Serum Metabolomic Profiling and Incident CKD among African Americans

Bing Yu,* Yan Zheng,* Jennifer A. Nettleton,* Danny Alexander,† Josef Coresh,‡§ and Eric Boerwinkle*†

Abstract

Background and objectives Novel biomarkers that more accurately reflect kidney function and predict future CKD are needed. The human metabolome is the product of multiple physiologic or pathophysiologic processes and may provide novel insight into disease etiology and progression. This study investigated whether estimated kidney function would be associated with multiple metabolites and whether selected metabolomic factors would be independent risk factors for incident CKD.

Design, setting, participants, & measurements In total, 1921 African Americans free of CKD with a median of 19.6 years follow-up among the Atherosclerosis Risk in Communities Study were included. A total of 204 serum metabolites quantified by untargeted gas chromatography–mass spectrometry and liquid chromatography–mass spectrometry was analyzed by both linear regression for the cross-sectional associations with eGFR (specified by the Chronic Kidney Disease Epidemiology Collaboration equation) and Cox proportional hazards model for the longitudinal associations with incident CKD.

Results Forty named and 34 unnamed metabolites were found to be associated with eGFR specified by the Chronic Kidney Disease Epidemiology Collaboration equation with creatine and 3-indoxyl sulfate showing the strongest positive (2.8 ml/min per 1.73 m² per +1 SD; 95% confidence interval, 2.1 to 3.5) and negative association (−14.2 ml/min per 1.73 m² per +1 SD; 95% confidence interval, −17.0 to −11.3), respectively. Two hundred four incident CKD events with a median follow-up time of 19.6 years were included in the survival analyses. Higher levels of 5-oxoproline (hazard ratio, 0.70; 95% confidence interval, 0.60 to 0.82) and 1,5-anhydroglucitol (hazard ratio, 0.68; 95% confidence interval, 0.58 to 0.80) were significantly related to lower risk of incident CKD, and the associations did not appreciably change when mutually adjusted.

Conclusions These data identify a large number of metabolites associated with kidney function as well as two metabolites that are candidate risk factors for CKD and may provide new insights into CKD biomarker identification.


Introduction

CKD, defined as the presence of damage or reduced function of the kidneys for at least 3 months, is a global public health problem that affects approximately 13% of the adult population in the United States, and the prevalence is increasing (1,2). CKD is associated with risk of cardiovascular disease (3,4), peripheral vascular disease (5), and all-cause mortality (6,7). Hypertension and diabetes are two key risk factors for the development of CKD (8). Serum creatinine, a low molecular weight muscle breakdown product, is the most commonly used biomarker to detect and diagnose CKD in clinical practice (9). Ongoing studies strive to identify novel biomarkers associated with CKD and ESRD as targets for future therapies or diagnostic tests. A priori, it is expected that many low molecular weight metabolites that are cleared by the kidney would accumulate when GFR is reduced (10,11).

CKD and kidney function are influenced by both genetic and environmental factors (12). The human metabolome is a reflection of the interaction between genes and the environment, and studies relating metabolomic profiles with kidney function may enhance our understanding of the physiology and pathophysiology underlying development and progression of CKD (13–15). Untargeted metabolomic technologies combining the separation of molecules by size and chemical characteristics with sensitive and quantitative detection provide an unbiased method to simultaneously assay a large number of metabolites in a sample (16). These technologies create an opportunity to investigate both the effect of kidney function on the metabolome and the ability of metabolomic measures to predict the onset of incident CKD (17). Several studies have shown that metabolic profiles can uncover potential biomarkers related to eGFR and different stages of CKD in multiple ethnicities (18–21). A recent study has reported that adding metabolomic predictors of renal function independent of eGFR to clinical data may significantly improve the ability to...
predict onset of CKD among European Americans (17). However, there are no metabolomics studies in African Americans, a race group with excess risk of CKD (22), and no longitudinal data have been reported to clarify whether an abnormal metabolic profile precedes the onset of CKD in this population.

We conducted both cross-sectional and longitudinal analyses of the relationship of CKD to the metabolic profile of a large number of African-American participants in the Atherosclerosis Risk in Communities (ARIC) cohort to better characterize the temporal relationship between human metabolome, kidney function, and CKD among African Americans. We hypothesized that estimated kidney function would be associated with multiple metabolites and that selected metabolomic factors would be independent risk factors for incident CKD.

Materials and Methods

Study Population

The ARIC Study is a longitudinal cohort study designed to ascertain the etiology and predictors of cardiovascular disease. The ARIC Study enrolled 15,792 middle-aged adults from four United States communities (Forsyth County, NC; Jackson, MS; suburbs of Minneapolis, MN; and Washington County, MD) between 1987 and 1989. A detailed description of the ARIC Study design and methods was published elsewhere (23). Study participants had three follow-up examinations, each approximately 3 years apart. Metabolomic profiles were measured in 1977 African Americans randomly selected from the Jackson, Mississippi field center. Participants with prevalent CKD were excluded, and 1921 participants with complete data for kidney function and incident CKD follow-up information were included in this study.

Demographic information was obtained during an in-home interview. Seated systolic BP (SBP) was measured three times by a random zero-mercury sphygmomanometer, and the average was used in this analysis. Antihypertensive medication use was recorded by self-report and ascertained by inventory. Diabetes was defined as fasting glucose ≥126 mg/dl, nonfasting glucose ≥200 mg/dl, or self-report of diagnosis by a physician of diabetes or sugar in the blood or use of oral diabetes medication or insulin. Smoking status was self-reported and categorized as current and noncurrent smoker. Prevalent coronary heart disease was defined by evidence of prior myocardial infarction by electrocardiogram, history of physician-diagnosed heart attack, or previous coronary reperfusion procedure. Plasma HDL cholesterol concentrations were determined by standardized enzymatic methods, and LDL cholesterol concentrations were calculated by the Friedewald equation.

The Chronic Kidney Disease Epidemiology Collaboration equation was applied to eGFR (eGFR_{CKD-EPI}) (24) using serum creatinine measured by a modified kinetic Jaffé method at baseline (1987–1989). Incident CKD was defined previously (25) in those participants with a baseline eGFR_{CKD-EPI} ≥ 60 ml/min per 1.73 m² as either at least one follow-up eGFR_{CKD-EPI} ≤ 60 ml/min per 1.73 m² and <75% of baseline (i.e., ≥25% decrease) or the individual had at least one surveillance-based CKD event (hospitalization or CKD-related death) after baseline up to 2008.

Assessment of Metabolites

Metabolite profiling was measured using baseline fasting serum samples that had been stored at −80°C since collection. In total, detection and quantification of 602 metabolites were completed by Metabolon, Inc. (Durham, NC) using an untargeted gas chromatography–mass spectrometry and liquid chromatography–mass spectrometry-based metabolomic quantification protocol (26,27). A schematic overview of Metabolon, Inc.’s measurement processes is shown in Supplemental Figure 1, and for the liquid chromatography–mass spectrometry, total ion count scan is shown in Supplemental Figure 2. Before the analysis of CKD, a rigorous assessment of the metabolomic data was done as described previously (28,29). Briefly, metabolites were excluded if (1) more than 80% of the samples had values below the detection limit and (2) the reliability coefficient on the basis of a comparison of two samples from 60 individuals collected 4–6 weeks apart was less than 0.60.

Among 204 metabolites meeting the criteria provided above (a full list is provided in Supplemental Table 1, including masses and coefficients of variation), 187 metabolites had ≥50% above the detection limit, and they were analyzed as continuous variables, where below-detection values were imputed with the lowest detected value for that metabolite in all samples. The other 17 reproducible metabolites with 50%–80% below detection were analyzed as ordinal variables with three levels: (1) below-detection limit values, (2) detected values below the median, and (3) detected values above the median.

Of 204 metabolites, 118 metabolites were named compounds within eight super-pathways, and the remaining 86 metabolites were unknown compounds. The unknown chemical identities were tagged beginning with X and followed by numbers (e.g., X-12345).

Statistical Analyses

Each of 187 metabolites treated continuously was centered by its mean and scaled by its SD. Pearson and Spearman partial correlation coefficients were calculated for continuous and ordinal metabolite levels, respectively, adjusting for age and sex. The Pearson correlation between creatinine measured within the metabolomics panel and by the Jaffé method was 0.89, and the metabolomic-measured creatinine is shown in Figure 1 but not considered further in this study.

The cross-sectional associations of eGFR_{CKD-EPI} with individual metabolite levels were assessed using linear regression. In model 1, we minimally adjusted for age and sex. In model 2, we further adjusted for SBP, antihypertensive medication use, and diabetes status. In model 3, we added additional risk factors of CKD to model 2 (i.e., current smoking status, prevalent coronary heart disease status, HDL cholesterol levels, and LDL cholesterol levels) (30). Multiple comparisons were taken into account using Bonferroni correction, and statistical significance at this stage was defined as P < 0.001.

We assessed the longitudinal associations of metabolite levels with incident CKD using Cox proportional hazards
The same covariates described above for model 3 in the cross-sectional analyses were considered with additional adjustment of eGFRCKD-EPI from baseline measurement. 

P values for individual metabolites were corrected using the Westfall–Young step-down min-P procedure (31), a resampling-based method that takes into account the dependence in the test statistics. The proportional hazards assumption was examined using the methods developed by Grambsch and Therneau (32). All statistical analyses were performed using R (http://www.r-project.org).

**Results**

**Characteristics of Participants**

Baseline characteristics by incident CKD status for 1921 eligible ARIC African Americans are shown in Table 1. Higher SBP, higher prevalence of diabetes, lower HDL cholesterol levels, and lower eGFRCKD-EPI levels were associated with development of CKD. The average baseline eGFRCKD-EPI was 105 ml/min per 1.73 m², and approximately 20% of the analyzed cohort had an eGFRCKD-EPI between 60 and 90 ml/min per 1.73 m².

**Metabolomic Association with eGFRCKD-EPI**

In linear regression analyses, models 1–3 showed consistent results; however, the majority of metabolite effect sizes was diminished in models 2 and 3 compared with model 1. We identified 74 metabolites (40 named and 34 unnamed) after adjustment for the full model (Figure 1, model 3) that were significantly associated with eGFRCKD-EPI. In general, higher metabolite levels were
associated with lower eGFR_{\text{CKD-EPI}}; the average difference in eGFR_{\text{CKD-EPI}} was $-3.7\text{ ml/min per 1.73 m}^2$ per +1-SD difference in metabolite levels in model 3. The largest measures of association were observed in the amino acid pathway, where a 1-SD increase of 3-indoxyl sulfate levels corresponded to 14.2 ml/min per 1.73 m$^2$ decrease in eGFR_{\text{CKD-EPI}} (95% confidence interval [95% CI], $-17.0$ to $-11.3$) (Figure 1). Three metabolites, creatinine (2.8 ml/min per 1.73 m$^2$ per +1 SD; 95% CI, 2.1 to 3.5) and carnitine (1.73 ml/min per 1.73 m$^2$ per +1 SD; 95% CI, 1.0 to 2.4), which are known to be synthesized by the kidney, and betaine (1.94 ml/min per 1.73 m$^2$ per +1 SD; 95% CI, 1.3 to 2.6), which is known to have increased excretion in CKD, were positively associated with eGFR_{\text{CKD-EPI}}; higher metabolite levels were related to higher eGFR_{\text{CKD-EPI}} (Figure 1). For the unknown metabolites that were significantly associated with baseline eGFR_{\text{CKD-EPI}}, the relationships were generally in the same direction as the named metabolites; higher metabolite levels were associated with lower eGFR_{\text{CKD-EPI}} (Supplemental Table 2).

### Metabolomic Risk Factors of Incident CKD

A total of 204 incident CKD events occurred during a median follow-up of 19.6 years. The median follow-up time for incident CKD patients was 12.5 years, with a range from 1.2 to 21.4 years. Participants in the lowest eGFR_{\text{CKD-EPI}} quintile at the baseline examination had the highest cumulative incidence at 19.7%. Using a principal component analysis, no metabolic pattern was identified to differentiate participants who later developed incident CKD compared with participants who did not (Supplemental Figure 3). In longitudinal analyses, we identified two baseline metabolites, 5-oxoproline and 1,5-anhydroglucitol (1,5-AG), that were significantly associated with incident CKD after correction of multiple testing (both inversely). Pearson correlation coefficients between 5-oxoproline 1,5-AG and eGFR_{\text{CKD-EPI}} were $-0.1$ and $-0.05$, respectively (the box plots of these two metabolites levels stratified by CKD case status are shown in Supplemental Figure 4). The lowest quartile of both of these metabolites had the highest cumulative hazards for future onset of CKD, and the quartiles of 5-oxoproline showed an inverse linear trend in cumulative hazards (Figure 2). A +1 SD higher 5-oxoproline corresponded to 30% lower hazard for incident CKD with adjustment for traditional risk factors and eGFR_{\text{CKD-EPI}} (hazard ratio [HR], 0.70; 95% CI, 0.60 to 0.82). The highest quartile had 51% lower hazard compared with the lowest quartile (P for trend<0.001) (Table 2). For 1,5-AG, the hazard for incident CKD was 32% lower per +1-SD difference when treated continuously (HR, 0.68; 95% CI, 0.58 to 0.80) (Table 2). Unlike 5-oxoproline, with higher levels that conferred lower risks of incident CKD, the lowest quartile of 1,5-AG showed an obvious increased risk, whereas the other upper three quartiles showed a similar pattern (Figure 2) (HR of lowest quartile versus others, 0.53; 95% CI, 0.39 to 0.74). The predictive ability for incident CKD was marginally improved by adding 5-oxoproline and 1,5-AG over the traditional risk factors (area under the curve=0.76 versus 0.75; P=0.29) (Figure 3). Additionally, the strength of their associations did not markedly change when conditioning on each other, and there is no interaction between 5-oxoproline, 1,5-AG, and eGFR_{\text{CKD-EPI}} (data not shown).

### Discussion

Using unbiased high-throughput metabolomic technologies, we identified multiple small molecule metabolites that were correlated with eGFR_{\text{CKD-EPI}} in African-American adults. We further identified two metabolites, 5-oxoproline and 1,5-AG, that were associated with risk of developing incident CKD after correcting for multiple testing. Interestingly, both associations were inverse, with high levels of these metabolites measured at baseline being associated with lower risk of developing CKD after 19 years of average follow-up. These associations were independent of other CKD risk factors and eGFR_{\text{CKD-EPI}}. These data underscore the challenges and opportunities in identifying novel biomarkers of kidney function and future kidney disease.

Previous metabolomic data related to kidney function are limited. A cross-sectional case-cohort study in 41 Japanese patients with CKD showed a crude correlation between 52 plasma metabolites and eGFR (19). However,
the study did not take into account the possible confounding by several key factors, such as BP and diabetes status, and the study also did not adjust for multiple hypothesis tests. Of 52 metabolites reported by Toyohara et al. (19), 12 metabolites were measured in our study, with seven of them found to be statistically, significantly, and consistently associated with eGFRCKD-EPI in the same direction in the data reported here. For example, N-acetylalanine was negatively associated with eGFRCKD-EPI, having the lowest \( P \) value in this study, and Toyohara et al. (19) reported N-acetylalanine to be a uremic retention anion. Another cross-sectional study of 30 participants, with 10 patients each in CKD stages 2–4, showed significant differences of plasma metabolites among various stages of CKD (20). In the study, 3-carboxy-4-methyl-5-propyl-2-furanpropanoate, a known uremic toxin, was higher in CKD stages 3 and 4 compared with CKD stage 2. Here, we reported that the same metabolite (3-carboxy-4-methyl-5-propyl-2-furanpropanoate) was significantly inversely related to eGFRCKD-EPI. A recent population-based study estimated the cross-sectional association of 151 serum metabolites and their ratios with kidney function in two independent samples, finding 22 metabolites and 516 ratios to be significant (33); five of 22 metabolites were measured in this study, and all of them were significantly associated with kidney function (\( P<0.001 \)). Four acylcarnitines were concordant with the findings in this study, where reduced kidney function leads to impaired excretion of acylcarnitines (34). Interestingly, carnitine was found to be inversely associated with eGFR in the study by Goek et al.
*effect of eGFR.*

The development of future CKD after accounting for the lower levels of 5-oxoproline are a potential risk factor of maintaining adequate GSH levels (43). Here, we report that oxidative stress, the levels of 5-oxoproline are important to key role in the pathogenesis of many vascular and metabolic diseases, including CKD (41,42). Under conditions of diabetic heart disease status, HDL cholesterol levels, and LDL medication use, diabetes status, current smoking status, prevalent coronary heart disease status, HDL cholesterol levels, and LDL cholesterol levels.

Both of the previous studies used targeted metabolomics technologies, which limited the number of shared metabolites with this study. Therefore, future studies using untargeted metabolomics technologies are needed to replicate our findings.

After 19 years of longitudinal follow-up, we identified 5-oxoproline and 1,5-AG measured at baseline as being inversely associated with risk of incident CKD. Case reports have shown that 5-oxoproline accumulation is a rare cause for high anion gap metabolic acidosis (36,37). One of the reasons for 5-oxoproline accumulation is renal failure, which causes diminished clearance of 5-oxoproline. 5-Oxoproline belongs to the glutathione metabolism pathway and is involved in both synthesis and degradation of glutathione (γ-glutamyl-cysteinyl-glycine [GSH]) (39). 5-Oxoproline is the substrate of glutamate, and can be used as an index of glycine insufficiency (40). GSH is synthesized by glutamate, cysteine, and glycine, and GSH deficiency contributes to oxidative stress, which plays a key role in the pathogenesis of many vascular and metabolic diseases, including CKD (41,42). Under conditions of oxidative stress, the levels of 5-oxoproline are important to maintain adequate GSH levels (43). Here, we report that lower levels of 5-oxoproline are a potential risk factor of the development of future CKD after accounting for the effect of eGFR.

1,5-AG is a carbohydrate found in nearly all foods, and after it is ingested, it is not metabolized; it is filtered by the glomerulus but reabsorbed back into the blood in the renal proximal tubule (44). High levels of glucose inhibit the reabsorption of 1,5-AG (44), and 1,5-AG is used clinically to distinguish excess glycemic variability postprandially (45). In this study, 1,5-AG has a moderate inverse correlation with fasting glucose levels. After adjusting for baseline glucose levels and diabetes status, 1,5-AG levels at the baseline examination persisted to show an inverse relationship with higher risk of incident CKD. These results are supported by a recent study suggesting that an increase of urinary carbohydrates, including 1,5-AG, can be an early indicator of renal toxicity (46). Taken together, 1,5-AG is a promising candidate biomarker for incidence CKD whether as a result of hyperglycemia excursions or other pathways.

Limitations of this study warrant consideration. In our study, GFR was estimated by the CKD-EPI formula without direct measurement. Through our internal quality-control process, we limited our analyses to those metabolites with few missing values; possible associations between the excluded metabolites and kidney diseases need additional investigation. The participants for this study were African Americans, who have a higher risk of CKD compared with European Americans (22); it is unknown if our findings could be generalized to other ethnic groups. Both 5-oxoproline and 1,5-AG show potential use in clinical practice, but routine application will require validated and standardized assays and replication of our findings. Despite limitations, the data presented here have important strengths. To our knowledge, this study is the first population-based untargeted metabolomics study of kidney function in African Americans. Moreover, this study was done in a well characterized population with long follow-up time and reliably measured metabolites.

In summary, we report 74 metabolites that are associated with eGFRCKD-EPI. Given the importance of kidney function on metabolite filtrations and disease prediction, future metabolic studies should consider kidney function when elucidating biologic mechanisms of disease. In addition, lower levels of 5-oxoproline and 1,5-AG were related to higher risk of incident CKD independently of other risk factors over 19 years of follow-up, which provides new insights into CKD biomarker identification.

**Acknowledgments**

The authors thank the staff and participants of the Atherosclerosis Risk in Communities (ARIC) Study for their important contributions.

The ARIC Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute Contracts HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C. The metabolomics measurements were sponsored by National Human Genome Research Institute Grant 3U01HG004402-03S1. B.Y. and Y.Z. are supported, in part, by a training fellowship from Burroughs Welcome Fund–The Houston Laboratory and Population Science Training Program in Gene–Environment Interaction Burroughs Wellcome Fund Grant 1008200. J.A.N. is supported by National Institutes of Health–National Institute of Diabetes and Digestive and Kidney Diseases Grant 5K01DK082729-04.

**Disclosures**

None.


Received: November 27, 2013 Accepted: April 8, 2014

Published online ahead of print. Publication date available at www.cjasn.org.

This article contains supplemental material online at http://cjasn.asnjournals.org/lookup/suppl/doi:10.2215/CJN.11971113/-/DCSupplemental.