Detection of autosomal dominant polycystic kidney disease by NMR spectroscopic fingerprinting of urine

Wolfram Gronwald[1], Matthias S. Klein[1], Raoul Zeltner[2], Bernd-Detlef Schulze[2], Stephan W. Reinhold[3], Markus Deutschmann[1], Ann-Kathrin Immervoll[1], Carsten A. Böger[3], Bernhard Banas[3], Kai-Uwe Eckardt[2] and Peter J. Oefner[1]

[1]Institute of Functional Genomics, University of Regensburg, Regensburg, Germany; [2]Department of Nephrology and Hypertension, University of Erlangen-Nürnberg, Erlangen, Germany and [3]Department of Internal Medicine II, University Hospital Regensburg, Regensburg, Germany

Autosomal dominant polycystic kidney disease (ADPKD) is a frequent cause of kidney failure; however, urinary biomarkers for the disease are lacking. In a step towards identifying such markers, we used multidimensional-multinuclear nuclear magnetic resonance (NMR) spectroscopy with support vector machine-based classification and analyzed urine specimens of 54 patients with ADPKD and slightly reduced estimated glomerular filtration rates. Within this cohort, 35 received medication for arterial hypertension and 19 did not. The results were compared with NMR profiles of 46 healthy volunteers, 10 ADPKD patients on hemodialysis with residual renal function, 16 kidney transplant patients, and 52 type 2 diabetic patients with chronic kidney disease. Based on the average of 51 out of 701 NMR features, we could reliably discriminate ADPKD patients with moderately advanced disease from ADPKD patients with end-stage renal disease, patients with chronic kidney disease of other etiologies, and healthy probands with an accuracy of >80%. Of the 35 patients with ADPKD receiving medication for hypertension, most showed increased excretion of proteins and also methanol. In contrast, elevated urinary methanol was not found in any of the control and other patient groups. Thus, we found that NMR fingerprinting of urine differentiates ADPKD from several other kidney diseases and individuals with normal kidney function. The diagnostic and prognostic potential of these profiles requires further evaluation.

Kidney International (2011) **79,** 1244–1253; doi:10.1038/ki.2011.30; published online 9 March 2011

KEYWORDS: autosomal dominant polycystic kidney disease; NMR spectroscopy; support vector machine; urine

Correspondence: *Wolfram Gronwald, Institute of Functional Genomics, University of Regensburg, Josef-Engert-Strasse 9, 93053 Regensburg, Germany. E-mail: Wolfram.Gronwald@klinik.uni-regensburg.de*

Received 21 April 2010; revised 21 December 2010; accepted 4 January 2011; published online 9 March 2011

Autosomal dominant polycystic kidney disease (ADPKD; Mendelian Inheritance in Man (MIM) 173900) is the most common inherited renal disorder, affecting 1:400–1:1000 live births. It is caused by mutations in the *PKD1* (*polycystic kidney disease 1 (autosomal dominant)*) and *PKD2* (*polycystic kidney disease 2 (autosomal dominant)*) genes that are responsible for 85 and 15% of cases, respectively.[1] ADPKD is characterized by multiple cysts in both kidneys, destruction of the normal kidney architecture, and a progressive deterioration of renal function, with ~50% of patients suffering from end-stage renal disease (ESRD) at the age of 60 years.[2] Hypertension and cysts in other organs, particularly in the liver, are frequent complications.

Diagnosis of ADPKD is based on the detection of multiple cysts by renal ultrasound and a positive family history. The cystic transformation of the kidneys may precede the decline of renal function by several decades.[3] Thus, biomarkers that aid the diagnosis and/or prognostication are of interest. In this context, several candidate urinary protein biomarkers have been evaluated, including the proteins MCP1 (MIM 158105), KIM1 (MIM 606518), NGAL (MIM 600181), and SFRP4 (MIM 606570).[4–7] In addition, proteomic approaches have been applied to the unbiased detection of urinary proteins that might serve as diagnostic biomarkers for ADPKD.[8]

Because of the kidney's role in maintaining homeostasis, the chemical composition of urine may vary greatly as a function of endogenous and exogenous factors. Distinct changes in composition may shed valuable insights into the pathophysiology of disease and provide novel biomarkers for the purposes of diagnosis, prognostication, and monitoring of therapeutic response.[9] To that end, nuclear magnetic resonance (NMR) is a powerful tool, because it allows the nonderivatized and nondestructive determination of free metabolites and other compounds in biological fluids with little sample pretreatment.[10] Signal volumes scale linearly with concentration and, in most cases, are independent of the chemical properties of the investigated molecules. NMR spectroscopy has previously been used in the context of cystic kidney disease in a rat model,[11] the analysis of dietary

effects,[12] and the examination of cyst aspirates,[13] but a systematic analysis of urine from ADPKD patients is lacking.

In this study, we applied one-dimensional (1D) NMR spectroscopy to urine specimens obtained from patients with ADPKD and other renal pathologies as well as apparently healthy controls. The NMR data were subjected to supervised learning with a support vector machine (SVM) to identify spectral features that predict ADPKD. Finally, we attempted to elucidate the chemical nature of the discriminatory features by two-dimensional (2D) NMR spectroscopy.

RESULTS

Patient characteristics

We included 54 patients with ADPKD (group 1) and a mean estimated glomerular filtration rate (eGFR) of 95.5 ± 27.7 ml/min per 1.73 m^2, indicating that most patients had well-preserved renal function. Nevertheless, there was significant disease burden as the mean total kidney volume, determined by magnetic resonance imaging-based volumetry,[14] was significantly enlarged over that of nonaffected individuals: 1869 ± 1244 ml vs 334 ml (range between minimum and maximum volume 194–614 ml).[15] The clinical characteristics along with those of the other patient groups investigated are summarized in Table 1. Of the 54 ADPKD patients, 35 received blood pressure-lowering medication, including angiotensin-converting enzyme inhibitors and angiotensin II type 1 blockers such as enalapril, lisinopril, ramipril, and candesartan, with or without the co-administration of furosemide and β-blockers (group 1A). The other 19 ADPKD patients received no medication in the absence of hypertension (group 1B). A total of 46 healthy volunteers with normal renal function, 23 men and women each, with a mean age of 38.7 ± 10.4 years served as the control group (group 2).

To gain insight into the specificity of the identified metabolite biomarkers for ADPKD, we included four additional groups of patients with chronic kidney diseases. The first of these groups (group 3) included 10 stable ADPKD patients with ESRD on hemodialysis with residual urine excretion. The second cohort (group 4) comprised 16 patients 3 months after renal transplantation, who had not experienced acute organ rejection; their mean eGFR was 53.3 ± 20.9 ml/min per 1.73 m^2. Group 5 comprised 30 patients

Table 1 | Clinical characteristics of the diseased patients enrolled: (a) group 1 ADPKD patients; and (b) patients suffering from other kidney ailments

(a)

ADPKD	All (*N*=54) (group 1)	Male (*N*=30) (group 1)	Female (*N*=24) (group 1)	Medication (*N*=35) (group 1A)	No medication (*N*=19) (group 1B)
Sex, male (%)	30 (56%)	—	—	22 (63%)	9 (47%)
Age, years	40.6 ± 5.9	40.6 ± 5.0	40.6 ± 7.0	41.2 ± 4.5	39.6 ± 7.8
BMI[a], kg/m^2	25.9 ± 4.3	26.1 ± 4.0	25.7 ± 4.7	26.7 ± 4.6	24.6 ± 3.3
Systolic BP[a], mm Hg	129 ± 15	134 ± 16	123 ± 11	129 ± 14	130 ± 17
Diastolic BP[a], mm Hg	86 ± 10	88 ± 10	83 ± 8	86 ± 9	84 ± 10
Serum creatinine[a], mg/dl	1.0 ± 0.3	1.2 ± 0.3	0.8 ± 0.2	1.1 ± 0.4	0.9 ± 0.2
eGFR[a], ml/min per 1.73 m^2	95.5 ± 27.7	87.9 ± 24.3	106.3 ± 29.3	88.5 ± 29.0	108.4 ± 20.4
ACR[a], mg/g	32.4 ± 37.3	34.4 ± 35.7	29.7 ± 39.9	38.3 ± 43.9	29.7 ± 34.2
Total kidney volume[a], ml	1869 ± 1244, *N*=38[b]	2152 ± 1477, *N*=24[b]	1383 ± 383, *N*=14[b]	1949 ± 1293, *N*=34[b]	1184 ± 138, *N*=4[b]

(b)

Control groups with reduced kidney function	ADPKD with ESRD (*N*=10) (group 3)	Renal transplant (*N*=16) (group 4)	Diabetes type 2 with microalbuminuria (*N*=30) (group 5)	Diabetes type 2 w/o microalbuminuria (*N*=22) (group 6)
Sex, male (%)	7 (70.0%)	11 (69%)	13 (43.3%)	8 (36.4%)
Age[a], years	60.0 ± 10.0	53 ± 15.4	66.0 ± 10.7	71.5 ± 6.5
BMI[a], kg/m^2	25.8 ± 3.5	26.3 ± 3.4	31.6 ± 5.6	32.5 ± 5.4
Systolic BP[a], mm Hg	133 ± 17.3	133.8 ± 21.6	140 ± 21.0	136 ± 14.0
Diastolic BP[a], mm Hg	76 ± 13.3	77.2 ± 8.9	76 ± 14.0	78 ± 8.0
Serum creatinine[a], mg/dl	NA	1.6 ± 0.6	0.88 ± 0.1	1.35 ± 0.2
eGFR[a], ml/min per 1.73 m^2	NA	53.3 ± 20.9	74.2 ± 7.2	44.0 ± 5.6
ACR[a], mg/g	1581.9 ± 1535.8	53.7 ± 125.3	74.0 ± 40.0	8.9 ± 8.1
Treatment with ACE-I or ARB	10 (100.0%)[c]	—	22 (73.3%)	19 (86.4%)

Abbreviations: ACE-I, angiotensin-converting enzyme inhibitor; ACR, albumin/creatinine ratio; ADPKD, autosomal dominant polycystic kidney disease; ARB, angiotensin receptor blocker; BMI, body mass index; BP, blood pressure; eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; NA, not available.
[a]Given are the mean value and the s.d.
[b]Number of patients, for whom magnetic resonance imaging (MRI) volumetric data were available.
[c]ADPKD patients with ESRD received individual combinations of drugs to treat hypertension.
Patients suffering from other kidney ailments were grouped as follows: group 3, stable renal failure ADPKD patients on hemodialysis with residual urine excretion; group 4, renal transplant recipients without acute organ rejection; group 5, diabetes mellitus type 2 patients with reduced eGFR and microalbuminuria; group 6, diabetes mellitus type 2 patients with severely reduced eGFR but no microalbuminuria. Details for the group of 46 apparently healthy volunteers (group 2) are given in the text of the article.

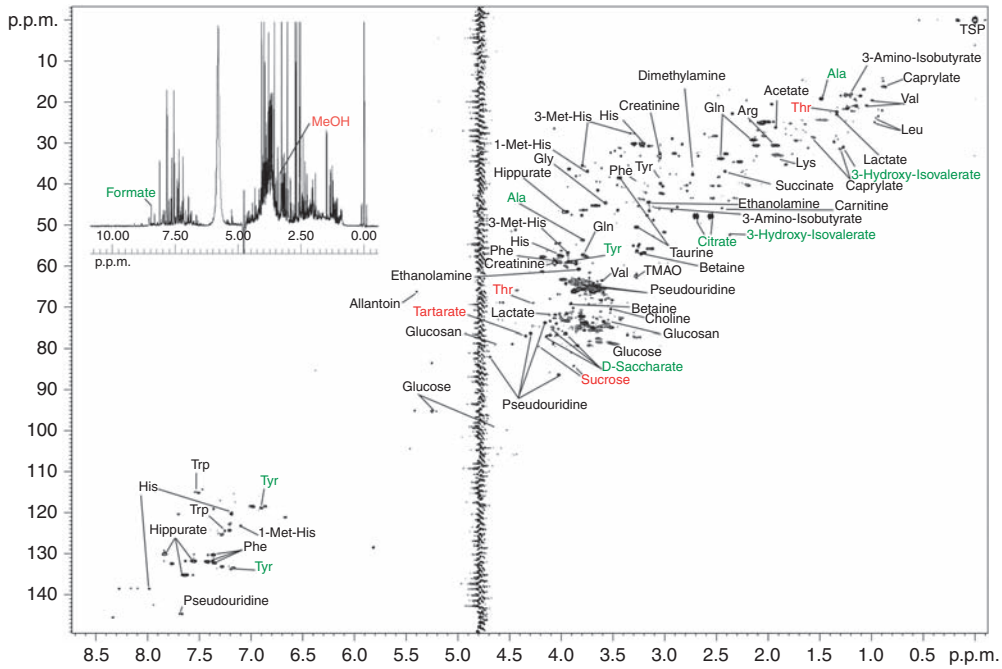

Figure 1 | The two-dimensional (2D) ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) spectrum of a healthy control urine specimen. In total, the signals of close to 120 metabolites could be unambiguously assigned in this spectrum. For reasons of clarity, only a subset is marked. Metabolites that are significantly up- or down-regulated in the specimens of the autosomal dominant polycystic kidney disease (ADPKD) patients of group 1 are color coded in red and green, respectively. The complete list of discriminating features is given in Table 3 and the full assignment of this spectrum is given in Supplementary Table S1 online. The insert shows the corresponding one-dimensional (1D) ^{1}H spectrum.

with diabetes mellitus type 2 with an eGFR ranging from 60 to 89 ml/min per 1.73 m^2 and an albumin/creatinine ratio of 30–300 mg/g, whereas the 22 patients of group 6 suffered from diabetes mellitus type 2 with an eGFR $<$50 ml/min per 1.73 m^2, although no microalbuminuria.

Discrimination of ADPKD patients with early disease from healthy volunteers by urine NMR fingerprinting

For each individual, both 1D ^{1}H and 2D ^{1}H-^{13}C HSQC (heteronuclear single quantum coherence) NMR spectra of spot urine specimens were acquired. Figure 1 shows exemplary 1D ^{1}H and 2D ^{1}H-^{13}C HSQC NMR spectra obtained for the same urine specimen of a healthy proband. The displayed 2D spectrum served as a reference for metabolite assignment (Figure 1); close to 120 different metabolites were assigned unambiguously based on these spectra (Supplementary Table S1 online). Metabolites significantly up- or down-regulated between healthy probands and group 1 ADPKD patients are color coded in red and green, respectively. Note that for reasons of clarity, only a subset of the metabolites has been marked.

Classification of the urine specimens of the 54 group 1 ADPKD patients against the 46 healthy volunteers was performed using the SVM algorithm on the 1D ^{1}H NMR data. To that end, each 1D NMR spectrum was split in 701 evenly spaced subsections (buckets or features), each of them ideally corresponding to a distinct urinary compound. Classification results were obtained by a nested cross-validation approach, whereby the patients and healthy volunteers were iteratively split into a training set and a test set. The classification algorithm was then trained on the training data to predict the test data. This procedure ensured that the training procedure was not biased by the test data. To obtain an optimal prediction accuracy of the SVM, the selection of discriminatory features is of prime importance. Therefore, in a nested loop, an inner cross-validation was performed for parameter optimization, thus obtaining an almost unbiased estimate of the true classification error.[16] A detailed description of the used nested cross-validation approach is given in the Supplementary Materials online.

Applying the nested cross-validation procedure, the number of discriminating features was on average optimized to 51. This allowed the prediction of ADPKD with an average accuracy of 85.0 ± 3.1%. Classification results of all tests are summarized in Table 2. Receiver operating curves were then employed to evaluate the performance of the classification. The area under the receiver operating curve (AUC; Figure 2a) amounted to 0.91, indicating a high probability that the classifier, an SVM in this case, would rank a randomly chosen ADPKD case higher than a randomly chosen control.

Table 3 lists the positions of the 51 predictive features in the 1D spectra and their chemical identity. Their significance is indicated by the corresponding *P*-values calculated for the comparison between ADPKD patients and healthy volunteers.

Table 2 | Summary of classification results obtained by the SVM approach

Groups compared	Prediction accuracy (arith. mean ± s.d.)	Area under ROC curve	Statistical approach	No. of features
1 vs 2	85.0 ± 3.1%	0.91	Nested CV	Average of 51 (Table 3)
1 vs 3	100.0 ± 0%	1.00	CV	Fixed set of 51 (Table 3)
1 vs 4	95.4 ± 0.6%	0.99	CV	Fixed set of 51 (Table 3)
1 vs 5	89.5 ± 1.0%	0.92	CV	Fixed set of 51 (Table 3)
1 vs 6	92.6 ± 1.4%	0.97	CV	Fixed set of 51 (Table 3)
1 vs 2–6	81.0 ± 1.7%	0.89	CV	Fixed set of 51 (Table 3)
1A vs 2	86.6 ± 2.3%	0.92	Nested CV	Average of 44 (Supplementary Table S2 online)
1B vs 2	82.0 ± 2.0%	0.79	Nested CV	Average of 23 (Table 4)

Abbreviations: CV, cross validation; ROC, receiver operating curve; SVM, support vector machine.

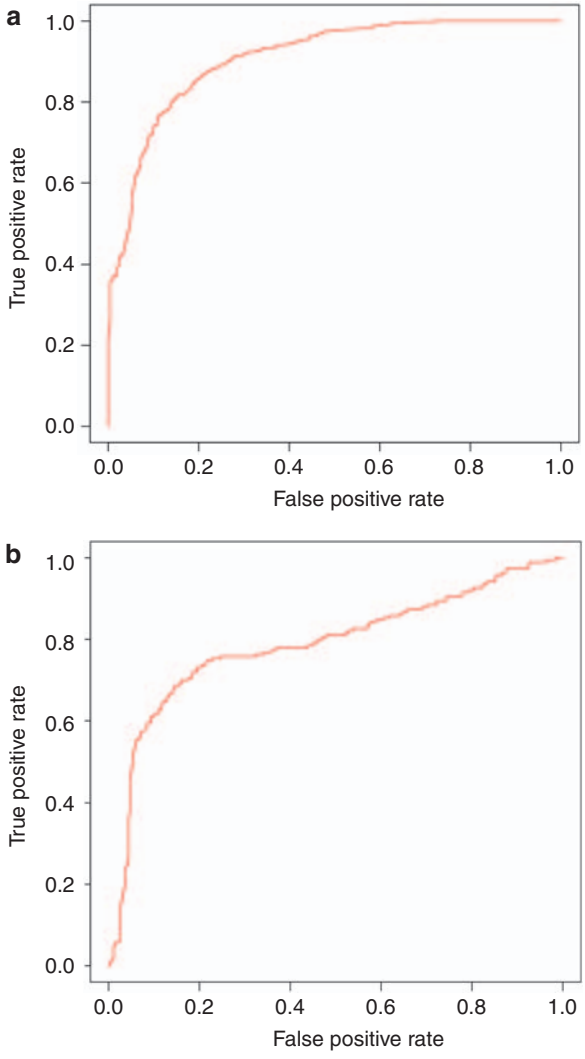

Figure 2 | Performance of classification. (**a**) Receiver operating curve (ROC) analysis of healthy volunteers versus all diseased autosomal dominant polycystic kidney disease (ADPKD) patients. ROC curve obtained from the analysis of 10 nested cross-validation runs using an inner cross-validation for parameter optimization. The number of features was optimized to 51. (**b**) ROC analysis of healthy volunteers versus nonmedicated ADPKD patients. ROC curve obtained from the analysis of 10 nested cross-validation runs using an inner cross-validation for parameter optimization. The number of features was optimized to 23.

Both raw *P*-values based on a two-sided *t*-test assuming Gaussian distribution of the data, which was confirmed using the Kolmogorov–Smirnov test, and *P*-values corrected for multiple testing are given. The latter values were adjusted for the false discovery rate according to Benjamini and Hochberg.[17] For some of these features, it was additionally possible to determine accurate concentration levels (Supplementary Table S2 online). Note that several of the discriminating features were only present in concentrations below their lower limits of quantification and, therefore, could not be quantified. Concentrations of discriminating features that could be quantified plus values for other highly abundant metabolites for the healthy control group and ADPKD patients are given. The discriminating features listed in Table 3 are also displayed in a heat map representation in Figure 3. Their upregulation and downregulation are color coded for all analyzed urine specimens in yellow and blue, respectively. Unambiguously identified discriminatory metabolites in decreasing order of their corrected *P*-values included formate ($P = 3.5e{-}07$), proteins ($P = 5.6e{-}07$), tartaric acid ($P = 2.9e{-}06$), 3-hydroxy-isovalerate ($P = 9.8e{-}06$), citrate ($P = 1.3e{-}05$), threonine ($P = 2.7e{-}05$), methanol ($P = 3.1e{-}05$), carbohydrates ($P = 3.3e{-}05$), sucrose ($P = 3.4e{-}05$), alanine ($P = 4.6e{-}05$), 6-hydroxynicotinic acid ($P = 5.9e{-}05$), D-saccharate ($P = 7.8e{-}05$), and tyrosine ($P = 1.4e{-}04$). In case that more than one NMR signal contributed to a given metabolite, the corresponding bucket showing the smallest *P*-value is shown. For the buckets attributed to proteins the assignment was additionally verified by acid hydrolysis that showed a clear decrease in the protein signals and a corresponding increase in free amino acids.

Discrimination of ADPKD patients with early disease from other kidney ailments

Next, the question was addressed whether the 51 NMR features (Table 3) that had allowed the successful discrimination of group 1 ADPKD patients from healthy controls would also distinguish the former from patients suffering from chronic kidney disease of other etiologies. To that end, the NMR fingerprints of the 54 ADPKD patients of group 1 were first compared with those of the 10 ADPKD patients with ESRD (group 3). A perfect separation between these two

Table 3 | Spectral positions and *P*-values of predictive features used by the SVM for classification of ADPKD patients irrespective of medication versus healthy controls

ID	Spectral position (p.p.m.)	*P*-value unadjusted	*P*-value BH adjusted[a]	Metabolite
104	8.465	4.99e−10	3.50e−07	Formate
327	4.245	1.90e−09	5.63e−07	Threonine/guadinosuccinic acid[b]
691	0.605	2.63e−09	5.63e−07	Proteins
690	0.615	3.21e−09	5.63e−07	Proteins
699	0.525	6.87e−09	8.07e−07	Proteins
342	4.095	6.90e−09	8.07e−07	D-saccharate/glucosan[b]
698	0.535	2.40e−08	2.40e−06	Proteins
693	0.585	3.33e−08	2.80e−06	Proteins
642	1.095	3.59e−08	2.80e−06	2-Methylbutyrylglycine/3-methyl-2-oxovaleric acid/2-methylbutyroylcarnitine[b]
318	4.335	4.06e−08	2.85e−06	Tartaric acid
700	0.515	4.96e−08	3.05e−06	Proteins
692	0.595	5.22e−08	3.05e−06	Proteins
687	0.645	6.07e−08	3.16e−06	Proteins
701	0.505	6.32e−08	3.16e−06	Proteins
689	0.625	2.03e−07	9.13e−06	Proteins
696	0.555	2.08e−07	9.13e−06	Proteins
514	2.375	2.38e−07	9.83e−06	3-Hydroxy-isovalerate
482	2.695	3.37e−07	1.25e−05	Citrate
483	2.685	3.37e−07	1.25e−05	Citrate
694	0.575	4.61e−07	1.61e−05	Proteins
640	1.115	4.97e−07	1.66e−05	2-Methylbutyrylglycine/3-methyl-2-oxovaleric acid/2-methylbutyroylcarnitine[b]
683	0.685	8.11e−07	2.58e−05	Proteins/bile acids
326	4.255	8.83e−07	2.69e−05	Threonine
678	0.735	9.54e−07	2.78e−05	Proteins/bile acids[b]
688	0.635	9.90e−07	2.78e−05	Proteins
415	3.365	1.17e−06	3.14e−05	Methanol
314	4.375	1.27e−6	3.30e−05	Hydroxyacetone/*N*-acetyltyrosine[b]
684	0.675	1.34e−06	3.33e−05	Proteins
256	6.945	1.38e−06	3.33e−05	Catechol/salicylic acid[b]
360	3.915	1.43e−06	3.34e−05	Carbohydrates
328	4.235	1.50e−06	3.39e−05	Sucrose
672	0.795	1.77e−06	3.88e−05	Proteins
602	1.495	2.15e−06	4.56e−05	Alanine
330	4.215	2.23e−06	4.58e−05	Sucrose
686	0.655	2.29e−06	4.58e−05	Proteins
329	4.225	2.57e−06	5.00e−05	Sucrose
670	0.815	2.72e−06	5.16e−05	Proteins
283	6.675	3.21e−06	5.93e−05	6-Hydroxynicotinic acid
257	6.935	3.54e−06	6.36e−05	Catechol/salicylic acid[b]
337	4.145	4.46e−06	7.81e−05	D-saccharate
315	4.365	5.62e−06	9.60e−05	Hydoxyacetone/*N*-acetyltyrosine[b]
671	0.805	6.68e−06	1.12e−04	Proteins
641	1.105	7.03e−06	1.15e−04	2-Methylbutyrylglycine/3-methyl-2-oxovaleric acid/2-methylbutyroylcarnitine[b]
682	0.695	7.38e−06	1.18e−04	Proteins
285	6.655	8.04e−06	1.25e−04	6-Hydroxynicotinic acid
478	2.735	8.25e−06	1.26e−04	Dimethylamine/citrate[b]
262	6.885	9.43e−06	1.41e−04	Tyrosine
477	2.745	1.09e−05	1.59e−04	Citrate
309	4.425	1.12e−05	1.60e−04	Dihydroxyacetone/*N*-acetyltyrosine[b]
455	2.965	1.22e−05	1.71e−04	Isocitrate[b]
680	0.715	1.49e−0.5	2.05e−04	Proteins

Abbreviations: ADPKD, autosomal dominant polycystic kidney disease; SVM, support vector machine.
[a]Adjusted for the false discovery rate according to Benjamini and Hochberg (BH).[17]
[b]Assignments where a signal could be attributed to more than one metabolite or where in case of very weak signals an unambiguous assignment was not possible. In case that a feature could be assigned to more than one metabolite, all possible assignments are given.

groups was achieved with an average prediction accuracy of $100 \pm 0.0\%$ and an AUC of 1.0. Next, the NMR data of the ADPKD patients of group 1 were compared with those of the 16 patients 3 months after renal transplantation without a history of acute allograft rejection (group 4). It proved feasible to distinguish the ADPKD patients from kidney transplant recipients with an average prediction accuracy of $95.4 \pm 0.6\%$ and an AUC of 0.99. Visual inspection of the heat

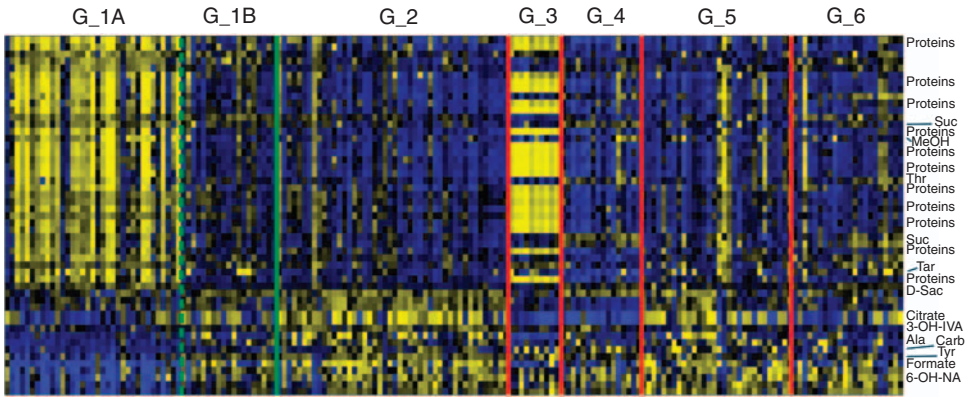

Figure 3 | Heat map representation of the six groups of investigated specimens. Displayed is the variation in nuclear magnetic resonance (NMR) signal volume for the 51 features that were identified for the discrimination between all autosomal dominant polycystic kidney disease (ADPKD) patients and the healthy controls. These features are also summarized in Table 3. The upregulated features are indicated in yellow, whereas downregulated ones are marked in blue. Rows are ordered in correlation with disease status. Rows that are mostly upregulated in ADPKD patients and downregulated in all other groups are shown in the top part of the figure and vice versa. Patients were grouped as follows: group 1A, ADPKD patients who received medication; group 1B, patients who received no medication; group 2, apparently healthy volunteers; group 3, stable renal failure ADPKD patients on hemodialysis with residual urine excretion; group 4, renal transplant recipients without acute organ rejection; group 5, diabetes mellitus type 2 patients with reduced eGFR and microalbuminuria; group 6, diabetes mellitus type 2 patients with severely reduced eGFR but no microalbuminuria. Ala, alanine; Carb, carbohydrates; D-Sac, D-saccharic acid; MeOH, methanol; Suc, sucrose; Tar, tartaric acid; Thr, threonine; Tyr, tyrosine; 3-OH-IVA, 3-hydroxyisovaleric acid; 6-OH-NA, 6-hydroxynicotinic acid.

map presented in Figure 3 shows that for most of the 51 features marked differences existed between these two groups. One of the few exceptions included citrate that was downregulated in both groups. After that, we investigated the discrimination of ADPKD patients from group 5 (30 microalbuminuric patients from the GENetics of DIAbetic Nephropathy (GENDIAN) study[18] with an eGFR of 60–89 ml/min per 1.73 m^2): an average prediction accuracy of 89.5 ± 1.0% and an AUC of 0.92 were obtained. For the discrimination of group 6 (22 patients selected from the GENDIAN study, who had a significantly reduced eGFR $<$50 ml/min per 1.73 m^2 but no microalbuminuria) from the ADPKD group, an average prediction accuracy of 92.6 + 1.4% with an AUC of 0.97 was obtained. In addition, we investigated whether it was possible to discriminate between the ADPKD patients of group 1 and all other groups, including diseased and healthy subjects, in a single step. For this purpose, the data of all 6 groups were combined, yielding a single data set of 178 samples. In this data set, the ADPKD samples of group 1 were assigned to one group and all other samples including ADPKD patients with ESRD were combined into a second group. Using the same set of 51 features as above, an average prediction accuracy of 81.0 + 1.7% and an AUC of 0.89 were obtained for the distinction between these two groups.

Investigation of medication and disease stage

To investigate whether pharmaceuticals administered to ADPKD patients, or metabolites thereof, or metabolic changes induced by these pharmaceuticals, influenced the classification results, the 35 ADPKD patients belonging to group 1A were compared with the healthy control group. Of the 35 ADPKD patients receiving medication for arterial hypertension, the vast majority was treated with ACE inhibitors or angiotensin II type 1 blockers.[19] A nested cross-validation run with inner-loop parameter optimization yielded on average 44 predictive features and an overall prediction accuracy of 86.6 ± 2.3% with an AUC of 0.92. The corresponding model (Supplementary Table S3 online) was very similar to the model obtained for all group 1 ADPKD patients versus the healthy controls.

To further elucidate the impact of medication and non-medication, respectively, on the urinary NMR fingerprints, the predictive outcome for the 19 nonmedicated patients (group 1B) was analyzed against the healthy control group (group 2). Arterial hypertension is an early indication of ADPKD present in ~60% of all cases before significant impairment of renal function.[20] As the average total renal volume of these 19 non-medicated patients was considerably smaller than that of the medicated group 1A, this group presumably included subjects at an earlier stage of the disease. As above, a nested cross-validation run with inner-loop parameter optimization was performed. The number of predictive features was optimized within the inner cross-validation to an average number of 23, and an overall prediction accuracy of 82.0 ± 2.0% was obtained with an AUC of 0.79 (Figure 2b). The corresponding model is shown in Table 4. Some of the compounds present in this model, such as citrate and 3-hydroxyisovaleric acid, were identical to the model presented in Table 3, whereas other compounds, including 1-methylnicotinamide, trigonelline, and nicotinamide riboside, constituted new additions.

Among the compounds not included in the new model were methanol and formate, as well as features representing proteins, peptides, and bile acids. This indicated that reabsorption of filtered proteins was still intact in the nonmedicated ADPKD patients.

DISCUSSION

In this study, we have characterized the urinary NMR fingerprints of ADPKD patients with well-preserved renal

Table 4 | Spectral positions and *P*-values of predictive features used by the SVM for the classification of nonmedicated ADPKD patients versus healthy controls

ID	Spectral position (p.p.m.)	*P*-value unadjusted	*P*-value BH adjusted[a]	Metabolite
62	8.885	6.08e−5	0.028	1-Methylnicotinamide
14	9.365	1.34e−4	0.028	Nicotinamide riboside
640	1.115	1.52e−4	0.028	2-Methylbutyrylglycine/3-methyl-2-oxovaleric acid/2-methylbutyroyl-carnitine[b]
548	2.035	1.62e−4	0.028	*N*-acetyl-L-glutamine/pyroglutamic acid[b]
23	9.275	2.12e−4	0.030	Nicotinuric acid[b]
404	3.475	3.47e−4	0.041	Carbohydrates/theobromine[b]
643	1.085	7.50e−4	0.075	2-Methylbutyrylglycine/3-methyl-2-oxovaleric acid/2-methylbutyroyl-carnitine[b]
514	2.375	1.18e−3	0.078	3-OH-isovaleric acid
309	4.425	1.18e−3	0.078	Dihydroxyacetone/*N*-acetyltyrosine[b]
13	9.375	1.23e−3	0.078	1-Methylnicotinamide
494	2.575	1.34e−3	0.078	Citrate
495	2.565	1.34e−3	0.078	Citrate
641	1.105	2.22e−3	0.113	2-Methylbutyrylglycine/3-methyl-2-oxovaleric acid/2-methylbutyroyl-carnitine[b]
55	8.955	2.25e−3	0.113	1-Methylnicotinamide
304	4.475	2.45e−3	0.115	1-Methylnicotinamide
644	1.075	2.75e−3	0.121	2-Methyl-3-ketovaleric acid/3-methyl-2-oxovaleric acid[b]
148	8.025	3.34e−3	0.124	3-Methylxanthine[b]
61	8.895	3.38e−3	0.124	1-Methylnicotinamide
483	2.685	3.54e−3	0.124	Citrate
482	2.695	3.54e−3	0.124	Citrate
37	9.135	3.75e−3	0.125	Trigonelline
132	8.185	4.24e−3	0.135	1-Methylnicotinamide
129	8.215	4.85e−3	0.148	1-Methylnicotinamide/hypoxanthine[b]

Abbreviations: ADPKD, autosomal dominant polycystic kidney disease; SVM, support vector machine.
[a]Adjusted for the false discovery rate according to Benjamini and Hochberg (BH).[17]
[b]Assignments where a signal could be attributed to more than one metabolite or where in case of very weak signals an unambiguous assignment was not possible. In case that a feature could be assigned to more than one metabolite, all possible assignments are given.

function. We could demonstrate that SVM-based classification of urinary NMR fingerprints allowed the reliable discrimination between ADPKD patients at a relatively early stage of kidney disease and those with ESRD, as well as patients suffering from chronic kidney disease for reasons other than ADPKD and apparently healthy controls. These data suggest the existence of a set of urinary compounds specific to early stages of ADPKD.

For the discrimination of ADPKD patients with early disease from apparently healthy volunteers, a set of 51 distinct urinary features was identified (Table 3). Here, elevated urinary protein levels are mainly observed for ADPKD patients, hinting at an endocytosis defect that disturbs the reabsorption of low-molecular-weight proteins by proximal tubular cells.[21] For the ESRD patients, this effect is much more pronounced as can be seen in Figure 3. Other discriminating metabolites included citrate and formate. Hypocitraturia is a common metabolic abnormality found in ADPKD patients, which is believed to contribute to the risk of stone formation aside from the anatomic obstruction of the renal collecting system by the cysts.[22] One probable explanation for the reduced urinary concentration of formate is that it reflects changes in the folate-mediated one-carbon metabolism, which generates one-carbon units in mitochondria in the form of formate by conversion of serine to glycine and incorporates formate after its entry into the cytoplasm into tetrahydrofolate to produce 10-formyl-tetrahydrofolate, which is utilized for the biosynthesis of purines and thymidine, and the remethylation of homocysteine to methionine. The latter, in turn, can be converted to *S*-adenosylmethionine, a cofactor for many methylation reactions, including the methylation of proteins, phospholipids, neurotransmitters, and DNA.[23,24]

The finding of elevated levels of urinary methanol in ADPKD patients not on hemodialysis came as a surprise. To the best of our knowledge, this is the first report on increased excretion of methanol for reasons other than acute methanol poisoning, chronic alcohol abuse, and occupational exposure to methanol vapor or methyl formate. The possibility of an erroneous assignment of the NMR signal (see Supplementary Materials online) to methanol was excluded by the independent confirmation of the quantitative NMR measurements by headspace-gas chromatography (data not shown). Elevated levels of urinary methanol were not observed in the 10 ADPKD patients on hemodialysis. A possible explanation for the observed increase in urinary methanol levels in ADPKD patients at a progressed stage, but not yet in need of dialysis, is an increased activity of the enzyme protein methylesterase 1 (PME1, MIM 611117), which produces methanol by hydrolyzing the protein-methyl esters formed by *S*-adenosylmethionine-dependent protein-carboxyl methyltransferases.[25] In most tissues, the activity of

the latter is at least tenfold greater than that of PME1.[26] In the proximal tubules of the rat kidneys, however, the activity of PME1 surpasses that of the protein-carboxyl methyltransferases by a factor of 1.5. Protein methylation–demethylation plays an important role in the regulation of protein function.[25] However, the significance of the unique ratio of methylating and demethylating enzyme activities in the kidney is still not understood. The high levels of PME1 activity in the proximal tubules suggest that this enzyme may be involved in the transmembrane transport of compounds such as sodium, phosphate, bicarbonate, amino acids, organic acids, and proteins. Indeed, it has been shown recently that phosphorylation of PME1 by the salt-inducible kinase-1 (SIK1, MIM 605705) in response to elevated intracellular sodium promotes the dissociation of PME1 from the catalytic subunit of protein phosphatase-2A (PP2A, MIM 603113), thereby activating the latter enzyme, which increases in turn the catalytic activity of Na^+,K^+-ATPase upon dephosphorylation of its α-subunit.[27] Interestingly, mislocation of functionally active Na^+,K^+-ATPase to apical, luminal plasma membranes of ADPKD epithelia is one of the cell biological hallmarks of ADPKD, which constitutes a complete reversal of the normal renal tubule polarized location in basolateral membranes.[28] Consequently, mislocation of Na^+,K^+-ATPase might disturb the above regulation process and in turn this could lead to an increased methanol production by PME1.

Another discriminating metabolite was alanine. Increased excretion of alanine had been previously identified as a marker of impaired proximal tubular function in a study on mercury-induced nephrotoxicity[23] and in an experimental model of ADPKD in rats.[11] Interestingly, in the ADPKD patients investigated here, alanine excretion was reduced (Figure 3). Among other discriminating features observed were threonine and carbohydrates (Table 3 and Figure 3), which had been also identified in a recent study of juvenile cystic mice.[29] Because of the presence of signal overlap in both the 1D and 2D spectra, unambiguous signal assignment of the carbohydrates to a specific sugar was in some cases not feasible. For the discriminating metabolite tyrosine, it is known that the excretion of tyrosine is reduced in chronic kidney disease, and so is the rate of tyrosine synthesis.[30] Other metabolites included 3-hydroxyisovalerate, sucrose, 6-hydroxy-nicotinic acid, and D-saccharate. The increased urinary levels of 3-hydroxyisovalerate in ADPKD patients might be an indicator of biotin deficiency due to impaired renal reclamation.[31] 6-Hydroxynicotinic acid is an intermediate in the oxidation of nicotinic acid by *Pseudomonas* and its measurement in urine by ^{1}H NMR spectroscopy has been applied to the diagnosis of *Pseudomonas aeruginosa* in urinary tract infection.[32] Finally, tartaric acid was identified as a discriminating feature. Dietary intake, in particular of grapes and grape products, is usually the major contributor to urinary tartrate.[33,34] In the case of ADPKD patients the administration of metoprolol tartrate might have served as an additional source of urinary tartrate.

Our findings have several implications for a better understanding and monitoring of the disease process, which deserve exploration in future studies. First, they imply that the set of distinguishing features or single components might support an early diagnosis of the disease. Ultrasound criteria have been developed for individuals with a positive family history,[35] but in patients with a negative family history diagnostic criteria are less well defined. The present study included patients with normal eGFR, in whom the urine metabolite profile was distinguishable, but long-term follow-up studies in patients with uncertain diagnosis are needed to define the diagnostic value of these parameters. Second, urinary compounds specific for ADPKD might help to monitor disease progression and could serve as surrogate markers for the success of interventions that aim to slow disease progression. There is uniform agreement that such markers could significantly facilitate the testing of therapeutic interventions, all the more as recent data suggest that great variability exists in the correlation between change in kidney function and cyst volume increase.[36] The present study was designed as a cross-sectional proof-of-concept study and, therefore, does not allow firm conclusions on changes of urinary fingerprints with disease progression. It is noteworthy, however, that the pattern in residual urine of ADPKD patients on hemodialysis was grossly different. Third, the urinary fingerprint may point to some important aspects in the pathogenesis of the disease and its progression.

In addition to the NMR-based metabolomic approach pursued in this study, urinary proteomics has been successfully applied to the prediction of ADPKD.[8] The coupling of capillary electrophoresis to mass spectrometry allowed the identification of a unique set of proteins serving as reliable biomarkers for the prediction of ADPKD. The comparison of these two studies clearly shows that ADPKD is reflected in both the urinary proteome and metabolome. The use of NMR-based metabolomics offers the additional advantage of only minimal required sample pretreatment and easy sample handling, enabling fast and fully automatic data collection.

In conclusion, this study demonstrates that an unbiased urinary fingerprint analysis of metabolites and proteins clearly differentiates ADPKD from several other kidney diseases and individuals with normal kidney function and that there is a significant potential for the identification of clinically useful biomarkers of ADPKD.

PATIENTS AND METHODS

Patient selection

The $n = 54$ ADPKD patients (group 1) with sonographically confirmed multiple renal cysts, an eGFR >60 ml/min per 1.73 m^2, and a positive family history according to the criteria published by Ravine *et al.*[35] were recruited at the outpatient clinic of the University of Erlangen-Nuremberg. Group 1 could be subdivided into a subset of $n = 35$ patients treated for arterial hypertension (group 1A) and a set of $n = 19$ nonmedicated patients without hypertension (group 1B). The $n = 46$ healthy volunteers (group 2) were recruited at the

University of Regensburg and the University Clinic of Erlangen, respectively. The group included 23 men and women each with a mean age ± s.d. of 38.7 ± 10.4 years and a mean urinary creatinine concentration ± s.d. of 8.517 ± 5.053 mmol/l. We further included 10 stable ADPKD patients with ESRD on hemodialysis with residual urine excretion (group 3), 16 patients 3 months after kidney transplantation with an uneventful postoperative course (group 4),[37] and 52 patients from the GENDIAN study:[18] selected were 30 patients in GENDIAN with CKD stage 2 (eGFR 60–89 ml/min per 1.73 m^2 and microalbuminuria; group 5), and 22 GENDIAN patients with an eGFR <50 ml/min per 1.73 m^2 but no micro-albuminuria (group 6). Urine samples and anthropometric measures for the GENDIAN patients included in the present study were obtained at the 6-year follow-up examination.

Written declarations of consent had been obtained from all study participants before inclusion. eGFR was calculated using the four-variable Modification of Diet in Renal Disease equation for traceable serum creatinine assays.[38] Spot urine samples were collected from all subjects, immediately frozen, and stored at $-80\,^{\circ}C$ until NMR analysis.

NMR spectroscopy

A total of 400 µl of urine was mixed with 200 µl of phosphate buffer, pH 7.4, and 50 µl of 29.02 mmol/l 3-trimethylsilyl-2,2,3,3-tetradeuteropropionate (TSP; Sigma-Aldrich, Taufkirchen, Germany) in deuterium oxide as internal standard. NMR experiments were carried out on a 600 MHz Bruker Avance III (Bruker BioSpin GmbH, Rheinstetten, Germany) employing a triple-resonance (^1H, ^{13}C ^{31}P, ^2H lock) cryogenic probe equipped with z-gradients and an automatic cooled sample changer. For each sample, a 1D ^1H NMR spectrum and a 2D ^1H-^{13}C HSQC spectrum were acquired following established protocols.[10] As described previously, NMR signals were assigned by comparison with reference spectra of pure compounds.[10]

Data analysis

Details about data analysis including the used software routines are given in the Supplementary Materials online.

Acid hydrolysis

Selected urine samples of group 1 were treated for 24 h with 6 N HCl at 120 °C. For NMR measurements, the dried residuals were dissolved in 400 µl H_2O, the pH was adjusted, and 200 µl of phosphate buffer, pH 7.4, plus 50 µl of 29.02 mmol/l TSP in deuterium oxide were added.

DISCLOSURE

All the authors declared no competing interests.

ACKNOWLEDGMENTS

This study was supported in part by BayGene, the BMBF, the Robert-Pfleger-Stiftung, the Else Kröner-Fresenius-Stiftung, and the intramural funding program of the Regensburg School of Medicine. We thank Professors Liane Paul and Thomas Gilg at the Institute of Forensic Medicine, Ludwigs-Maximilians-University Munich, Germany, for independent validation of methanol concentrations by headspace-gas chromatography. We are grateful to Stefan Bentink, Claudio Lottaz, Inka Appel, and Professor Rainer Spang for helpful discussions regarding statistics. In addition, we thank Professors Frank Schweda and Armin Kurtz for their critical reading of the article.

SUPPLEMENTARY MATERIAL

Table S1. Compound assignment for the high-resolution 2D ^1H-^{13}C HSQC spectrum of a healthy control urine specimen that is also shown in Figure 1.
Table S2. List of metabolites quantified by NMR.
Table S3. Predictive features used by the SVM for the classification of the medicated ADPKD patients of group 1 versus the healthy control group.
Figure S1. Assignment of methanol.
Supplementary material is linked to the online version of the paper at http://www.nature.com/ki

REFERENCES

1. Igarashi P, Somlo S. Genetics and pathogenesis of polycystic kidney disease. *J Am Soc Nephrol* 2002; **13**: 2384–2398.
2. Gabow PA. Autosomal dominant polycystic kidney disease. *Am J Kidney Dis* 1993; **22**: 511–512.
3. Grantham JJ, Chapman AB, Torres VE. Volume progression in autosomal dominant polycystic kidney disease: the major factor determining clinical outcomes. *Clin J Am Soc Nephrol* 2006; **1**: 148–157.
4. Zheng D, Wolfe M, Cowley BDJ *et al.* Urinary excretion of monocyte chemoattractant protein-1 in autosomal dominant polycystic kidney disease. *J Am Soc Nephrol* 2003; **14**: 2588–2595.
5. Kuehn EW, Park M, Somlo S *et al.* Kidney injury molecule-1 expression in murine polycystic kidney disease. *Am J Physiol Renal Physiol* 2002; **283**: F1326–F1336.
6. Bolignano D, Donato V, Coppolino G *et al.* Neutrophil gelatinase-associated lipocalin (NGAL) as a marker of kidney damage. *Am J Kidney Dis* 2008; **52**: 595–605.
7. Romaker D, Puetz M, Teschner S *et al.* increased expression of secreted frizzled-related protein 4 in polycystic kidneys. *J Am Soc Nephrol* 2009; **20**: 48–56.
8. Kistler AD, Mischak H, Poster D *et al.* Identification of a unique urinary biomarker profile in patients with autosomal dominant polycystic kidney disease. *Kidney Int* 2009; **76**: 89–96.
9. Holmes E, Foxall PJD, Spraul M *et al.* 750 MHz ^1H NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease. *J Pharm Biomed Anal* 1997; **15**: 1647–1659.
10. Gronwald W, Klein MS, Kaspar H *et al.* Urinary metabolite quantification employing 2D NMR spectroscopy. *Anal Chem* 2008; **80**: 9288–9297.
11. Ogborn MR, Sareen S, Prychitko J *et al.* Altered organic anion and osmolyte content and excretion in rat polycystic kidney disease: an NMR study. *Am J Physiol Renal Physiol* 1997; **272**: F63–F69.
12. Ogborn MR, Bankovic-Calic N, Shoesmith C *et al.* Soy protein modification of rat polycystic kidney disease. *Am J Physiol Renal Physiol* 1998; **274**: 541–549.
13. Foxall PJD, Price RG, Jones JK *et al.* High resolution magnetic resonance spectroscopy of cyst fluids from patients with polycystic kidney disease. *Biochim Biophys Acta* 1992; **1138**: 305–314.
14. van den Dool SW, Wasser MN, de Fijter JW *et al.* Functional renal volume: quantitative analysis at gadolinium-enhanced MR angiography—feasibility study in healthy potential kidney donors. *Radiology* 2005; **236**: 189–195.
15. Cohen EI, Kelly SA, Edye M *et al.* MRI estimation of total renal volume demonstrates significant association with healthy donor weight. *Eur J Radiol* 2009; **71**: 283–287.
16. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC-Bioinformatics* 2006; **7**: 91.
17. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B* 1995; **57**: 289–300.
18. Böger CA, Haak T, Götz AK *et al.* Effect of ACE and AT-2 inhibitors on mortality and progression to microalbuminuria in a nested case-control study of diabetic nephropathy in diabetes mellitus type 2: results from the GENDIAN study. *Int J Clin Pharmacol Ther* 2006; **44**: 364–374.

19. Gansevoort RT, de Zeeuw D, de Jong PE. ACE inhibitors and proteinuria. *Pharm World Sci* 1996; **18**: 204–210.
20. Ecder T, Schrier RW. Cardiovascular abnormalities in autosomal-dominant polycystic kidney disease. *Nat Rev Nephrol* 2009; **5**: 221–228.
21. Obermüller N, Kränzlin B, Blum WF. An endocytosis defect as a possible cause of proteinuria in polycystic kidney disease. *Am J Physiol Renal Physiol* 2001; **280**: F244–F253.
22. Umbreit EC, Childs MA, Patterson DE *et al.* Percutaneous nephrolithotomy for large or multiple upper tract calculi and autosomal dominant polycystic kidney disease. *J Urol* 2010; **183**: 183–187.
23. Nicholson JK, Timbrell JA, Sadler PJ. Proton NMR spectra of urine as indicators of renal damage mercury-induced nephrotoxicity in rats. *Mol Pharmacol* 1985; **27**: 644–651.
24. Herbig K, Chiang E-P, Lee L-R *et al.* Cytoplasmic serine hydroxymerhyltransferase mediates competition between folate-dependent deoxyribonucleotide and S-adenosylmethionine biosyntheses. *J Biol Chem* 2002; **277**: 38381–38389.
25. Veeraragavan K, Gagnon C. Mammalian protein methylesterase. *Biochem J* 1989; **260**: 11–17.
26. Chene L, Bourget L, Vinay P *et al.* Preferential localization of a protein methylating-demethylating system in proximal tubules of rat kidney. *Arch Biochem Biophys* 1982; **213**: 299–305.
27. Sjöström M, Stenstrom K, Eneling K *et al.* SIK1 is part of a cell sodium-sensing network that regulates active sodium transport through a calcium-dependent process. *Proc Natl Acad Sci USA* 2007; **104**: 16922–16927.
28. Wilson PD. Polycystic kidney disease. *N Engl J Med* 2004; **350**: 151–163.
29. Taylor SL, Ganti S, Bukanov NO *et al.* A metabolomics approach using juvenile cystic mice to identify urinary biomarkers and altered pathways in polycystic kidney disease. *Am J Physiol Renal Physiol* 2010; **298**: F909–F922.
30. Kopple JD. Phenylalanine and tyrosine metabolism in chronic kidney failure. *J Nutr* 2007; **137**: 1586S–1590S.
31. Mardach R, Zempleni J, Wolf B *et al.* Biotin dependency due to a defect in biotin transport. *J Clin Invest* 2002; **109**: 1617–1623.
32. Gupta A, Dwivedi M, Mahdi AA *et al.* ^{1}H-nuclear magnetic resonance spectroscopy for identifying and quantifying common uropathogenes: a metabolic approach to the urinary tract infection. *BJUI* 2009; **104**: 236–244.
33. Petrarulo M, Marangella M, Bianco O *et al.* Ion-chromatographic determination of l-tartrate in urine samples. *Clin Chem* 1991; **37**: 90–93.
34. Lord RS, Burdette CK, Bralley JA. Significance of urinary tartaric acid. *Clin Chem* 2005; **51**: 672–673.
35. Ravine D, Gibson RN, Walker RG *et al.* Evaluation of ultrasonographic diagnostic criteria for autosomal dominant polycystic kidney disease. *Lancet* 1994; **343**: 824–827.
36. Gunay-Aygun M, Font-Montgomery E, Lukose L *et al.* Correlation of kidney function, volume and imaging findings, and PKHD1 mutations in 73 patients with autosomal recessive polycystic kidney disease. *Clin J Am Soc Nephrol* 2010; **5**: 972–984.
37. Banas MC, Banas B, Wolf J *et al.* Smoking behaviour of patients before and after renal transplantation. *Nephrol Dial Transplant* 2008; **23**: 1442–1446.
38. Levey AS, Coresh J, Greene T *et al.* Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Ann Intern Med* 2008; **145**: 247–254.