Identification of a unique urinary biomarker profile in patients with autosomal dominant polycystic kidney disease

Andreas D. Kistler[1], Harald Mischak[2], Diane Poster[1], Mohammed Dakna[2], Rudolf P. Wüthrich[1] and Andreas L. Serra[1]

[1]*Department of Nephrology, University Hospital, Zürich, Switzerland and* [2]*Mosaiques Diagnostics and Therapeutics AG, Hannover, Germany*

To gain some insight into early disease progression in human autosomal dominant polycystic kidney disease (ADPKD), we analyzed the urine proteome of 41 young patients with ADPKD whose renal function was relatively preserved. Using capillary electrophoresis and mass spectrometry, we compared these results to those from age-matched healthy controls and patients with other renal diseases. There were 197 proteins with significantly altered urinary excretion; and 38 of them could be sequenced, most of which were collagen fragments. This suggests that there is high turnover of extracellular matrix proteins. Uromodulin peptides, previously implicated in tubular injury, were also found in the urine specimens. These marker proteins were found to distinguish patients from controls with a high degree of accuracy. The sensitivity and specificity of this marker set remained high in an independent validation cohort of 24 patients with ADPKD and 35 healthy controls, and even in comparisons of patients with a variety of other renal diseases or patients with kidney or bladder cancer. These findings present a potential hypothesis for the mechanisms of disease progression in ADPKD which will need to be confirmed by further studies.

Kidney International (2009) **76,** 89–96; doi:10.1038/ki.2009.93; published online 1 April 2009

KEYWORDS: autosomal dominant polycystic kidney disease (ADPKD); biomarkers; capillary electrophoresis; mass spectrometry; proteomics urine analysis

Correspondence: *Andreas D. Kistler, Department of Nephrology, University Hospital, Rämistrasse 100, Zürich 8091, Switzerland.*
E-mail: andreas.kistler@usz.ch

Received 23 October 2008; revised 8 January 2009; accepted 17 February 2009; published online 1 April 2009

Autosomal-dominant polycystic kidney disease (ADPKD) is the most frequent hereditary kidney disease with a prevalence between 1:400 and 1:1000,[1,2] and accounts for 7–10% of all patients requiring renal replacement therapy.[3] The continuous development and growth of innumerable cysts in both kidneys starts early in childhood, but most patients remain asymptomatic until their third or fourth decade of life. After the occurrence of the symptoms, such as flank pain and hematuria, the glomerular filtration rate usually starts to decline at a rapid rate leading to end-stage renal disease within a decade.[4] The disease is caused by mutations in the PKD1 (85% of cases) or PKD2 gene (15% of cases). Although these genes were characterized more than a decade ago,[5–7] the precise processes leading to cyst formation and loss of renal function remain incompletely understood. Several mechanisms contributing to cyst formation have been identified, including a disbalance in epithelial cell proliferation and apoptosis, secretory defects, altered cell–matrix interactions, cell polarity, ciliary dysfunction, and altered intracellular signaling (reviewed by Wilson[3]).

In human ADPKD with preserved kidney function, renal tissue is rarely available because kidney biopsy is contraindicated. Consequently, our knowledge of molecular mechanisms in the early course of ADPKD is scarce and mainly based on studies of rodent models for polycystic kidney disease. Although these animal models are of great value, none of them fully displays all the features of human ADPKD. The application of proteomic tools to screen for differentially excreted proteins in the urine represents a noninvasive approach to gain insight into pathophysiological processes and is recently becoming more important.[8] Analysis of body fluids with capillary electrophoresis coupled online to mass spectrometry (CE-MS) allows the simultaneous detection of more than 1000 different proteins and peptides in a sample within a short examination time.[9,10] These proteins and peptides are characterized by their CE migration time and the mass, while the signal intensity gives a measure for their relative abundance.[11] CE-MS of human urine has been used to establish the diagnosis or to predict prognosis of IgA nephropathy,[12] diabetic nephropathy,[13] ureteropelvic

junction obstruction,[14] and prostate and urothelial cancer.[15,16]

The purpose of this investigation was to identify specific urinary protein biomarkers that enable to distinguish ADPKD patients from healthy controls and from patients with other renal diseases. Sequencing of identified biomarkers was carried out to gain insight into ADPKD-specific molecular mechanisms of disease progression.

RESULTS

Clinical characteristics of all ADPKD patients are summarized in Table 1, and demographic data of the control patients in Table 2. The compiled data of the CE-MS analysis of urinary samples from 17 ADPKD and its comparison with 86 samples from age- and sex-matched apparently healthy controls are graphically depicted in Figure 1a and b. To avoid artifacts based on sporadic signals because of differences in food, lifestyle, medication, or bacterial contamination, only peptides that were observed in >50% of either controls or ADPKD patients were examined, all others were excluded. The analysis of this data set for significant differences of single biomarkers (adjusted *P*-value <0,05) resulted in the identification of 197 potential urinary biomarkers, which are listed in Supplementary Table S1. As the number of potential biomarkers by far exceeded the number of samples studied, we defined the most consistent biomarkers by an iterative algorithm leaving out 30% of cases and controls from the analysis at random. This procedure was repeated 10 times, and only biomarkers that were found to be significant in at least 7 of 10 repeats were accepted, resulting in a reduction to 38 biomarkers, which are listed in Table 3. The distribution of the urinary biomarkers that enabled the discrimination between ADPKD and the control groups is shown in Figure 1c and d.

A support vector machine (SVM)-based model was created by combining these 38 biomarkers and including additional controls to enable high specificity. This model applied to the 17 cases and 86 controls in the training set discriminated ADPKD from controls with 100% sensitivity and 98.8% specificity. Upon complete take-one-out cross-validation, sensitivity was 94.1% and specificity was 98.8% (area under the receiver operating characteristic curve (AUC): 0.998).

The support vector machine-based biomarker model established in the training set was subsequently validated in an independent masked data set of 24 ADPKD patients and 35 healthy controls. To avoid the introduction of bias because of differences in the training and the validation set, these two groups were selected to be similar with respect to clinical and demographic parameters. The analysis revealed a sensitivity of 87.5% and a specificity of 97.5% (AUC: 0.95), similar to the cross-validation of the training set. The ROC curves of both training and validation sets are shown in Figure 2.

Table 2 | Demographic characteristics of control patients

	N	Age, years (mean ± s.d.)	Gender (% female)
Normal controls			
Training set	86	30 ± 10	52
Validation set	35	30 ± 5	60
Additional controls	189	37 ± 15	49
Elderly patients	127	68 ± 10	55
Chronic kidney diseases	150	45 ± 19	26
Renal cell carcinoma	113	69 ± 10	53
Bladder carcinoma	112	64 ± 12	23

Table 1 | Clinical characteristics of patients with ADPKD

	Training set	Validation set	All
n	17	24	41
Demographic characteristics			
Male, *n* (%)	10 (48.8)	16 (66.7)	26 (63.4)
Female, *n* (%)	7 (41.2)	8 (33.3)	15 (36.6)
Age, years	31.4 ± 6.8	31.1 ± 5.1	31.2 ± 5.8
SBP (mm Hg)	135.7 ± 17.9	133.9 ± 15.7	134.7 ± 16.5
DBP (mm Hg)	83.2 ± 10.8	86.8 ± 8.5	85.3 ± 9.5
Complications of ADPKD			
Arterial hypertension, *n* (%)	13 (76.5)	18 (75)	31 (75.6)
Antihypertensive treatment, *n* (%)	7 (41.2)	10 (41.7)	17 (41.5)
ACEi and/or ARB therapy, *n* (%)	7 (41.2)	10 (41.7)	17 (41.5)
Recurrent flank pain, *n* (%)	9 (53)	7 (29.2)	16 (39.0)
History of macrohematuria, *n* (%)	1 (5.9)	4 (16.7)	5 (12.2)
History of UTI, *n* (%)	2 (11.8)	1 (4.2)	3 (7.3)
History of ICB, *n* (%)	0	0	0
Renal parameters			
Serum creatinine (μmol/l)	85.8 ± 16.6	88.0 ± 14.2	87.1 ± 15.1
Creatinine clearance (C-G; ml/min)	117.7 ± 21.3	111.9 ± 20.3	114.3 ± 20.6
Urinary protein/creatinine (g/mmol)	0.010 ± 0.005	0.008 ± 0.004	0.009 ± 0.004
Total kidney volume (cm^3)	1038 ± 673	1063 ± 640	1053 ± 645

ACEi, angiotensin-converting enzyme inhibitor; ADPKD, autosomal-dominant polycystic kidney disease; ARB, angiotensin receptor blocker; C-G, estimated according to Cockcroft–Gault; DBP, diastolic blood pressure; ICB, intracranial bleeding; SBP, systolic blood pressure; UTI, urinary tract infection.
Data are mean ± s.d. if not otherwise depicted.

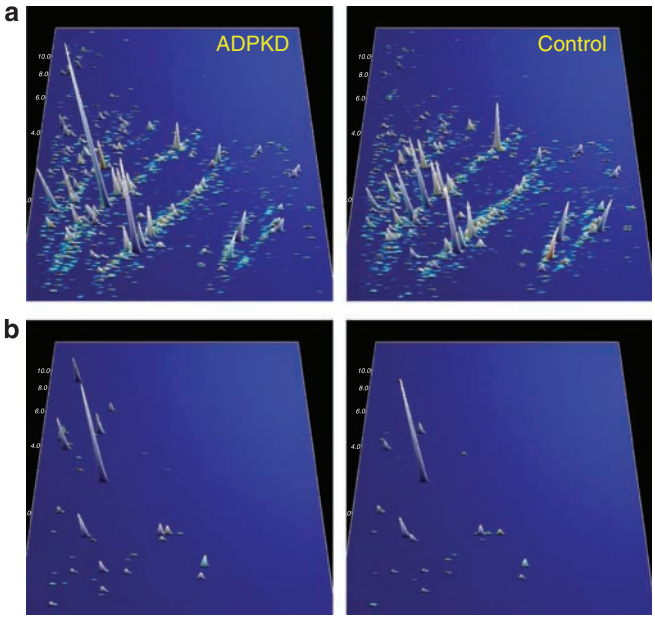

Figure 1 | Urine proteomic pattern in ADPKD patients and healthy individuals. (**a**) Compiled (average) protein patterns of all urine samples from patients with ADPKD ($n = 17$; left panel) and the control group in the training set ($n = 86$; right panel). The molecular mass (0.8–15 kDa, on a logarithmic scale) is plotted against normalized migration time (18–45 min). Signal intensity is encoded by peak height and color. (**b**) Distribution of the 38 potential diagnostic biomarkers for ADPKD. All statistically significant biomarkers from Table 3 are shown.

To further evaluate the specificity of the ADPKD urinary proteomic pattern, the model was additionally tested in normal controls (16 of 189 positive, 91.5% specificity, AUC: 0.923), patients with different chronic renal diseases (7 of 150 positive, 95.3% specificity, AUC: 0.955), with bladder cancer (17 of 112 positive, 84.4% specificity, AUC: 0.919), with renal cell cancer (21 of 113 positive, 81.4% specificity, AUC: 0.895), and healthy individuals aged >60 years (40 of 127 positive, 68.5% specificity, AUC: 0.838). When evaluating all data sets combined (24 cases from the validation cohort and 691 controls), the AUC was 0.910. These results reveal a high specificity, which is reduced in individuals of advanced age. It can be noted that both groups of patients with renal cell cancer and bladder cancer, in which the specificity of the model was slightly reduced, were aged >60 years.

To test, whether disease severity or antihypertensive treatment influences the accuracy of the biomarker model, all ADPKD patients from the training and the validation sets were combined. Their score in the 38 biomarker model did not correlate with total kidney volume, age-adjusted total kidney volume, proteinuria, or medication with either angiotensin-converting enzyme inhibitors or angiotensin receptor blockers (data not shown). Thus, the biomarkers that have been included in the model are relatively robust and not affected by these covariates.

We were able to obtain sequences from 38 of the 197 potential biomarkers, as shown in Table 3. The 38 sequenced peptides represented 13 fragments of collagen α1(I) (6 upregulated and 7 downregulated), one fragment of collagen α2(I) (upregulated), 11 fragments of collagen α1(III) (8 upregulated and 3 downregulated), one fragment of collagen α1(XVIII) (upregulated), 6 fragments of uromodulin (2 upregulated and 4 downregulated), 2 fragments of α1-antitrypsin (both upregulated), and one fragment of each cystatin B (upregulated), polymeric immunoglobulin receptor (upregulated), Na^+/K^+–ATPase γ-chain (upregulated), and ProSAAS (downregulated).

Of all sequenced biomarkers, 6 were present among the 38 biomarkers that were used for classification. We subsequently examined an ADPKD-specific biomarker model based on only these six sequenced peptides. Their combination enabled classification of the training set with a sensitivity of 100% and a specificity of 95.5%. Upon complete cross-validation, a sensitivity of 94.1% and a specificity of 94.2% (AUC: 0.97) could be obtained. This biomarker model was again tested using the same independent masked validation set as above. Here, a sensitivity of 91.7% and a specificity of 80.0% could be obtained (AUC: 0.89). These results indicate a high diagnostic value of these six urinary biomarkers. However, the additional, yet-unidentified 32 biomarkers enhance the performance of the classification.

DISCUSSION

We have identified and validated a distinct pattern of polypeptides excreted in the urine that characterizes patients with ADPKD. This is the first study that shows a unique urine proteome profile, which distinguishes ADPKD patients from healthy controls and from patients with various other forms of renal diseases or pathologies of the excretory urinary tract.

In several recently published reports on biomarkers for different renal diseases, the potential biomarkers were mostly defined on small cohorts of patients, and no statistical correction for multiple testing was used (see Candiano *et al.*,[17] Zhou *et al.*,[18] Varghese *et al.*,[19] and Weissinger *et al.*[20]). As a consequence, most of the initially identified potential biomarkers failed in an independent validation set, if validation was even performed. These shortcomings have resulted in the proposition for guidelines for clinical proteomics.[21] Therefore, and in accordance with these guidelines, we have used a large number of data sets, statistics adjusted for multiple testing, and a masked independent validation set to define urinary biomarkers for ADPKD.

A biomarker model was developed that distinguishes ADPKD patients from age-matched healthy controls with a high degree of reliability. Validation in an independent cohort confirmed the high sensitivity and high specificity of the model. We have included only relatively young ADPKD patients with well-preserved renal function. Furthermore, the urinary proteomic model was capable to distinguish between ADPKD and a large number of different renal diseases with a high degree of specificity. Therefore, the identified urinary markers likely reflect ADPKD-specific pathophysiological

Table 3 | Potential biomarkers used in the SVM-based model

Peptide			ADPKD		Control		*P*-value		ROC		
ID	Mass	CE time	Frequency	Amplitude	Frequency	Amplitude	Unadjusted	BH	AUC	Sequence	Protein
4976	902.41	20.85	1	2.97 (3.02)	0.82	2.08 (2.59)	4.20E−10	2.23E−08	0.8570451		
11705	984.46	24.92	0.76	1.66 (2.17)	0.22	0.40 (1.85)	5.30E−05	8.67E−04	0.8269494		
13188	1013.37	25.17	1	3.09 (3.16)	0.93	2.66 (2.88)	3.27E−04	3.56E−03	0.7250342		
16320	1075.49	20.61	0.88	2.20 (2.47)	0.22	0.43 (1.93)	2.52E−06	6.31E−05	0.9015048		
16875	1082.52	21.69	0.76	1.57 (1.94)	0.34	0.65 (1.86)	1.66E−03	1.25E−02	0.7346101		
18939	1114.48	24.21	0.82	1.83 (2.14)	0.39	0.87 (2.24)	5.06E−03	2.77E−02	0.6928865		
19773	1128.39	33.59	1	3.12 (3.06)	0.92	2.77 (3.01)	3.44E−03	2.03E−02	0.6155951		
23224	1178.39	20.71	0.59	1.40 (2.38)	0.21	0.48 (2.26)	6.77E−03	3.41E−02	0.7024624		
23360	1179.52	27.11	1	3.09 (3.11)	0.86	2.39 (2.81)	1.07E−08	4.56E−07	0.8720930		
23697	1186.53	22.39	0.94	2.76 (2.88)	0.89	2.54 (2.88)	6.23E−03	3.17E−02	0.7455540	DDGEAG KpGRpG	Collagen alpha-1 (I) chain [231–242]
24502	1200.54	25.03	0.53	1.70 (3.86)	0.15	0.40 (2.07)	1.01E−02	4.63E−02	0.6949384		
28103	1257.44	33.92	1	3.40 (3.39)	0.97	3.09 (3.25)	2.19E−03	1.52E−02	0.6839945		
32471	1326.56	27.11	0.12	0.36 (3.05)	0.64	1.81 (2.77)	2.32E−05	4.30E−04	0.7335841		
34017	1354.64	22.14	0.71	1.60 (2.24)	0.14	0.25 (1.78)	8.28E−05	1.19E−03	0.8283174		
35535	1383.59	27.63	1	2.60 (2.65)	0.63	1.53 (2.37)	7.82E−12	5.11E−10	0.8112175	GSpGGpGS DGKpGPpG	Collagen alpha-1 (III) chain [540–555]
48176	1580.88	23.87	0.41	1.14 (2.81)	0.85	2.64 (3.14)	7.06E−04	6.39E−03	0.7735978	IDQSRVL NLGPITR	Uromodulin [593–606]
48580	1588.71	30.15	1	2.88 (2.90)	0.51	1.07 (2.12)	1.61E−06	4.55E−05	0.9227086	TGLSMDGG GSPKGDVDP	Na/K-ATPase gamma chain [2–18]
52189	1640.58	23.24	1	3.69 (3.67)	0.92	3.31 (3.65)	2.00E−03	1.44E−02	0.6067031		
54525	1680.75	30.03	0.94	3.22 (3.41)	0.7	2.43 (3.50)	4.40E−03	2.46E−02	0.5153899	GLpGTGGP pGENGK pGEp	Collagen alpha-1 (III) chain [642–659]
54688	1684.67	30.66	0.88	2.92 (3.37)	0.25	0.83 (3.30)	1.66E−07	5.87E−06	0.8324213		
54687	1684.67	31.75	0.24	0.62 (2.58)	0.74	2.28 (3.17)	5.81E−05	9.32E−04	0.7602599		
54703	1684.71	29.65	0.12	0.26 (2.20)	0.57	1.43 (2.50)	2.69E−05	4.67E−04	0.7356361		
56493	1716.66	20.18	0.71	1.81 (2.60)	0.54	1.36 (2.46)	5.41E−03	2.90E−02	0.7127223		
73434	2067.82	20.62	1	3.04 (3.12)	0.89	2.73 (3.16)	5.83E−03	3.02E−02	0.5916553		
111001	2825.27	24.49	1	4.18 (4.21)	0.98	4.12 (4.22)	2.20E−03	1.52E−02	0.5745554	ERGEAGIpG VpGAKGE DGKDGSpG EpGANG	Collagen alpha-1 (III) chain [448–477]
124193	3166.27	22.06	0.71	1.96 (2.85)	0.29	0.64 (2.13)	8.73E−04	7.42E−03	0.7831737		
125103	3202.43	30.6	0.47	1.09 (2.41)	0.07	0.14 (2.00)	9.06E−03	4.28E−02	0.6792066		
139455	3630.44	21.78	0.71	1.83 (2.57)	0.24	0.53 (2.32)	6.10E−04	6.03E−03	0.7821477		
143344	3774.81	28.23	0.24	0.64 (2.77)	0.51	1.38 (2.75)	7.26E−03	3.61E−02	0.7551300		
148086	3986.65	20.6	1	3.69 (3.72)	0.71	2.25 (3.24)	9.68E−13	7.48E−11	0.8734610		
149624	4043.64	20.39	0.94	2.83 (3.10)	0.44	1.20 (2.70)	6.60E−07	1.93E−05	0.8098495		
153832	4196.75	20.84	0.71	1.89 (2.78)	0.32	0.72 (2.28)	2.17E−03	1.52E−02	0.7619699		
159396	4409.89	20	0.82	2.56 (3.19)	0.38	1.02 (2.73)	2.16E−04	2.58E−03	0.7835157		
165781	4671.82	23.28	0.82	2.12 (2.65)	0.66	1.57 (2.45)	1.15E−02	4.95E−02	0.6853625		
167786	4771.07	20.2	0.76	2.38 (3.08)	0.3	0.76 (2.51)	2.71E−04	3.11E−03	0.8043776		
168314	4799.96	23.81	1	3.48 (3.53)	0.87	2.85 (3.24)	1.39E−05	2.81E−04	0.7387141		
177664	6169.57	24.77	0.76	2.46 (3.11)	0.18	0.47 (2.67)	2.38E−05	4.31E−04	0.8467852		
184058	8837.41	21.06	1	3.48 (3.68)	0.41	1.18 (2.86)	1.28E−21	5.46E−19	0.9350205		

ADPKD, autosomal-dominant polycystic kidney disease; AUC, area under the curve; BH, adjusted according to Benjamini and Hochberg; CE, capillary electrophoresis; ROC, receiver operating characteristic; SVM, support vector machine.
Given are the internal ID, mass, capillary electrophoresis migration time, frequency (i.e., the proportion of patients in whose urine the denoted peptide was detectable), and mean (s.d.) amplitude (i.e., the peak intensity of the denoted peptide in the mass spectrum) in ADPKD patients and controls, *P*-value for the comparison of cases and controls (unadjusted and adjusted according to the Benjamini and Hochberg[49]), ROC value in the training set, and, if applicable, sequence, parental protein, and position of the peptide within the parental protein sequence.

processes associated with cyst formation and expansion, rather than just an impaired renal function. The score derived from the 38 biomarker model did not correlate with clinical markers of disease severity or treatment with angiotensin-converting enzyme inhibitors or angiotensin II receptor blockers, indicating that the performance of the model is not affected by these covariates.

Although the diagnosis of ADPKD is in most cases easily established based on an age-dependent cystic renal phenotype and a positive family history,[22] current markers predicting disease progression are less reliable. There is considerable inter- and intrafamilial variability in the rate of progression to kidney failure, and renal prognosis often cannot be determined by the individual's phenotype during early stages of the disease.[23] In light of recently emerging therapeutic options for ADPKD, it will be of paramount importance to select patients with rapidly progressive disease for treatment and avoid exposing patients with mild disease

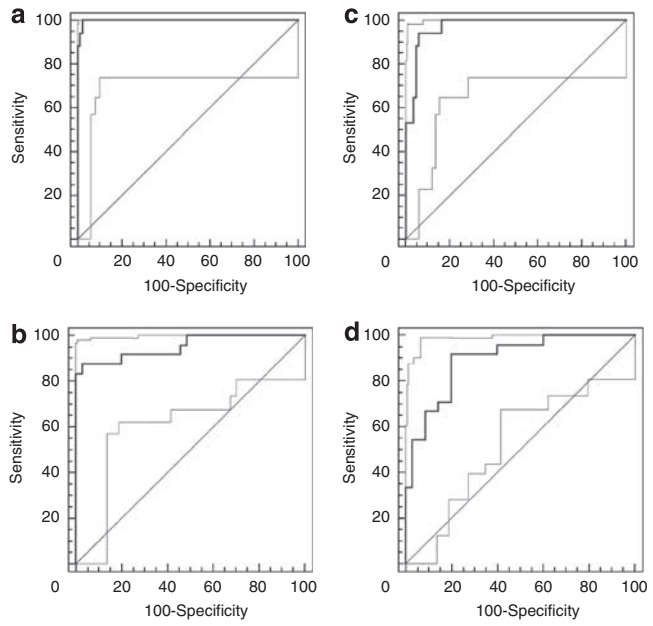

Figure 2 | Performance of the biomarker models using ROC analysis (with 95% CI). (**a**) Classification results in the cross-validations based on 38 biomarkers specific for ADPKD in the training set consisting of 17 cases and 86 controls. (**b**) Classification results based on the same 38 biomarkers in the validation set consisting of 59 blinded samples (24 cases and 35 controls). (**c**) Classification results in the cross-validations based on the six sequenced biomarkers specific for ADPKD in the training set. (**d**) Classification results based on the same six biomarkers in the validation set.

to expensive therapies with potential side effects. Our finding of an ADPKD-specific urinary proteomic pattern suggests that the progression rate of the disease might also be reflected by the urinary proteome. The scores of the reported biomarker model did not correlate with age-adjusted total kidney volume, because these biomarkers have been selected for their high consistency among different ADPKD patients and therefore unlikely reflect disease severity. To define a prognostic model, the complete proteomic patterns derived by CE-MS will have to be correlated with longitudinal data on kidney volume growth in ADPKD patients. However, as proteomic differences among patients with ADPKD are likely to be smaller than those between ADPKD and control patients, a larger sample size will be required to define reliable progression markers.

The data presented here and elsewhere (Rossing *et al.*,[13] Decramer *et al.*,[14] and Julian *et al.*[24]) indicate that a combination of several markers to a model results in a much higher test accuracy than the use of single biomarkers. The capability to simultaneously detect a large number of small proteins and peptides represents a major advantage of CE-MS. It is virtually impossible to quantify all the 38 peptides of our model appropriately by alternative means. As pointed out recently,[25,26] it is unlikely that the biomarkers can be accurately displayed by an immunological assay, as such assays generally do not enable assessment of post-translational modifications (including processing of C and N termini). Alternative techniques to display the identified biomarkers would include multiple reaction monitoring or immunocapture, followed by matrix-assisted laser desorption/ionization-time of flight mass spectrometry. However, the establishment of such assays is still a considerable challenge, and may not be successful. Furthermore, both approaches would in addition require the assessment of internal calibrants.[26] Therefore, and as CE-MS is a rapid and robust method, we propose the clinical use of CE-MS (in case the above-mentioned prognostic biomarkers can be defined) rather than the development of an alternative test for the detection of the described fragments.

We were able to determine the amino acid sequence of 38 of the total 197 differentially excreted polypeptides. As outlined in detail in several recent reports, although sequencing of tryptic digests can be considered routine, sequencing of naturally occurring peptides still represents a challenge that often cannot be met, to a large degree, likely because of post-translational modifications.[27,28] Most of the identified peptides represent collagen type I or type III fragments. The expansion of cysts in ADPKD requires adaptive changes of the extracellular matrix (ECM), and several studies on polycystic kidney disease have revealed ECM abnormalities.[29,30] Serum levels of matrix metalloproteinases have been found to be increased in ADPKD patients,[31] and the matrix metalloproteinase inhibitor, batimastat, reduced cyst growth in (cy/+) rats, a rodent model of polycystic kidney disease.[32] We could recently show that sirolimus, which was effective to slow cyst growth and preserve renal function in (cy/+) rats, did also ameliorate the enhanced expression of matrix metalloproteinases in these animals.[33] Our finding that many collagen fragments are differentially excreted in ADPKD underscores the importance of ECM remodeling during disease progression. Not only upregulation but also downregulation of certain collagen fragments is observed in ADPKD. However, the majority of collagen fragments are upregulated, indicating increased protease activity. It is tempting to speculate that disease-specific changes in protease activity are responsible for this observation.

It is interesting that, the proteomic pattern of young ADPKD patients overlapped with that of apparently healthy elderly patients. The aging of the kidney is associated with an extensive remodeling of the ECM as reflected by the increased expression of ECM components.[34–36] Although the significance of ECM turnover during aging is not clear, these processes do likely contribute to interstitial fibrosis in the aging kidney. Our finding might describe a common pathway in ADPKD and in aging of the kidney.

Several uromodulin fragments were also differentially excreted in the urine of ADPKD patients. Uromodulin, which is identical to Tamm–Horsfall protein, is the most abundant protein in the urine of healthy individuals. Urinary secretion occurs by proteolytic cleavage of the large N-terminal ectodomain from the glycosylphosphatidylinositol-anchored counterpart exposed at the luminal cell surface of the thick

ascending limb of Henle's loop.[37] Although the physiological role of uromodulin is uncertain, it is believed to be involved in the pathogenesis of various renal tubular disorders, such as cast nephropathy, urolithiasis, and tubulointerstitial nephritis (reviewed by Serafini-Cessi *et al.*[38]). Moreover, reduced urinary Tamm–Horsfall protein excretion is considered a reliable index of distal tubular cell damage.[39–41] A recent study showed that familial juvenile hyperuricemic nephropathy and autosomal-dominant medullary cystic kidney disease 2 arise from mutations in the uromodulin gene.[42] There has been no report on altered expression, processing, or urinary excretion of uromodulin in ADPKD so far. In can be noted that, all fragments differentially excreted in ADPKD are located in the C-terminal of the putative cleavage site, thus representing further proteolytic processing of the glycosylphosphatidylinositol-anchored residual protein moiety after proteolytic release of the N-terminal ectodomain. The amount of intact N-terminal ectodomain excreted in the urine of ADPKD patients cannot be estimated on the basis of our data because of its large size and the limitation of CE-MS to small-molecular-weight proteins.

Taken together, we were able to identify a urinary proteomic 'footprint' of ADPKD, which distinguishes the disease from health and from other renal diseases with a high degree of reliability. The ADPKD-specific urinary proteomic pattern most likely mirrors alterations in the ECM, which occur early in the course of the disease when the excretory renal function is still preserved. Further studies are needed to elucidate the pathophysiological significance of the alterations in the excretion of collagen and uromodulin fragments and to establish prognostic biomarkers for ADPKD.

MATERIALS AND METHODS

Patients and procedures

The ADPKD subjects for this study were participants of an earlier described registry.[43] Patients affected by ADPKD, aged between 19 and 41 years with preserved renal function (estimated creatinine clearance $\geqslant$70 ml/min according to the Cockcroft–Gault formula) were included. The diagnosis of ADPKD was based on a positive family history and ultrasonographic diagnostic criteria.[22] A detailed medical history was obtained, including ADPKD-related symptoms, previous hospitalization, and medication. Urine samples were collected in the morning, after voiding the first urine of the day. Blood pressure measurements and blood analyses were performed as described earlier.[43] In addition, all ADPKD patients underwent renal magnetic resonance imaging without contrast media. Kidney volumes were determined from magnetic resonance imaging sequences with high accuracy and reliability as reported earlier.[43]

Urine samples of 17 ADPKD patients and 86 age- and sex-matched apparently healthy individuals (mean ± s.d. age 30.4 ± 4.9 years; range 18–40; 47.7% males) were analyzed to identify an ADPKD-specific proteomic pattern. The identified pattern was validated by analyzing urine samples from another 24 ADPKD patients and 35 healthy controls (mean ± s.d. age 30.5 ± 9.6 years; range 18–40; 31.4% males).

Additional urine samples from healthy controls ($n=189$), elderly healthy controls ($n=127$), patients with chronic kidney disease ($n=150$; 42 IgA nephropathy; 32 membranous glomerulonephritis; 30 diabetic nephropathy; 27 focal segmental glomerulosclerosis; and 17 ANCA vasculitis) and from patients with bladder ($n=112$) or renal cell cancer ($n=113$) were collected as described earlier.[13,16,24,44] Shortly, second or third urine (mid-stream) of the morning was collected and frozen within 1 h below −20°C, as suggested in recently published recommendations.[45] Also, following these recommendations, no protease inhibitors were added. The clinical data on the additional controls were described in several recent manuscripts.[12,16,20,24] Their demographic data are summarized in Table 2.

Informed consent was obtained from all patients and healthy controls after local ethics committee approval. These studies were performed in accordance with the Declaration of Helsinki Principles.

Sample preparation and CE-MS analysis

All urine samples for CE-MS analyses were stored at −20°C until analysis and underwent a maximum of two freeze/thaw cycles. We found no significant differences when samples went through up to three freeze/thaw cycles (Mischak, unpublished observation). A 0.7-ml aliquot was thawed immediately before use and diluted with 0.7 ml of 2 M urea, 10 mM NH_4OH containing 0.02% sodium dodecyl sulfate. To remove higher molecular mass proteins, the sample was filtered using Centrisart ultracentrifugation filter devices (20 kDa molecular weight cut-off (MWCO); Sartorius, Goettingen, Germany) at 3000 *g* until 1.1 ml of filtrate was obtained. This filtrate was then applied onto a PD-10 desalting column (Amersham Bioscience, Uppsala, Sweden) equilibrated in 0.01% NH_4OH in high-performance liquid chromatography-grade H_2O (Roth, Karlsruhe, Germany) to decrease matrix effects by removing urea, electrolytes, salts, and to enrich polypeptides. Finally, all samples were lyophilized, stored at 4 °C, and suspended in high-performance liquid chromatography-grade H_2O shortly before CE-MS analysis, as described.[16]

CE-MS analysis was performed as described[16,46] using a P/angiotensin-converting enzyme MDQ Capillary Electrophoresis System (Beckman Coulter, Fullerton, CA, USA) on-line coupled to a Micro-TOF MS (Bruker Daltonic, Bremen, Germany). The ESI sprayer (Agilent Technologies, Palo Alto, CA, USA) was grounded, and the ion-spray interface potential was set between −4 and −4.5 kV. Data acquisition and MS acquisition methods were automatically controlled by the CE through contact-close-relays. Spectra were accumulated every 3 s, over a range of *m/z* 350–3000. Accuracy, precision, selectivity, sensitivity, reproducibility, and stability of the analytical platform were indicated and described in great detail elsewhere.[16]

Data processing and cluster analysis

Mass spectral ion peaks representing identical molecules at different charge states were deconvoluted into single masses using the MosaiquesVisu software.[47] Migration time and ion signal intensity (amplitude) were normalized using internal polypeptide standards,[15] which seem to be unaffected by any disease state studied to date.[27] The resulting peak list characterizes each polypeptide by its molecular mass (kDa), normalized migration time (min), and normalized signal intensity. All detected polypeptides were deposited, matched, and annotated in a Microsoft SQL database, allowing comparison of multiple samples (patient groups). To determine whether polypeptides observed in different samples were identical, a linear function was used that provided an allowed 50 p.p.m. absolute mass deviation for peptides of 800 Da, which increased linearly to 100 p.p.m. for peptides with a maximum mass of 20 kDa. A similar linear function was used when comparing CE migration times, with

small peptides being allowed a 2% and higher mass peptides a 5% absolute time deviation to account for the decreased resolution of large species. CE-MS data of all individual samples can be accessed in Supplementary Table 2.

Disease-specific polypeptide patterns were generated using support vector machine-based MosaCluster software.[14]

Statistical methods, definition of biomarkers, and sample classification

Sensitivity and specificity were calculated on the basis of tabulating the number of correctly classified samples. Confidence intervals (95% CI) based on exact binomial calculations were carried out in MedCalc (version 8.1.1.0, MedCalc Software, Mariakerke, Belgium; http://www.medcalc.be). The receiver operating characteristic (ROC) plot was obtained by plotting all sensitivity values on the *y* axis against their equivalent 1-specificity values on the *x* axis for all available thresholds (MedCalc Software). The AUC was evaluated, as it provides a single measure of overall accuracy that is not dependent on a particular threshold.[48] The reported unadjusted *P*-values were calculated using the natural logarithm-transformed intensities and the Gaussian approximation to the *t*-distribution. Statistical adjustment for multiple testing was carried out by the method described by Benjamini and Hochberg.[49]

Sequencing of polypeptides

Candidate biomarkers were sequenced using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) analysis as recently described in detail.[50]. Further analysis was performed using instruments with electron transfer dissociation (ETD) capability.[51–53] Immediately before the analysis, samples were re-suspended (50 µl in 100 mM acetic acid), bomb-loaded onto a 360×75 µm microcapillary pre-column, and connected to a 360×50 µm analytical column with an $\sim$1 µm tip pulled with a laser puller (both columns were packed in-house with approximately 5–8 cm of C18 resin). Peptides were separated by nanoflow reversed-phase high-performance liquid chromatography (Agilent 1100; flow split by tee to $\sim$60nl/min) and introduced into either an ETD-enabled Thermo Fisher linear trap quadrupole (LTQ) ion trap (Thermo Fisher Scientific, San Jose, CA, USA) or Thermo Fisher LTQ-Orbitrap (Thermo Fisher Scientific, Bremen, Germany) mass spectrometer, also capable of ETD analysis, through nanoelectrospray ionization (nano-ESI). Samples were analyzed using the LTQ in a data-dependent manner, with the five most abundant species subjected to both ETD and collision-activated dissociation (CAD) fragmentation (in separate alternating scans). Ion trap instrumental parameters most conducive to achieving the highest fragmentation efficiencies for both CAD and ETD were used as described recently.[54] Spectral data was converted into .dta files (DTA's) and searched against the IPI human non-redundant database using the Open Mass Spectrometry Search Algorithm (OMSSA; free from NCBI, http://pubchem.ncbi.nlm.nih.gov/omssa/); *e*-value cutoff was 0.01. All matched sequences were manually validated.

DISCLOSURE

HM is the founder and co-owner of Mosaiques Diagnostics, who developed the CE-MS technology. MD is an employee of Mosaiques Diagnostics.

ACKNOWLEDGMENTS

This study was supported by the Swiss National Science Foundation (No 310000-118166), by the PKD Foundation (Kansas, MO, USA), and by an unrestricted research grant by Wyeth Europe. We thank Julia Hofmann, Ruth Russi, and Marian Struker for the excellent technical assistance. HM was supported in part by EUROTRANS-BIO Grant ETB-2006-016.

SUPPLEMENTARY MATERIAL

Table S1. List of 197 potential biomarkers found to be significantly different between ADPKD and controls.

Table S2. Pivot Table consisting of 1 worksheet that includes all CE-MS data for the samples used in the study for biomarker definition. Supplementary material is linked to the online version of the paper at http://www.nature.com/ki

REFERENCES

1. Dalgaard OZ. Bilateral polycystic disease of the kidneys; a follow-up of two hundred and eighty-four patients and their families. *Acta Med Scand Suppl* 1957; **328**: 1–255.
2. Iglesias CG, Torres VE, Offord KP *et al.* Epidemiology of adult polycystic kidney disease, Olmsted County, Minnesota: 1935–1980. *Am J Kidney Dis* 1983; **2**: 630–639.
3. Wilson PD. Polycystic kidney disease. *N Engl J Med* 2004; **350**: 151–164.
4. Hateboer N, v Dijk MA, Bogdanova N *et al.* Comparison of phenotypes of polycystic kidney disease types 1 and 2. European PKD1-PKD2 Study Group. *Lancet* 1999; **353**: 103–107.
5. The European Polycystic Kidney Disease Consortium. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* 1994; **78**: 725.
6. The International Polycystic Kidney Disease Consortium. Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. *Cell* 1995; **81**: 289–298.
7. Mochizuki T, Wu G, Hayashi T *et al.* PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* 1996; **272**: 1339–1342.
8. Fliser D, Novak J, Thongboonkerd V *et al.* Advances in urinary proteome analysis and biomarker discovery. *J Am Soc Nephrol* 2007; **18**: 1057–1071.
9. Fliser D, Wittke S, Mischak H. Capillary electrophoresis coupled to mass spectrometry for clinical diagnostic purposes. *Electrophoresis* 2005; **26**: 2708–2716.
10. Weissinger EM, Hertenstein B, Mischak H *et al.* Online coupling of capillary electrophoresis with mass spectrometry for the identification of biomarkers for clinical diagnosis. *Expert Rev Proteomics* 2005; **2**: 639–647.
11. Kaiser T, Wittke S, Just I *et al.* Capillary electrophoresis coupled to mass spectrometer for automated and robust polypeptide determination in body fluids for clinical use. *Electrophoresis* 2004; **25**: 2044–2055.
12. Haubitz M, Wittke S, Weissinger EM *et al.* Urine protein patterns can serve as diagnostic tools in patients with IgA nephropathy. *Kidney Int* 2005; **67**: 2313–2320.
13. Rossing K, Mischak H, Dakna M *et al.* Urinary proteomics in diabetes and CKD. *J Am Soc Nephrol* 2008; **19**: 1283–1290.
14. Decramer S, Wittke S, Mischak H *et al.* Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis. *Nat Med* 2006; **12**: 398–400.
15. Theodorescu D, Fliser D, Wittke S *et al.* Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine. *Electrophoresis* 2005; **26**: 2797–2808.
16. Theodorescu D, Wittke S, Ross MM *et al.* Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis. *Lancet Oncol* 2006; **7**: 230–240.
17. Candiano G, Musante L, Bruschi M *et al.* Repetitive fragmentation products of albumin and alpha1-antitrypsin in glomerular diseases associated with nephrotic syndrome. *J Am Soc Nephrol* 2006; **17**: 3139–3148.
18. Zhou H, Pisitkun T, Aponte A *et al.* Exosomal Fetuin-A identified by proteomics: a novel urinary biomarker for detecting acute kidney injury. *Kidney Int* 2006; **70**: 1847–1857.
19. Varghese SA, Powell TB, Budisavljevic MN *et al.* Urine biomarkers predict the cause of glomerular disease. *J Am Soc Nephrol* 2007; **18**: 913–922.
20. Weissinger EM, Wittke S, Kaiser T *et al.* Proteomic patterns established with capillary electrophoresis and mass spectrometry for diagnostic purposes. *Kidney Int* 2004; **65**: 2426–2434.
21. Mischak H, Apweiler R, Banks R *et al.* Clinical Proteomics: a need to define the field and to begin to set adequate standards. *Proteomics Clin Appl* 2007; **1**: 148–156.

22. Ravine D, Gibson RN, Walker RG *et al.* Evaluation of ultrasonographic diagnostic criteria for autosomal dominant polycystic kidney disease 1. *Lancet* 1994; **343**: 824–827.
23. Reed BY, McFann K, Reza Bekheirnia M *et al.* Variation in age at ESRD in autosomal dominant polycystic kidney disease. *Am J Kidney Dis* 2008; **51**: 173–183.
24. Julian BA, Wittke S, Novak J *et al.* Electrophoretic methods for analysis of urinary polypeptides in IgA-associated renal diseases. *Electrophoresis* 2007; **28**: 4469–4483.
25. Good DM, Thongboonkerd V, Novak J *et al.* Body fluid proteomics for biomarker discovery: lessons from the past hold the key to success in the future. *J Proteome Res* 2007; **6**: 4549–4555.
26. Jantos-Siwy J, Schiffer E, Brand K *et al.* Quantitative urinary proteome analysis for biomarker evaluation in chronic kidney disease. *J Proteome Res* 2008.
27. Coon J, Zurbig P, Dakna M *et al.* CE-MS analysis of the human urinary proteome for biomarker discovery and disease diagnostics. *Proteomics Clin Appl* 2008 (in press).
28. Mischak H, Coon JJ, Novak J *et al.* Capillary electrophoresis-mass spectrometry as a powerful tool in biomarker discovery and clinical diagnosis: an update of recent developments. *Mass Spectrom Rev* 2008.
29. Ebihara I, Nakamura T, Takahashi T *et al.* Altered extracellular matrix component gene expression in murine polycystic kidney. *Ren Physiol Biochem* 1995; **18**: 73–80.
30. Ramasubbu K, Gretz N, Bachmann S. Increased epithelial cell proliferation and abnormal extracellular matrix in rat polycystic kidney disease. *J Am Soc Nephrol* 1998; **9**: 937–945.
31. Nakamura T, Ushiyama C, Suzuki S *et al.* Elevation of serum levels of metalloproteinase-1, tissue inhibitor of metalloproteinase-1 and type IV collagen, and plasma levels of metalloproteinase-9 in polycystic kidney disease. *Am J Nephrol* 2000; **20**: 32–36.
32. Obermuller N, Morente N, Kranzlin B *et al.* A possible role for metalloproteinases in renal cyst development. *Am J Physiol Renal Physiol* 2001; **280**: F540–F550.
33. Berthier CC, Wahl PR, Le Hir M *et al.* Sirolimus ameliorates the enhanced expression of metalloproteinases in a rat model of autosomal dominant polycystic kidney disease. *Nephrol Dial Transplant* 2008; **23**: 880–889.
34. Peleg I, Greenfeld Z, Cooperman H *et al.* Type I and type III collagen mRNA levels in kidney regions of old and young rats. *Matrix* 1993; **13**: 281–287.
35. Eikmans M, Baelde HJ, de Heer E *et al.* Effect of age and biopsy site on extracellular matrix mRNA and protein levels in human kidney biopsies. *Kidney Int* 2001; **60**: 974–981.
36. Melk A, Mansfield ES, Hsieh SC *et al.* Transcriptional analysis of the molecular basis of human kidney aging using cDNA microarray profiling. *Kidney Int* 2005; **68**: 2667–2679.
37. Fukuoka S, Kobayashi K. Analysis of the C-terminal structure of urinary Tamm–Horsfall protein reveals that the release of the glycosyl phosphatidylinositol-anchored counterpart from the kidney occurs by phenylalanine-specific proteolysis. *Biochem Biophys Res Commun* 2001; **289**: 1044–1048.
38. Serafini-Cessi F, Malagolini N, Cavallone D. Tamm–Horsfall glycoprotein: biology and clinical relevance. *Am J Kidney Dis* 2003; **42**: 658–676.
39. Thornley C, Dawnay A, Cattell WR. Human Tamm–Horsfall glycoprotein: urinary and plasma levels in normal subjects and patients with renal disease determined by a fully validated radioimmunoassay. *Clin Sci (Lond)* 1985; **68**: 529–535.
40. Lynn KL, Marshall RD. Excretion of Tamm–Horsfall glycoprotein in renal disease. *Clin Nephrol* 1984; **22**: 253–257.
41. Glauser A, Hochreiter W, Jaeger P *et al.* Determinants of urinary excretion of Tamm–Horsfall protein in non-selected kidney stone formers and healthy subjects. *Nephrol Dial Transplant* 2000; **15**: 1580–1587.
42. Hart TC, Gorry MC, Hart PS *et al.* Mutations of the UMOD gene are responsible for medullary cystic kidney disease 2 and familial juvenile hyperuricaemic nephropathy. *J Med Genet* 2002; **39**: 882–892.
43. Kistler AD, Poster D, Krauer F *et al.* Kidney volume progression in ADPKD is detectable within 6 months by magnetic resonance imaging without gadolinium. *Kidney Int* 2009; **75**: 235–241.
44. Kaiser T, Hermann A, Kielstein JT *et al.* Capillary electrophoresis coupled to mass spectrometry to establish polypeptide patterns in dialysis fluids. *J Chromatogr A* 2003; **1013**: 157–171.
45. Thongboonkerd V. Practical points in urinary proteomics. *J Proteome Res* 2007; **6**: 3881–3890.
46. Wittke S, Mischak H, Walden M *et al.* Discovery of biomarkers in human urine and cerebrospinal fluid by capillary electrophoresis coupled to mass spectrometry: towards new diagnostic and therapeutic approaches. *Electrophoresis* 2005; **26**: 1476–1487.
47. Neuhoff N, Kaiser T, Wittke S *et al.* Mass spectrometry for the detection of differentially expressed proteins: a comparison of surface-enhanced laser desorption/ionization and capillary electrophoresis/mass spectrometry. *Rapid Commun Mass Spectrom* 2004; **18**: 149–156.
48. DeLeo J. Receiver operating characteristic laboratory (ROCLAB): software for developing decision strategies that account for uncertainty. In: Ayyub BM (ed). *Proceedings Second International Symposium on Uncertainty Modeling and Analysis (College Park, MD, USA)*. IEEE Computer Society Press, 1993, pp 318–325.
49. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 2003; **19**: 368–375.
50. Zurbig P, Renfrow MB, Schiffer E *et al.* Biomarker discovery by CE-MS enables sequence analysis via MS/MS with platform-independent separation. *Electrophoresis* 2006; **27**: 2111–2125.
51. Coon JJ, Shabanowitz J, Hunt DF *et al.* Electron transfer dissociation of peptide anions. *J Am Soc Mass Spectrom* 2005; **16**: 880–882.
52. Good DM, Coon JJ. Advancing proteomics with ion/ion chemistry. *Biotechniques* 2006; **40**: 783–789.
53. Syka JE, Coon JJ, Schroeder MJ *et al.* Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA* 2004; **101**: 9528–9533.
54. Good DM, Wirtala M, McAlister GC *et al.* Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* 2007; **6**: 1942–1951.