

A metabolomic study of low estimated GFR in non-proteinuric type 2 diabetes mellitus

D. P. K. Ng · A. Salim · Y. Liu · L. Zou · F. G. Xu ·
S. Huang · H. Leong · C. N. Ong

Received: 14 July 2011 / Accepted: 19 September 2011 / Published online: 25 October 2011
© Springer-Verlag 2011

Abstract

Aims/hypothesis We carried out a urinary metabolomic study to gain insight into low estimated GFR (eGFR) in patients with non-proteinuric type 2 diabetes.

Methods Patients were identified as being non-proteinuric using multiple urinalyses. Cases ($n=44$) with low eGFR and controls ($n=46$) had eGFR values <60 and ≥ 60 ml min⁻¹ 1.73 m⁻², respectively, as calculated using the Modification of Diet in Renal Disease formula. Urine samples were analysed by liquid chromatography/mass spectrometry (LC/MS) and GC/MS. False discovery rates were used to adjust for multiple hypotheses testing, and selection of metabolites that best predicted low eGFR status was achieved using least absolute shrinkage and selection operator logistic regression. **Results** Eleven GC/MS metabolites were strongly associated with low eGFR after correction for multiple hypotheses

testing (smallest adjusted p value= 2.62×10^{-14} , largest adjusted p value= 3.84×10^{-2}). In regression analysis, octanol, oxalic acid, phosphoric acid, benzamide, creatinine, 3,5-dimethoxymandelic amide and *N*-acetylglutamine were selected as the best subset for prediction and allowed excellent classification of low eGFR (AUC=0.996). In LC/MS, 19 metabolites remained significant after multiple hypotheses testing had been taken into account (smallest adjusted p value= 2.04×10^{-4} , largest adjusted p value= 4.48×10^{-2}), and several metabolites showed stronger evidence of association relative to the uraemic toxin, indoxyl sulphate (adjusted p value= 3.03×10^{-2}). The potential effect of confounding on the association between metabolites was excluded.

Conclusions/interpretation Our study has yielded substantial new insight into low eGFR and provided a collection of potential urinary biomarkers for its detection.

Electronic supplementary material The online version of this article (doi:10.1007/s00125-011-2339-6) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

Keywords Low eGFR · Metabolomics · Non-proteinuric · Type 2 diabetes · Uraemic toxin

D. P. K. Ng (✉) · A. Salim · Y. Liu · L. Zou · F. G. Xu ·
S. Huang · C. N. Ong

Department of Epidemiology and Public Health,
Yong Loo Lin School of Medicine,
National University of Singapore,
16 Medical Drive MD3,
Singapore 117597, Republic of Singapore
e-mail: ephnpkd@nus.edu.sg

H. Leong
Clinical Services, National Healthcare Group Polyclinics,
Singapore, Republic of Singapore

C. N. Ong (✉)
NUS Environmental Research Institute,
National University of Singapore,
5A, Engineering Drive,
Singapore 117411, Republic of Singapore
e-mail: ephocn@nus.edu.sg

Abbreviations

ACR	Albumin/creatinine ratio
eGFR	Estimated GFR
FDR	False discovery rate
FMOC	9-Fluorenylmethoxycarbonyl
LASSO	Least absolute shrinkage and selection operator
LC	Liquid chromatography
MDRD	Modification of Diet in Renal Disease
OPLS-DA	Orthogonal partial least-squares discriminant analysis
PARP-1	Poly(ADP-ribose) polymerase-1
PC	Principal component
r _{cf}	Relative centrifugal force

ROC Receiver operating characteristic
SDCS Singapore Diabetes Cohort Study

Introduction

It has been variously reported that some diabetic patients have low renal function (typically expressed as estimated GFR [eGFR]) even in the absence of proteinuria [1–5]. However, little else is known about the risk factors or mechanisms associated with low eGFR in patients with either type 1 [5] or type 2 diabetes [1–4]. Metabolomics is a technological platform for the identification and quantification of the metabolome, the collection of all small molecules present in an organism or biological sample [6]. Metabolomics has been greatly facilitated by recent developments in mass detectors, allowing techniques such as liquid chromatography/MS (LC/MS) and GC/MS to support analysis of the metabolome [6].

Interrogation of the metabolome potentially offers unprecedented insights into a disease or phenotype and provides initial access to biomarker and pathway discovery [7–9]. We applied these improved techniques to probe the metabolomes of Chinese type 2 diabetic patients with low eGFR. We observed striking associations between low eGFR and several urinary metabolites, which extend beyond the few known uraemic toxins. Besides offering new leads on the possible mechanisms underlying low eGFR, these metabolites could serve as novel biomarkers for the detection of chronic kidney disease.

Methods

Patients and urine samples All patients for this study were from the Singapore Diabetes Cohort Study (SDCS). Briefly, the recruitment process of SDCS was as follows. Since 2004, all patients previously diagnosed as having type 2 diabetes and treated at primary care facilities of the National Healthcare Group Polyclinics in Singapore were invited to join SDCS. Patients with a history of mental illness were excluded. Of the patients approached, 91% agreed to participate in the study and formed part of the cohort. Consenting patients completed a questionnaire to elicit information on demographics, lifestyle factors and medical family history and also had their physical measurements taken. Random (not first morning) spot urine specimens were typically collected in the morning at the outpatient polyclinic and used for laboratory analyses. Medical records were reviewed to obtain information on their metabolic control and the presence of co-morbidities and complications including any history of non-diabetic

kidney disease. Lipid measurements were performed on fasting blood samples.

The research protocol was approved by both the National University of Singapore Institutional Review Board and the National Healthcare Group Domain-Specific Review Board, and patients participating in this cohort gave informed consent.

Definitions of non-proteinuria and low eGFR Patients in this metabolomic study were identified as being non-proteinuric using multiple spot urine samples. To thoroughly exclude the presence of proteinuria, urine samples were required to test negative on Labstix (Bayer Corporation, Elkhart, IN, USA) or Micral-Test (Boehringer Mannheim, Mannheim, Germany) or have an albumin/creatinine ratio (ACR) <3.5 µg/µmol (Exocell, Philadelphia, PA, USA) on at least two of the last three urinalyses. Most of the patients were therefore likely to be normoalbuminuric, although it was possible for some to have microalbuminuria especially if this was transient. eGFR was calculated using the simplified Modification of Diet in Renal Disease (MDRD) equation, where $eGFR (\text{ml min}^{-1} 1.73 \text{ m}^{-2}) = 186.3 \times (\text{plasma creatinine in } \mu\text{mol/l} \times 0.011)^{-1.154} \times (\text{age in years})^{-0.203} \times (0.742 \text{ for women}) \times (1.21 \text{ if subject is black})$ [10]. Cases ($n=44$) were defined as patients with $eGFR < 60 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$, and controls ($n=46$) had eGFR values $\geq 60 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$. As a history of cataract was strongly associated with low eGFR in SDCS (data not shown), presence of this complication was used as an exclusion criterion to eliminate potential confounding.

Metabolomic analysis using GC/MS Urine samples (20 µl) were incubated with 20 µl (10 mg/ml) urease enzyme for 30 min at 37°C. Then urease and other proteins were precipitated with 180 µl ice-cold methanol, which contained 10 µg/ml 9-fluorenylmethoxycarbonyl (Fmoc)-glycine as an internal standard. After separation by centrifugation (16 relative centrifugal force [rcf] × 10 min, 4°C), 100 µl supernatant fraction was dried under nitrogen and derivatised with 150 µl methoxamine (50 µg/ml in pyridine, 37°C × 2 h) followed by 150 µl MSTFA (37°C × 16 h). After centrifugation (4°C, 6 rcf × 1 min), the supernatant fraction was injected into GC/MS. The derivatised sample (1.0 µl) was introduced by splitless injection with an Agilent 7683 Series autosampler into an Agilent 6890 GC System (both from Agilent Technologies, Santa Clara, CA, USA) equipped with a fused-silica capillary column HP-5MSI (30 m × 0.25 mm i.d., 0.25 µm film thickness) as reported previously [11]. The inlet temperature was set at 250°C. Helium was used as the carrier gas at a constant flow rate of 1.0 ml/min. The column effluent was introduced into the ion source of an Agilent 5973 Mass Selective Detector (Agilent Technologies). The transfer line temperature was set at 280°C and the ion source temperature

at 230°C. The mass spectrometer was operated in electron impact mode (70 eV). Data acquisition was performed in full scan mode from m/z 50 to 550 with a scan time of 0.5 s. The compounds were identified by comparison of mass spectra and retention time with those of reference standards, and those available in libraries (NIST 0.5). A total of 106 peaks with specific retention times in GC/MS analyses were detected in this study.

Metabolomic analysis using LC/MS The urine samples were diluted 1:1 with methanol (containing 10 µg/ml FMOG-glycine as an internal standard) before being vortex-mixed for 3 min. After separation by centrifugation (16 rcf × 10 min, 4°C), the supernatant fraction was injected for LC/MS analysis. LC/MS analysis was performed on an Agilent 1200 HPLC system (Agilent Technologies, Waldbronn, Germany) equipped with a 6410 QQQ triple quadrupole mass detector and managed by a MassHunter workstation. The column used for the separation was an Agilent rapid resolution HT Zorbax SB-C18 (2.1 × 50 mm, 1.8 µm; Agilent Technologies, Santa Clara, CA, USA). The oven temperature was set at 50°C. The gradient elution involved a mobile phase consisting of (A) 0.1% formic acid in water and (B) 0.1% formic acid in methanol. The initial condition was set at 5% of B. The following solvent gradient was applied: from 5% B to 100% B within 20 min, then hold for 2 min. Flow rate was set at 0.2 ml/min, and 5 µl of samples was injected. The electrospray ionisation mass spectra were acquired in positive and negative ion mode. The ion spray voltage was set at 4,000 V. The heated capillary temperature was maintained at 350°C. The drying gas and nebuliser nitrogen gas flow rates were 10 l/min and 207×10^3 Pa, respectively. For full scan mode analysis, spectra were stored from m/z 100 to 1,000. A total of 144 peaks with specific retention times in LC/MS analyses were detected in this study. The compounds were searched for using the Human Metabolome Database (www.hmdb.ca) using ion mass and further identified by either MS/MS fragmentation pattern or reference standards.

Metabolomic data preprocessing Each chromatogram obtained from GC/MS and LC/MS analysis was processed for baseline correction and peak area calculation manually. The data were combined into a single matrix by aligning peaks with the same mass and retention time for GC/MS and LC/MS data, respectively. The area of each peak was normalised to that of the internal standard in each dataset. There was no further normalisation of the metabolites with respect to creatinine because the level of this metabolite was different between cases and controls (Tables 2 and 3).

Statistical analysis Statistical comparison of clinical characteristics between cases and controls was performed using

two-sample t tests in the case of quantitative traits. In the event where the data distribution deviated from normal distribution, Mann–Whitney tests were used. For qualitative traits, comparison between cases and controls was performed using Fisher's exact tests.

The preprocessed metabolomic data were exported into Soft Independent Modeling of Class Analysis (SIMCA)-P (version 11.0; Umetrics AB, Umea, Sweden) for orthogonal partial least-squares discriminant analysis (OPLS-DA). To compare median signal intensities of the metabolites between cases and controls, the Mann–Whitney test was applied to each metabolite separately. The resultant p values for all metabolites were subsequently adjusted to account for multiple hypotheses testing. The false discovery rate (FDR) method of Benjamini and Yekutieli [12] was used to perform the adjustment.

To determine how well these metabolites performed in separating cases from controls, we used two approaches. In the first approach, principal component (PC) analysis was performed for metabolites with a statistically significant association in the univariate analysis. The first and second PC scores were then used as predictors in the logistic regression model, with case status as the outcome. To ensure that the PC estimates were robust, any outlying observations that lay more than four standard deviations from the mean were removed. The receiver operating characteristic (ROC) curve for the logistic model was calculated, and the AUC was used to assess the quality of prediction, with AUC closer to 1 indicating better performance.

The second approach involved selecting the best subset of metabolites that can be used to predict case status. Least absolute shrinkage and selection operator (LASSO) logistic regression [13] was used, with the optimal LASSO estimate determined using leave-one-out cross-validation. Briefly, LASSO logistic regression is very similar to the ordinary logistic regression, except that LASSO places restriction on the number of metabolites with non-zero regression coefficients. The LASSO algorithm optimally selects this subset of metabolites with non-zero coefficients. In our study, leave-one-out cross-validation was used to select this subset.

Results

Patient characteristics Cases and controls were comparable in most clinical characteristics except that cases were older at the time of recruitment and had higher serum creatinine values, as would be expected as these variables were directly used to compute eGFR values in the MDRD equation (Table 1). Cases were also older than controls at the time of diabetes diagnosis, but this difference was

Table 1 Clinical characteristics of cases and controls

Variable	Cases (n=44)	Controls (n=46)	p value
Age (years)	67.93±8.96	60.80±9.39	0.0004
Male	19 (43.2)	13 (28.3)	0.1390
Age at diabetes diagnosis (years)	58.27±10.90	52.18±10.58	0.0129
Duration of diabetes (years)	9.46±9.60	8.50±7.72	0.8975
Current modality for diabetes treatment ^a			0.2240
Diet and exercise only	13 (30.2)	9 (21.4)	
Oral hypoglycaemic agent with/without diet and exercise	25 (58.1)	32 (76.2)	
Oral hypoglycaemic agent + insulin	4 (9.3)	1 (2.3)	
Insulin alone	1 (2.3)	0 (0.0)	
HbA _{1c} (%)	7.4±0.96	7.07±0.67	0.0701
HbA _{1c} (mmol/mol)	57.37±7.34	53.77±10.50	
Systolic blood pressure (mmHg)	136.48±14.98	133.33±11.74	0.2963
Diastolic blood pressure (mmHg)	74.82±8.82	76.17±6.86	0.4167
Mean arterial pressure (mmHg)	95.387±8.90	95.22±6.54	0.9857
Triacylglycerol (mmol/l)	1.64±0.81	1.40±0.50	0.0941
Cholesterol (mmol/l)	4.61±0.76	4.68±0.89	0.7459
LDL-cholesterol (mmol/l)	2.60±0.60	2.77±0.77	0.2560
HDL-cholesterol (mmol/l)	1.27±0.29	1.28±0.30	0.8860
Serum creatinine (μmol/l)	113.75±23.31	74.37±16.36	<0.0001
Smoking			0.817
Yes	4 (9.01)	3 (6.5)	
Never	33 (75.0)	37 (80.4)	
Ex-smoker	7 (15.9)	6 (13.0)	
Waist hip ratio	0.90±0.06	0.87±0.06	0.0705
BMI (kg/m ²)	24.92±3.69	24.86±3.68	0.9426
eGFR (ml min ⁻¹ 1.73 m ⁻²)	51.66±6.88	83.61±16.06	NA
ACR (μg/μmol)			
Mean ± SD	2.45±5.69 ^b	2.45±3.64	0.999
Median (IQR)	1.06 (1.63)	1.31 (2.14)	0.168

Continuous variables are presented as mean±SD, categorical variables as n (%)

^aTreatment information is missing for one case and four controls

^bMean and subsequent SD was computed after removal of one outlier

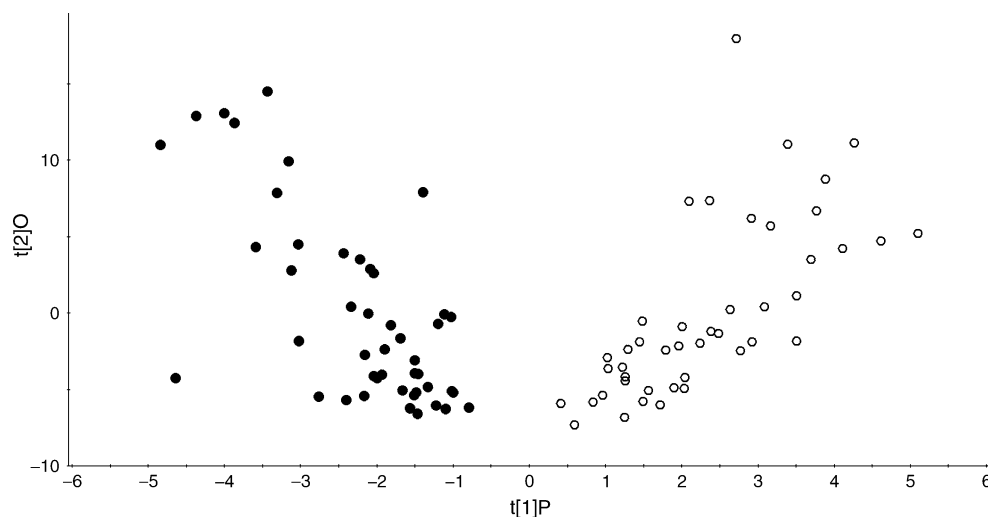
NA, not applicable

borderline significant (unadjusted $p=0.0129$). ACR values were similar between cases and controls (Table 1).

GC/MS analyses OPLS-DA revealed that cases could be clearly segregated from controls on the basis of the 106 peaks detected by GC/MS (Fig. 1). Univariate analyses of the metabolite signal intensities revealed striking associations between 24 metabolites and low eGFR (unadjusted $p<0.05$; Table 2). Of these 24 associations, 11 remained statistically significant after correction to account for multiple hypotheses testing (adjusted $p<0.05$). In particular, the p values associated with six metabolites (oxalic acid, octanol, *N*-acetylglutamine, 3,5-dimethoxymandelic amide, benzamide and phosphoric acid) were very small, with the largest p value= 7.28×10^{-5} (phosphoric acid) and the smallest p value= 2.62×10^{-14} (oxalic acid). Urinary creatinine was also higher in cases than controls ($p=6.46\times 10^{-8}$; Table 2).

PC analysis was next performed to determine the clustering of the 24 metabolites. The first two PCs of GC metabolites explained 56.5% of correlations between these metabolites. The first PC axis is a contrast between phosphoric acid (metabolite 4) and 3,5-dimethoxymandelic amide (metabolite 13) on the one hand and the rest of the metabolites (Electronic supplementary material [ESM] Fig. 1). The second axis was largely responsible for the separation of cases and controls, and could be characterised by a group of metabolites that tended to have higher signal intensities in controls (phosphoric acid [metabolite 4], 3,5-dimethoxymandelic amide [metabolite 13], benzamide [metabolite 7], L-serine [metabolite 6], D-glucuronic [metabolite 20], oxalic acid [metabolite 3], succinic acid [metabolite 5] and uric acid [metabolite 22]) and metabolites that were of higher signal intensity among cases, including creatinine (metabolite 10), *N*-acetylglutamine (metabolite 23) and octanol (metabolite 2). On the basis

Fig. 1 OPLS-DA score plot obtained from GC/MS data based on 106 peaks. Black circles, controls ($n=46$); white circles, cases ($n=44$). The x -axis $t[1]$ and y -axis $t[2]$ indicate the first PC and second PC, respectively



of these first and second PCs, the cases could be separated quite well from controls. With the diagonal line in ESM Fig. 1 used as a simple rule for separating cases from controls, all controls were placed below the line and all but five cases were located above it. In logistic regression where the first and second PC scores were used to classify participants into cases and controls, the ROC curve revealed very good discriminatory power ($AUC=0.999$).

The above classification was achieved by using all 24 metabolites. However, from the univariate analysis (Table 2), it is clear that only a subset of metabolites was significantly associated after adjustment for multiple testing, and, furthermore, metabolites tended to show clustering, i.e., their signal intensities tended to vary together across the different samples, as was evident in the PC plots. We therefore next determined whether the same level of classification could be achieved with a smaller select group of metabolites. With the use of LASSO logistic regression, the following metabolites were selected as the best subset for case prediction: octanol, oxalic acid, phosphoric acid, benzamide, creatinine, 3,5-dimethoxymandelic amide and *N*-acetylglutamine.

Using these seven metabolites to predict case status, a classification ($AUC=0.995$) was achieved that was as good as that previously derived using all 24 metabolites (ESM Fig. 2). This led to the conclusion that higher signal intensities of octanol, creatinine and *N*-acetylglutamine and lower signal intensities of oxalic acid, phosphoric acid, benzamide and 3,5-dimethoxymandelic amide were all independent predictors of low eGFR. Among the clinical variables, only age at diabetes diagnosis, age at recruitment and serum creatinine were statistically significant in univariate analyses (Table 1). In LASSO regression, only serum creatinine was selected into the model when the significant metabolites listed above were included. The AUC of the model with serum creatinine added is 0.996, which is very similar to the model without serum creatinine ($AUC=0.995$).

LC/MS analyses OPLS-DA revealed clear segregation of cases from controls on the basis of 144 peaks detected in LC/MS (Fig. 2). Univariate analyses revealed a total of 32 metabolites that were significantly associated with low eGFR (unadjusted $p<0.05$). Of these, 19 remained significantly associated after multiple hypotheses testing had been taken into account (adjusted $p<0.05$; Table 3). Of these, 17 were detected in positive ion mode, while two were determined under the negative ion mode, including indoxyl sulphate, a well-established uraemic toxin [14]. Relative to indoxyl sulphate (adjusted $p=3.03\times 10^{-2}$), several metabolites clearly showed stronger evidence of statistical association with low eGFR, with p values that were at least a magnitude smaller. These included 4-methoxyphenylacetic acid, *N*⁶-acetyl-L-lysine, chondroitin sulphate, citric acid, phenylacetyl-L-glutamine, 2-deoxyuridine and deoxyypyridinoline (Table 3).

PC analyses of the LC/MS metabolites using the first two PCs explained 53.5% of correlations. The metabolites had higher signal intensities among cases compared with controls, with three clusters of metabolites being evident (ESM Fig. 3). Compared with the GC/MS results, the separation of cases from controls was less optimal, with only moderate discriminatory power ($AUC=0.777$), as observed using the ROC curve (data not shown). LASSO regression revealed a subset of seven metabolites (*N*⁶-acetyl-L-lysine, caffeine, 4-methoxyphenylacetic acid, chondroitin sulphate, hyocholic acid/cholic acid/ursocholic acid, phenyl sulphate and α -hydroxyhippuric acid) that best predicted case status with an AUC of 0.870, an improvement on that achieved using PC scores (ESM Fig. 4). The potential effect of confounding on the association between metabolites and low eGFR by patient clinical variables was excluded. In univariate analyses, age at diabetes diagnosis, age at examination and serum creatinine were statistically significant at the 5% significance level (Table 1). However,

Table 2 Univariate analysis of metabolite signal intensities measured by GC/MS

Metabolite	Median (SD)		Unadjusted <i>p</i> value ^a	Adjusted <i>p</i> value ^b
	Control	Cases		
Oxalic acid	0.263 (0.457)	0.000 (0.000)	4.71×10^{-17}	2.62×10^{-14}
Octanol	0.287 (0.204)	0.899 (0.596)	9.65×10^{-13}	2.68×10^{-10}
3,5-Dimethoxymandelic amide	0.764 (0.464)	0.292 (0.181)	4.75×10^{-11}	6.98×10^{-9}
<i>N</i> -Acetylglutamine	0.246 (0.735)	1.811 (2.325)	5.02×10^{-11}	6.98×10^{-9}
Creatinine	2.639 (3.237)	11.224 (5.723)	6.97×10^{-10}	6.46×10^{-8}
Benzamide	0.102 (0.026)	0.079 (0.020)	7.63×10^{-8}	5.80×10^{-6}
Phosphoric acid	43.706 (18.812)	29.391 (12.001)	1.18×10^{-6}	7.28×10^{-5}
2-Hydroxyadipic acid	0.034 (0.025)	0.056 (0.039)	1.75×10^{-5}	8.86×10^{-4}
Ribonic acid	0.353 (0.610)	1.168 (1.163)	1.80×10^{-4}	7.71×10^{-3}
Hydroxyphenylacetic acid	0.123 (0.216)	0.261 (0.380)	6.98×10^{-4}	2.43×10^{-2}
Sarcosine	0.144 (0.247)	0.183 (0.069)	1.18×10^{-3}	3.84×10^{-2}
Salicyluric acid	0.849 (2.176)	2.120 (4.442)	2.10×10^{-3}	6.50×10^{-2}
Uric acid	1.847 (1.598)	1.250 (1.435)	2.34×10^{-3}	6.85×10^{-2}
β -Hydroxybutyric acid	0.035 (0.061)	0.070 (0.068)	5.22×10^{-3}	1.45×10^{-1}
<i>cis</i> -Aconitic acid	0.128 (0.258)	0.233 (0.518)	7.02×10^{-3}	1.86×10^{-1}
2-Ketogluconic acid	0.073 (0.203)	0.164 (0.258)	9.35×10^{-3}	2.36×10^{-1}
Dodecanoic acid (C12:0)	0.346 (0.086)	0.397 (0.236)	1.48×10^{-2}	3.42×10^{-1}
Threitol	0.340 (0.451)	0.689 (0.751)	1.55×10^{-2}	3.44×10^{-1}
3-Hydroxyhippuric acid	0.308 (0.707)	0.565 (0.719)	1.97×10^{-2}	4.20×10^{-1}
Succinic acid	0.180 (0.177)	0.137 (0.092)	2.10×10^{-2}	4.32×10^{-1}
D-Glucuronic acid	0.088 (0.481)	0.124 (0.267)	3.05×10^{-2}	5.47×10^{-1}
Xylitol	0.762 (1.817)	1.758 (1.782)	3.73×10^{-2}	6.21×10^{-1}
Pseudouridine	0.903 (2.307)	1.944 (2.482)	3.66×10^{-2}	6.21×10^{-1}
L-Serine	0.106 (0.439)	0.184 (0.446)	4.71×10^{-2}	7.27×10^{-1}

Only metabolites with unadjusted *p* values of <0.05 are shown

^a*p* value from Mann–Whitney test comparing the medians

^bAdjusted for multiple hypotheses testing by controlling for FDR [12]

only age at recruitment and serum creatinine were selected by LASSO regression when the significant metabolites

listed above were included. The AUC for this model where age at recruitment and serum creatinine were added to the

Fig. 2 OPLS-DA score plot obtained from LC/MS data on 144 peaks. Black circles, controls ($n=46$); white circles, cases ($n=44$). The *x*-axis *t*[1] and *y*-axis *t*[2] indicate the first PC and second PC, respectively

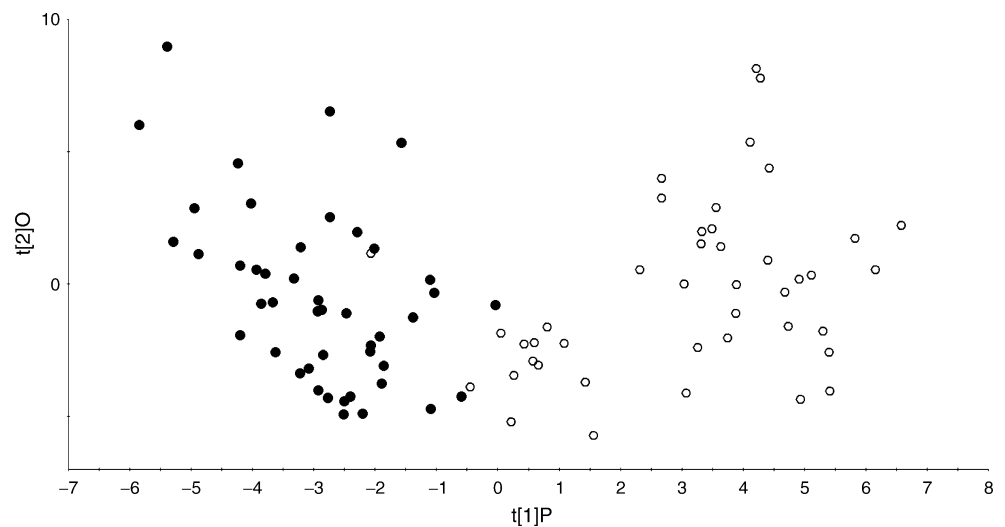


Table 3 Univariate analysis of metabolite signal intensities measured by LC/MS

Metabolite	Median (SD)		Unadjusted <i>p</i> value ^a	Adjusted <i>p</i> value ^b
	Control	Cases		
Positive ion mode				
4-Methoxyphenylacetic acid ^c	0.022 (0.031)	0.080 (0.076)	5.11×10^{-7}	2.04×10^{-4}
<i>N</i> ⁶ -Acetyl-L-lysine ^c	0.16 (0.069)	0.258 (0.249)	2.16×10^{-6}	3.82×10^{-4}
Chondroitin sulphate ^c	0.057 (0.099)	0.162 (0.213)	2.39×10^{-6}	3.82×10^{-4}
Citric acid ^d	0.261 (0.241)	0.441 (0.435)	3.60×10^{-6}	4.80×10^{-4}
Phenylacetyl-L-glutamine ^c	0.827 (0.993)	1.693 (2.117)	4.91×10^{-6}	4.91×10^{-4}
2-Deoxyuridine ^d	0.792 (0.565)	1.554 (1.378)	6.18×10^{-6}	5.49×10^{-4}
Deoxyypyridinoline ^c	0.052 (0.059)	0.119 (0.140)	1.52×10^{-5}	8.07×10^{-4}
Dehydrotestosterone glucuronide/retinyl-β-glucuronide ^c	0.091 (0.105)	0.157 (0.162)	3.28×10^{-4}	1.09×10^{-2}
<i>N</i> -Acetylspermine ^c	0.056 (0.080)	0.168 (0.185)	3.92×10^{-4}	1.21×10^{-2}
Creatinine ^d	1.105 (0.753)	1.581 (1.260)	5.19×10^{-4}	1.47×10^{-2}
Sphingosine ^c	0.112 (0.187)	0.252 (0.671)	5.51×10^{-4}	1.47×10^{-2}
3,7-Dimethyluric acid ^c	0.131 (0.252)	0.408 (0.462)	6.75×10^{-4}	1.69×10^{-2}
10-Nitrolinoleic acid ^c	0.045 (0.097)	0.098 (0.145)	8.34×10^{-4}	1.85×10^{-2}
2,6-Dimethylheptanoyl carnitine nonanoylcarnitine ^c	0.085 (0.092)	0.288 (0.202)	8.52×10^{-4}	1.85×10^{-2}
Hyochoic acid/cholic acid/ursochoic acid ^c	0.371 (0.418)	0.501 (0.952)	8.09×10^{-4}	1.85×10^{-2}
Phosphoribosyl-formylglycine amidine ^c	0.065 (0.124)	0.128 (0.342)	1.42×10^{-3}	2.87×10^{-2}
IDP ^c	0.000 (0.183)	0.052 (0.161)	2.52×10^{-3}	4.48×10^{-2}
Caffeine ^d	0.139 (0.189)	0.307 (0.275)	3.74×10^{-3}	5.86×10^{-2}
α-Hippuric acid	2.186 (2.540)	3.543 (4.859)	6.06×10^{-3}	9.14×10^{-2}
Pregnanediol-3-glucuronide ^c	0.000 (0.067)	0.045 (0.168)	8.23×10^{-3}	1.15×10^{-1}
Suberylglycine ^c	0.180 (0.161)	0.284 (0.347)	8.92×10^{-3}	1.23×10^{-1}
Vinylacetylglycine ^c	0.525 (0.467)	0.715 (1.211)	1.65×10^{-2}	1.97×10^{-1}
Phytyl diphosphate ^c	0.023 (0.076)	0.073 (0.150)	1.71×10^{-2}	2.01×10^{-1}
Deoxycorticosterone/docosapentaenoic acid ^c	0.000 (0.123)	0.000 (0.206)	2.40×10^{-2}	2.69×10^{-1}
Galactose ^d	0.210 (0.964)	0.305 (0.355)	4.28×10^{-2}	4.44×10^{-1}
Negative ion mode				
Androsterone glucuronide/etiocholanolone glucuronide ^c	0.345 (0.650)	0.597 (0.685)	1.44×10^{-3}	2.87×10^{-2}
Indoxyl sulphate ^c	10.484 (17.112)	18.120 (20.359)	1.56×10^{-3}	3.03×10^{-2}
α-Hydroxyhippuric acid ^d	0.606 (2.977)	1.688 (4.975)	3.31×10^{-3}	5.50×10^{-2}
Urothion ^c	1.162 (2.281)	1.594 (2.050)	5.76×10^{-3}	8.86×10^{-2}
Salicyluric acid ^c	0.297 (0.863)	0.794 (1.221)	6.27×10^{-3}	9.29×10^{-2}
3-Oxochoic acid ^c	0.603 (0.822)	0.890 (1.091)	2.42×10^{-2}	2.69×10^{-1}
Phenyl sulphate	2.592 (4.019)	3.142 (7.253)	4.80×10^{-2}	4.84×10^{-1}

Only metabolites with unadjusted *p* values of <0.05 are shown

^a *p* value from Mann–Whitney test comparing the medians

^b Adjusted for multiple hypotheses testing by controlling for FDR [12]

^c Identified by MS/MS fragmentation pattern

^d Identified by reference standard

IDP, inosine diphosphate

model was 0.978, which represents a significant improvement over the model with metabolites only (AUC=0.870).

Validation In an attempt to provide some kind of validation for the above GC/MS and LC/MS results, 45 individuals (23 controls, 22 cases) were randomly selected from the 90

participants and used to discover the important metabolites. These metabolites were then validated by calculating the AUC based on the 45 remaining unselected participants. This random selection was repeated ten times, yielding a range of AUC values. For GC/MS metabolites, the AUC for the validation set was consistent and ranged from 0.934

to 1.000. For the LC/MS metabolites, the range was less optimal, with AUC values of 0.477–1.000.

Discussion

Our study has succeeded in unveiling a wealth of information linking a number of urinary metabolites with low eGFR. Moreover, these novel and statistically robust associations were found in diabetic patients who were persistently non-proteinuric and thus would conventionally have been regarded to be at low risk of chronic kidney disease.

The candidate metabolites for low eGFR extended well beyond the few widely acknowledged uraemic toxins. Indeed, while we did find significant associations with urinary levels of uraemic toxins such as indoxyl sulphate, creatinine and the methoxylated form of phenylacetic acid, at least 13 other metabolites exhibited much stronger evidence of an association with low eGFR, with *p* values that were at least a magnitude smaller when compared with indoxyl sulphate (Tables 2 and 3). These new candidate biomarkers include oxalic acid, octanol, 3,5-dimethoxymandelic amide, *N*-acetylglutamine, benzamide, phosphoric acid, 2-hydroxyadipic acid, *N*⁶-acetyl-L-lysine, chondroitin sulphate, citric acid, phenylacetyl-L-glutamine, 2-deoxyuridine and deoxyypyridinoline.

Our literature review suggested that only a few of the metabolites, such as creatinine, oxalic acid and phosphoric acid, had been earlier linked to renal disease in humans [15, 16]. While raised serum creatinine serves as an indicator of renal function, urinary creatinine is primarily used in the context of a 24 h creatinine clearance test to measure GFR. Regarding oxalic and phosphoric acids, high urinary levels of these metabolites predisposed to the formation of kidney stones [17], but the effect of kidney stones on reducing renal function remained unclear [18]. It was noteworthy that the urinary levels of these metabolites were actually lower in the cases than controls in our study.

In particular, oxalic acid was detected in 44 out of 46 control participants (95.6%) but in none of the cases (ESM Fig. 5). This striking observation may be consistent with the systemic retention of this metabolite in the presence of low eGFR. In this connection, it is interesting to note that increased plasma oxalic acid levels had been previously linked with chronic kidney disease [19].

Chondroitin sulphate is a type of glycosaminoglycan that is covalently linked to proteins, forming proteoglycans. It is reportedly absent in the normal glomerular basement membrane, but its content increases soon after the onset of experimental diabetes [20, 21]. Semi-quantitative analysis suggested a reduction in total urinary glycosaminoglycan

content in diabetic animal models compared with non-diabetic controls with increasing diabetes duration and this appeared to correlate with albuminuria [22]. In normoalbuminuric type 1 diabetic patients, increased urinary excretion of glycosaminoglycans, including chondroitin sulphate, was related to poorer glycaemic control and longer diabetes duration, although the association between the excretion of these metabolites and eGFR was not reported [23, 24].

An important highlight of our study is the previously unrecognised roles of the remaining metabolites in the modulation of eGFR. Even so, there may be some underlying biological plausibility. For example, cases with low eGFR had lower levels of benzamide compared with controls. Benzamide has been shown to be an endogenous inhibitor of poly(ADP-ribose) polymerase-1 (PARP-1) [25, 26], a key enzyme that has been strongly implicated in causing diabetes-associated endothelial dysfunction [27]. Pharmacological inhibition of PARP-1 ameliorated various features associated with nephropathy in both experimental models of type 1 and type 2 diabetes, including albuminuria [28, 29]. Urinary albumin and protein excretion were reduced in the streptozotocin-diabetic mouse model with the endogenous *Parp1* gene constitutively ablated by gene knockout [30]. While these previous animal studies concentrated primarily on the effect of PARP-1 in ameliorating albuminuria, our present finding on benzamide provides intriguing fresh evidence of a role for PARP-1 in the modulation of eGFR in diabetic patients.

Despite the salient findings, certain limitations of our study should be acknowledged. First, our study was based on a case–control study design, and therefore we could not draw conclusions regarding causation as in a cohort. This case–control study design was, however, appropriate in the current instance because the comprehensive metabolomic profiling of a large cohort would have been prohibitive in terms of both logistics and costs. The selected metabolites here identified may now serve as prospective leads in future cohort studies. These same logistic and cost considerations also placed a necessary limit on our sample size, consequently reducing the power of our study. However, this decreased power did not appear to substantially affect our study, as we successfully detected a number of associations that remained robust after correction for multiple hypotheses testing. Third, GFR was estimated using a serum creatinine-based equation rather than being directly measured. This can be expected to have caused some degree of disease misclassification and, conceivably, made it harder to identify potential metabolites that had weaker associations with low eGFR. Another potential shortcoming was that our metabolomic study was confined to urine and not the corresponding blood specimens from the patients. This was due to the inherent technical

difficulties in metabolomic analysis of this complex biofluid. Characterisation of the blood metabolome may have provided a more holistic context within which our findings could be interpreted.

To balance this, attention should be drawn to several strengths of our study. First, the metabolomic approach used was comprehensive by leveraging both GC/MS and LC/MS platforms. Second, positive and statistically robust associations were uncovered. In addition, the detection of associations with known uraemic toxins provided some degree of validation of our study design including the characterisation of our patient sample. Third, the chance of false positive findings was minimised by careful adjustment for multiple hypotheses testing. Fourth, our study focused on low eGFR in non-proteinuric patients. Aside from yielding new insight into the renal aspects of these patients, a significant benefit of our study design was the minimisation of any potential confounding due to the presence of proteinuria on the associations between the metabolomic profiles and low eGFR. Finally, to our knowledge, this is the first study of the metabolomics of eGFR in human type 2 diabetes. Previous reports have attempted to study changes in the urinary metabolome associated with type 2 diabetes per se, without any attention being paid to their possible association with the accompanying renal traits. Critically, few of these studies were on human specimens [31], with the majority being performed on experimental models of diabetes [32–35].

In conclusion, our investigation of the metabolomics of low eGFR in non-proteinuric type 2 diabetic patients has yielded substantial biological insight. In addition, we have identified several potential candidate biomarkers that may prove useful for detecting and monitoring chronic kidney disease. Individually, these potential biomarkers, especially the seven GC/MS metabolites, showed highly significant associations with low eGFR and had good discriminatory power. When used collectively, the metabolomic signatures appeared to be promising for disease classification.

Acknowledgements This study was supported by a NUS secondment fund to C. N. Ong, with additional support from the NUS Life Sciences (LSI) and Environmental Research Institute (NERI). We thank the staff of the National University of Singapore, as well as doctors, nurses and patients of the National Healthcare Group Polyclinics, for contributing to the establishment of the SDCS.

Contribution statement DPKN and CNO contributed to the conception and design. DPKN, AS, YL, LZ, FGX, SH, HL and CNO analysed and interpreted the data. All authors contributed to the drafting of the article and gave final approval of the version to be published.

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

References

- Kramer CK, Leitão CB, Pinto LC, Silveiro SP, Gross JL, Canani LH (2007) Clinical and laboratory profile of patients with type 2 diabetes with low glomerular filtration rate and normoalbuminuria. *Diabetes Care* 30:1998–2000
- Retnakaran R, Cull CA, Thorne KI, Adler AI, Holman RR, Study Group UKPDS (2006) Risk factors for renal dysfunction in type 2 diabetes. *UKPDS 74. Diabetes* 55:1832–1839
- MacIsaac RJ, Tsalamandris C, Panagiotopoulos S, Smith TJ, McNeil KJ, Jerums G (2004) Nonalbuminuric renal insufficiency in type 2 diabetes. *Diabetes Care* 27:195–200
- MacIsaac RJ, Panagiotopoulos S, McNeil KJ et al (2006) Is nonalbuminuric renal insufficiency in type 2 diabetes related to an increase in intrarenal vascular disease? *Diabetes Care* 29:1560–1566
- Niewczasz MA, Ficociello LH, Johnson AC et al (2009) Serum concentrations of markers of TNF α and Fas-mediated pathways and renal function in nonproteinuric patients with type 1 diabetes. *Clin J Am Soc Nephrol* 4:62–70
- Weckwerth W, Morgenthal K (2005) Metabolomics: from pattern recognition to biological interpretation. *Drug Discov Today* 10:1551–1558
- Niwa T (2009) Recent progress in the analysis of uremic toxins by mass spectrometry. *J Chrom B* 877:2600–2606
- Koulman A, Lane GA, Harrison SJ (2009) From differentiating metabolites to biomarkers. *Anal Bioanal Chem* 394:663–670
- Lewis GD, Asnani A, Gerszten RE (2008) Application of metabolomics to cardiovascular biomarker and pathway discovery. *J Am Coll Cardiol* 52:117–123
- National Kidney Foundation (2002) K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Am J Kidney Dis* 39:S1–S266
- Xu F, Zou L, Ong CN (2010) Experiment-oriented variations, multi-peak and multi-origination phenomena in derivatization-based GC-MS metabolomics. *Trends Anal Chem* 29:269–275
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29:1165–1188
- Tibshirani R (1997) The lasso method for variable selection in the Cox model. *Stat Med* 16:385–395
- Niwa T (2010) Indoxyl sulphate is a nephro-vascular toxin. *J Ren Nutr* 20:S2–S6
- Wills MR (1985) Uremic toxins, and their effect on intermediary metabolism. *Clin Chem* 31:5–13
- Mydlík M, Derzsiová K (2010) Vitamin B6 and oxalic acid in clinical nephrology. *J Ren Nutr* 20:S95–S102
- Worcester EM, Coe FL (2010) Calcium kidney stone. *N Engl J Med* 363:954–963
- Gillen DL, Worcester EM, Coe FL (2005) Decreased renal function among adults with a history of nephrolithiasis: a study of NHANES III. *Kidney Int* 67:685–690
- Mydlík M, Derzsiová K (2008) Oxalic acid as a uremic toxin. *J Ren Nutr* 18:33–39
- Karasawa R, Nishi S, Suzuki Y, Imai N, Arakawa M (1997) Early increase of chondroitin sulphate glycosaminoglycan in the basement membrane of rats with diabetic glomerulopathy. *Nephron* 76:62–71
- McCarthy KJ, Abrahamson DR, Bynum KR, St John PL, Couchman JR (1994) Basement membrane-specific chondroitin sulphate proteoglycan is abnormally associated with the glomerular capillary basement membrane of diabetic rats. *J Histochem Cytochem* 42:473–484
- Cadaval RAM, Kohlman O, Michelacci YM (2000) Urinary excretion of glycosaminoglycans and albumin in experimental diabetes mellitus. *Glycobiology* 10:185–192

23. De Muro P, Fresu P, Tonolo G et al (2006) A longitudinal evaluation of urinary glycosaminoglycan excretion in normoalbuminuric type 1 diabetic patients. *Clin Chem Lab Med* 44:561–567
24. De Muro P, Fresu P, Formato M et al (2002) Urinary glycosaminoglycan and proteoglycan excretion in normoalbuminuric patients with type 1 diabetes mellitus. *J Nephrol* 15:290–296
25. Martin DR, Lewington AJP, Hammerman MR, Padanilam BJ (2000) Inhibition of poly(ADP-ribose) polymerase attenuates ischemic renal injury in rats. *Am J Physiol Regul Integr Comp Physiol* 279:R1834–R1840
26. Banasik M, Komura H, Shimoyama M, Ueda K (1992) Specific inhibitors of poly(ADP-ribose) synthetase and mono(ADP-ribose)transferase. *J Biol Chem* 267:1569–1575
27. Soriano FG, Virag L, Jagtap P et al (2001) Diabetic endothelial dysfunction: the role of poly(ADP-ribose) polymerase activation. *Nat Med* 7:108–113
28. Shevalye H, Stavniichuk R, Xu W et al (2010) Poly(ADP-ribose) polymerase (PARP) inhibition counteracts multiple manifestations of kidney disease in long-term streptozotocin-diabetic rat model. *Biochem Pharmacol* 79:1007–1014
29. Szabó C, Biser A, Benko R, Böttinger E, Suszták K (2006) Poly(ADP-ribose) polymerase inhibitors ameliorate nephropathy of type 2 diabetic *Lepi^{db/db}* mice. *Diabetes* 55:3004–3012
30. Shevalye H, Maksimchyk Y, Watcho P, Obrosova IG (2010) Poly(ADP-ribose) polymerase-1 (PARP-1) gene deficiency alleviates diabetic nephropathy. *Biochim Biophys Acta* 1802:1020–1027
31. Zhao X, Fritsche J, Wang J et al (2010) Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits. *Metabolomics* 6:362–374
32. Cai S, Huo T, Xu J, Lui X, Zheng S, Li F (2009) Effect of mitiglinide on streptozotocin-induced experimental type 2 diabetic rats: a urinary metabonomics study based on ultra-performance liquid chromatography–tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 877:3619–3624
33. Zhao LC, Zhang XD, Liao SX, Gao HC, Wang HY, Lin DH (2010) A metabonomic comparison of urinary changes in Zucker and GK rats. *J Biomed Biotechnol* 2010:431894.
34. Zhao LC, Liu X, Xie L, Gao H, Lin D (2010) ¹H NMR-based metabonomic analyses of metabolic changes in streptozotocin-induced diabetic rats. *Anal Sci* 26:1277–1282
35. Patterson AD, Bonzo JA, Li F et al (2011) Metabolomics reveals attenuation of the SLC6A20 kidney transporter in nonhuman primate and mouse models of type 2 diabetes mellitus. *J Biol Chem* 286:19511–19522