Genome-Wide Association Studies of Metabolite Concentrations (mGWAS): Relevance for Nephrology

Anna Köttgen, MD, MPH,[*] Johannes Raffler, PhD,[†] Peggy Sekula, PhD,[*] and Gabi Kastenmüller, PhD[†,‡]

Summary: Metabolites are small molecules that are intermediates or products of metabolism, many of which are freely filtered by the kidneys. In addition, the kidneys have a central role in metabolite anabolism and catabolism, as well as in active metabolite reabsorption and/or secretion during tubular passage. This review article illustrates how the coupling of genomics and metabolomics in genome-wide association analyses of metabolites can be used to illuminate mechanisms underlying human metabolism, with a special focus on insights relevant to nephrology. First, genetic susceptibility loci for reduced kidney function and chronic kidney disease (CKD) were reviewed systematically for their associations with metabolite concentrations in metabolomics studies of blood and urine. Second, kidney function and CKD-associated metabolites reported from observational studies were interrogated for metabolite-associated genetic variants to generate and discuss complementary insights. Finally, insights originating from the simultaneous study of both blood and urine or by modeling intermetabolite relationships are summarized. We also discuss methodologic questions related to the study of metabolite concentrations in urine as well as among CKD patients. In summary, genome-wide association analyses of metabolites using metabolite concentrations quantified from blood and/or urine are a promising avenue of research to illuminate physiological and pathophysiological functions of the kidney.

Keywords: Genome-wide association studies, kidney, filtration, CKD, metabolomics

The kidneys have a profound impact on the human metabolome, the collection of small molecules that represent intermediates or end products of metabolism. Not only do the kidneys have a central role in the filtration of metabolites present in blood, but also in metabolite generation (anabolism), breakdown (catabolism), as well as active secretion and reabsorption along the nephron. The prominent role of the kidneys in clearing blood metabolites is illustrated by the routine use of serum creatinine concentrations to estimate the filtration function of the kidneys, the estimated glomerular filtration rate (eGFR).

Studies of the metabolome are of interest to the field of nephrology for several reasons: first, many of these small molecules are freely filtered. Metabolites therefore represent attractive filtration markers in addition to creatinine,[1] which has limitations that can lead to imprecise estimates of GFR and hence the definition and staging of chronic kidney disease (CKD).[2,3] Second, studies of the human metabolome can reveal indicators of the nonfiltration functions of the kidney in health and disease. Third, metabolites may be causal for the etiology and/or progression of kidney diseases, such as chronically increased glucose for the development of diabetes and subsequent diabetic kidney disease. Fourth, metabolites may represent useful markers of CKD severity and progression that contain prognostic information in addition to serum creatinine.

The coupling of genetics to metabolomics is a powerful tool to gain insights into molecular mechanisms that link human genetic variation to interindividual differences in metabolite concentrations, a concept termed *genetically influenced metabotypes* (GIMs).[4] The GIM concept stipulates that the presence of metabolic differences introduced by genetic variation may influence an individual's ability to respond to challenges as well as its susceptibility to disease. The genome and the metabolome are closely linked, as opposed to the more distal relationship between genetic variants and disease end points. This close, proximal relationship along with the long-standing knowledge of human physiology and biochemistry facilitates the biological interpretation of identified GIMs. Genetic variants associated with metabolite concentrations often map into genes encoding for enzymes that directly or indirectly metabolize the respective or a closely related metabolite, or into genes encoding for their transport proteins.[4–9]

[*]*Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany.*
[†]*Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.*
[‡]*German Center for Diabetes Research (DZD), Neuherberg, Germany.*

Financial support: Supported by German Research Foundation grant KO 3598/4-1 (P.S. and A.K.), and grants CRC 1140 and KO 3598/3-1 (A.K.).

Conflict of interest statement: none.

Address reprint requests to Anna Köttgen, MD, Institute of Genetic Epidemiology, Medical Center, University of Freiburg. E-mail: anna.koettgen@uniklinik-freiburg.de (A. Köttgen)

0270-9295/ - see front matter

https://doi.org/10.1016/j.semnephrol.2018.01.009

With respect to nephrology, mGWAS therefore can deliver insights into physiologic functions of the kidney such as differences in the ability to excrete waste products, or to gain insights into mechanisms relevant to the etiology or progression of kidney diseases. For example, genetic studies can implicate renal metabolite transporters through the identification of genetic variants in the corresponding genes that are associated with concentrations of the transported metabolite. Because the substrates for many renal transporters are only partially identified and their relevance in human beings in vivo often is unclear,[10] such studies have the potential to uncover novel aspects of renal transport physiology. Genetic studies of metabolite concentrations in urine additionally may deliver insights about metabolites that are generated or metabolized specifically within the kidney and released into urine. Finally, genetic studies of human metabolism among patients with CKD may identify processes that are specifically up-regulated in the disease state, for example, catabolic functions of the kidneys such as detoxification reactions or increased tubular secretion in the setting of high blood metabolite concentrations.

In this article, we address several complementary aspects of how genetic studies of metabolites can provide insights relevant to nephrology and provide examples where relevant. Nongenetic studies of human metabolism with implications for nephrology are covered by another article in this issue.

OVERVIEW OF PUBLISHED MGWAS STUDIES

We conducted a literature review to systematically capture mGWAS articles (Table 1). Most mGWAS published to date examined blood, followed by studies that examined urine. The studies used proton nuclear magnetic resonance spectroscopy (NMR)- or mass spectrometry (MS)-based methods for metabolite quantification, and were conducted predominantly in populations of European ancestry. The high number of citations despite the relatively recent publication date of many of the articles illustrates the impact of these studies and the initiation of multiple follow-up projects (Table 1).

INSIGHTS INTO (PATHO-)PHYSIOLOGIC KIDNEY FUNCTIONS: MGWAS OF BLOOD METABOLITE CONCENTRATIONS

Genetic studies of blood metabolite concentrations in the general population can implicate processes related to renal metabolite handling and hence kidney function physiology. If the identified genetic variants additionally are associated with kidney function and CKD end points, the implicated metabolites may represent causes (intermediate phenotypes) or consequences of CKD (reverse causation). Although insights into causes of CKD improve our understanding of the molecular processes of disease and may provide a basis for the development of targeted therapies, the identification of metabolic consequences of CKD may be useful for understanding comorbidities and disease progression.

mGWAS of Single Metabolites

In a low-dimensional setting, GWAS of individual metabolites already have been conducted before the advent of metabolomics techniques that enable the simultaneous quantification of a wide range of metabolites. In fact, many biomarkers used in routine clinical biochemistry are metabolites such as lipids, glucose, or uric acid (serum urate). As an example, a GWAS of serum urate concentrations in 2008 identified significantly associated genetic variants that mapped into the *ABCG2* locus.[11] *ABCG2* encodes for a multidrug resistance transporter that had not been linked previously to urate transport and whose physiological function was unclear. Follow-up studies showed that ABCG2 is a human urate transporter expressed at the apical membrane of renal tubular cells and the intestine, where it mediates urate secretion. A common ABCG2 loss-of-function variant, Q141K, was identified and confirmed experimentally.[12] This example illustrates the potential of mGWAS to not only provide new insights into renal physiology, but to also be of direct pathophysiological relevance: increased serum urate concentrations can cause gout, the most common inflammatory arthritis. More generally, studying the genetics of metabolite concentrations in the general population enables insights into diseases that result when the respective metabolites exceed their physiological ranges. In nephrology, this could be of interest for the study of diabetic kidney disease and kidney stones, among others.

mGWAS Using Metabolomics

An example of insights into tubular transport mechanisms from a high-dimensional setting was reported by Suhre et al.[7] This study of hundreds of blood metabolites quantified via a nontargeted MS-based approach identified, among many other loci, genetic variants in *SLC16A9* that were associated with free carnitine concentrations. The investigators experimentally confirmed that human SLC16A9 (monocarboxylate transporter 9) functioned as a carnitine efflux transporter using the *Xenopus* oocyte system. Antibody staining for monocarboxylate transporter 9 protein in human tissues shows expression in the basolateral membrane of intestinal epithelial cells and the apical membrane of renal tubular epithelial cells,[13] suggesting that the

Table 1. Published Genome-Wide Association Studies of Metabolomic Traits

Study	Measurement Technique	Matrix	Frequency		Study Source(s)	Population Size	Citations, n
			Metabolic Traits	Pairwise Ratios			
Analyses of blood traits (serum or plasma)							
Gieger et al,[5] 2008	MS, Biocrates	Serum	363	All	1 study, German, male	284	318
Illig et al,[6] 2010	MS, Biocrates	Serum	163	26,406	Discovery: 1 study, German	1,809	317
					Replication: 1 study, British	422	
Mittlestrass et al,[62] 2011*	MS, Biocrates	Serum	131	–	Discovery: 1 study, German	1,797	133
					Replication: 2 studies, German	1,218/328	
Suhre et al,[7] 2011	MS, Metabolon	Serum	276	37,179	2 studies, German/ British	2,820	417
Demirkan et al,[66] 2012	MS, targeted	Plasma	153	Selected	5 studies, European	4,034	74
Inouye et al,[60] 2012	NMR, nontargeted	Serum	130	–	2 studies, Finn	6,600	60
Kettunen et al,[67] 2012	NMR, nontargeted	Serum	117	Selected	5 studies, Finn	8,330	246
Krumsiek et al,[56] 2012†	MS, Metabolon	Serum	517	–	1 study, German	1,768	58
Tukiainen et al,[68] 2012	NMR, nontargeted	Serum	117	Selected	5 studies, Finn	8,330	43
Hong et al,[69] 2013	MS, nontargeted	Serum	6,138	–	Discovery: 1 study, Swedish, case-control setting	402 (including 214 cases with prostate cancer)	16
					Replication: 1 study, Swedish, case-only setting	489 (only cases with prostate cancer)	
Raffler et al,[70] 2013†	NMR, nontargeted	Plasma	8,600	124,750	1 study, German	1,757	8
Rhee et al,[71] 2013	MS, targeted	Plasma	217	–	1 study, US Europeans	2,076	66
Ried et al,[61] 2014	MS, Biocrates (B)	Serum	B: 151	–	Discovery: 1 study, German	1,809	10
	MS, Metabolon (M)		M: 193		Replication: 1 study, British	843	
Shin et al,[8] 2014	MS, Metabolon	Plasma or serum	486	98,346	2 studies, German/ British	7,824	213
Yu et al,[18] 2014	MS, Metabolon	Serum	308	–	1 study, US African American	1,260	24
Demirkan et al,[72] 2015	NMR, nontargeted	Serum	42	–	1 study, Dutch	2,118	23
Draisma et al,[40] 2015	MS, Biocrates	Serum	129	–	Discovery: 7 studies, European	7,478	32
					Replication: 1 study, German	1,182	
Kraus et al,[73] 2015	MS, targeted	Plasma	63	–	Discovery: 1 study, US, case-control setting	1,490 (including 745 cases of ca stenosis)	14
					Replication: 1 study, US	2,022	

Table 1 (*continued*)

Study	Measurement Technique	Matrix	Frequency		Study Source(s)	Population Size	Citations, n
			Metabolic Traits	Pairwise Ratios			
Krumsiek et al,[63] 2015*	MS, Metabolon	Serum	507	–	Discovery: 1 study, German Replication: 1 study, German	1,756 1,000	9
Kettunen et al,[74] 2016	NMR, nontargeted	Plasma or serum	123	–	14 studies, European	24,925	30
Rhee et al,[75] 2016	MS, nontargeted (discovery) MS, Metabolon (replication)	Plasma or serum	217	–	Discovery: 1 study, US Replication: 1 study, US European	2,076 1,528	7
Yet et al,[59] 2016	MS, Biocrates (B) MS, Metabolon (M)	Serum	B: 160 M: 488	–	1 study, British	1,001	11
Yu et al,[76] 2016	MS, Metabolon	Serum	308	–	Discovery: 1 study, US African American Replication: 1 study, US African American	1,361 508	1
Davis et al,[77] 2017	NMR, nontargeted	Serum	72	–	1 study, Finn, male	8,372	0
Long et al,[9] 2017	MS, Metabolon	Serum	644	–	1 study, British	1,960	6
Analyses of urine traits							
Suhre et al,[24] 2011	NMR, nontargeted	Nonfasting or overnight fasting, spontaneous urine samples	59	1,661	Discovery: 1 study, German, male Replication: 2 studies (1) German, female (2) German, both sexes	862 870 992	127
Montoliu et al,[25] 2013	NMR, nontargeted	Urine†	Unknown	–	1 study, Brazil	265	4
Rueedi et al,[26] 2014†	NMR, nontargeted	Urine	1,276	–	Discovery: 1 study, European Replication: 1 study, Brazil	835 601	25
Raffler et al,[27] 2015†	NMR, targeted (T) and nontargeted (N)	Nonfasting or overnight fasting, spontaneous urine samples	T: 55 N: 166	T: 1,463 N: 13,695	Discovery: 1 study, German Replication: 1 study, German	3,861 1,691	14
McMahon et al,[28] 2017	MS, targeted	Fasting morning urine samples	151	–	1 study, US, case-control setting	386 (including 193 cases of incident CKD)	2

Analyses of blood (serum or plasma) and urine traits							
Nicholson et al,[23] 2011	NMR, nontargeted (N) MS, Biocrates (B; only plasma)	Plasma and spot urine samples	N: 512 B: 163	N: – B: all	Discovery: 1 study, British, female, twin Replication: 1 study, British, case-control setting	142 69 (including 27 metabolic syndrome cases)	59

All assessed study populations represent population-based samples, often drawn as subsamples from parent studies if not stated otherwise. The literature search was as follows: mGWAS publications were identified via: (source 1) look-up http://www.metabolomix.com/, or (source 2) PubMed/Medline search based on medical subject heading (MeSH) terms (December 2017). To set up the MeSH term search, we reviewed which Mesh terms had been used in conjunction with studies listed in source 1 (N = 29). We constructed a search algorithm based on a selection of the most relevant terms. The Medline search resulted in 2,879 entries (1,859 published in the past 5 years). Titles and, if requested, abstracts of these references were reviewed manually, leading to the identification of 2 additionally studies not listed in source 1. Any study following a candidate approach or focused on the evaluation of few metabolites of 1 biochemical class were excluded. Finally, the list additionally was manually curated. The number of how many times an mGWAS was cited was obtained via Web of Science (including citations of an mGWAS by another one).

Abbreviations: MS, mass spectrometry–type technique; ca stenosis, coronary artery stenosis.

*mGWAS was stratified by sex.

†Study also had a focus on metabolite identification in nontargeted metabolomics data.

transporter may be important in the intestinal absorption and renal reabsorption of free carnitine. Interestingly, genetic variants in *SLC16A9* also have been identified in association with serum urate.[14] Most human transport proteins studied to date do not exclusively transport a single substrate, even if affinity across substrates varies. Because the coverage of the human metabolome in metabolomics experiments constantly is expanding, it is likely that mGWAS will continue to deliver important insights into renal metabolite handling in human beings (eg, by identifying new transporter substrates), even if the study population size remains constant.

mGWAS: Insights into Genetic Risk Loci for CKD

The association of genetic variants in loci detected in GWAS of kidney function and CKD with metabolite concentrations is particularly interesting because it may illuminate the pathophysiology of causes and/or consequences of CKD. We therefore performed a systematic search of metabolite associations for known genetic loci linked to kidney function or CKD. Table 2 represents an overview of 16 loci identified through this search, all of which contained genetic variants associated with metabolite concentrations in blood.

Several of these loci, including *GCKR*, *NAT8*, and *CPS1*, represent hubs of metabolism, as illustrated by the numerous associated metabolites. Moreover, *GCKR* especially was identified in GWAS of many other complex traits and diseases not listed in Table 1. Of special interest to nephrology is *NAT8*. It encodes for N-acetyl-transferase 8, an enzyme specifically expressed in liver and kidney. NAT8 has been proposed as important for the acetylation of the α-amino group of cysteine S-conjugates to form mercapturic acids that then can be excreted in bile and urine. The pathway therefore represents a major route of detoxification.[15] Interestingly, the major A allele at the missense index variant rs13538 in *NAT8* is associated with a lower eGFR,[16,17] higher blood concentrations of few metabolites (N-δ-acetylornithine, 2-aminooctanoate, methionine sulfone), but lower blood concentrations of many N-α-acetylated amino acids (Table 2). The latter could be considered products of the N-acetylation reaction catalyzed by NAT8. Consistent with the mGWAS findings, higher concentrations of N-δ-acetyl-ornithine in blood were associated with lower eGFR in the general population,[1,7] and N-δ-acetyl-ornithine–associated variants in *NAT8* were associated with new-onset CKD in another study.[18] These observations could be explained by a genetically encoded lower ability to detoxify waste products that leads to the accumulation of toxic substances and subsequent kidney dysfunction and disease. An alternative

Table 2. CKD-Associated Loci With Metabolite Associations in Blood or Urine

Genetic Locus	SNP	Kidney-Related Trait* (Study, Effect Allele, Effect Direction)	Metabolite (Study, Effect Allele, Effect Direction)		Functional Interpretation
			Serum†	Urine	
CASP9/ DNAJC16	rs12124078 chr1:15869899	$eGFR_{crea}$ (G-[78])	4-guanidinobutanoate (G-[9])		*AGMAT* also is contained in the LD interval and is an eQTL for the index SNP in some GTEx tissues. Agmatinase hydrolyses agmatine to putrescine and urea in human beings; it is catabolized by way of 4-guanidinobutyrate to 4-aminobutyrate into the Krebs cycle in plants.[79] This matches the function of the associated metabolite from human mGWAS. In the human GWAS of eGFR, morpholino knockdown of casp9 in zebrafish embryos showed podocyte and tubular abnormalities with altered dextran clearance.
GCKR	rs1260326 chr2:27730940	$eGFR_{crea}$ (C-[17])	Triglycerides (T+,[29] A-,[80] NR[81]) Total cholesterol (T+[29]) PC (40:5) (T+[40]) PC (32:2) (T+[40]) PC (34:4) (T+[40]) Threonine (C-[9]) Alanine (T+,[8] C-[74]) 3-methyl-2-oxovalerate (T+[8]) γ-glutamylglutamine (T-[8]) Mannose (T-[8]) Lactate (T+[82]) X-03094 (T+[8]) X-22822 (C-[9])		The glucokinase regulatory protein encoded by *GCKR* is synthesized in liver. It regulates glucokinase, a key enzyme in glucose metabolism. The allele associated with lower kidney function is associated with lower concentrations of blood lipids, threonine, alanine, and lactate, and with higher concentrations of mannose in blood. Observationally, however, higher triglyceride concentrations are associated with lower kidney function. The index variant has been identified in dozens of GWAS of anthropometric and metabolic diseases or traits, underscoring the central role of GCKR in metabolism. Whether the association with kidney function can be ascribed to direct effects in the kidney versus indirect effects of unfavorable metabolic profiles on kidney function is unclear.
NAT8	rs13538‡ chr2:73868328 rs10206899 chr2:73900900	$eGFR_{crea}$ (A-[17]) Creatinine (A+[16])	N-δ-acetylornithine (G-,[9] A+[8]) N-acetylcitrulline (G+[9]) N-acetyl-1-methylhistidine (G+[9]) N-acetyl-3-methylhistidine (G+[9]) N-acetylglutamine (G+[9]) N-acetyltyrosine (G+[9]) N-acetyllysine (A-[8]) N-acetylphenylalanine (G+[9]) N-acetylhistidine (G+[9]) N-acetylasparagine (G+[9]) 2-aminooctanoate (G-,[9] A+[8]) 2-hydroxyoctanoate (G-[9]) methionine sulfone (G-[9]) X-11787 (A+,[8] A+[18])	N-acetylated compound (G+[27])	*NAT8* encodes for N-acetyl-transferase 8, which is expressed specifically in liver and kidney. Its function in human beings is not well studied; it has been proposed as important for the acetylation of the α-amino group of cysteine S-conjugates to form mercapturic acids that can be excreted in bile and urine. The pathway therefore represents a major route of detoxification.[15] Alternatively, or in addition, it may catalyze peptidyl-lysine N6-acetylation of various proteins. The allele associated with lower kidney function is associated with lower levels of N-α-acetylated metabolites in blood and urine and higher concentrations of some other metabolites. This can be reconciled with genetically reduced NAT8 function, leading to accumulation of toxic substances and resulting kidney damage, or by the fact that both NAT8 function and glomerular filtration represent complementary but correlated functions of the kidney.

			X-12511 (G+[9]) X-13477 (A+[8]) X-21411 (G-[9])		
CPS1	rs1047891 (alias rs7422339) chr2:211540507	$eGFR_{crea}$ (A-[17]) CKD (A+[83])	Glycine (A-[9]) Serine ([9]A+, NR[40]) N-acetylglycine (A+[9]) Propionylglycine (A-[9]) N-palmitoylglycine (A-[9]) Hexanoylglycine (A-[9]) Cinnamoylglycine (A-[9]) Betaine (A+[9]) Homoarginine (A-[9]) Guanidinosuccinate (A-[9]) Pyroglutamine (A-[9]) γ-glutamylglycine (A+[9]) γ-glutamylthreonine (A-[9]) 3-methylglutaconate (A+[9]) 3-methylglutarylcarnitine (2) (A-[9]) Glutarylcarnitine (C5) (A-[9]) HDL cholesterol (C+[29]) X-16570 (A+[9])	Glycine ([27]A+) Creatine ([27]A+)	*CPS1* encodes a mitochondrial carbamoyl phosphate synthetase, which synthesizes carbamoyl phosphate from ammonia and bicarbonate. It is the rate-limiting enzyme in the urea cycle and highly expressed in liver. Rare mutations in CPS1 cause hyperammonemia (MIM 237300). Glycine could be linked to *CPS1* via excess ammonia that is converted to glycine, a precursor of the osmolyte betaine. The GWAS index variant is a missense SNP. The allele associated with lower eGFR is associated with lower glycine and higher betaine in blood and higher glycine in urine. Genetic associations with kidney function could result from a direct effect connection of CPS1 function to the kidney, or because low eGFR and hyperammonemia are correlated.
SLC34A1	rs6420094 chr5:176817636	$eGFR_{crea}$ (G-[17]) CKD (G+[83])	(Iso)Leucylphenylalanine (G+[9]) Glycylphenylalanine (G+[9]) Prolylproline (G+[9])		*SLC34A1* encodes for a Na+ phosphate cotransporter that is highly expressed in the kidney; it has not been linked to the transport of phenylalanine to date. The index variant is also an eQTL for the neighboring *RGS14* transcript in several tissues; RGS14 has been linked experimentally to kidney disease via altered salt sensitivity.[84]
RSPO3	rs1936800 chr6:127436064	BUN (T+[19])	HDL cholesterol (C+[29]) Triglycerides (T+[29])		The index variant is upstream of and in LD with several eQTLs of *RSPO3.* The allele associated with higher BUN shows association with higher blood triglycerides and lower HDL cholesterol. Because the variant was not reported in association with eGFR or CKD, it may reflect mechanisms in the liver and be associated with BUN because of increased BUN production rather than reduced excretion. RSPO3 plays a role in Wnt signaling.

Table 2 (*continued*)

Genetic Locus	SNP	Kidney-Related Trait* (Study, Effect Allele, Effect Direction)	Metabolite (Study, Effect Allele, Effect Direction)		Functional Interpretation
			Serum†	Urine	
ATXN2	rs653178 chr12:112007756	$eGFR_{cys}$ (C-[17])	Kynurenine (T-[8])		The lead variant is a cis-eQTL for *ALDH2* and a trans-eQTL for variants on multiple other chromosomes. *ALDH2* encodes the aldehyde dehydrogenase 2 family (mitochondrial), which is important in alcohol degradation. Kynurenine has been reported to inhibit ALDH2 activity.[85] The connection to genes in the locus to CKD is unclear.
ALDH2	rs671§ chr12:112241766	Creatinine (A+[19])	N-acetylaspartate (NR[9]) Kynurenine (NR[9])		rs617 encodes for a K to E substitution in *ALDH2* and is associated with multiple traits. The variant is not found in EUR study populations but is common in East Asians. It is in LD with the *ATXN2* locus in East Asians; variants in high LD extend into multiple other genes in the locus. ALDH2 detoxifies both endogenous and exogenous aldehydes, but precise mechanisms linking it to CKD are unknown.
	rs2074356§ chr12:112645401	BUN (A+[19])			
GATM	rs2467853 chr15:45698793	$eGFR_{crea}$ (G-[22]) CKD (G+[83])	Homoarginine (G+[9])		*GATM* encodes for arginine:glycine amidinotransferase, an enzyme involved in creatine biosynthesis. It catalyzes the transfer of a guanido group from L-arginine to glycine, resulting in the immediate precursor of creatine, guanidinoacetic acid. The association with eGFR estimated from creatinine but not cystatin C supports the role of genetic variation in *GATM* in creatinine synthesis.[22] The enzyme's function also matches the observed association with homoarginine in blood.[86]
CYP1A1	rs2472297 chr15:75027880	$eGFR_{crea}$ (C-[87])	Caffeine (T-[8])		*CYP1A1* encodes for cytochrome P450 family 1 subfamily A member 1, which is involved in phase I drug metabolism. Genetic variants in *CYP1A1* are associated with coffee consumption,[88] matching the observed metabolite association. CYP1A1 has been linked to kidney function through several pathways, including detoxification mechanisms. The allele associated with lower kidney function is associated with higher levels of caffeine.

SLC7A6	rs1111571 chr16:68363181	$eGFR_{crea}$ (G-[89])	Glutaroylcarnitine (A-[8]) Lysine (A+[8])		SLC7A6 has diverse functions in amino acid transport, including sodium-independent uptake of dibasic amino acids and sodium-dependent uptake of some neutral amino acids. The allele associated with lower eGFR is associated with lower lysine levels in blood. An association between lower urinary lysine levels and risk of CKD has been reported from observational studies.[28]
CDK12/ PNMT	rs11078903 chr17:37631924	$eGFR_{crea}$ (A-[78])	Methionine sulfone (A-[9])	Histidine (G-[27]) Alanine (G-[27]) Tyrosine (G-[27])	*PNMT* encodes phenylethanolamine N-methyltransferase, which catalyzes the last step of catecholamine biosynthesis. Catecholamine biosynthesis is connected via several steps to tyrosine metabolism, potentially explaining the mGWAS association in urine. One previous study reported an association between SNPs in *PNMT* and acute kidney injury.[90] No mechanistic link between *CDK12* and kidney disease is known. Multiple other genes map into this region of extensive LD.
SLC7A9	rs4805834 chr19:33453659	Creatinine (C+[16]) $eGFR_{crea}$ (C-[17])	Homocitrulline (T-[8])	Lysine (T+[27])	*SLC7A9* encodes for transport protein expressed in the apical membrane of renal tubular cells that exchange dibasic amino acids in urine such as lysine and cystine against intracellular neutral amino acids. The allele associated with lower kidney function is associated with lower urinary lysine concentrations. Conversely, mutations in *SLC7A9* can cause autosomal-recessive cystinuria, featuring higher levels of lysine in urine. Kidney disease in these patients results from cystine urolithiasis; the mechanisms therefore may differ from the one observed for common variants.
SLC6A13	rs10774021 chr12:349298	$eGFR_{crea}$ (T-[17])	3-aminoisobutyrate (T-[9]) 1-methylimidazoleacetate (T-[9]) Imidazole lactate (T+[9]) Pyroglutamine (T-,[9] T+[8]) Deoxycarnitine (T-[9]) X-11334(NR[8])	3-aminoisobutyrate (C-[27])	*SLC6A13* is known to encode the GABA transporter GAT2, which is highly expressed in kidney. The associations identified in blood and urine mGWAS suggest that it also may transport the structurally related 3-aminoisobutyrate. In GWAS of kidney function, the index variant was associated with creatinine, but not cystatin-C–based eGFR, suggesting that the association with eGFR may result from a relation to creatinine transport rather than reduced kidney function. In support, SLC6A13 shows high homology with a known creatine transporter, SLC6A8.

Table 2 (*continued*)

Genetic Locus	SNP	Kidney-Related Trait* (Study, Effect Allele, Effect Direction)	Metabolite (Study, Effect Allele, Effect Direction)		Functional Interpretation
			Serum†	Urine	
INHBC	rs1106766 chr12:57809456	$eGFR_{crea}$ (C-[83])	Triglycerides (C+[29]) HDL cholesterol (T+[29])		*INHBC* encodes the inhibin β C chain, which belongs to the TGF-β superfamily. Inhibins act as hormones with a wide range of actions; the mechanism behind the association with kidney function is unclear.
DPEP1	rs164748‖ chr16:89708292	$eGFR_{crea}$ (G-[83])	Homocysteine (NR[91]) Methionine sulfone (NR[9])		*DPEP1* encodes for the renal enzyme dipeptidase 1, which hydrolyzes a wide range of dipeptides. Genetic variants in *DPEP1* have been reported to associate with homocysteine levels in several studies, providing a plausible link to the observed mGWAS finding. There is extensive LD and multiple genes in this genomic region, it is therefore unclear whether *DPEP1* represents the most likely candidate gene for association with CKD.

The list of genetic loci associated with kidney-function–related traits is based on Wuttke and Kottgen[92] and was updated manually to include additional traits and recent publications. For each locus, we list the SNP that shows the strongest association signal for the given trait. For these SNPs and their proxies in strong linkage disequilibrium ($r^2 >= 0.5$), we used the variant annotation and block annotation tools of SNiPA (v3.2 March 2017, www.snipa.org[50]) to identify associations ($P < 5\times10^{-8}$) reported in GWAS with metabolic traits.
Abbreviations: BUN, blood urea nitrogen; $eGFR_{crea}$, eGFR based on serum creatinine; $eGFR_{cys}$, eGFR based on cystatin C; eQTL, expression Quantitative Trait Locus/Loci; EUR, European; GTEx, Genotype Tissue Expression[52]; HDL, high-density lipoprotein; LD, linkage disequilibrium; TGF, transforming growth factor; NR, not reported.
*Many of the loci were found in subsequent GWAS of kidney function and disease. Only the first study is cited for a given trait.
†Creatinine was detected as an associated metabolite for several loci (*SLC7A9*, *TBX2*, *SLC22A2*, *CPS1*, and *NAT8*), but is not presented because of its use to estimate the outcome eGFR.
‡Effect alleles and directions for metabolite associations are shown in respect to this SNP.
§SNP is not in SNiPA's variant panel. Significant metabolite associations within a 500-kb window are shown instead.
‖Metabolite associations were not detected with the index SNP itself, but with a variant/variants in LD. NR, not reported.

explanation could be that NAT8 function reflects a different but correlated aspect of kidney health than does glomerular filtration, leading to the identification of the same associated variants with both eGFR and NAT8-related metabolites. This shared genetic association with complementary but correlated read-outs of kidney function is exemplified in the large overlap of genetic loci associated with serum creatinine and blood urea nitrogen.[19] Future studies that could further evaluate the aspect of cause versus consequence of mGWAS findings include studies of the corresponding metabolites among CKD patients, in prospective studies, through the use of Mendelian randomization,[20] as well as experimental studies.

Another theme among the loci in Table 2 is that many of them contain genes encoding for transport proteins expressed in the kidney, including *SLC7A6*, *SLC7A9*, and *SLC34A1*. Both *SLC7A6* and *SLC7A9* encode for proteins that transport cationic amino acids[21] and showed associations with lysine concentrations in mGWAS. Although the *SLC7A6* variant associated with lower eGFR was associated with lower blood lysine concentrations, the variant in *SLC7A9* that was associated with lower eGFR was associated with higher blood concentrations of homocitrulline and lower concentrations of lysine concentrations in urine. A direct comparison and interpretation of these findings is complicated by the nonoverlapping metabolite panels across mGWAS, the different biofluids studied, and by our limited knowledge of the precise tissue distribution and localization, as well as of the unknown substrates of many transport proteins. The observations, however, are consistent with the hypothesis that the function of transport proteins as quantified by their transported metabolites may represent a read-out of kidney health.

Finally, some loci in Table 2 such as gene (*GATM*) and *SLC6A13* may illustrate that the use of eGFR as a creatinine-based measure of kidney function can lead to the identification of genetic variants related to creatinine metabolism per se, but not to kidney filtration function. *GATM* encodes for an enzyme involved in the generation of guanidinoacetic acid, the immediate precursor of creatine. The association of genetic variants with eGFR estimated from creatinine but not from cystatin C, an alternative peptide marker of kidney function, supports a role of *GATM* in creatinine synthesis rather than reduced eGFR.[22] Likewise, the index variant in *SLC6A13* was associated with creatinine-based, but not cystatin-C–based, eGFR in a large GWAS meta-analysis of kidney function.[17] Although SLC6A13 is known as a γ-aminobutyric acid (GABA) transporter, it shows high homology with a known creatine transporter, SLC6A8. This opens up the possibility that SLC6A13 may have a wider range of substrates than currently appreciated.

INSIGHTS INTO (PATHO-)PHYSIOLOGIC KIDNEY FUNCTIONS: MGWAS OF URINE METABOLITE CONCENTRATIONS

The study of genetic determinants of metabolite concentrations in urine has the potential to show additional insights beyond those that can be gained by studying blood. For example, urine metabolite concentrations may reflect kidney-specific processes such as active tubular secretion or reabsorption, or may be more amenable to the detection of GIMs because of higher concentrations of some metabolites in urine compared with blood.

Five mGWAS in urine have been published to date,[23–27] all of which quantified metabolites via NMR. Across studies, genetic variants in 26 genetic loci were identified in association with urinary metabolite concentrations at genome-wide significance. In addition, one study reported on the association between genetic variants on the exome chip and the concentration of urinary metabolites quantified via a MS-based method.[28]

Some of the 26 published GIMs in urine have not been detected in mGWAS studies of blood, but all of the kidney function and CKD-associated loci highlighted in Table 2 that contain associations with urine metabolites also show metabolite associations in blood. Raffler et al[27] reported that two thirds of 22 urine GIMs detected in their study also were identified when studying the respective metabolites' blood concentrations (or those of a related metabolite) with concordant effect directions. This suggests that the majority of associations with metabolites detected in urine is owing to their glomerular filtration. The ability to detect genetic loci that translate from blood to urine because of a metabolite's active reabsorption or secretion, rather than filtration, also is conceivable. The direction of such associations in blood and urine will be determined by the presence and localization of metabolite-specific transport proteins. In addition, Raffler et al[27] also reported instances in which a metabolite identified in urine could be considered unrelated to the one(s) previously detected in blood mGWAS. For example, although genetic variants in *PNMT* are associated significantly with high-density lipoprotein cholesterol in blood,[29] the associated metabolites in urine were histidine and tyrosine.[27]

As shown in Table 2, significant associations were identified between genetic variants in *SLC7A9* and urinary lysine concentrations in urine,[27] but not blood, despite the ability to detect other GIMs for lysine in blood such as *SLC7A6*.[8] This observation can be explained plausibly by the localization of SLC7A9 at the apical membrane of tubular epithelial cells, where it reabsorbs basic amino acids including lysine from urine in exchange for neutral intracellular amino acids.

Genetically encoded differences in lysine reabsorption therefore may be detected more readily in urine. Interestingly, the *SLC7A9* allele associated with lower kidney function in GWAS of eGFR is associated with lower urinary lysine concentrations.[27] This observation would be consistent with unfavorable effects of altered concentrations of extracellular or intracellular SLC7A9 substrate(s) on kidney function. Alternatively, as described for NAT8 earlier, SLC7A9 activity may capture a complementary read-out of kidney function correlated with eGFR, leading to the identification of the same genetic loci.

The study of metabolite concentrations in the urine raises an additional methodologic question, how to best account for interindividual differences in urine dilution. Different approaches have been used, the most common ones being indexing of individual metabolite concentrations to urinary creatinine,[24,27,28] or to the probabilistic quotient.[23,30] Some studies in which metabolites were quantified via nontargeted NMR or MS-based approaches also used normalization to the complete metabolite content as represented by the sum of peak integrals or through standardizing the spectrum of each individual by Z-transformation based on all spectral features after binning of the spectra to achieve a mean of zero and standard deviation of one.[26,31] Future efforts should include population-based and CKD patient-based studies that systematically use known positive genetic controls to compare and contrast different normalization methods.

LOCI ASSOCIATED WITH CONCENTRATIONS OF THE SAME METABOLITE IN BOTH BLOOD AND URINE MGWAS

GIMs that are associated with concentrations of the same metabolite in both blood and urine may be especially interesting because their effect directions and sizes can be directly compared. Because of the small number of urine mGWAS and the difference in platforms used (NMR versus MS), the number of such cases still is small but can be expected to grow in the future. The comparison of effects can be particularly instructive when effects occur in opposing directions. For example, although the allele associated with lower eGFR at *SLC6A13* is associated with lower concentrations of 3-aminoisobutyrate in blood,[9] it is associated with higher concentrations of 3-aminoisobutyrate in urine.[27] *SLC6A13* is known to encode GABA transporter 2 (GAT2), a transporter of GABA and betaine that is highly expressed in kidney.[32] Both γ-aminobutyric acid (GABA) and betaine are structurally related to 3-aminoisobutyrate. Together, the associations identified in blood and urine mGWAS suggest that SLC6A13 also may transport 3-aminoisobutyrate across kidney cells, which could now be tested experimentally. Associations with opposing genetic effect directions also were observed for glycine concentrations and *CPS1*. Here, the allele associated with lower eGFR was associated with lower glycine concentrations in blood and higher ones in urine (Table 2). The biological interpretation, however, is not as straightforward because the encoded carbamoyl phosphate synthetase I operates in the urea cycle and is highly expressed in liver and intestine. In addition, an observational study reported a positive association between higher urinary glycine concentrations and lower odds of new-onset CKD,[28] which is not consistent with the effect direction from genetic studies.

INSIGHTS FROM MAPPING KIDNEY FUNCTION–ASSOCIATED METABOLITES ONTO THEIR RESPECTIVE GENES

The investigation of gene regions encoding for metabolites that were linked to kidney function and disease in nongenetic studies also provided useful insights. This is illustrated by various examples in Table 3, which lists genome-wide significant genetic associations ($P < 5.0\times10^{-8}$) with metabolites that were found to correlate with incident CKD in the recent metabolome-wide study in urine by McMahon et al[28] or that have been linked repeatedly to eGFR, eGFR decrease, or incident CKD in studies of blood metabolites.[1,33,34,35] Additional studies in specific settings, such as studies of the urine metabolome among patients with diabetic kidney disease, also exist but are not the focus of Table 3.[36]

For instance, serum inositol levels were associated positively with incident CKD,[30] and myo-inositol, which is the most abundant naturally occurring inositol stereoisomer, was associated negatively with eGFR[1] in population-based studies. Moreover, myo-inositol also showed associations with genetic variants at the *SLC5A11* locus: the minor allele at the index variant rs17702912 was correlated with lower myo-inositol concentrations in blood,[8] while the same variant was associated significantly with higher myo-inositol concentrations in urine.[27] The index variant is located downstream of the gene and was annotated as a regulatory variant in numerous tissues and cell types. SLC5A11 is a Na^+-myo-inositol cotransporter expressed at the apical brush-border membrane of renal tubular cells,[37] where it mediates myo-inositol uptake from urine into tubular cells.[38] Reduced function of SLC5A11 would result in higher urinary and lower blood myo-inositol concentrations, as observed for the reported index variant. Because genetic variants at the *SLC5A11* locus have not been identified in association with kidney function or CKD and

Table 3. Selected CKD-Associated Metabolites With Genetic Associations to Metabolites in Blood and Urine.

Metabolite	Association Study With eGFR/iCKD	Levels for Lower eGFR	Locus	Chr	SNP, Blood Association	SNP, Urine Association	Biological Plausability
Citrulline	eGFR (serum)[1] iCKD (serum)[30]	+	*ALDH18A1* (hre) *TCTN3* (hre) *ENTPD1* (hre) *C10orf131* (hre)	10	rs1935815[8]		*ALDH18A1* (and *CPS1)* encode mitochondrial enzymes involved in the urea cycle. ALDH18A1 is critical in the de novo biosynthesis of proline, ornithine, and arginine.
			CPS1 (hre) *LANCL1* (hre)	2	rs10164524[8]		CPS1 catalyzes the first step of urea cycle. In newborn screening, abnormal citrulline levels are used as indicators of urea cycle disorders. *CPS1* is a known CKD risk locus (see Table 2).
			ARL15 (h)	5	rs682103[8]		
Glutaryl-carnitine (C5)	eGFR (serum)[1,34]	+	*GCDH* (hem) *HOOK2* (hem) *CALR* (hem) *PRDX2* (hre)	19	rs8012[8,9] rs2974754[7]		Various loci associated with glutarylcarnitine are linked to mitochondrial respiration or they are involved in lysine, hydroxytryptophan, and tryptophan degradation or transport. The protein encoded by *GCDH* belongs to the acyl-CoA dehydrogenase family. It catalyzes the oxidative decarboxylation of glutaryl-CoA to crotonyl-CoA in the degradation pathway of lysine, hydroxylysine, and tryptophan. Glutaric aciduria type 1 (GA1) is a rare mitochondrial disease known for this locus, which presents with high levels of glutaric acid.
			SLC7A6[*] (he) *PRMT7* (hrem) *SMPD3* (hrem) *CCDC79* (hem) *NAE1* (hem) *ELMO3* (hem) *ACD* (hem) *SLC7A6OS* (hrm) *CMTM3* (hre) *DYNC1LI2* (hre) *TPPP3* (hre) *GFOD2* (hre) *NFATC3* (hre) *PLA2G15* (hre)	16	rs8056893[8]		*SLC7A6* encodes a transporter of di-basic amino acids including lysine. The locus has been identified in GWAS of eGFR at the suggestive significance level.

Table 3 (*continued*)

Metabolite	Association Study With eGFR/iCKD	Levels for Lower eGFR	Locus	Chr	SNP, Blood Association	SNP, Urine Association	Biological Plausability
			CPT2 (hrem) *LRP8* (hrep)	1	rs13375749[8] rs11581518[9]		Together with carnitine palmitoyltransferase I, the protein encoded by *CPT2* oxidizes LCFAs in the mitochondria. Defects in *CPT2* are associated with mitochondrial LCFA oxidation disorders.
			EFCAB1 (hre) *C8orf22* (hre)	8	rs17641971[8] rs4873099[9]		
			CPS1 (hm)	2	rs715[8,9]		Gene description: see and Table 2.
			ABCC1 (h)	16	rs246234[8]		*ABCC1* encodes MRP1, multidrug resistance–associated protein 1, a transporter for a variety of metabolites.
			ETFA (hem)	15	rs2291449[8]		Mutations in *ETFA* cause MADD (also known as glutaric acidemia IIA). MADD is an autosomal-recessive disorder of fatty acid and amino acid oxidation and features excretion of large amounts of glutaric acid MADD has been associated with cases of fetal nephromegaly as well as postnatally diagnosed polycystic disease.[93]
			ETFDH˙ (he) *PPID* (hepm)	4	rs55936281[9]		*ETFDH* encodes a component of the electron-transfer system in mitochondria and is essential for electron transfer from a number of mitochondrial flavin-containing dehydrogenases to the main respiratory chain. Mutations in this gene are another cause of MADD, also known as glutaric acidemia IIC. As in type IIA, patients excrete large amounts of glutaric acid.

			SLC22A4 (hrem) *C5orf56* (hrem) *P4HA2* (hre) *SLC22A5* (hre) *IRF1* (hre)	5	rs1981524[8]	Both *SLC22A4* and *SLC22A5* encode organic zwitterion/cation transporters for which L-carnitine is a known substrate.[94] Another gene in the same locus, *P4HA2*, encodes for a component of prolyl 4-hydroxylase, which forms 4-hydroxyproline and is important in collagen synthesis.
Kynurenine	eGFR (serum)[1] iCKD (serum)[30]	+	*SLC7A5* (hre)	16	rs8051149[8] rs28582913[9]	*SLC7A5* encodes LAT1, which transports and exchanges a wide variety of substrates, including tryptophan. It has been reported to mediate tryptophan/kynurenine exchange.[95]
			ATXN2 (hrem) *SH2B3* (hrm) *MAPKAPK5* (hem) *TMEM116* (hem)	12	rs3184504[8]	This locus has been identified in many GWAS, including eGFR GWAS.[17] *ATXN2* encodes a cytoplasmic protein that localizes to the endoplasmic reticulum and plasma membrane. It has been reported to have a role in endocytosis, mTOR signaling, ribosomal translation, and mitochondrial function.
			IDO1 (hre)	8	rs10085935[8]	*IDO1* encodes an indoleamine 2,3-dioxygenase, which catalyzes the degradation of the essential amino acid L-tryptophan to N-formyl-kynurenine. This enzyme contributes to several pathophysiological processes such as antimicrobial and antitumor defense, immunoregulation, and antioxidant activity.
Myo-inositol[†]	eGFR (serum)[1] iCKD (serum)[30]	+	*ISYNA1*[*] (hr) *SSBP4* (hre)	10 19	rs16924894[8] rs4808136[8]	*ISYNA1* encodes an inositol-3-phosphate synthase enzyme, which plays an important role in the myo-inositol biosynthesis pathway by catalyzing the rate-limiting conversion of glucose 6-phosphate to myo-inositol-1-phosphate.

Table 3 (*continued*)

Metabolite	Association Study With eGFR/iCKD	Levels for Lower eGFR	Locus	Chr	SNP, Blood Association	SNP, Urine Association	Biological Plausability
			SLC5A11 (he)	16	rs4788439[8]	rs17702912[27]	*SLC5A11* encodes the sodium myo-inositol cotransporter 2.
C-glycosyl-tryptophan	eGFR (serum)[1] Δ eGFR (serum)[1]	+		5	rs6867478[8]		
O-sulfo-L-tyrosine	eGFR (serum)[1] Δ eGFR (serum)[1]	+	*ARSA* (hem)	22	rs6151429[8]		*ARSA* encodes arylsulfatase A hydrolyzing sulfatides to cerebrosides and sulfate. The enzyme also acts on O-sulfo-L-tyrosine at similar rates.[96]
			ABO (hre)	9	rs651007[8]		The *ABO* gene product has glycosyltransferase activity in certain blood groups; strongest association at the same SNP: ADpSGEGDFXAEGGGVR, dipeptides.
Glycine	iCKD (urine)[28]	- (urine)	*CPS1* (hrem) *LANCL1* (hre)	2	rs715[8,41] rs1047891[9] rs16844839[40] rs4673553[7]	rs715[27]	See above and Table 2.
			SLC36A2 (he) *SLC36A1* (he) *GM2A* (re) *FAT2* (re)	5		rs3846710[27]	*SLC36A2* encodes for a pH-dependent proton-coupled amino acid transporter that belongs to the amino acid auxin permease 1 protein family. The encoded protein primarily transports small amino acids such as glycine, alanine, and proline. Mutations in this gene are associated with iminoglycinuria and hyperglycinuria.
			GCSH (hrem) *C16orf46* (hrem) *ATMIN* (hem)	16	rs10459871[40]		The glycine cleavage system is composed of 4 protein components and catalyzes the degradation of glycine. The H protein (GCSH) shuttles the methylamine group

			CENPN (hre)				of glycine from the P protein (GLDC) to the T protein (GCST). Defects in this gene are a cause of nonketotic hyperglycinemia (NKH, MIM 605899).
			ALDH1L1 (hre)	3	rs1992854[40]		The protein encoded by *ALDH1L1*, 10-formyltetrahydrofolate dehydrogenase, catalyzes the conversion of 10-formyltetrahydrofolate, NADP+, and water to tetrahydrofolate, NADPH, and carbon dioxide. Glycine is a key component of the folate pathway.
Histidine	iCKD (urine)[28]	- (urine)	*PNMT** (hre) *IZPBP2* (hrem) *MED1* (hrem) *STARD3* (hrem) *GSDMB* (hrem) *ERBB2* (hem) *PGAP3* (hem) *STAC2* (hre) *FBXL20* (hre) *PPP1R1B* (hre) *KZF3* (hre) *ORMDL3* (hre)	17		rs12947281[27]	*PNMT* encodes a phenylethanolamin N-methyltransferase and is part of catecholamine and tyrosine metabolism. Norharmane (9H-pyrido[3,4-b]indole), a compound with a nitrogen-containing heterocycle, is also a substrate of the enzyme Histidine also includes a nitrogen-containing heterocycle and might be another substrate.
			HAL (hm)	12	rs61937878[9]		*HAL* encodes histidine ammonia-lyase, a cytosolic enzyme catalyzing the first reaction in histidine catabolism, the nonoxidative deamination of L-histidine to trans-urocanic acid. Histidine ammonia-lyase defects cause histidinemia, which is characterized by increased histidine and histamine and decreased urocanic acid in body fluids.
			SLC6A19 (he)	5		rs11133665[27]	*SLC6A18* and *SLC6A19* encode transporters for neutral amino acids expressed in intestine and kidney. SLC6A18 knockout mice not only differed in their urinary glycine excretion but also showed significantly higher fractional excretion of histidine,[97] suggesting that SLC6A18 may be important in the renal tubular handling of histidine.

Table 3 (*continued*)

Metabolite	Association Study With eGFR/iCKD	Levels for Lower eGFR	Locus	Chr	SNP, Blood Association	SNP, Urine Association	Biological Plausability
			SLC2A9 (hrem) *WDR1* (hre)	4	rs16868246[8] rs13131257[7]		*SLC2A9* encodes a transport protein for urate, glucose and fructose. It is a known GWAS susceptibility locus for hyperuricemia and gout. The transporter is expressed in the proximal tubules, where it mediates reabsorption of urate. It has not been related to histidine transport to date.
				5	rs13164277[9]		

We used summary statistics from GWAS with metabolic traits by Draisma et al,[40] Long et al,[9] Raffler et al,[27] Shin et al,[8] and Suhre et al.[7] From these, we extracted all SNPs that showed association signals ($P < 5\times10^{-8}$) with at least one metabolite reported to be significantly associated with iCKD in the recent metabolome-wide study in urine by McMahon et al[28] or that had been linked repeatedly to eGFR, eGFR decrease, or iCKD in studies of blood metabolites.[1,30,34,35] SNPs within a genomic distance less than 1 Mb were grouped into a locus. All SNPs within a locus as well as their proxies in linkage disequilibrium ($r^2 >= 0.5$, based on 1000 Genomes Project phase 3 v5 data, European population) were used as input for SNiPA's block annotation tool (v3.2 March 2017, www.snipa.org[50]). SNiPA results were used to assign plausible candidate genes to a given locus. For each gene within the boundaries of a locus, we collected evidence, in which we noted whether at least one SNP hit the transcribed region (h), hit a regulatory element (r), was an eQTL (e) or pQTL (p), or was a missense variant (m). We regard the gene with the highest number of different evidences to be the best candidate.

Abbreviations: iCKD, incident chronic kidney disease; Δ eGFR, annual change in estimated glomerular filtration rate; LCFA, long-chain fatty acid; MADD, multiple acyl-CoA dehydrogenation deficiency; mTOR, mechanistic Target Of Rampamycin; NADP, nicotinamide adenine dinucleotide phosphate; NADPH, Nicotinamide adenine dinucleotide phosphate.

*Genes that were ranked better manually because of a functional match with the associated metabolite.

†Scyllo-inositol also associates with variants in the *SLC5A11* locus with $P < 5\times10$. The index SNP for this association is rs4787294.

consequently are not highlighted in Table 2, this example illustrates the added potential of integrating information from kidney function trait to metabolite and then to gene (Table 3).

Another example of integrating CKD-associated metabolites in light of their genetic associations to identify processes that potentially are implicated in renal disease was shown by McMahon et al.[28] In a nested case-control study within the Framingham Offspring cohort with 193 participants who developed CKD and 193 sex- and age-matched controls, the investigators found lower urinary glycine and histidine concentrations to be associated with a higher risk of incident CKD. Subsequent testing of the association of exome chip variants with all urine metabolites showed, among others, an association of a rare missense variant in *SLC36A2* (rs77010315; MAF, 0.01) with higher urinary glycine levels at a *P* value of less than 1.0×10^{-5}. This variant is listed in ClinVar[98] as a cause of autosomal-dominant hyperglycinuria. An association of a common variant at the same locus (rs3846710) with urinary glycine has been reported previously in a GWAS of urine metabolites in a population-based cohort (Table 3),[27] *SLC36A2* encodes a solute transporter for small amino acids including glycine and is expressed in the apical membrane of the proximal tubule.[39] Although localization, monogenic disease, and gene transport function support the gene's role in glycine reabsorption from the urine, genetic variation in the locus has not been reported in association with kidney function or CKD. This observation supports the notion that urinary glycine concentrations are not simply a complementary and correlated read-out of kidney function to eGFR that would lead to the identification of the same genetic loci.

To gain further insights into genes that are potentially involved in the regulation of CKD-associated glycine, we screened large mGWAS in population-based cohorts[7–9,27,40] that are searchable through http://snipa.org and http://gwas.eu for associations with glycine at a genome-wide significance level (Table 3). In addition to the *SLC36A2* locus,[27] serum and urinary glycine levels also were reported in association with *CPS1*,[7–9,27,40,41] *GCSH*,[40] and *ALDH1L1*,[40] all of which can plausibly be connected to glycine metabolism. Similar observations were made for urinary histidine concentrations (Table 3 and McMahon et al[28]).

A common theme arising from the examples in Table 3 is a link between genes encoding for mitochondrial proteins, metabolites indicative of mitochondrial function such as glutarylcarnitine, and kidney function. The kidneys are highly metabolic organs that contain many mitochondria for energy generation. These small organs consume approximately 8% of the body's resting energy expenditure to perform their various functions such as active reabsorption, making the kidneys along with the heart the organs with the highest resting metabolic rate in human beings.[42] Mitochondrial dysfunction causes reduced energy production, morphologic and metabolic changes on a cellular level, increased generation of reactive oxygen species, and, ultimately, reduced kidney function.[43] Such dysfunction has been linked to early as well as advanced stages of kidney diseases, especially for diabetic kidney disease[36] and acute kidney injury.[43] The restoration of mitochondrial homeostasis therefore has been discussed as a potential therapeutic approach to improve kidney function and/or prevent CKD progression.[44]

INSIGHTS GENERATED FROM THE MODELING OF KIDNEY-SPECIFIC FUNCTIONS FROM METABOLITES

The renal generation, breakdown, re-absorption, and secretion of metabolites actively influences their concentrations beyond what would be expected from the pure filtration of blood to urine. These functions of the kidney can be modeled from metabolite concentrations, a concept illustrated in Figure 1. The potential of modeling kidney-specific functions from metabolite concentrations is exemplified by Raffler et al[27] for the association between myo-inositol and *SLC5A11* variants (discussed earlier). When the researchers directly modeled the transport function of SLC5A11 as the ratio of urinary myo-inositol to blood myo-inositol concentrations, the strength of the genetic association increased by several orders of magnitude.[27] These findings suggest that the ratio of a metabolite's blood and urine concentrations can capture specific transport functions of tubular cells.

An additional measure of transport mechanisms across the nephron is a metabolite's fractional excretion (FE). The FE is defined as the percentage of a metabolite filtered by the kidney that is excreted in the urine. In theory, genetic associations with a metabolite's FE should be a complementary approach to highlight active transport processes of a metabolite along the nephron. To date, however, systematic studies of the genetics of multifluid metabolite concentrations have not been reported and represent an attractive avenue for further research.

In addition to the ratio of a metabolite's concentration across urine and blood, the ratio of the concentrations of two metabolites within the same body fluid (eg, urine) can provide additional information. Of the 22 replicated loci reported to be associated with the urinary concentrations of at least one metabolite,[27] 8 were detected only through the use of ratios. Two potential explanations for this observation are that the ratios more closely reflect physiological mechanisms

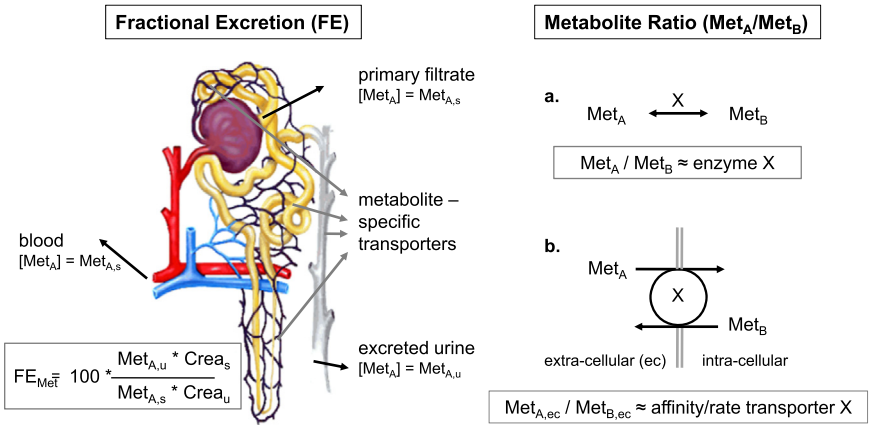

Figure 1. Schematic figure illustrating the modeling of metabolites to reflect kidney-specific functions beyond glomerular filtration. Metabolites can be modeled as their FE in studies that have paired measurements of a metabolite's concentration (Met) in both blood (b) and urine (u). Metabolite ratios (Met_A/Met_B) can be informative because they may reflect substrate and product of an enzymatic reaction (a), or represent a read-out of a metabolite exchanger's affinity or transport rate (b). The FE or metabolite ratios then can be related to genetic variation to gain insights into the proteins potentially involved in the modeled processes. Nephron courtesy of Michael Köttgen, MD.

such as enzymatic conversions of one metabolite into another or genetically encoded changes in substrate specificity, or that the use of ratios reduces noise because of the cancellation of systematic errors.

Finally, an elegant example of a direct read-out of intrarenal functions was described by Rhee et al.[30] The investigators directly sampled blood from patients' renal artery and renal vein, and applied metabolomics to evaluate the metabolites' arteriovenous gradients across the kidney. They identified several metabolites with significant concentration changes during renal passage, which is discussed in more detail in Rhee et al[45] in this issue. Because the study only examined nine such patients and did not have genetic information, it could not use genetics as a tool to facilitate the identification of mechanism underlying human metabolism. However, it may not be likely that future larger studies will be conducted in an analogous fashion because of the invasiveness and need for a clear clinical indication justifying catherization.

MGWAS IN KIDNEY DISEASE PATIENTS

The conduct of mGWAS specifically among CKD patients could be particularly informative to identify processes that are up-regulated once kidney function declines. It is conceivable that detoxification reactions of accumulating metabolites or active secretion of metabolites along the nephron increase as kidney filtration function declines. The accumulation of metabolites, many of which have been described as uremic toxins,[46,47] could represent a "challenge model" that aids in the identification of genetic determinants for these metabolites. This hypothesis has not been tested to date because data from large-scale CKD populations with both metabolomics and genomics measurements was lacking. A comparison of effect sizes between patients with CKD and individuals without CKD may additionally be complicated because of differences in metabolite panels/coverage and/or data analysis, such as the use of different transformations across studies. Therefore, studies dedicated to test the hypothesis that CKD populations allow for the detection of genetic effects on metabolites that are more easily detected in the presence of kidney disease ideally should be designed in a way that allows the systematic comparison of effect sizes with population-based studies.

CURRENT LIMITATIONS AND OPEN QUESTIONS

A general challenge for the study of genetic determinants of the human metabolome is how to integrate these large data sets and to synthesize, interpret, and display the generated wealth of information. Current genetic data sets based on standard genome-wide arrays followed by state-of-the-art genotype imputation contain approximately 10 million common and well-imputed variants for individuals of European ancestry. Current nontargeted MS-based metabolomics platforms such as the services offered by Metabolon (Durham, North Carolina) return quantitative information for more than 1,500 metabolites in biological fluids. Hence, the study of the genetic basis of the individual metabolites and their pair-wise ratios requires the conduct of 1,500 and 1,124,250 GWAS, respectively, in one biofluid alone. Various methods such as *OmicABLE* and *MatrixEQTL* have been developed to accelerate the computation of linear models in a GWAS setting.[48,49]

Because the number of genome-wide significant results from well-powered mGWAS typically is high,

the processing, aggregation, and annotation of results for further interpretation requires automated procedures. In recent mGWAS by Shin et al[8] and Raffler et al,[27] gene assignment within a locus and determination of the overlap of associated single-nucleotide polymorphisms (SNPs) with previous mGWAS and GWAS with clinical end points, was facilitated through automated combination of complementary sources of evidence. For this task, Raffler et al[27] made use of the SNiPA resource,[50] which provides SNP-centric information integrated from various gene- and SNP-based annotation data sets such as EnsEMBL,[51] GTEx,[52] GWAS Catalog,[53] and others together with precalculated linkage disequilibrium data for all superpopulations from the 1000 Genomes Project.[54] Because SNiPA allows for easy access to 500,000 metabolite quantitative trait loci (QTLs), 15,000 protein QTLs, 20 million expression QTLs, 400,000 regulatory clusters, 600,000 microRNA target sites, 1 million gene-associated enhancers and promoters, and 330,000 disease and phenotype associations, we also used the resource for extracting information for Tables 2 and 3 in this review.

Another challenge for the interpretation of findings from mGWAS is that a large part of the metabolome is still uncharted territory, with new metabolites discovered on a weekly basis. This is illustrated by the services offered by Metabolon: while Suhre et al[7] reported on 276 blood metabolites in 2011, a recent study by Long et al[9] using an update of the platform already reported on 486 identified metabolites. In addition, the studied metabolites included 158 unnamed molecules that can be recognized and quantified (named X-*) but whose identity has yet to be confirmed using isotope-labeled standards. These unknown metabolites can represent hitherto-unknown intermediates or products of human metabolism, or can be of xenobiotic origin, including medications, toxins, or products from microbiota or food. Despite major advances in metabolite identification over the past years through the development of chemoinformatic tools that systematically use and integrate spectral properties of unknown metabolites, annotation of as yet unidentified signals in nontargeted metabolomics data sets remains a huge challenge in the field.[55] As shown for both NMR- and MS-based nontargeted metabolomics, the functions of genes that are associated with unnamed compounds can provide complementary biological insights that can aid in metabolite identification.[26,56,57] By combining bioinformatics and chemoinformatic approaches (eg, using knowledge about the implicated gene's function together with molecule-specific spectral characteristics such as mass and retention time),[58] it may be possible to identify novel und previously unsuspected molecules of importance in the etiology and/or progression of kidney diseases.

Published mGWAS used data from a variety of platforms. Although many GIMs could be replicated across platforms, other GIMs were uniquely identified on particular platforms. Even for the same metabolite, differences in metabolite quantification as in targeted versus nontargeted methods can result in a subset of associations that are unique to each method. As an example, Yet et al[59] investigated how a targeted (Biocrates (Innsbruck, Austria) panel) and an nontargeted (Metabolon) MS-based method for metabolite quantification compared in mGWAS performed in the same study population. Genetic associations for 43 known metabolites present on both platforms generally correlated well. There were 18 loci unique to the Metabolon platform and 1 locus unique to the Biocrates platform, which underscores the complementary nature of different platforms and the potential of nontargeted metabolomics. The latter also was shown in the urine mGWAS by Raffler et al,[27] who described genetic loci uniquely identified using a nontargeted approach or a targeted approach based on the same measured spectra.

Even if the identity of a metabolite is known and it is quantified from the same biofluid and was measured on the same metabolomics platform, the lack of standardization across published mGWAS studies complicates a comparison of reported effect sizes. Differences result from the methods used for data cleaning, imputation, trait transformations, and data analysis.

Studies of patients with CKD raise additional questions: in addition to low eGFR, these patients also often manifest considerable proteinuria. Depending on the metabolomics approach used, this may influence results from mGWAS studies in urine. Although NMR-based methods quantify the amount of free metabolite concentrations, the quantified fraction of total metabolite concentration may vary depending on the specific procedure of sample preparation (eg, solvent used, and so forth) that is used in an MS-based method. This may affect the ability to identify genetic associations with metabolites that display a high degree of protein binding such as tryptophan. It is conceivable that free metabolite concentrations are more relevant for the identification of metabolite transporters along the nephron, although a direct comparison of genetic studies of metabolites quantified via both NMR and MS in the same samples is lacking.

In most mGWAS, interrelation between metabolites within pathways are ignored; only a few studies have used the correlation between metabolites within pathways for their analysis.[60,61] In some studies, the investigators inferred metabolic networks based on partial correlations of metabolite levels, which have been shown to reconstruct true biochemical relationships, to embed the identified genetic association into

their metabolic context for aiding biological interpretation.[8,56,62,63] Extending this concept to infer higher-dimensional relationships between metabolites (in multiple fluids), proteins, and genes across omics layers is currently limited by the availability of such multi-omics data for the same samples, but might provide useful additional insights for kidney function in the future.[64,65]

OUTLOOK AND CONCLUSIONS

Future efforts in the mGWAS area should aim to improve comparability across studies through standardization of analytic approaches. In the absence of a unified technology that allows for a comprehensive and standardized quantification of all metabolites in a sample, imputation of missing metabolites may be a fruitful area of future research. Additional areas that would benefit from further automatization and research include computational approaches to assign identity to unknown molecules, and to identify the most likely causal gene in a genomic interval identified through mGWAS. Besides translating unknown but quantifiable molecules to known ones, the study of additional human biofluids and from cell- and animal-based experimental systems also will deliver additional novel molecules that are not even quantified to date. Improvements in the connection of metabolites into networks and in the modeling of higher-dimensional relationships among metabolites and between metabolites, proteins, transcripts, and so forth should provide additional insights into metabolic pathways when associated with genetic information. Our understanding of the influence of disease states and environmental influences, that is, the dynamic aspect of the metabolome, also should be a major focus of future studies. Finally, the central question of cause versus consequence for observed metabolite-disease associations can be addressed experimentally or through large-scale Mendelian randomization experiments. In summary, the integration of genomic and metabolomic data is an exciting research area for the field of nephrology, which we expect to grow significantly in the coming years.

REFERENCES

1. Sekula P, Goek ON, Quaye L, et al. A metabolome-wide association study of kidney function and disease in the general population. J Am Soc Nephrol. 2016;27:1175-88.
2. Stevens LA, Coresh J, Greene T, et al. Assessing kidney function–measured and estimated glomerular filtration rate. N Engl J Med. 2006;354:2473-83.
3. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. N Engl J Med. 2012;367:20-9.
4. Suhre K, Raffler J, Kastenmuller G. Biochemical insights from population studies with genetics and metabolomics. Arch Biochem Biophys. 2016;589:168-76.
5. Gieger C, Geistlinger L, Altmaier E, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genet. 2008;4:e1000282.
6. Illig T, Gieger C, Zhai G, et al. A genome-wide perspective of genetic variation in human metabolism. Nat Genet. 2010;42:137-41.
7. Suhre K, Shin SY, Petersen AK, et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature. 2011;477:54-60.
8. Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. Nat Genet. 2014;46:543-50.
9. Long T, Hicks M, Yu HC, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nat Genet. 2017;49:568-78.
10. Hediger MA, Clemencon B, Burrier RE, et al. The ABCs of membrane transporters in health and disease (SLC series): introduction. Mol Aspects Med. 2013;34:95-107.
11. Dehghan A, Kottgen A, Yang Q, et al. Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. Lancet. 2008;372:1953-61.
12. Woodward OM, Kottgen A, Coresh J, et al. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. Proc Natl Acad Sci U S A. 2009;106:10338-42.
13. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347:1260419.
14. Kolz M, Johnson T, Sanna S, et al. Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. PLoS Genet. 2009;5:e1000504.
15. Veiga-da-Cunha M, Tyteca D, Stroobant V, et al. Molecular identification of NAT8 as the enzyme that acetylates cysteine S-conjugates to mercapturic acids. J Biol Chem. 2010;285:18888-98.
16. Chambers JC, Zhang W, Lord GM, et al. Genetic loci influencing kidney function and chronic kidney disease. Nat Genet. 2010;42:373-5.
17. Kottgen A, Pattaro C, Boger CA, et al. New loci associated with kidney function and chronic kidney disease. Nat Genet. 2010;42:376-84.
18. Yu B, Zheng Y, Alexander D, et al. Genetic determinants influencing human serum metabolome among African Americans. PLoS Genet. 2014;10:e1004212.
19. Okada Y, Sim X, Go MJ, et al. Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. Nat Genet. 2012;44:904-9.
20. Sekula P, Del Greco MF, Pattaro C, et al. Mendelian randomization as an approach to assess causality using observational data. J Am Soc Nephrol. 2016;27:3253-65.
21. Fotiadis D, Kanai Y, Palacin M. The SLC3 and SLC7 families of amino acid transporters. Mol Aspects Med. 2013;34:139-58.
22. Kottgen A, Glazer NL, Dehghan A, et al. Multiple loci associated with indices of renal function and chronic kidney disease. Nat Genet. 2009;41:712-7.
23. Nicholson G, Rantalainen M, Li JV, et al. A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. PLoS Genet. 2011;7:e1002270.
24. Suhre K, Wallaschofski H, Raffler J, et al. A genome-wide association study of metabolic traits in human urine. Nat Genet. 2011;43:565-9.

25. Montoliu I, Genick U, Ledda M, et al. Current status on genome-metabolome-wide associations: an opportunity in nutrition research. Genes Nutr. 2013;8:19-27.
26. Rueedi R, Ledda M, Nicholls AW, et al. Genome-wide association study of metabolic traits reveals novel gene-metabolite-disease links. PLoS Genet. 2014;10:e1004132.
27. Raffler J, Friedrich N, Arnold M, et al. Genome-wide association study with targeted and non-targeted NMR metabolomics identifies 15 novel loci of urinary human metabolic individuality. PLoS Genet. 2015;11:e1005487.
28. McMahon GM, Hwang SJ, Clish CB, et al. Urinary metabolites along with common and rare genetic variations are associated with incident chronic kidney disease. Kidney Int. 2017;91:1426-35.
29. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45:1274-83.
30. Rhee EP, Clish CB, Ghorbani A, et al. A combined epidemiologic and metabolomic approach improves CKD prediction. J Am Soc Nephrol. 2013;24:1330-8.
31. Dieterle F, Ross A, Schlotterbeck G, et al. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Anal Chem. 2006;78:4281-90.
32. Vogl FC, Mehrl S, Heizinger L, et al. Evaluation of dilution and normalization strategies to correct for urinary output in HPLC-HRTOFMS metabolomics. Anal Bioanal Chem. 2016;408:8483-93.
33. Christiansen B, Meinild AK, Jensen AA, et al. Cloning and characterization of a functional human gamma-aminobutyric acid (GABA) transporter, human GAT-2. J Biol Chem. 2007;282:19331-41.
34. Goek ON, Doring A, Gieger C, et al. Serum metabolite concentrations and decreased GFR in the general population. Am J Kidney Dis. 2012;60:197-206.
35. Goek ON, Prehn C, Sekula P, et al. Metabolites associate with kidney function decline and incident chronic kidney disease in the general population. Nephrol Dial Transplant. 2013;28:2131-8.
36. Sharma K, Karl B, Mathew AV, et al. Metabolomics reveals signature of mitochondrial dysfunction in diabetic kidney disease. J Am Soc Nephrol. 2013;24:1901-12.
37. Coady MJ, Wallendorff B, Gagnon DG, et al. Identification of a novel Na+/myo-inositol cotransporter. J Biol Chem. 2002;277:35219-24.
38. Hammerman MR, Sacktor B, Daughaday WH. myo-Inositol transport in renal brush border vesicles and it inhibition by D-glucose. Am J Physiol. 1980;239:F113-20.
39. Boll M, Foltz M, Rubio-Aliaga I, et al. Functional characterization of two novel mammalian electrogenic proton-dependent amino acid cotransporters. J Biol Chem. 2002;277:22966-73.
40. Draisma HHM, Pool R, Kobl M, et al. Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. Nat Commun. 2015;6:7208.
41. Xie W, Wood AR, Lyssenko V, et al. Genetic variants associated with glycine metabolism and their role in insulin sensitivity and type 2 diabetes. Diabetes. 2013;62:2141-50.
42. Wang Z, Ying Z, Bosy-Westphal A, et al. Specific metabolic rates of major organs and tissues across adulthood: evaluation by mechanistic model of resting energy expenditure. Am J Clin Nutr. 2010;92:1369-77.
43. Bhargava P, Schnellmann RG. Mitochondrial energetics in the kidney. Nat Rev Nephrol. 2017;13:629-46.
44. Granata S, Dalla Gassa A, Tomei P, et al. Mitochondria: a new therapeutic target in chronic kidney disease. Nutr Metab (Lond). 2015;12:49.
45. Rhee EP. A systems-level view of renal metabolomics. Semin Nephrol. 2018;38:142-50.
46. Kikuchi K, Itoh Y, Tateoka R, et al. Metabolomic analysis of uremic toxins by liquid chromatography/electrospray ionization-tandem mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci. 2010;878:1662-8.
47. Tanaka H, Sirich TL, Plummer NS, et al. An enlarged profile of uremic solutes. PLoS One. 2015;10:e0135657.
48. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28:1353-8.
49. Fabregat-Traver D, Sharapov S, Hayward C, et al. High-performance mixed models based genome-wide association analysis with omicABEL software. F1000Res. 2014;3:200.
50. Arnold M, Raffler J, Pfeufer A, et al. SNiPA: an interactive, genetic variant-centered annotation browser. Bioinformatics. 2015;31:1334-6.
51. Flicek P, Amode MR, Barrell D, et al. Ensembl 2014. Nucleic Acids Res. 2014;42:D749-55.
52. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648-60.
53. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009;106:9362-7.
54. 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56-65.
55. Schymanski EL, Ruttkies C, Krauss M, et al. Critical assessment of small molecule identification 2016: automated methods. J Cheminform. 2017;9:22.
56. Krumsiek J, Suhre K, Evans AM, et al. Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. PLoS Genet. 2012;8:e1003005.
57. Rueedi R, Mallol R, Raffler J, et al. Metabomatching: using genetic association to identify metabolites in proton NMR spectroscopy. PLoS Comput Biol. 2017;13:e1005839.
58. Quell JD, Romisch-Margl W, Colombo M, et al. Automated pathway and reaction prediction facilitates in silico identification of unknown metabolites in human cohort studies. J Chromatogr B Analyt Technol Biomed Life Sci. 2017;1071:58-67.
59. Yet I, Menni C, Shin SY, et al. Genetic influences on metabolite levels: a comparison across metabolomic platforms. PLoS One. 2016;11:e0153672.
60. Inouye M, Ripatti S, Kettunen J, et al. Novel loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. PLoS Genet. 2012;8:e1002907.
61. Ried JS, Shin SY, Krumsiek J, et al. Novel genetic associations with serum level metabolites identified by phenotype set enrichment analyses. Hum Mol Genet. 2014;23:5847-57.
62. Mittelstrass K, Ried JS, Yu Z, et al. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. PLoS Genet. 2011;7:e1002215.
63. Krumsiek J, Mittelstrass K, Do KT, et al. Gender-specific pathway differences in the human serum metabolome. Metabolomics. 2015;11:1815-33.
64. Do KT, Kastenmuller G, Mook-Kanamori DO, et al. Network-based approach for analyzing intra- and interfluid metabolite associations in human blood, urine, and saliva. J Proteome Res. 2015;14:1183-94.
65. Zierer J, Pallister T, Tsai PC, et al. Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model. Sci Rep. 2016;6:37646.
66. Demirkan A, van Duijn CM, Ugocsai P, et al. Genome-wide association study identifies novel loci associated with

circulating phospho- and sphingolipid concentrations. PLoS Genet. 2012;8:e1002490.
67. Kettunen J, Tukiainen T, Sarin AP, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat Genet. 2012;44:269-76.
68. Tukiainen T, Kettunen J, Kangas AJ, et al. Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci. Hum Mol Genet. 2012;21:1444-55.
69. Hong MG, Karlsson R, Magnusson PK, et al. A genome-wide assessment of variability in human serum metabolism. Hum Mutat. 2013;34:515-24.
70. Raffler J, Romisch-Margl W, Petersen AK, et al. Identification and MS-assisted interpretation of genetically influenced NMR signals in human plasma. Genome Med. 2013;5:13.
71. Rhee EP, Ho JE, Chen MH, et al. A genome-wide association study of the human metabolome in a community-based cohort. Cell Metab. 2013;18:130-43.
72. Demirkan A, Henneman P, Verhoeven A, et al. Insight in genome-wide association of metabolite quantitative traits by exome sequence analyses. PLoS Genet. 2015;11:e1004835.
73. Kraus WE, Muoio DM, Stevens R, et al. Metabolomic quantitative trait loci (mQTL) mapping implicates the ubiquitin proteasome system in cardiovascular disease pathogenesis. PLoS Genet. 2015;11:e1005553.
74. Kettunen J, Demirkan A, Wurtz P, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat Commun. 2016;7:11122.
75. Rhee EP, Yang Q, Yu B, et al. An exome array study of the plasma metabolome. Nat Commun. 2016;7:12360.
76. Yu B, Li AH, Metcalf GA, et al. Loss-of-function variants influence the human serum metabolome. Sci Adv. 2016;2:e1600800.
77. Davis JP, Huyghe JR, Locke AE, et al. Common, low-frequency, and rare genetic variants associated with lipoprotein subclasses and triglyceride measures in Finnish men from the METSIM study. PLoS Genet. 2017;13:e1007079.
78. Pattaro C, Kottgen A, Teumer A, et al. Genome-wide association and functional follow-up reveals new loci for kidney function. PLoS Genet. 2012;8:e1002584.
79. Kumar S, Saragadam T, Punekar NS. Novel route for agmatine catabolism in Aspergillus niger involves 4-guanidinobutyrase. Appl Environ Microbiol. 2015;81:5593-603.
80. Sabatti C, Service SK, Hartikainen AL, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet. 2009;41:35-46.
81. Chasman DI, Pare G, Mora S, et al. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. PLoS Genet. 2009;5:e1000730.
82. Tin A, Balakrishnan P, Beaty TH, et al. GCKR and PPP1R3B identified as genome-wide significant loci for plasma lactate: the Atherosclerosis Risk in Communities (ARIC) study. Diabet Med. 2016;33:968-75.
83. Pattaro C, Teumer A, Gorski M, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun. 2016;7:10023.
84. Mahajan A, Rodan AR, Le TH, et al. Trans-ethnic fine mapping highlights kidney-function genes linked to salt sensitivity. Am J Hum Genet. 2016;99:636-46.
85. Badawy AA, Bano S, Steptoe A. Tryptophan in alcoholism treatment I: kynurenine metabolites inhibit the rat liver mitochondrial low Km aldehyde dehydrogenase activity, elevate blood acetaldehyde concentration and induce aversion to alcohol. Alcohol Alcohol. 2011;46:651-60.
86. Hanff E, Kayacelebi AA, Yanchev GR, et al. Simultaneous stable-isotope dilution GC-MS measurement of homoarginine, guanidinoacetate and their common precursor arginine in plasma and their interrelationships in healthy and diseased humans. Amino Acids. 2016;48:721-32.
87. Li M, Li Y, Weeks O, et al. SOS2 and ACP1 loci identified through large-scale exome chip analysis regulate kidney development and function. J Am Soc Nephrol. 2017;28:981-94.
88. Sulem P, Gudbjartsson DF, Geller F, et al. Sequence variants at CYP1A1-CYP1A2 and AHR associate with coffee consumption. Hum Mol Genet. 2011;20:2071-7.
89. Gorski M, van der Most PJ, Teumer A, et al. 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. Sci Rep. 2017;7:45040.
90. Alam A, O'Connor DT, Perianayagam MC, et al. Phenylethanolamine N-methyltransferase gene polymorphisms and adverse outcomes in acute kidney injury. Nephron Clin Pract. 2010;114:c253-9.
91. van Meurs JB, Pare G, Schwartz SM, et al. Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. Am J Clin Nutr. 2013;98:668-76.
92. Wuttke M, Kottgen A. Insights into kidney diseases from genome-wide association studies. Nat Rev Nephrol. 2016;12:549-62.
93. Whitfield J, Hurst D, Bennett MJ, et al. Fetal polycystic kidney disease associated with glutaric aciduria type II: an inborn error of energy metabolism. Am J Perinatol. 1996;13:131-4.
94. Koepsell H. The SLC22 family with transporters of organic cations, anions and zwitterions. Mol Aspects Med. 2013;34:413-35.
95. Kaper T, Looger LL, Takanaga H, et al. Nanosensor detection of an immunoregulatory tryptophan influx/kynurenine efflux cycle. PLoS Biol. 2007;5:e257.
96. Fluharty AL, Stevens RL, Goldstein EB, et al. The activity of arylsulfatase A and B on tyrosine O-sulfates. Biochim Biophys Acta. 1979;566:321-6.
97. Singer D, Camargo SM, Huggel K, et al. Orphan transporter SLC6A18 is renal neutral amino acid transporter B0AT3. J Biol Chem. 2009;284:19953-60.
98. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016;44:(D1):D862-8. http://dx.doi.org/10.1093/nar/gkv1222. PubMed PMID:26582918.