Genome-Wide Association Studies of Metabolites in Patients with CKD Identify Multiple Loci and Illuminate Tubular Transport Mechanisms

Yong Li,1 Peggy Sekula,1 Matthias Wuttke,1 Judith Wahrheit,2 Birgit Hausknecht,3 Ulla T. Schultheiss,1 Wolfram Gronwald,4 Pascal Schlosser,1 Sara Tucci,5 Arif B. Ekici,6 Ute Spiekerkoetter,5 Florian Kronenberg,7 Kai-Uwe Eckardt,3 Peter J. Oefner,4 and Anna Köttgen1, the GCKD Investigators

Due to the number of contributing authors, the affiliations are listed at the end of this article.

ABSTRACT
Background The kidneys have a central role in the generation, turnover, transport, and excretion of metabolites, and these functions can be altered in CKD. Genetic studies of metabolite concentrations can identify proteins performing these functions.

Methods We conducted genome-wide association studies and aggregate rare variant tests of the concentrations of 139 serum metabolites and 41 urine metabolites, as well as their pairwise ratios and fractional excretions in up to 1168 patients with CKD.

Results After correction for multiple testing, genome-wide significant associations were detected for 25 serum metabolites, two urine metabolites, and 259 serum and 14 urinary metabolite ratios. These included associations already known from population-based studies. Additional findings included an association for the uremic toxin putrescine and variants upstream of an enzyme catalyzing the oxidative deamination of polyamines (AOC1, P-min=2.4×10^{-12}), a relatively high carrier frequency (2%) for rare deleterious mis-sense variants in ACADM that are collectively associated with serum ratios of medium-chain acylcarnitines (P-burden=6.6×10^{-16}), and associations of a common variant in SLC7A9 with several ratios of lysine to neutral amino acids in urine, including the lysine/glutamine ratio (P=2.2×10^{-23}). The associations of this SLC7A9 variant with ratios of lysine to specific neutral amino acids were much stronger than the association with lysine concentration alone. This finding is consistent with SLC7A9 functioning as an exchanger of urinary cationic amino acids against specific intracellular neutral amino acids at the apical membrane of proximal tubular cells.

Conclusions Metabolomic indices of specific kidney functions in genetic studies may provide insight into human renal physiology.


Metabolites are small molecules that represent intermediates or products of metabolic processes. They play important roles in energy generation, signaling, and the regulation of enzymatic reactions. The concentrations of these small molecules in cells and body fluids result from a balance of their intake and generation, their transport across compartments, and their breakdown and excretion. The kidneys play a central role in all of these processes, providing a rationale for metabolomics research in...
nephrology. Previous studies have linked blood metabolite concentrations to common genetic markers across the genome and found a strong genetic component to the measured concentrations of many metabolites. The implicated genes are often involved in balancing blood metabolite concentrations through metabolite generation (e.g., encoding the rate-limiting enzyme), turnover, or excretion (e.g., encoding metabolite transporters). In CKD, reduced GFR leads to elevated concentrations of many metabolites in blood. We therefore hypothesized that the presence of CKD represents a “challenge” model for metabolite handling. Under such a model, metabolites may be quantifiable that are usually below the limit of detection, and renal mechanisms that facilitate active metabolite reabsorption or excretion may be altered and could inform about tubular functions. Studying metabolite concentrations in patients with CKD may therefore allow for the detection of genetic loci influencing such processes.

Only a few previous studies have investigated genetic influences on metabolite concentrations in urine, which are of particular interest for the field of nephrology. Not only does the urine contain metabolites that are exclusively or predominantly generated in the kidneys and secreted into urine, but urinary metabolite concentrations also allow for the modeling of specific renal functions that may be affected in the presence of CKD. For example, genetic screens for the metabolite’s fractional excretion (FE) could identify tubular transport proteins for this metabolite, or genetic variants associated with the ratio of metabolites in the urine could identify substrates that are counter-transported across the apical tubular cell membrane.

We therefore set out to test the following hypotheses: first, genetic investigations of metabolite concentrations in serum and urine of patients with CKD can replicate findings from previous population-based studies and identify additional genetic loci that may reflect the metabolic challenge posed by CKD. Second, modeling of kidney-specific functions on the basis of metabolite concentrations can be used to gain insights into tubular transport mechanisms and metabolic reactions of importance in CKD or detectable in its presence. To test these hypotheses in a proof-of-principle study, we carried out genome-wide association studies (GWAS) of metabolite concentrations in serum and urine as well as their FEs and pairwise ratios in up to 1168 patients with CKD participating in the German Chronic Kidney Disease (GCKD) study.

METHODS

Study Design and Participants

The GCKD study is a prospective cohort study of patients with CKD treated by nephrologists. It was approved by the local ethics committees and registered in the national registry for clinical studies (DRKS 00003971). Between 2010 and 2012, 5217 eligible adult patients provided written consent and were enrolled into the study. Patients were included if they had an eGFR between 30 and 60 ml/min per 1.73 m² or an eGFR>60 ml/min per 1.73 m² and a urinary albumin-to-creatinine ratio (UACR) > 300 mg/g, albuminuria > 300 mg/d, a urinary protein-to-creatinine ratio > 500 mg/g, or proteinuria > 500 mg/d. Trained personnel obtained information on clinical data, socio-demographic factors, medical and family history, medications, and health-related quality of life. The leading cause of CKD was ascertained from the treating nephrologist. Moreover, biospecimens (plasma, serum, whole blood, spot urine) were collected in a standardized way at the enrollment visit, processed, and shipped frozen to a central laboratory for routine clinical chemistry and to a central biobank for future analyses following standard operating procedures. A complete description of study design and the recruited study population can be found elsewhere. For the current analysis, serum and urine specimens collected at baseline were selected for metabolite measurements from a subset of GCKD participants: all participants recruited from the Freiburg study center and additionally all GCKD patients from other study centers with autosomal-dominant polycystic kidney disease, focal segmental glomerulosclerosis, membranous nephropathy, membranoproliferative GN, rapid progressive GN, pauci-immune, and anti–glomerular basement membrane GN as the leading cause of CKD (Supplemental Material).

Genotyping and Imputation

Genomic DNA was extracted from whole blood using an automated magnetic bead–based technology, quantified and normalized on a pipetting robot platform, and available for 5123 GCKD participants. Genotyping was conducted for 2,612,357 markers at the Helmholtz Center Munich using Omni2.5Exome BeadChip arrays (Illumina, GenomeStudio, Genotyping Module Version 1.9.4). Data cleaning was carried out separately for the Omni2.5 content and the Exome Chip content of the array. For the Omni2.5 content, data QC and cleaning followed the protocol of Anderson et al. Per-individual QC steps included evaluation of call rate, sex check, heterozygosity, cryptic relatedness, and genetic ancestry. Altogether, 89 individuals were removed during the per-individual QC. In the per–single nucleotide polymorphism (SNP) QC
CLINICAL RESEARCH

Exome Chip association analyses. Specified markers that were subsequently used for GWAS.

J Am Soc Nephrol

Samples were shipped frozen to BIOCRATES Life Sciences AG.

Metabolite Measurements

1000 Genomes Project ALL haplotypes.

Filtering for imputation quality (info) value of $>0.8$. More detailed information of the Exome Chip QC is found in the Supplemental Material. The cleaned Exome Chip dataset contained genotypes of 226,233 variants from 23,377,794 SNPs. The scripts from Anderson et al. which are on the basis of Plink v1.90 and R programming language were used in the genotyping data cleaning steps.

Cleaned Omni2.5 genotype data were then imputed using IMPUTE2 (v2.3.1) following the Best Practices for Imputation on the IMPUTE2 website (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#best_practices).

Similarly, per-individual QC and per-SNP QC steps were performed for the Exome Chip content, with additional checks specific to exome chip data as in Guo et al. Overall, serum measurements from 1200 individuals and urinary measurements from 1187 individuals could be obtained. All concentration values given are in micromolar (micromoles per liter). A detailed overview of the measured metabolites is provided in Supplemental Table 1. Measurements were further subjected to detailed quality control and data cleaning procedures (Supplemental Material).

Additional Variables

Serum creatinine was measured in the GCKD study in a central laboratory using an IDMS traceable enzymatic assay (Creatinine plus, Roche). eGFR values were calculated using the creatinine-based Chronic Kidney Disease Epidemiology Collaboration formula. The UACR was calculated from urinary albumin and creatinine measurements, where creatinine was measured using the same assay as in serum and albumin with the ALBU-XS assay (Roche/Hitachi Diagnostics GmbH, Mannheim, Germany).

GWAS and Exome Chip Analyses

Cleaned metabolite data were merged with genotype data yielding a dataset for 139 serum metabolites from 1168 individuals and 41 urinary metabolites from 1155 individuals for subsequent genetic analyses (Supplemental Material). The concentrations of the urinary metabolites were normalized using the probabilistic quotient to account for dilution. Within each matrix, pairwise ratios were calculated: Met A/Met B. For 37 of the metabolites, both serum and urinary concentrations were available, and the FE for each of these metabolites was calculated as $(100 \times \frac{\text{Met}_u \times \text{Cre}_a}{\text{Met}_a \times \text{Cre}_u})$ where Met$_u$ and Cre$_u$ are the urinary concentrations of the metabolite and of creatinine, and Met$_a$ and Cre$_a$ their serum concentrations. FEs and ratios were subjected to analytic level QC on the basis of evaluations of completeness, variability, and outliers, resulting in 34 FEs, 9591 serum metabolite ratios, and 820 urinary metabolite ratios for analyses. Before GWAS, all values (single metabolites, FEs, ratios) were transformed using rank-based inverse normal transformation.

An in-house GWAS analysis pipeline was set up using SNPTTEST (v2.5.2; https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) for association testing and GWA-toolbox (V2.2.4) for QC of the GWAS results. This pipeline was used for the GWAS of single serum and urinary metabolite concentrations and the FEs, using filtered imputed data of genotype dosages and assuming an additive genetic model. As in previous GWAS of metabolite concentrations, age and sex

steps, SNPs with $<96\%$ call rate, deviating from Hardy-Weinberg equilibrium ($P$ value $<1E^{-5}$), and those with duplicate positions were removed. Detailed steps of data cleaning are available in the Supplemental Material. The final dataset for the Omni2.5 genotypes contained 5034 individuals and 2,337,794 SNPs. The scripts from Anderson et al. which are on the basis of Plink v1.90 and R programming language were used in the genotyping data cleaning steps.

Cleaned Omni2.5 genotype data were then imputed using IMPUTE2 (v2.3.1) following the Best Practices for Imputation on the IMPUTE2 website (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#best_practices). SHAPEIT v2 was used for strand alignment as well as for prephasing. The 1000 Genomes Project ALL haplotypes - Phase 3 integrated variant set was used as the reference. Imputation was performed chunk-wise with 5-Mbp sequences in each chunk. Initial imputation yielded genotypes for 81,698,995 variants. Filtering for imputation quality (info) value of $>0.8$ and minor allele frequency (MAF) $\geq 1\%$ led to 9,281,895 high-quality imputed markers that were subsequently used for GWAS.

Similarly, per-individual QC and per-SNP QC steps were performed for the Exome Chip content, with additional checks specific to exome chip data as in Guo et al. In the per-individual QC steps, 96 individuals were removed, and in the per-SNP QC, 3818 SNPs were removed on the basis of call rate $>95\%$ and Hardy-Weinberg equilibrium $P$ value $\geq 1E^{-5}$. More detailed information of the Exome Chip QC is found in the Supplemental Material. The cleaned Exome Chip dataset contained genotypes of 226,233 variants from 5027 individuals. These cleaned data were postprocessed by zCall$^3$ with a z-score threshold of six and then used in the Exome Chip association analyses.

**Metabolite Measurements**

Samples were shipped frozen to BIOCRATES Life Sciences AG in Innsbruck (Austria) and stored at $-80{\degree}C$ until the time of analysis in 2014/2015 when aliquots were thawed, centrifuged, and the supernatant was used for analysis. Following the manufacturer’s protocol of the commercially available AbsoluteIDQ p180 Kit, serum and urine samples were handled in a uniform and standard way that includes a precipitation/extraction step. Using that kit, metabolites was calculated as (100 $\times$ Met$_u$ $\times$ Cre$_a$/ (Met$_a$ $\times$ Cre$_u$)) where Met$_u$ and Cre$_u$ are the urinary concentrations of the metabolite and of creatinine, and Met$_a$ and Cre$_a$ their serum concentrations. FEs and ratios were subjected to analytic level QC on the basis of evaluations of completeness, variability, and outliers, resulting in 34 FEs, 9591 serum metabolite ratios, and 820 urinary metabolite ratios for analyses. Before GWAS, all values (single metabolites, FEs, ratios) were transformed using rank-based inverse normal transformation.

An in-house GWAS analysis pipeline was set up using SNPTTEST (v2.5.2; https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) for association testing and GWA-toolbox (V2.2.4) for QC of the GWAS results. This pipeline was used for the GWAS of single serum and urinary metabolite concentrations and the FEs, using filtered imputed data of genotype dosages and assuming an additive genetic model. As in previous GWAS of metabolite concentrations, age and sex
were included as covariates in all analyses; analyses of serum metabolite concentrations additionally included adjustment for log(eGFR). After correcting for multiple testing using a Bonferroni procedure, statistical significance was defined as P<3.6E−10 for 139 serum metabolites, P<1.2E−09 for 41 urinary metabolites, and P<1.5E−09 for 34 FEs; suggestive significance was defined as P value <5.0E−08. Regional association plots of loci with significant associations were generated by LotusZoom (v1.3).27

The proportion of variance in inverse-normal transformed metabolite trait (y) explained by an index SNP was calculated as R²=β²×var(SNP)/var(y), where β is the estimated effect of the SNP on y, var(SNP)=2×MAF²SNP×(1−MAF²SNP) according to the binomial distribution of the alleles.28

Because of the large number of evaluated ratios, a two-step analytic approach was used, with the first step consisting of GWAS of the 9591 serum and 820 urinary metabolite ratios using the software OmicABEL29 on the basis of cleaned, genotyped markers. Adjustment variables are the same as in the single variant analysis by SNPTEST. The OmicABEL results were filtered for associations with chi-square-association statistics >30, which is equivalent to a P value <4.3E−8. Serum and urinary metabolite ratios with at least one SNP that passed this filter were then subjected in a second step to the SNPTEST-based GWAS pipeline using imputed genotype data, as for single metabolites and FEs. The Bonferroni-corrected significance threshold was set at 5E−8/9591=5.2E−12 for serum metabolite ratios to account for the conduct of 9591 GWAS, and 5E−8/820=6.1E−11 for urinary metabolite ratios. Additionally, the GWAS results of ratios were filtered on the “P-gain,” defined as the minimum of the two P values of the single metabolites composing the ratio divided by the P value of the ratio.30 Only associations with a P-gain>9591 in serum and P-gain>820 in urine were further considered.

For the analysis of the exome chip data, two kinds of gene-based tests of single metabolites, FEs, and metabolite ratios were carried out using the R package seqMeta (v1.6.5) and using the same phenotypes and adjustment variables as for the single variant associations: a burden test and the sequence kernel association test. For both tests, only variants with MAF<1% and likely to be functional (splicing, nonsynonymous, stopgain, and stoploss variants on the basis of the SNPInfo annotation file used in the CHARGE Consortium) were included.31 The test results were filtered and only genes with cumulative minor allele count ≥10 and with ≥2 variants per gene were retained. In total, 6924 genes met these filter criteria. To account for multiple testing, the overall significance threshold (0.05) was adjusted for the number of respective outcomes (per matrix), the number of genes (n=6924), and the number of performed tests (n=2). The significance thresholds were thus P<2.6E−08 for 139 serum metabolites, P<8.8E−08 for 41 urinary metabolites, P<3.8E−10 for 9591 serum metabolite ratios, P<4.4E−09 for 820 urinary metabolite ratios, and P<1.1E−07 for 34 FEs. Suggestive statistical significance was defined as P<1E−06.

Power Considerations
The power to detect associations between metabolites and genetic variants in a study of 1150 unrelated individuals such as ours is provided across a range of MAFs and effect sizes in Supplemental Table 10. For example, the power to detect an association at the genome-wide significance level of 5.0E−08 for genetic markers with MAF>0.1 and an absolute effect size >0.5 is >80%.

RESULTS
Among the 1143 patients with CKD for whom both genetic information as well as serum and urinary metabolite measurements were available, the mean eGFR and the median UACR were 51.6 ml/min per 1.73 m² and 160.8 mg/g, respectively (Table 1). The mean age was 56 years, and 63% of the patients were male. An overview of all metabolites studied, including their abbreviations, limits of detection, and coefficients of variation, is provided in Supplemental Table 1.

Genetic Association Studies of Metabolites in Patients with CKD: Model Selection
As a first step, we tested whether known genetic associations with metabolites could be replicated in a CKD population, and evaluated the potential influence of covariate selection and, specifically for urinary metabolites, the method used to account for dilution. We systematically selected variants with large genetic effects on metabolites from early publications of metabolite GWAS3–12 where both variant and metabolite were available in our data. These known genetic associations also achieved genome-wide significance among patients with CKD, as illustrated by associations between serum nonanoylcarnitine concentrations and an SNP in ACADL or serum butyrylcarnitine concentrations and an SNP in ACAD5, as well as urinary lysine concentrations and an SNP in SLC7A9 (Table 2). Adjustment of serum metabolite associations for eGFR led to increased statistical significance, mainly due to smaller SEM. Thus, genetic screens with metabolite concentrations in serum were adjusted for eGFR in subsequent analyses. In urine, the association between rs7247977 in SLC7A9 and lysine concentrations was best detected when probabilistic quotient–normalization25 was used to account for urine dilution, rather than normalization by urinary creatinine concentrations. The association changed little upon further adjustment for UACR. Hence, subsequent GWAS for urinary metabolite concentrations were adjusted for age and sex.

GWAS and Burden Tests Identify Significant Associations with 25 Serum Metabolites
We carried out GWAS of common SNPs (MAF>0.01) with concentrations of 139 metabolites in serum. The genomic inflation factors ranged from 0.99 to 1.03, consistent with the absence of inflation from systematic errors. Multiple previously reported associations were identified at genome-wide
The number of GCKD patients with genetic data and either serum measurements or urinary measurements is slightly higher: n=1168 and n=1155, respectively. BMI, body mass index; IQR, interquartile range.

*These variables were not available for some patients (proportion<2%).
Table 2. Evaluation of known genetic associations in a CKD population, exemplified for nonanoylcarnitine (serum) and lysine (urine)

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Model</th>
<th>SNP</th>
<th>Chr Position</th>
<th>Ref/Effect</th>
<th>Effect</th>
<th>SEM</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serum</td>
<td>Nonanoylcarnitine, C9</td>
<td>n2286963</td>
<td>2</td>
<td>211060050</td>
<td>T/G</td>
<td>0.34</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>Butyrylcarnitine, C4</td>
<td>n2014355</td>
<td>12</td>
<td>121175524</td>
<td>T/C</td>
<td>0.28</td>
<td>0.41</td>
</tr>
<tr>
<td>Urine</td>
<td>Lysine creatinine-normalized</td>
<td>n7427977</td>
<td>19</td>
<td>33358355</td>
<td>T/C</td>
<td>0.39</td>
<td>0.22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Normalization</th>
<th>Chr Position</th>
<th>Ref/Effect</th>
<th>Effect</th>
<th>SEM</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serum</td>
<td>Normalization</td>
<td>Chr Position</td>
<td>Ref/Effect</td>
<td>Effect</td>
<td>SEM</td>
<td>P Value</td>
</tr>
<tr>
<td></td>
<td>adjustment for age, sex</td>
<td>2</td>
<td>211060050</td>
<td>T/G</td>
<td>0.34</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>adjustment for age, sex, eGFR</td>
<td>12</td>
<td>121175524</td>
<td>T/C</td>
<td>0.28</td>
<td>0.41</td>
</tr>
<tr>
<td>Urine</td>
<td>creatinine-normalized adjustment for age, sex, lnUACR</td>
<td>19</td>
<td>33358355</td>
<td>T/C</td>
<td>0.39</td>
<td>0.22</td>
</tr>
</tbody>
</table>

**GWAS and Burden Tests of Serum Metabolite Ratios Highlight Variants Central to Lipid and Acylcarnitine Metabolism**

Metabolite concentrations can be used to model physiologic functions of the kidney. Specifically, ratios of metabolites can reflect not only the activity of enzymatic reactions, but also co- or counter-transport at the apical or basolateral membrane of tubular cells, resulting in synchronized changes of these metabolites in urine and blood, respectively. We therefore first evaluated the association between common genetic variants and metabolite ratios in serum. Genome-wide significant associations were identified between genetic variants and 259 metabolite ratios mapping into 11 loci (Supplemental Table 5), many of which implicated known genes central to lipid biosynthesis. These analyses confirmed FADS1 as a major player in the metabolism of phosphatidylcholines, ACADM and ACADS in metabolism of medium-chain acylcarnitines, SPTLC3 in the synthesis of sphingomyelins, as well as PDXDC1 in phospholipid metabolism.3,35

In the gene-based tests of rare variants on serum metabolite ratios, we identified an exome-wide significant association (P<2.4E−10) for one gene, ACADM. Significant associations were observed for several metabolite ratios, with the lowest P value observed for the decanoylcarnitine/dodecanoylcarnitine ratio from the burden test (P=6.6E−16, Figure 2). The low burden test P value, despite the modest sample size and low frequency of the evaluated deleterious variants (MAF<0.01), suggested large and concordant effects of the contributing variants affecting the metabolite concentrations. Indeed, the four variants in ACADM that contributed to the test were all associated with increases in the decanoylcarnitine/dodecanoylcarnitine ratio (Figure 2). ACADM encodes medium-chain acyl-CoA dehydrogenase, which catalyzes the initial reaction in the β-oxidation of C4 to C12 straight-chain acyl-CoAs. Mutations in the ACADM gene cause MCAD deficiency (MIM #201450), a condition that is part of newborn screening on the basis of acylcarnitine analysis in blood. Three of the four variants detected in our screen have been reported as cause of classic MCAD deficiency if present in the homozygous state (p.E43K, p.Y67H, p.K329E), confirming the functional and clinical relevance of these mutations. The mutation p.K329E gives rise to a severe disease phenotype, whereas the clinical picture is milder for the mutations p.E43K and p.Y67H, suggesting residual enzyme function. These observations from monogenic manifestations of the disease are consistent with the effect sizes observed for heterozygous carriers in our study, with p.K329E showing the largest effect on the decanoylcarnitine/dodecanoylcarnitine ratio. The last variant identified in our study, p.I364T, is yet of uncertain clinical
Table 3. Common variants (MAF≥0.05) associated with serum metabolites at genome-wide significance (P<3.6E–10)

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>SNP</th>
<th>Chr</th>
<th>Position</th>
<th>NEA</th>
<th>EA</th>
<th>Effect</th>
<th>SEM</th>
<th>P Value</th>
<th>AF_EA</th>
<th>Proportion of Variance Explained</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>C10</td>
<td>rs12134854</td>
<td>1</td>
<td>76137606</td>
<td>T</td>
<td>C</td>
<td>-0.30</td>
<td>0.04</td>
<td>1.43E-12</td>
<td>0.31</td>
<td>0.04</td>
<td>SLC44A5(dist=60807),ACADM(dist=52437)</td>
</tr>
<tr>
<td>Ac-Orn</td>
<td>rs375811360</td>
<td>2</td>
<td>73870324</td>
<td>C</td>
<td>G</td>
<td>-0.84</td>
<td>0.05</td>
<td>1.04E-62</td>
<td>0.21</td>
<td>0.24</td>
<td>NAT8</td>
</tr>
<tr>
<td>C9</td>
<td>rs3764913</td>
<td>2</td>
<td>21107409</td>
<td>T</td>
<td>C</td>
<td>0.46</td>
<td>0.04</td>
<td>1.34E-35</td>
<td>0.34</td>
<td>0.09</td>
<td>ACADL</td>
</tr>
<tr>
<td>Gly</td>
<td>rs1047891</td>
<td>2</td>
<td>211540507</td>
<td>C</td>
<td>A</td>
<td>0.39</td>
<td>0.04</td>
<td>2.43E-20</td>
<td>0.31</td>
<td>0.07</td>
<td>CPS1</td>
</tr>
<tr>
<td>C18:2</td>
<td>rs12194000</td>
<td>6</td>
<td>110776012</td>
<td>A</td>
<td>G</td>
<td>-0.31</td>
<td>0.05</td>
<td>3.44E-11</td>
<td>0.24</td>
<td>0.03</td>
<td>SLC22A16</td>
</tr>
<tr>
<td>C4</td>
<td>rs146203232</td>
<td>6</td>
<td>160521855</td>
<td>C</td>
<td>T</td>
<td>-0.54</td>
<td>0.06</td>
<td>1.57E-18</td>
<td>0.09</td>
<td>0.05</td>
<td>IGF2R</td>
</tr>
<tr>
<td>Putrescine</td>
<td>rs4725366</td>
<td>7</td>
<td>150522054</td>
<td>A</td>
<td>G</td>
<td>-0.36</td>
<td>0.05</td>
<td>2.42E-12</td>
<td>0.81</td>
<td>0.04</td>
<td>TMEM176A(dist=19846),AOC1(dist=27511)</td>
</tr>
<tr>
<td>C0</td>
<td>rs1171616</td>
<td>10</td>
<td>61468589</td>
<td>G</td>
<td>T</td>
<td>0.38</td>
<td>0.05</td>
<td>2.84E-15</td>
<td>0.77</td>
<td>0.05</td>
<td>SLC16A9</td>
</tr>
<tr>
<td>PC ae C38:4</td>
<td>rs11320420</td>
<td>11</td>
<td>61542006</td>
<td>GA</td>
<td>G</td>
<td>-0.28</td>
<td>0.04</td>
<td>2.35E-10</td>
<td>0.33</td>
<td>0.04</td>
<td>MYRF</td>
</tr>
<tr>
<td>PC ae C36:5</td>
<td>rs11320420</td>
<td>11</td>
<td>61542006</td>
<td>GA</td>
<td>G</td>
<td>-0.33</td>
<td>0.04</td>
<td>2.79E-13</td>
<td>0.33</td>
<td>0.05</td>
<td>MYRF</td>
</tr>
<tr>
<td>PC ae C36:5</td>
<td>rs11320420</td>
<td>11</td>
<td>61542006</td>
<td>GA</td>
<td>G</td>
<td>-0.34</td>
<td>0.04</td>
<td>1.47E-14</td>
<td>0.33</td>
<td>0.05</td>
<td>MYRF</td>
</tr>
<tr>
<td>PC ae C38:5</td>
<td>rs174537</td>
<td>14</td>
<td>61552680</td>
<td>G</td>
<td>T</td>
<td>-0.32</td>
<td>0.04</td>
<td>9.45E-13</td>
<td>0.33</td>
<td>0.04</td>
<td>MYRF</td>
</tr>
<tr>
<td>PC ae C36:5</td>
<td>rs174537</td>
<td>14</td>
<td>61552680</td>
<td>G</td>
<td>T</td>
<td>-0.34</td>
<td>0.04</td>
<td>2.47E-14</td>
<td>0.33</td>
<td>0.05</td>
<td>MYRF</td>
</tr>
<tr>
<td>PC ae C38:5</td>
<td>rs174537</td>
<td>14</td>
<td>61552680</td>
<td>G</td>
<td>T</td>
<td>-0.36</td>
<td>0.04</td>
<td>1.37E-16</td>
<td>0.33</td>
<td>0.06</td>
<td>MYRF</td>
</tr>
<tr>
<td>PC ae C38:4</td>
<td>rs102275</td>
<td>14</td>
<td>61557803</td>
<td>T</td>
<td>C</td>
<td>-0.40</td>
<td>0.04</td>
<td>9.76E-21</td>
<td>0.33</td>
<td>0.07</td>
<td>TMEM258</td>
</tr>
<tr>
<td>lysoPC a C20:4</td>
<td>rs102275</td>
<td>14</td>
<td>61557803</td>
<td>T</td>
<td>C</td>
<td>-0.50</td>
<td>0.04</td>
<td>2.63E-32</td>
<td>0.33</td>
<td>0.11</td>
<td>TMEM258</td>
</tr>
<tr>
<td>PC aa C36:4</td>
<td>rs102274</td>
<td>14</td>
<td>61557826</td>
<td>T</td>
<td>C</td>
<td>-0.41</td>
<td>0.04</td>
<td>8.07E-21</td>
<td>0.33</td>
<td>0.07</td>
<td>TMEM258</td>
</tr>
<tr>
<td>PC aa C36:3</td>
<td>rs174560</td>
<td>14</td>
<td>61581764</td>
<td>T</td>
<td>C</td>
<td>0.28</td>
<td>0.04</td>
<td>2.52E-10</td>
<td>0.29</td>
<td>0.03</td>
<td>FADS1</td>
</tr>
<tr>
<td>PC aa C36:5</td>
<td>rs174566</td>
<td>14</td>
<td>61592362</td>
<td>A</td>
<td>G</td>
<td>-0.32</td>
<td>0.04</td>
<td>4.66E-14</td>
<td>0.33</td>
<td>0.05</td>
<td>FADS1(dist=7833),FADS2(dist=3351)</td>
</tr>
<tr>
<td>C4</td>
<td>rs2014355</td>
<td>12</td>
<td>121175524</td>
<td>T</td>
<td>C</td>
<td>0.42</td>
<td>0.04</td>
<td>3.90E-27</td>
<td>0.28</td>
<td>0.07</td>
<td>ACADS</td>
</tr>
<tr>
<td>PC aa C28:1</td>
<td>rs7160525</td>
<td>14</td>
<td>64232220</td>
<td>G</td>
<td>A</td>
<td>0.42</td>
<td>0.05</td>
<td>6.57E-15</td>
<td>0.16</td>
<td>0.05</td>
<td>SGPP1(dist=37464),SYNE2(dist=87463)</td>
</tr>
<tr>
<td>SM (OH) C14:1</td>
<td>rs112329286</td>
<td>14</td>
<td>64239877</td>
<td>T</td>
<td>TTTTA</td>
<td>0.36</td>
<td>0.05</td>
<td>5.44E-12</td>
<td>0.16</td>
<td>0.04</td>
<td>SGPP1(dist=45121),SYNE2(dist=79806)</td>
</tr>
<tr>
<td>PC ae C32:1</td>
<td>rs1077989</td>
<td>14</td>
<td>67975822</td>
<td>A</td>
<td>C</td>
<td>-0.28</td>
<td>0.04</td>
<td>1.97E-12</td>
<td>0.48</td>
<td>0.04</td>
<td>TMEM229B</td>
</tr>
<tr>
<td>Asn</td>
<td>rs2282377</td>
<td>14</td>
<td>104571820</td>
<td>C</td>
<td>G</td>
<td>0.43</td>
<td>0.06</td>
<td>1.23E-11</td>
<td>0.13</td>
<td>0.04</td>
<td>ASPG</td>
</tr>
<tr>
<td>Pro</td>
<td>rs117935223</td>
<td>22</td>
<td>18911333</td>
<td>C</td>
<td>A</td>
<td>0.62</td>
<td>0.07</td>
<td>1.79E-17</td>
<td>0.09</td>
<td>0.06</td>
<td>PRODH</td>
</tr>
</tbody>
</table>

Sample size: n=1168. Chr, chromosome; NEA, noneffect allele; EA, effect allele; AF_EA, allele frequency of effect allele; Genes, gene the SNP maps into or genes neighboring the SNP and the corresponding distances in base pairs.
significance (Figure 2). Our findings reveal a relatively high carrier frequency of potentially deleterious ACADM mutations with large effects on metabolite concentrations in an adult population of European ancestry (23 of 1157=2.0%; 1.9% without p.I364T). This relatively high frequency is also observed among European ancestry individuals in the gnomAD database, suggesting that this finding is more widely applicable. Full results of all serum metabolite ratios associated at suggestive significance with rare variants in one or more genes are listed in Supplemental Table 6.

GWAS of Urine Metabolite Ratios Highlight Amino Acids Counter-Transported by SLC7A9

There were several genetic loci with significant SNP associations with urinary metabolite ratios (Supplemental Table 7). Interestingly, these associations delivered valuable insights into renal physiology: genetic variants in SLC7A9 showed significant associations with lysine-containing ratios that were many orders of magnitude stronger than the association of the respective SNP with urinary lysine concentrations alone (Figure 3). This is exemplified by the lysine/glutamine ratio (P=2.2E−23, P-gain=2.1E+11; see Methods), the lysine/tyrosine ratio (P=3.1E−22, P-gain=1.5E+10), and the lysine/threonine ratio (P=5.8E−22, P-gain=1.5E+10). Other significant lysine-containing ratios contained alanine, glycine, serine, asparagine, phenylalanine, and tryptophan. Of note, all of these amino acids carry no net electric charge at physiologic pH, consistent with the known role of SCL7A9 as an exchanger of cationic amino acids (lysine, arginine, and ornithine) and cystine in the renal tubular lumen against intracellular neutral amino acids (AA) described in in vitro studies. This finding illustrates the potential of modeling metabolite concentrations to uncover specific physiologic functions of the kidney. The concept is illustrated in Figure 3A: individuals homozygous for the minor C allele at the intronic index SNP rs12460876 show higher urinary lysine/AA ratios than those homozygous for the major T allele, indicating less efficient function of the encoded SLC7A9 transport protein. Modeling of the association between rs12460876 and all pairwise amino acid ratios in urine revealed specific lysine/AA ratios with strong associations (Figure 3B), representing a readout of human SLC7A9 function in vivo. Other amino acid ratios showed little (mostly ornithine-containing ratios) or no association, such as lysine/AA− ratios (aspartate, glutamate).

The evaluation of the aggregate effect of rare variants on urinary metabolite ratios did not yield findings that achieved exome-wide significance (P<4.4E−09). There were, however, biologically plausible findings of suggestive significance (P<1E−06, Supplemental Table 8). For example, rare variants in SLC36A2 were associated with the urinary glycine-to-serine ratio (P=4.0E−07). This matches the known transport capacity of the encoded transporter for glycine and its localization to the S1 segment of the proximal tubule. Transport capacity for serine has only been described for murine SLC36A2 so far and could now be tested experimentally for human SLC36A2. Another example is the association between rare variants in PAH and the phenylalanine/tyrosine ratio (P=5.9E−07), which matches the function of the encoded phenylalanine hydroxylase that catalyzes the rate-limiting step in phenylalanine catabolism, the hydroxylation of phenylalanine to tyrosine. Although an earlier study had reported the association between rare variants in PAH and blood phenylalanine concentrations, our findings highlight that modeling of enzyme function as the ratio of the enzyme’s substrate and product adds additional information. In addition, the examination of urine adds information, because the gene is exclusively expressed in liver and kidney, and urinary concentrations integrate the information of reactions occurring in the kidney. More generally, these findings suggest that the study of metabolite ratios in urine of patients with CKD in a larger study sample is a promising avenue for future research.

GWAS and Burden Tests of Metabolite FEs

Lastly, GWAS of common SNPs with the FEs of 34 metabolites yielded one finding of genome-wide significance, the association between variants in the MLEC gene at the ACADS locus and the FE of butyrylcarnitine (C4, P=2.8E−14; Supplemental Table 9). However, these associations were not as significant as the ones with serum butyrylcarnitine concentration alone (P=3.9E−27, Table 3). No significant or suggestive associations

Table 4. Effect of G allele of rs4725366 (AOC1 locus) on serum putrescine concentrations stratified by eGFR

<table>
<thead>
<tr>
<th>eGFR Category (ml/min per 1.73 m²)</th>
<th>n</th>
<th>Effect</th>
<th>SEM</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥60</td>
<td>293</td>
<td>−0.28</td>
<td>0.10</td>
<td>6.37E−03</td>
</tr>
<tr>
<td>45–59</td>
<td>343</td>
<td>−0.41</td>
<td>0.09</td>
<td>4.16E−06</td>
</tr>
<tr>
<td>30–44</td>
<td>391</td>
<td>−0.30</td>
<td>0.09</td>
<td>1.34E−03</td>
</tr>
<tr>
<td>&lt;30</td>
<td>117</td>
<td>−0.60</td>
<td>0.15</td>
<td>1.12E−04</td>
</tr>
</tbody>
</table>

n, sample size.
(P<1E−06) were identified for the aggregate effect of rare deleterious variants in any gene and the FEs of the 34 evaluated metabolites. Thus, investigation of the FEs of metabolites did not provide additional insights.

**DISCUSSION**

Our study has several main findings: first, genetic influences on metabolite concentrations in serum and urine from previous population-based studies2–7,9,11,40,41 translate to patients with CKD. Second, the study of patients with CKD revealed a novel association between concentrations of a uremic toxin, putrescine,42 and variation near the gene encoding amine oxidase copper-containing 1. Third, metabolite ratios are useful composite indices that reflect the presence of dose-response effects with eGFR <30 compared with those with eGFR ≥60 ml/min per 1.73 m² may hint at the presence of dose-response effects with worsening kidney function and illustrate the value of studying “challenged” populations (such as disease cohorts) in general. Again, however, differences across studies complicate comparisons: the lack of reported associations with putrescine in previous population-based studies may be because it was not included on targeted assays or removed during quality control, because it was concentrated too lowly for quantification, because it was measured but not identified as putrescine, or because it showed no genetic associations and was therefore not reported. Although transcriptional activation of AOC1 has recently been linked to kidney morphogenesis in mice by affecting polyamine breakdown,44 it cannot be inferred from our study whether elevated putrescine concentrations occur as a consequence of CKD or contribute causally to CKD etiology and/or progression.

In terms of biologic insights, our study offers snapshots of renal physiology in humans and generates novel hypotheses for experimental confirmation. This former is exemplified by the identification of the neutral amino acids transported in exchange for urinary lysine in humans in vivo, and the latter by the potential transport capacity of human SLC36A2 for serine. Several previous studies reported associations between the same index or a proxy variant at the SLC7A9 locus with urinary lysine concentrations6,11 or some lysine-containing ratios,11 but without biologic interpretations in light of tubular counter-transport mechanisms. Interestingly, GWAS of eGFR in the general population could show that the rs12460876 C allele is associated not only with higher urinary lysine concentrations but also with higher eGFR and lower CKD prevalence and incidence,6,13,45 consistent with the association between higher urinary lysine concentrations and decreased risk for CKD.13 The molecular mechanisms connecting reduced SLC7A9-mediated transport, resulting in higher urinary lysine and cystine and higher intracellular AA concentrations, to better eGFR/lower CKD risk are presently unclear.
Studies of genetic influences on metabolite concentrations in patients with CKD have the potential to ultimately deliver insights of clinical importance; for example, by identifying genetic predispositions to unfavorable profiles of CKD risk factors or to CKD comorbidities. This may be especially relevant for rare variants with large effects on CKD risk factors, the presence and prevalence of which is illustrated by the discovery of rare coding ACADM variants in our study. Illuminating pathways that underlie processes such as hyperlipidemia or stone formation may eventually help to better treat or prevent these common CKD-related conditions. In addition, genetic discoveries related to metabolites that are considered uremic toxins such as putrescine can be used to test their causal involvement in disease progression once large-scale genetic studies of CKD progression become available.

Some limitations of our study warrant mention: the use of the BIOCRATES p180 assay for metabolite quantification results in a limited number of metabolites to study. Although nontargeted approaches to quantify metabolites result in a much greater number of metabolites for study and cover classic uremic toxins that are not part of the BIOCRATES assay, the latter was chosen in this study because it includes isotope-labeled standards for many metabolites that allow for comparison and integration of their concentrations across urine and blood. We could not comprehensively compare genetic effect sizes on metabolite concentrations in patients with CKD to those from previous population-based studies because of differences in methods to quantify metabolites and statistical data analysis. Our study was limited to individuals of European ancestry, and can therefore not necessarily be generalized. Twenty-four-hour urine collections were not available in the GCKD study, which may have led to increased variation of urinary metabolite concentrations because of differences in circadian rhythms resulting in reduced power to detect associations. Urate is one of few metabolites whose FE has been assessed

Figure 3. Counter-transport of amino acids by SLC7A9 in the renal proximal tubule cell is revealed by the analysis of urinary metabolite ratios. The upper part (A) contains a schematic representation of SLC7A9 function in the proximal tubule, reabsorbing dibasic urinary amino acids such as lysine in exchange for intracellular neutral amino acids. Individuals homozygous for the minor C allele at the index SNP rs12460876 show higher lysine/neutral amino acid (AA0) ratios in the urine, indicating less efficient lysine reuptake at the brush border. Listed amino acids are restricted to the ones measured in this study. The lower panel (B) represents a heatmap of the strength of the association between genotype at intronic rs12460876 (scaled chi square statistics) and the evaluated pairwise metabolite ratios in urine. It highlights the AA0 counter-transported against lysine in vivo. For comparison, lysine by itself has a scaled chi square statistic of 0.39, indicated by light orange color. Ala, alanine; Asn, asparagine; Asp, aspartate; Gln, glutamine; Glu, glutamate; Gly, glycine; His, histidine; Ile, isoleucine; Lys, lysine; Orn, ornithine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; Val, valine.
from different urine collection methods in studies of healthy individuals. In these studies, the FEIs for urate quantified from spot urine were similar to those from 24-hour urines, but this may not pertain to all metabolites.

Our study represents the first study of genetic influences on metabolite profiles in a large population of patients with CKD. We were able to examine up to 1168 patients with CKD and benefitted from the standardized data collection, sample preprocessing and storage procedures, DNA extraction, and genotyping in the GCKD study. Our study shows that genetic influences on metabolite profiles in the general population translate to patients with CKD and that novel genetic associations can be identified in this study population. Modeling of renal metabolite handling such as counter-transport mechanisms delivered novel insights into human renal (patho-)physiology and generated new hypotheses for experimental confirmation. Future extensions of sample size and metabolite coverage hold great promise to generate additional insights into human renal physiology in vivo.

ACKNOWLEDGMENTS

We are grateful for the willingness of the patients to participate in the German Chronic Kidney Disease (GCKD) study. The enormous effort of the study personnel of the various regional centers is highly appreciated. We thank the large number of nephrologists who provide routine care for the patients and collaborate with the GCKD study. The GCKD Investigators are listed in the Supplementary Information.

The work of P.S., Y.L., and A.K. and metabolite quantification was supported by grants KO 3598/2-1 and KO 3598/4-1 of the German Research Foundation (DFG). The work of M.W., U.T.S., and A.K. was supported by DFG CRC 1140. P. Schlosser was supported by DFG CRC 992. A.K. was additionally supported by a DFG Heisenberg Professorship (KO 3598(3)-1); U.T.S. was additionally supported by the Else Kroener Fresenius Kolleg Nierenfunktionsstörungen als Komplikation von Systemerkrankungen (NAKSYS). Genotyping was supported by Bayer Pharma Aktiengesellschaft (AG). Metabolite quantification was additionally supported by the Else Kroener Fresenius Foundation and the German Ministry of Education and Research (BMBF) Gerontosy II NephAge Project, 031 5896A. The GCKD study was funded by grants from the BMBF (grant number 01ER0804) and the Kuratorium fuer Heimodialyse (KiH) Foundation for Preventive Medicine.

DISCLOSURES

J.W. is employed by BIOCRATES Life Sciences Aktiengesellschaft (AG) (Innsbruck, Austria).

REFERENCES

Mieming, AT; Graber, A. (Fischnalerstrasse 14/Top3, A-6020 Innsbruck, AT); Stögl, W.M. (Hoher Weg 9/7, A-6020 Innsbruck, AT); Device for quantitative analysis of a metabolite profile. (BIOCRATES Life Sciences AG (Innrain 66, 6020 Innsbruck, AT), 2014).


See related article, “Insights into CKD from Metabolite GWAS,” on pages 1349–1351.

This article contains supplemental material online at http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2017101099/-/DCSupplemental.

AFFILIATIONS

1Institute of Genetic Epidemiology, Department of Biometry, Epidemiology, and Medical Bioinformatics, and 2Department of General Pediatrics, Center for Pediatrics and Adolescent Medicine, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany; 3BIOCRATES Life Sciences Aktiengesellschaft, Innsbruck, Austria; 4Department of Nephrology and Hypertension and 5Institute of Human Genetics, University of Erlangen-Nürnberg, Erlangen, Germany; 6Institute of Functional Genomics, University of Regensburg, Regensburg, Germany; and 7Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria.