Urinary metabolites along with common and rare genetic variations are associated with incident chronic kidney disease

Gearoid M. McMahon[1,2], Shih-Jen Hwang[1], Clary B. Clish[3], Adrienne Tin[4], Qiong Yang[5], Martin G. Larson[1,6], Eugene P. Rhee[7], Man Li[4,8,9], CKDGen Consortium, Daniel Levy[1,10], Christopher J. O'Donnell[1], Josef Coresh[4], J. Hunter Young[4,11], Robert E. Gerszten[3,12] and Caroline S. Fox[1,13]

[1]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA; [2]Division of Nephrology, Brigham and Women's Hospital, Boston, Massachusetts, USA; [3]Broad Institute, Cambridge, Massachusetts, USA; [4]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; [5]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA; [6]Department of Mathematics and Statistics, Boston University, Boston, Massachusetts, USA; [7]Nephrology Division, Massachusetts General Hospital, Boston, Massachusetts, USA; [8]Division of Nephrology, University of Utah, Salt Lake City, Utah, USA; [9]Department of Human Genetics, University of Utah, Salt Lake City, Utah, USA; [10]Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA; [11]Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland, USA; [12]Cardiology Division, Massachusetts General Hospital, Boston, Massachusetts, USA; and [13]Division of Endocrinology and Metabolism, Brigham and Women's Hospital, Boston, Massachusetts, USA

We assessed the association between urinary metabolites, genetic variants, and incident chronic kidney disease (CKD) in the Framingham Offspring cohort. Among the participants, 193 individuals developed CKD (estimated glomerular filtration rate under 60 ml/min/1.73m^2) between cohort examinations 6 (1995-1998) and 8 (2005-2008, mean follow-up 9.7 years). They were age- and sex-matched to 193 control individuals free of CKD. A total of 154 urinary metabolites were measured using mass spectrometry, and the association between metabolites and CKD was examined using logistic regression. Next, we tested the genetic associations of each metabolite with an Illumina exome chip. Urinary glycine and histidine were associated with a lower risk of incident CKD with an odds ratio of 0.59 (95% confidence interval [CI] 0.43-0.80) and 0.65 (0.50-0.85) respectively, per one standard deviation increase in metabolite concentration. Follow-up in the Atherosclerosis Risk in Communities cohort confirmed the association of urinary glycine with CKD. In exome chip analyses, 36 single nucleotide polymorphisms at 30 loci were significantly associated with 31 metabolites. We surveyed exome chip findings for associations with known renal function loci such as rs8101881 in *SLC7A9* coding for an amino acid transporter, which has been associated with a lower risk of CKD. We found this polymorphism was significantly associated with higher levels of lysine and NG-monomethyl-L-arginine (NMMA). Increased urinary lysine and NMMA were associated with a lower risk of CKD (0.73 [0.50-0.90] and 0.66 [0.53-0.83], respectively) in the univariate model. Thus, low urinary glycine and histidine are associated with incident CKD. Furthermore, genomic association of urinary metabolomics identified lysine and NMMA as being linked with CKD and provided additional evidence for the association of *SLC7A9* with kidney disease.

Chronic kidney disease (CKD) is an important public health problem that affects as many as 13% of adults in the United States.[1] CKD increases the risk of end-stage renal disease, is associated with a range of complications including anemia and bone disease, and is an independent risk factor for cardiovascular disease and all-cause mortality.[2–4] Currently, serum creatinine is the primary biomarker of CKD.[5] However, creatinine is an insensitive biomarker of kidney injury, particularly in the early stages of disease, and does not provide information as to the nature and cause of kidney injury.[6] As a result, there is a need for novel biomarkers that can more accurately identify individuals at risk of CKD before the development of overt disease while also potentially providing information regarding the site and nature of renal injury.

The urine is an important potential source of kidney biomarkers.[7] Targeted metabolomic profiling of the urine may identify such biomarkers. Differences in serum

Correspondence: *Caroline S. Fox, NHLBI's Framingham Heart Study, 73 Mt. Wayte Avenue, Suite 2, Framingham, Massachusetts 01702, USA. E-mail: foxca@nhlbi.nih.gov*

metabolite concentrations were shown to predict incident CKD,[8–10] while a genomewide association study of urinary biomarkers demonstrated the feasibility of this technique in a community-based population.[11]

Thus, the goal of this study was to identify novel urinary metabolites associated with incident CKD in the Framingham Heart Study. To provide further insight into the biological mechanisms associated with these urinary metabolites, we examined genetic associations of these metabolites using exome chip genotypes. We hypothesized that by combining phenotypic, metabolomic, and genetic data, we could better identify individuals at risk of CKD and uncover biology related to CKD.

RESULTS
Study sample characteristics
Baseline characteristics of the participants are shown in Table 1. The mean age was 65.2 years, and nearly half of the participants were women. The mean duration of follow-up was 9.7 years. Cases were more likely to have hypertension (64.3 vs. 46.6%, $P = 0.0005$) and dipstick proteinuria (27.6 vs. 16.2%, $P = 0.007$). Not surprisingly, the mean estimated glomerular filtration rate (eGFR) was lower in cases at baseline (78 vs. 86 ml/min per 1.73 m^2, $P < 0.001$) and at follow-up (50 vs. 79 ml/min 1.73 m^2, $P < 0.001$).

Urinary metabolites and incident CKD
Our first step was to test the association of the measured urinary metabolites with incident CKD case status. Of the 154 metabolites included in the panel, in total, 120 were measurable in all participants. Three metabolites (indoleacetate, benzoate, and phosphocreatine) were missing in >100

Table 1 | Characteristics of the participants at the baseline examination (offspring examination 6)

	Controls (N = 193)	Cases (N = 193)	P value
Age, y	64.7 (6.6)	65.6 (5.8)	0.16
Female sex	95 (49)	95 (49)	1.0
Systolic blood pressure, mm Hg	131 (17)	136 (19)	0.01
Diastolic blood pressure, mm Hg	74 (9)	76 (9)	0.07
Hypertension	90 (47)	124 (64)	0.0005
Hypertension treatment	55 (29)	91 (47)	0.0001
Current smoker	20 (10)	24 (12)	0.06
HDL cholesterol, mg/dl	51 (15)	49 (17)	0.21
Triglycerides, mg/dl	118 (32, 492)	136 (42, 833)	0.03
Lipid-lowering treatment	27 (14)	44 (23)	0.03
Body mass index, kg/m^2	27.6 (4.4)	28.2 (4.8)	0.23
Diabetes	20 (10)	33 (17)	0.06
Dipstick proteinuria	31 (16)	53 (27)	0.007
eGFR examination 6, ml/min per 1.73 m^2	86 (13)	78 (12)	<0.000001
eGFR examination 8, ml/min per 1.73 m^2	79 (9)	50 (8)	<0.000001

eGFR, estimated glomerular filtration rate; HDL, high-density lipoprotein.
Data are mean (SD) or median (min, max) for continuous variables and N (%) for categorical variables.

participants and were therefore excluded leaving 151 metabolites in the final analysis.

In total, 33 metabolites achieved the multiple testing significance threshold ($P < 0.05/24$) in the unadjusted model for incident CKD. In all cases, higher metabolite levels were associated with a decreased risk of incident CKD. Of these 33 metabolites, 18 were alpha amino acids, including 10 essential amino acids. Supplementary Table S1 shows the metabolic pathways and classes for the metabolites associated with CKD in the univariate analysis.

Table 2 displays the metabolites that reached the significance threshold in the unadjusted and multivariable-adjusted analyses. The results for all 151 metabolites are presented in Supplementary Table S2. After multivariable adjustment (baseline eGFR, diabetes, hypertension, and proteinuria), 2 metabolites remained associated with incident CKD ($P < 0.0021$). Increased levels of glycine (odds ratio [OR]: 0.59 per 1-SD increase, 95% confidence interval [CI] 0.43–0.80) and histidine (OR: 0.65, 95% CI 0.5–0.85) were associated with a reduced risk of incident CKD. Glycine is a simple, nonessential amino acid that is involved in the production of DNA, phospholipids, and collagen and plays an important role in 1-carbon metabolism. Histidine is an essential amino acid that is involved in protein synthesis and has important antioxidant and anti-inflammatory properties.[12]

A previous study examined the association of plasma metabolites with incident CKD (1265 controls, 117 cases) in participants from the Framingham Heart Study.[9] We tested the association of plasma glycine and histidine with incident CKD in 1382 individuals from this study. Neither plasma glycine (OR: 1.00, 95% CI 0.81–1.24, $P = 0.99$) nor plasma histidine was associated with incident CKD (OR: 0.84, 95% CI 0.69–1.02, $P = 0.08$) after multivariable adjustment.

To determine whether the addition of metabolites contributed to risk prediction above the usual clinical predictors, we calculated C statistics for models predicting CKD with and without the addition of glycine and histidine. The addition of the metabolites led to a 4% discrimination improvement relative to the clinical model alone (area under the curve, 0.72 and 95% CI 0.67–0.77 and 0.76 and 95% CI 0.71–0.81, respectively).

We followed up by testing the association between urinary glycine and incident CKD in a subcohort of individuals from the Atherosclerosis Risk in Communities (ARIC) study ($N = 998$). In total, incident CKD developed in 149 individuals (14.9%). Lower levels of urinary glycine were associated with incident CKD after full covariate adjustment (hazard ratio: 0.82, 95% CI 0.69–0.99).

Genetic associations of urinary glycine and histidine
Because urinary glycine and histidine were independently associated with incident CKD, we analyzed the association of these metabolites with exome chip variants. The lead single-nucleotide polymorphisms (SNPs) in association with glycine and histidine are shown in Table 3. rs77010315 in