80–1.72)	1.66 (1.09–2.52)[d]

eGFR, estimated glomerular filtration rate. Hazard ratios were computed excluding 39 participants with eGFR < 60 ml/min/1.73 m^2 at baseline. The letter n indicates the number of participants with eGFR ≥ 60 and < 60 ml/min/1.73 m^2 in the cross-sectional (odds ratio) and longitudinal (hazard ratio) analyses. The censoring date in Cox regression coincided with the follow-up visit. All analyses were adjusted for baseline variables, including mean arterial pressure, waist-to-hip ratio, smoking, plasma glucose, γ-glutamyltransferase, total-to–high-density-lipoprotein cholesterol ratio, 24-hour albuminuria, and use of diuretics, inhibitors of the renin-angiotensin system (β-blockers, angiotensin-converting enzyme inhibitors, and angiotensin type-1 receptor blockers) and vasodilators (calcium-channel blockers and α-blockers). *P* values reflect Bonferroni-corrected significance of the associations.
[a]$P \leq 0.0001$.
[b]$P \leq 0.001$.
[c]$P < 0.05$.
[d]$P \leq 0.01$.

Microalbuminuria is currently the standard for predicting progression to renal impairment. Mucin-1 was the strongest peptidomic predictor in our analysis (Figure 1, Table 3). Using urinary mucin-1 (p8342) instead of 24-hour microalbuminuria (Figure 4) to predict eGFR decline from ≥ 60 to < 60 ml/min/1.73 m^2 increased ($P = 0.011$) the AUC from 0.58 (95% CI = 0.50–0.66) to 0.72 (95% CI = 0.64–0.79). Replacing the urinary albumin-to-creatinine ratio by p8342 likewise increased

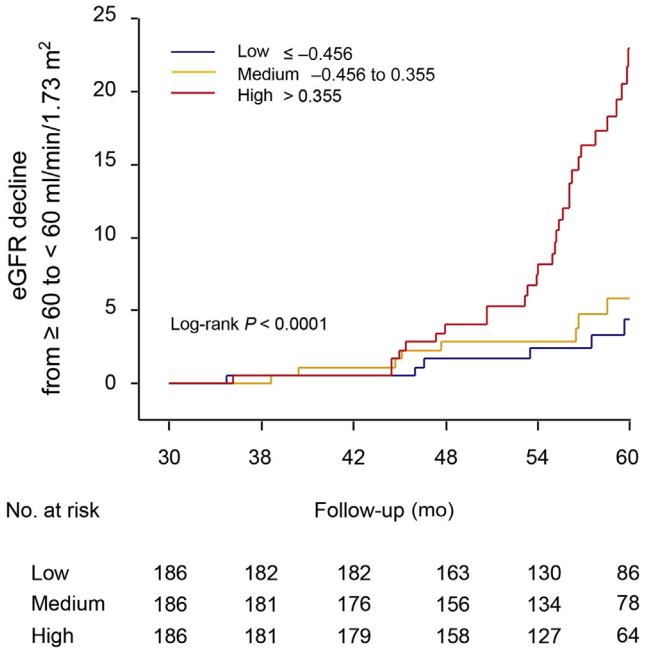

Figure 3. Kaplan–Meier survival function estimates for estimated glomerular filtration rate (eGFR) decline from ≥ 60 to < 60 ml/min per 1.73 m^2 by thirds of the mucin-1 subunit α (p8342) distribution. The *P* value for the between-group differences was derived by the log-rank test.

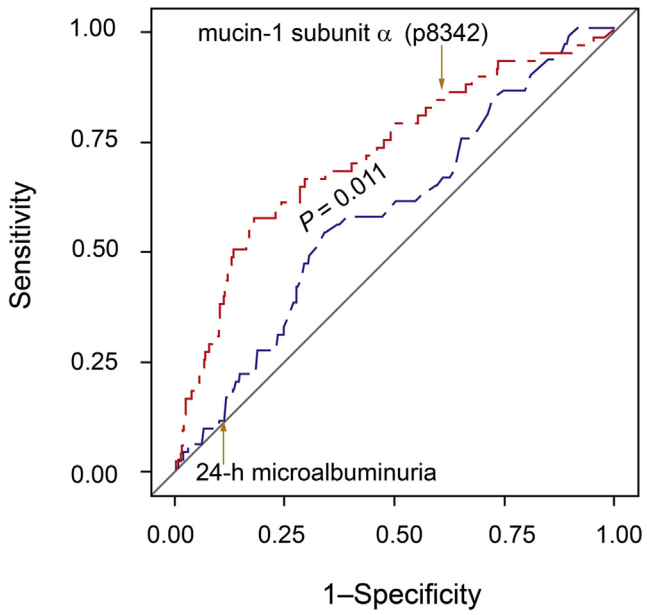

Figure 4. Receiver operating characteristic (ROC) curves for decline in estimated glomerular filtration rate (eGFR) from ≥ 60 to < 60 ml/min per 1.73 m^2. Using urinary mucin-1 (p8342) instead of 24-hour microalbuminuria to predict the decline of eGFR increased ($P = 0.011$) the AUC from 0.58 (95% CI = 0.50–0.66) to 0.72 (95% CI = 0.64–0.79).

($P = 0.010$) the AUC from 0.60 (95% CI = 0.53–0.68) to 0.72 (95% CI = 0.64–0.79). Adding urinary mucin-1 to 24-hour microalbuminuria or the albumin-to-creatinine ratio increased ($P \leq 0.0034$) the AUCs to 0.72 (95% CI = 0.64–0.79) and 0.72 (95% CI = 0.65–0.80), respectively. In a sensitivity analysis, we additionally considered baseline eGFR as predictor. For baseline eGFR alone and eGFR combined with either urinary mucin-1 or 24-hour microalbuminuria, the AUCs were 0.81 (95% CI = 0.76–0.87), 0.84 (95% CI = 0.79–0.89), and 0.82 (95% CI = 0.77–0.87), respectively. Adding mucin-1 to eGFR slightly but significantly ($P = 0.05$) increased the AUC, whereas this was not the case for 24-hour microalbuminuria ($P = 0.26$).

DISCUSSION

A comprehensive literature search (see Supplementary Material) did not reveal any population study in which renal function was correlated with individual sequenced urinary peptides. Our main findings can be summarized as follows: (i) eGFR and change in eGFR over time were inversely correlated with peptide fragments derived from the mucin-1 subunit α, collagen I and III, and fibrinogen α-chain (Figure 1, Table 2, and Supplementary Table S3); (ii) the relative risk of having or progressing to eGFR < 60 ml/min/1.73 m^2 was associated with fragments from mucin-1 subunit α and collagen I (Figure 1, Table 3, and Supplementary Table S4); (iii) and PLS confirmed the aforementioned findings and identified a mucin-1 subunit α fragment as

the strongest peptidomic correlate of a reduced eGFR with a VIP score of 2.47, compared with a score of 1.80 for collagen I fragment p77763 as the next contender (Figure 2). The most salient finding of our study was therefore the identification of the urinary mucin-1 subunit α fragment as a correlate and predictor of renal dysfunction in the general population.

Mucin-1, also known as Krebs von den Lungen−6 antigen (KL-6), is a high-molecular-weight (400 kDa), heavily *O*-glycosylated, type-I membrane-tethered glycoprotein that, under normal conditions, is expressed at the apical surface of epithelial cells.[29] The α-subunit of the protein consists of the N-terminus (104 amino acids), a 20−amino acid, variable-number, tandem-repeat segment, which is repeated 25 to 125 times depending on genetic variation and splicing, and a C-terminus of 170 amino acids.[29] The α-subunit is noncovalently bound to the transmembrane β-subunit, which consists of a small extracellular region of 58 amino acids, a transmembrane region of 28 amino acids, and a cytoplasmic tail of 72 amino acids.[29] Normal kidneys express mucin-1 in the thick segment of the loop of Henle and in the distal tubules and collecting ducts.[30–32] Diabetic kidneys also express mucin-1 in the thin segment of the loop of Henle and proximal tubules and in parietal cells lining the Bowman space.[30] The mucin-1 subunit α protrudes 200 to 500 nm above the plasma membrane, far above all other membrane-associated proteins within the 10-nm protective glycocalyx.[33]

The main function of mucin-1 is to shield cell surfaces by maintenance of a luminal epithelial muco-barrier.[34] Mucin-1 is chemotactic for human fibroblasts[35] but inhibits cellular binding to collagen I and IV.[29] The secreted mucin-1 isoform is a ligand for the mucin-1 receptor and, upon binding, initiates signal transduction.[29] In mice, renal ischemia-perfusion injury induced mucin-1 expression in all tubular epithelia.[36] The protein stimulates expression of hypoxia-inducible transcription factors, which orchestrate a protective response against acute kidney injury.[36] Our current observations are in line with these experimental findings,[36] the expression of mucin-1 in renal endothelium and tubular epithelium[29] as well as its upregulation in diabetic kidneys.[30] Recently several investigators discovered a frameshift mutation in the *MUC1* gene, located on chromosome 1 (1q21).[37] The mutation creates an abnormal sequence, encoding a new peptide that accumulates inside the MUC1-expressing renal tubular cells.[38,39] It causes autosomal dominant medullary cystic kidney disease type 1. The clinical characteristics include progression to CKD, with a widely variable age of onset of end-stage renal disease, probably dependent on gene−gene and gene−environment interactions,[39] a bland urinary sediment without blood and minimal protein, renal fibrosis, and cysts, but with no other organs expressing mucin-1 being affected. The observation that affected patients do not have proteinuria and the low prevalence of microalbuminuria in our study probably explains why p8432 was a better predictor of eGFR decline than the urinary albumin-to-creatinine ratio or 24-hour microalbuminuria. Mutations in the uromodulin gene (*UMOD*) cause medullary cystic kidney disease type 2,[34,40] which is characterized by juvenile hyperuricaemia, frequent gout, and progressive CKD. However, in our current study, eGFR was only weakly associated with 4 uromodulin fragments, with VIP scores ranging from 0.56 to 0.82.

Decline of renal function and renal fibrosis are often accompanied by a decrease in the urinary excretion of collagen fragments.[10,41] We observed that eGFR and change in eGFR were bi-directionally associated with 2 collagen I fragments and in addition were inversely associated with a collagen III fragment (Table 2, Figures 1 and 2). Several mechanisms might explain these bidirectional associations. Higher levels of tissue inhibitor of matrix metalloproteinase type 1, as observed in patients with renal dysfunction,[42] might inhibit the breakdown of collagen. If originating from extrarenal sites, collagen fragments, depending on their size, might either be insufficiently retained by the glomerular sieve or fail to be reabsorbed in the tubules.[5,10,41] Finally, reactive overexpression of mucin-1 in the tubular epithelium of patients with early renal impairment might lead to less retention of collagen fragments.[29] Both eGFR and change in eGFR were also inversely associated with the fibrinogen α chain (Table 2, Figures 1 and 2). Fibrinogen plays a role in the pathogenesis of fibrotic disorders by acting as a profibrotic ligand for a variety of cellular surface receptors. In a murine model of renal interstitial fibrosis induced by obstruction of the ureter,[43,44] pharmacological or genetic depletion of fibrinogen protected the kidneys against fibrosis. In selected patients with hypertensive nephropathy, the urinary excretion of the fibrinogen α chain was 15-fold elevated compared with that in healthy controls and was associated with a rapid decline in renal function (6.7 ml/min/1.73 m^2 per year).[45]

One potential limitation of the current study is that urinary peptidomic biomarkers derived from collagen originate not only from the kidney but from other organs as well, including bone tissue and the heart. However, urine is the matrix closest to the kidney.[4] In analyses adjusted for body height as an index of skeletal health[46] or left ventricular dysfunction,[47] our findings were confirmatory (data not shown). In our

view, the aforementioned limitation does not apply to p8342, because there is no association between mucin-1 levels in serum and urine,[30] because renal tubular cells express mucin-1[30–32] as a specific renoprotective agent,[30,36] and because genetic mutation of *MUC1* cause tubulointerstitial fibrosis.[38,39] Furthermore, we analyzed only the urinary peptides with a detectable signal in more than 95% of the participants. Ignoring biomarkers with missing values might waste potentially important information, explaining why, in previous studies of a more exploratory nature, this threshold was relaxed to 70%[47] or lower.[48] However, as a major objective of the present study was to deepen insight into the pathophysiological pathways leading to renal injury in the general population, we chose to apply a more stringent criterion, thereby avoiding the possibility of false-positive findings. Finally, the proportional hazard regression included only a subset of 558 participants (69.3%) who had both a baseline and follow-up measurement and did not have a baseline eGFR < 60 ml/min/1.73 m^2.

In conclusion, shedding of the mucin-1 subunit α, an extracellular protein expressed in renal tubular epithelium, is a predictor of renal impairment and, in the general population, outperforms other urinary biomarkers, including collagen fragments and microalbumin. Elucidation of the underlying molecular pathways in experimental studies might establish mucin-1 as a novel urinary biomarker allowing the early detection of renal dysfunction at a stage when prevention of CKD remains an achievable option.

DISCLOSURE

MP and PZ are employed by Mosaiques-Diagnostics GmbH. HM is the co-founder and co-owner of Mosaiques Diagnostics GmbH and Diapat GmbH, Hannover, Germany. All the other authors declared no competing interests.

ACKNOWLEDGMENTS

The European Union (HEALTH-FP7-278249-EUMASCARA, HEALTH-F7-305507 HOMAGE and the European Research Council (Advanced Researcher Grant 2011-294713-EPLORE and Proof-of-Concept Grant 713601-uPROPHET) and the Fonds voor Wetenschappelijk Onderzoek Vlaanderen, Ministry of the Flemish Community, Brussels, Belgium (G.0881.13, G.088013, and 11Z0916N) and The Ministry of Economy and Competitiveness, Madrid, Spain (Carlos III Institute of Health, CIBER-CV CB16/11/00483) currently support this study. The authors gratefully acknowledge the contribution of the nurses working at the examination center (Linda Custers, Marie-Jeanne Jehoul, Daisy Thijs, and Hanne Truyens) and the clerical staff at the Studies Coordinating Centre (Vera De Leebeeck and Renilde Wolfs).

SUPPLEMENTARY MATERIAL

Expanded Methods. Detailed description of the methods used for the urinary proteomic profiling, including sample preparation, capillary electrophoresis coupled with mass spectrometry, data processing, and peptide sequencing.

Literature Search. Describes how the current study moves beyond what was already known before and highlights the implications of the overall findings.

Table S1. Statistical analysis work flow.

Table S2. Baseline characteristics of 597 participants who underwent a follow-up assessment of estimated glomerular filtration rate (eGFR).

Table S3. Multivariable-adjusted associations of estimated glomerular filtration rate (eGFR) with urinary proteomic biomarkers.

Table S4. Multivariable-adjusted associations of estimated glomerular filtration rate (eGFR) category stage with urinary proteomic biomarkers.

Table S5. Urinary collagen fragments with known amino acid sequence.

Figure S1. Participant flow chart.

Figure S2. Percentage of participants according to the number of detected sequenced peptides.

Supplementary material is linked to the online version of the paper at www.kireports.org.

REFERENCES

1. Wang H, Dwyer-Lindgren L, Lofgren KT, et al. Age-specific and sex-specific mortality in 187 countries, 1970-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2071–2094.
2. Coresh J, Selvin E, Stevens LA, et al. Prevalence of chronic kidney disease in the United States. *JAMA*. 2007;298:2038–2047.
3. Editorial. The Lancet kidney campaign: an opportunity for partnerships. *Lancet*. 2016;387:1038–1039.
4. Klein J, Bascands JL, Mischak H, et al. The role of urinary peptidomics in kidney disease research. *Kidney Intern*. 2016;89:539–545.
5. Sumpio BE, Hayslett JP. Renal handling of proteins in normal and disease states. *Q J Med*. 1985;57:611–635.
6. Jantos-Siwy J, Schiffer E, Brand K, et al. Quantitative urinary proteome analysis for biomarker evaluation in chronic kidney disease. *J Proteome Res*. 2009;8:268–281.
7. Klein J, Papadopoulos T, Mischak H, et al. Comparison of CE-MS/MS and LC-MS/MS sequencing demonstrates significant complementarity in natural peptide identification in human urine. *Electrophoresis*. 2014;35:1060–1064.
8. Argilés A, Siwy J, Duranton F, et al. CKD273, a new proteomics classifier assessing CKD and its prognosis. *PLoS One*. 2013;8:e62837.

9. Critselis E, Lambers Heerspink H. Utility of the CKD273 peptide classifier in predicting chronic kidney disease progression. *Nephrol Dial Transplant*. 2016;31:249–254.
10. Schanstra JP, Zürbig P, Alkhalaf A, et al. Diagnosis and prediction of CKD progression by assessment of urinary peptides. *J Am Soc Nephrol*. 2015;26:1999–2010.
11. Zhang Z, Staessen JA, Thijs L, et al. Left ventricular diastolic dysfunction in relation to the urinary proteome: a proof-of-concept study in a general population. *Int J Cardiol*. 2014;176:158–165.
12. World Medical Association. Declaration of Helsinki. *JAMA*. 2013;227:184–189.
13. Jaffe M. Über den Niederschlag, welchen Pikrinsäure in normalen Harn erzeugt und über eine neue Reaction des Kreatinins. *Z Physiol Chem*. 1886;10:391–400.
14. Peake M, Whiting M. Measurement of serum creatinine—current status and future goals. *Clin Biochem Rev*. 2006;27:173–182.
15. Myers GL, Miller WG, Coresh J, et al. Recommendations for improving serum creatinine measurement: a report from the laboratory working group of the National Kidney Disease Education Program. *Clin Chem*. 2006;52:5–18.
16. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150:604–612.
17. Mancia G, Fagard R, Narkiewicz K, et al. 2013 ESH/ESC Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *Eur Heart J*. 2013;34:2159–2219.
18. KDIGO Board members. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl*. 2017;3:1–150.
19. Fox CS, Larson MG, Leip EP, et al. Predictors of new-onset kidney disease in a community-based population. *JAMA*. 2004;291:844–850.
20. van Blijderveen JC, Straus SM, Zietse R, et al. A population-based study on the prevalence and incidence of chronic kidney disease in the Netherlands. *Int Urol Nephrol*. 2014;46:583–592.
21. Halbesma N, Jansen DF, Heymans MW, et al. Development and validation of a general population renal risk score. *Clin J Am Soc Nephrol*. 2011;6:1731–1738.
22. Ma J, Yang Q, Hwang SJ, et al. Genetic risk score and risk of stage 3 chronic kidney disease. *BMC Nephrol*. 2017;18:32.
23. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabet Care*. 2003;26(suppl 1):S5–S20.
24. Meleth S, Deshane J, Kim H. The case for well-conducted experiments to validate statistical protocols for 2D gels: different pre-processing = different lists of significant proteins. *BMC Biotechnol*. 2005;5:7.
25. Lazar C, Gatto L, Ferro M, et al. Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies. *J Proteome Res*. 2016;15:1116–1125.
26. O'Brien E, Asmar R, Beilin L, et al. European Society of Hypertension recommendations for conventional, ambulatory and home blood pressure measurement. *J Hypertens*. 2003;21:821–848.
27. Blom G. Statistical estimates and transformed beta-variables. *Biomed J*. 1961;3:285.
28. Tobias RD. *An introduction to partial least squares regression*. Cary, NC: SAS Institute Inc.; 1997:1250–1257.
29. Apostolopoulos V, Stojanovska L, Gargosky SE. MUC1 (CD227): a multi-tasked molecule. *Cell Mol Life Sci*. 2015;72:4475–4500.
30. Takahashi T, Takamura K, Sakaue S, et al. Elevated serum KL-6 concentrations in patients with diabetes mellitus. *J Diabetes Complications*. 2002;16:352–358.
31. Leroy X, Copin MC, Devisme L, et al. Expression of human mucin genes in normal kidney and renal cell carcinoma. *Histopathology*. 2002;40:450–457.
32. Cao Y, Karsten U, Zerban H, et al. Expression of MUC1, Thomsen-Friedenreich-related antigens, and cytokeratin 19 in human renal cell carcinomas and tubular clear cell lesions. *Virchows Arch*. 2000;436:119–126.
33. Zhang D, Isaka Y, Imamura R, et al. Glycocalyx damage estimated using colloidal iron staining. *Cell Transplant*. 2008;17:159–163.
34. Eckardt KU, Alper SL, Antignac C, et al. Autosomal dominant tubulointerstitial kidney disease: diagnosis, classification, and management—a KDIGO consensus report. *Kidney Int*. 2015;88:676–683.
35. Hirasawa Y, Kohno N, Yokoyama A, et al. KL-6, a human MUC1 mucin, is chemotactic for human fibroblasts. *Am J Respir Cell Mol Biol*. 1997;17:501–507.
36. Pastor-Soler NM, Sutton TA, Mang HE, et al. Muc1 is protective during kidney ischemia-reperfusion injury. *Am J Physiol Renal Physiol*. 2015;308:F1452–F1462.
37. Kirby A, Gnirke A, Jaffe DB, 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.
38. Ekici AB, Hackenbeck T, Morinière V, et al. Renal fibrosis is the common feature of autosomal dominant tubulointerstitial kidney diseases caused by mutations in mucin 1 or uromodulin. *Kidney Int*. 2014;86:589–599.
39. Bleyer AJ, Kmoch S, Antignac C, et al. Variable clinical presentation of an MUC1 mutation causing medullary cystic kidney disease type 1. *Clin J Am Soc Nephrol*. 2014;9:527–535.
40. Vylet'al P, Kublová M, Kalbácová M, et al. Alterations of uromodulin biology: a common denominator of the genetically heterogeneous FJHN/MCKD syndrome. *Kidney Int*. 2006;70:1155–1169.
41. Maahs DM, Siwy J, Argilés A, et al. Urinary collagen fragments are significantly altered in diabetes: a link to pathophysiology. *PLoS One*. 2010;5:9.
42. Hörstrup JH, Gehrmann M, Schneider B, et al. Elevation of serum and urine levels of TIMP-1 and tenascin in patients

with renal disease. *Nephrol Dial Transplant*. 2002;17:1005–1013.

43. Sörensen I, Susnik N, Inhester T, et al. Fibrinogen, acting as a mitogen for tubulointerstitial fibroblasts, promotes renal fibrosis. *Kidney Int*. 2011;80:1035–1044.
44. Craciun FL, Ajay AK, Hoffmann D, et al. Pharmacological and genetic depletion of fibrinogen protects from kidney fibrosis. *Am J Physiol Renal Physiol*. 2014;307:F471–F484.
45. Øvrehus MA, Zürbig P, Vikse BE, et al. Urinary proteomics in chronic kidney disease: diagnosis and risk of progression beyond albuminuria. *Clin Proteomics*. 2015;12:21.
46. Staessen JA, Roels HA, Emelianov D, et al. Environmental exposure to cadmium, forearm bone-density, and risk of fractures: prospective population study. *Lancet*. 1999;353:1140–1144.
47. Rossing K, Bosselmann HS, Gustafsson F, et al. Urinary proteomics pilot study for biomarker discovery and diagnosis in heart failure with reduced ejection fraction. *PLoS One*. 2016;11:e0157167.
48. Good DM, Zürbig P, Argilés A, et al. Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. *Mol Cell Proteomics*. 2010;9:2424–2437.