university of groningen

University Medical Center Groningen

University of Groningen

Chronic kidney disease

Thio, C. H. L.

DOI:
10.33612/diss.133648108

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Thio, C. H. L. (2020). *Chronic kidney disease: Insights from social and genetic epidemiology*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. https://doi.org/10.33612/diss.133648108

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 30-09-2023

Genome-wide association scan of serum urea in European populations identifies two novel loci

CHL Thio, A Reznichenko, PJ van der Most, Z Kamali, A Vaez, Lifelines Cohort Study group, JH Smit, BW Penninx, T Haller, E Mihailov, A Metspalu, J Damman, MH de Borst, P van der Harst, N Verweij, GJ Navis, RT Gansevoort, IM Nolte, H Snieder

Am J Nephrol. 2019;49(3):193-202.

CHAPTER

6

ABSTRACT

Introduction. Serum urea level is a heritable trait commonly used as a diagnostic marker for kidney function. GWAS in East-Asian populations identified a number of genetic loci related to serum urea but there is a paucity of data for European populations.

Methods. We performed a two-stage meta-analysis of GWASs on serum urea in 13,312 participants, with independent replication in 7379 participants of European ancestry.

Results. We identified six genome-wide significant SNPs in or near six loci, of which two were novel (*POU2AF1* and *ADAMTS9-AS2*). Replication of East-Asian and Scottish data provided evidence for an additional eight loci. SNPs tag regions previously associated with anthropometric traits, serum magnesium, and urinary albumin-to-creatinine ratio, as well as expression quantitative trait loci for genes preferentially expressed in kidney and gastro-intestinal tissues.

Conclusions. Our findings provide insights in the genetic underpinnings of urea metabolism, with potential relevance to kidney function.

INTRODUCTION

Serum urea is a diagnostic marker of renal function widely used in clinical practice. Urea is eliminated by the kidneys into urine as waste product of protein metabolism. The net serum urea concentration therefore reflects the excretory capacity of the kidney and elevated values are interpreted as reduced kidney function. Serum urea (or blood urea nitrogen, BUN, when only the nitrogen part is assayed), along with creatinine, is the most frequently requested measurement of kidney function in the assessment of patients with kidney disease. These two markers are not equivalent in estimation of kidney function, and in some conditions (peritoneal dialysis, heart failure) serum urea is considered to be superior to creatinine[1-3]. Alternatively to single-marker use, urea-to-creatinine (or BUN-to-creatinine, respectively) ratio can be used for differential diagnosis of acute kidney injury (prerenal, postrenal, or renal) when one marker is disproportionally elevated or lowered relative to the other[4-6].

Serum urea concentration is highly variable (reference range 1.8-7.1 mmol/L), and besides kidney function, it also depends on hydration status, metabolic rate, dietary protein intake, medication use, liver and cardiac function[5, 6]. Genetic factors may also play a role: one twin study estimated heritability for serum urea concentration to be 44%[7], indicating a contribution of genetic factors to the inter-individual variability of this measure. Furthermore, genome-wide association studies (GWAS) on BUN in East-Asians reported SNP associations at 13 loci[8-11]. For Europeans, there is paucity of data. A recent single-cohort study in the UK did not find any significant associations with urea levels[12], while in a Scottish single-cohort study (N=19,293), five genetic variants were associated with urea[13]. These findings are yet to be replicated in other European cohorts. Concurrently, multiple GWASs in individuals of European descent identified a number of loci associated with serum creatinine and creatinine-based indices of kidney function[14-18]. The genetics underlying urea and creatinine are expected to overlap, because, to a large extent, the serum concentration of both are influenced by kidney function. The studies in East-Asians confirm this notion as they reported *MPPED2-DCDC5* to be associated with both urea and creatinine[10], thus suggesting involvement of this gene with regulation of kidney function. Furthermore, family data from the UK show a positive genetic correlation between urea and creatinine (r_g=0.56)[12]. The existence of exclusively urea-associated loci is also plausible, given that serum levels are not only dependent on kidney function. Identifying these loci will help explain a proportion of kidney function-independent inter-individual variability in urea levels

in the general population and ultimately will provide insight into pathways and regulating mechanisms involved in this metabolic compound.

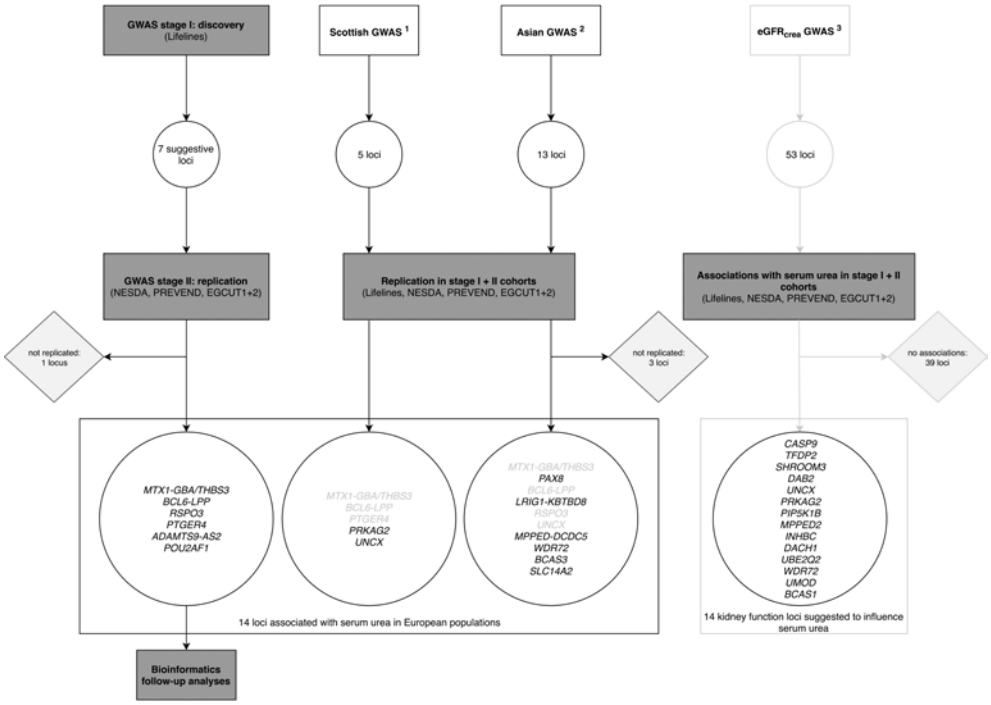
We therefore aimed to identify genetic loci influencing serum urea concentrations in populations of European ancestry. In addition, we compared our results with previous findings from East-Asian and Scottish studies to identify shared loci for serum urea.

METHODS

Study design

An overview of the study design is provided in **Figure 1**. Our strategy consisted of a number of steps. First, we performed a two-stage meta-analysis of GWAS to identify SNPs associated with serum urea. Second, we performed a replication study of loci identified in previous GWAS in East-Asian and Scottish populations. Third, we examined whether known eGFRcrea loci were also associated with serum urea. Furthermore, we conducted bioinformatics follow-up analyses on identified SNPs to identify candidate loci. Each step is detailed below.

Figure 1. Design and results of the present study. Genetic loci in GREY typefont indicate that these loci overlap between GWAS studies on serum urea/BUN.



Study population

Stage I discovery analyses were performed in 13,312 subjects from the Lifelines Cohort Study. Stage II replication testing was performed in 7379 subjects from the PREVEND (N=3387), NESDA (N=2523), EGCUT1 (N=712), and EGCUT2 (N=757) cohorts (**Supplementary Note 1**).

The Lifelines Cohort Study is a multidisciplinary prospective population-based cohort study with a unique three-generation design that examines health and health-related behavior of 165,729 participants living in the north-eastern region of the Netherlands (www.lifelines.nl/researcher). Participants were recruited from November 2006 to December 2013. Eligible individuals were invited through their general practitioner or through participating family members. Additionally, there was the option to self-register. The recruitment and data collection, as well as the representativeness of the data have been described in detail elsewhere[19,20]. Of the 165,729 participants, 15,368 presumably unrelated, oldest members of their respective families, were genotyped (details below). The Lifelines Cohort Study was conducted according to the guidelines in the Declaration of Helsinki and all procedures involving human subjects were approved by the Medical Ethical Committee of the University Medical Center Groningen. Written informed consent was obtained from all participants during their visit at one of the research centers.

Genotyping, quality control, and imputation

A total of 15,368 individuals of the Lifelines Cohort Study were genotyped using the Illumina HumanCytoSNP-12 array and called using GenomeStudio (San Diego, CA, USA). Only autosomal single nucleotide polymorphisms (SNPs) were used in this study. SNPs were excluded when the call rate was <95%, when the minor allele frequency (MAF) was <1%, or when the p-value of the Hardy-Weinberg equilibrium (HWE) test was $<10^{-6}$. Samples were removed when the call rate was <95%, when there was a sex mismatch between database and genotypes, when the heterozygosity deviated >4 SD from the mean heterozygosity over all samples, when it was a first-degree relative to a sample that had a higher call rate, or when non-Caucasian ancestry was likely. After quality control, a total of 268,407 SNPs and 13,385 samples remained. The resulting dataset was phased using MACH[21] and imputed using Minimac[22] with the HapMap Phase 2 CEU haplotypes[23] as reference set. SNPs with an imputation quality $r^2<0.3$ or a

MAF<1% were excluded after imputation. The resulting number of SNPs available for analysis was 1.99 x10^6. The procedure for genotyping, quality control, and imputation of the replication cohorts is described in **Supplementary Note S1**.

Phenotype measurement in Lifelines

At the baseline examination, the participants in the study were asked to fill in a questionnaire before the visit. During the visit, a number of investigations were conducted and blood and 24h-urine samples were taken. A total of 13,385 genotyped participants were included into the present study. The final number of individuals analyzed for serum urea was 13,312 after excluding subjects with extreme values of urea deviating >4 standard deviations (SD) from the mean. Serum urea measurements were performed with an ultraviolet kinetic assay on a Roche Modular. Serum creatinine was measured by an enzymatic method, IDMS traceable on a Roche Modular (Roche, Mannheim, Germany). We estimated eGFRcrea with the 4-variable Modification of Diet in Renal Disease (MDRD) Study equation[24]. Body-mass index (BMI, kg/m^2) was calculated by dividing weight(kg) by squared height (m^2).

Statistical analysis

Three GWASs on serum urea were performed. In the first GWAS, a linear regression for each SNP was performed using an additive SNP model adjusting for age, age^2, sex, body mass index (BMI), and the first ten principal components to adjust for population stratification using PLINK[25]. In the second GWAS, $\log_{10}$-transformed eGFRcrea, was added to the model. In a third GWAS, we adjusted for serum creatinine instead of *log*eGFRcrea. In addition to these three GWAS, we performed sex-stratified analyses. Next, the GWAS results were checked for quality using the QCGWAS package in R[26]. For each GWAS, suggestive SNPs (p-value <10^{-6} in Stage I analyses) were clumped for linkage disequilibrium (LD; r^2>0.1) using pairwise LD checking in SNAP[27] to identify independent index SNPs. These suggestive index SNPs were taken forward to Stage II replication.

The same linear regression analyses as described above were applied to the suggestive SNPs identified in the discovery sample in each of the four replication cohorts separately. The replication results of these SNPs were meta-analyzed using an inverse variance fixed-effects meta-analysis as implemented in the software package GWAMA[28]. A SNP was considered replicated with a one-sided

p-value <0.05 (i.e. same direction of effect), and with significance at the genome-wide level in combined Stage I+II samples ($p<5 \times 10^{-8}$).

Finally, we also sought to replicate 20 SNPs at 13 genetic loci previously identified in GWASs of East-Asian samples[8-11], as well as five SNPs at five loci identified in a Scottish sample[13]. The replication results of these 25 SNPs were meta-analyzed using an inverse variance fixed-effects meta-analysis as implemented in the software package GWAMA[28]. We used all five cohorts (i.e. Lifelines, NESDA, PREVEND, EGCUT1+2) for these analyses. We considered a SNP replicated at a one-sided p<0.05.

Secondary analyses and Bioinformatics

Associations with kidney function

We meta-analyzed associations of 53 known kidney function SNPs[17] with serum urea in all Stage I+II cohorts. Conversely, to examine associations of our six index SNPs with kidney function, we searched publicly available summary data from the same meta-analysis of GWAS on eGFRcrea[17]. At a one-sided p<0.05, we tested whether variants genome-wide significantly associated with lower eGFRcrea were associated with higher urea, and whether SNPs genome-wide significantly associated with higher urea were associated with lower eGFRcrea.

Proportion of phenotypic variance explained

We estimated the proportion of phenotypic variance explained in the NESDA cohort by regressing serum urea level on a weighted genetic risk score (GRS) comprising the effects of all six index SNPs, of the six index SNPs +11 independent SNPs from the Scottish and East-Asian studies, and of the 53 eGFRcrea SNPs. These analyses were performed using PLINK[25] and R[29] on independent SNPs (ldlink.nci.nih.gov) using the effect sizes from the discovery sample (our six index SNPs) or from literature as weights.

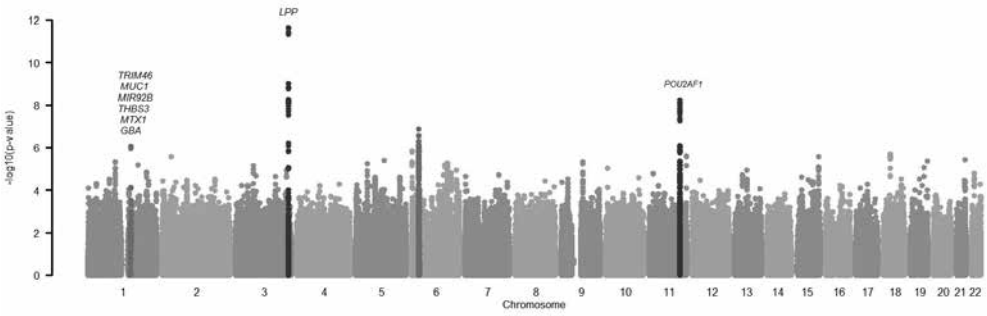
Bioinformatics characterization of the replicated SNPs

We examined functionality (i.e. non-synonymous SNPs and expression quantitative trait loci, eQTL) of the identified index SNPs. To this end, we first converted the positions of all replicated index SNPs to NCBI build 37. We then used the 1000 Genomes Project phase3 release[30] of variant calls to find proxy SNPs in moderate ($r^2>0.5$) and high LD ($r^2>0.8$) with our index SNPs. This dataset is based on the 2013-

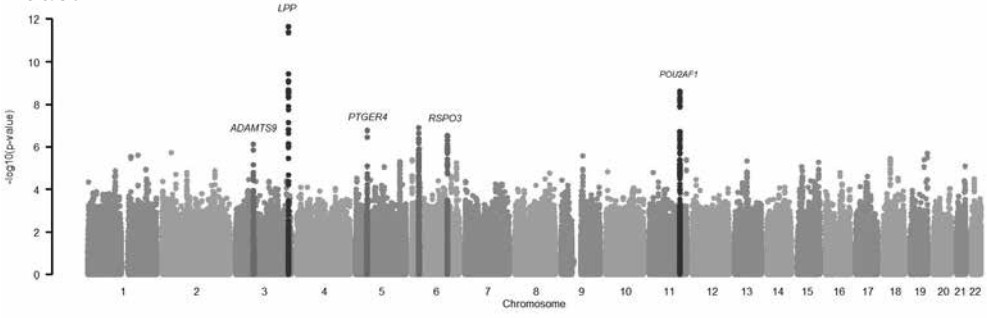
05-02 sequence freeze and alignments. We used version v5a (Feb. 20th, 2015), including the 503 subjects of European ancestry. We used ANNOVAR (version 16 July 2017) (annovar.openbioinformatics.org)[31] for annotation of the index SNPs. We queried PolyPhen-2 (genetics.bwh.harvard.edu/pph2/)[32] to assess whether effects of non-synonymous SNPs were predicted to be malignant. Furthermore, we performed a lookup of the index and proxy SNPs in the GWAS catalog[33] to ascertain whether these SNPs were previously associated with other phenotypes. Genes close to the six index SNPs were followed-up for local expression (*cis*eQTL) in various tissues based on publicly available transcriptomics data: Human Protein Atlas (www.proteinatlas.org)[34], GTEx Portal (www.gtexportal.org/)[35], and blood tissue (genenetwork.nl/bloodeqtlbrowser/)[36]. Furthermore, we examined eQTLs in donor kidney tissue in TransplantLines (detailed description of data and methods in **Supplementary Note 11**)[37, 38].

Figure 2. Manhattan plots of stage I GWAS for serum urea level. The x-axis represents chromosomal position. The y-axis represents two-sided significance on the –log10 scale. Dark grey indicates genome-wide significant hit ($p<5 \times 10^{-8}$), grey indicates suggestive hit ($5 \times 10^{-8} \leq p < 1 \times 10^{-6}$).
Model 1: adjusted for age, age^2, sex, BMI, principal components 1-10
Model 2: Model 1 + logeGFRcrea

Model 1



Model 2



RESULTS

Meta-analysis results

Manhattan plots of stage I for models 1 and 2 are shown in **Figure 2**. Regional association plots, showing location and significance of top hits for models 1 and 2 relative to known loci, are shown in **Supplementary Figure S3**. Risk of bias due to population stratification was assessed and considered acceptable (=1.05) (**Supplementary Figure S4**).

For models 1 and 2, seven index SNPs were at least suggestive ($p<1x10^{-6}$) in stage I. Of these seven SNPs, rs17586946 on chromosome 6 was only suggestive in the combined Stage I+II samples ($p=1.4x10^{-7}$) and hence not replicated. **Table 1** shows results of the remaining six SNPs. For model 1, we replicated three SNPs (rs914615, rs4686914, rs2003313) at three genomic loci, significantly associated with serum urea at the genome-wide level ($p<5x10^{-8}$) in the combined Stage I+II samples. In the second, *log*eGFRcrea-adjusted model, two SNPs from model 1 (rs4686914 and rs2003313) were again identified, while in addition three other SNPs (rs998394, rs11954639, rs2503107) were identified and replicated with genome-wide level significance. One SNP (rs914615) did not reach suggestive significance of $p<1x10^{-6}$ after *log*eGFRcrea adjustment ($p=2.9x10^{-6}$) and therefore deemed non-significant for this model. A third, serum creatinine adjusted model, yielded essentially the same results as the *log*eGFRcrea-adjusted model **(Supplementary Figure 2a, Supplementary Table S5**).

Sex-stratified analysis yielded no additional loci: 1) we found no significant associations in females-only models, and 2) in males-only models, we identified two additional SNPs (rs9860469 and rs9820812) in high linkage disequilibrium (LD) ($r^2=0.70$ and $r^2=1.0$, respectively) with a SNP already identified in models 1-2 (rs4686914) (**Supplementary Figure S2B**). Effects of rs4686914 and rs11954639 were stronger in men (Supplementary Table S6).

Replication of previously reported urea loci

We replicated 10 out of 13 East-Asian loci[8-11] at a one-sided $p<0.05$ (**Supplementary Table S7a**). SNPs at three loci (*MECOM, C12orf51, GNAS*) were not replicated in the present study. All five Scottish loci[13] were replicated (**Supplementary Table S7b**). In total, 14 loci are now confirmed for Europeans (**Figure 3**).

Table 1. Replicated SNP associations with serum urea

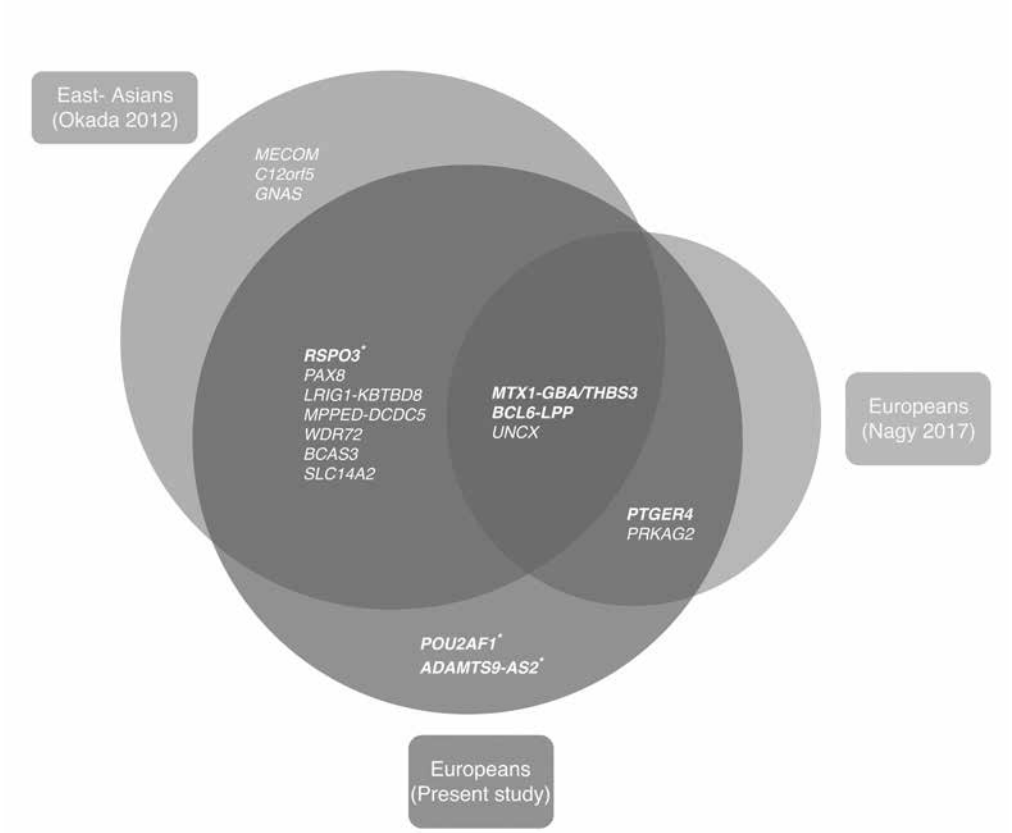
SNP ID	Chr	Position (bp)[a]	Type	Nearest gene	Effect/Non effect allele (EAF)[b]	Model	Stage I (LifeLines)				Stage II (PREVEND, NESDA, EGCUT1+2)				Stage I+II				I² %
							B	SE	p	N	B	SE	p	N	B	SE	p	N	
rs914615	1	153442516	intronic	*THBS3*	A/G (0.476)	1	0.070	0.014	8.9E-07	13312	0.065	0.020	1.3E-03	7379	0.068	0.012	**4.3E-09**	20689	0.0
						2	0.064	0.014	2.9E-06	13311	0.063	0.020	1.2E-03	7335	0.064	0.011	**1.3E-08**	20646	0.0
rs4686914	3	189200234	intergenic	*LPP*	T/C (0.308)	1	-0.110	0.016	2.4E-12	13312	-0.101	0.021	2.2E-06	7378	-0.107	0.013	**2.6E-17**	20690	0.0
						2	-0.106	0.015	2.3E-12	13311	-0.098	0.021	2.1E-06	7334	-0.103	0.012	**2.3E-17**	20645	0.0
rs998394	3	64776227	ncRNA/intronic	*ADAMTS9-AS2*	A/G (0.458)	1	-0.063	0.014	7.3E-06	13312	-0.049	0.020	1.4E-02	7379	-0.058	0.011	3.7E-07	20691	0.0
						2	-0.067	0.014	7.5E-07	13311	-0.058	0.019	2.2E-03	7335	-0.064	0.011	**7.1E-09**	20646	0.0
rs11954639	5	40710736	intergenic	*PTGER4*	T/C (0.071)	1	-0.165	0.037	5.8E-06	13312	-0.170	0.040	2.4E-05	7379	-0.168	0.027	**6.1E-10**	20691	0.0
						2	-0.185	0.035	1.8E-07	13311	-0.182	0.039	2.9E-06	7335	-0.183	0.026	**2.3E-12**	20646	0.0
rs2503107	6	127505069	intronic	*RSPO3*	C/A (0.449)	1	-0.075	0.017	8.6E-06	13312	-0.051	0.020	1.2E-02	7377	-0.065	0.013	4.9E-07	20689	0.0
						2	-0.084	0.016	2.9E-07	13311	-0.056	0.020	4.2E-03	7333	-0.072	0.013	**8.1E-09**	20644	18.0
rs2003313	11	110709203	intergenic	*POU2AF1*	T/A (0.448)	1	-0.088	0.015	6.0E-09	13312	-0.048	0.020	1.7E-02	7377	-0.073	0.012	**1.3E-09**	20691	60.6
						2	-0.087	0.015	2.5E-09	13311	-0.055	0.019	4.3E-03	7333	-0.075	0.012	**9.5E-11**	20644	43.2

Meta-analysis of associations obtained from linear regressions of replicated SNPs with serum urea level assuming additive effects of alleles. Estimates of B and se are presented in mmol/L.
Abbreviations: B, unstandardized regression coefficient; Chr, chromosome; bp, basepair; EAF, effect allele frequency; I^2, heterogeneity statistic; SE, standard error; SNP, single nucleotide polymorphism;

[a] position based on NCBI b36/hg18
[b] EAF in the complete sample (Stage I + II)
˙not suggestive (p≥1E-06) in Stage I for this model

Model 1: adjusted for age,age², sex, body-mass index, principal components 1-10
Model 2: model 1 + log_{10} eGFRcrea

Figure 3. Overview of all 17 currently identified genetic loci. Overlap indicates replication in present study. The six ***BOLD*** loci are genome-wide significant ($p<5 \times 10^{-8}$) in the present study; all other loci in overlapping areas were replicated in the present study at a one-sided p<0.05.
* Novel loci for European populations.



Secondary analyses and Bioinformatics

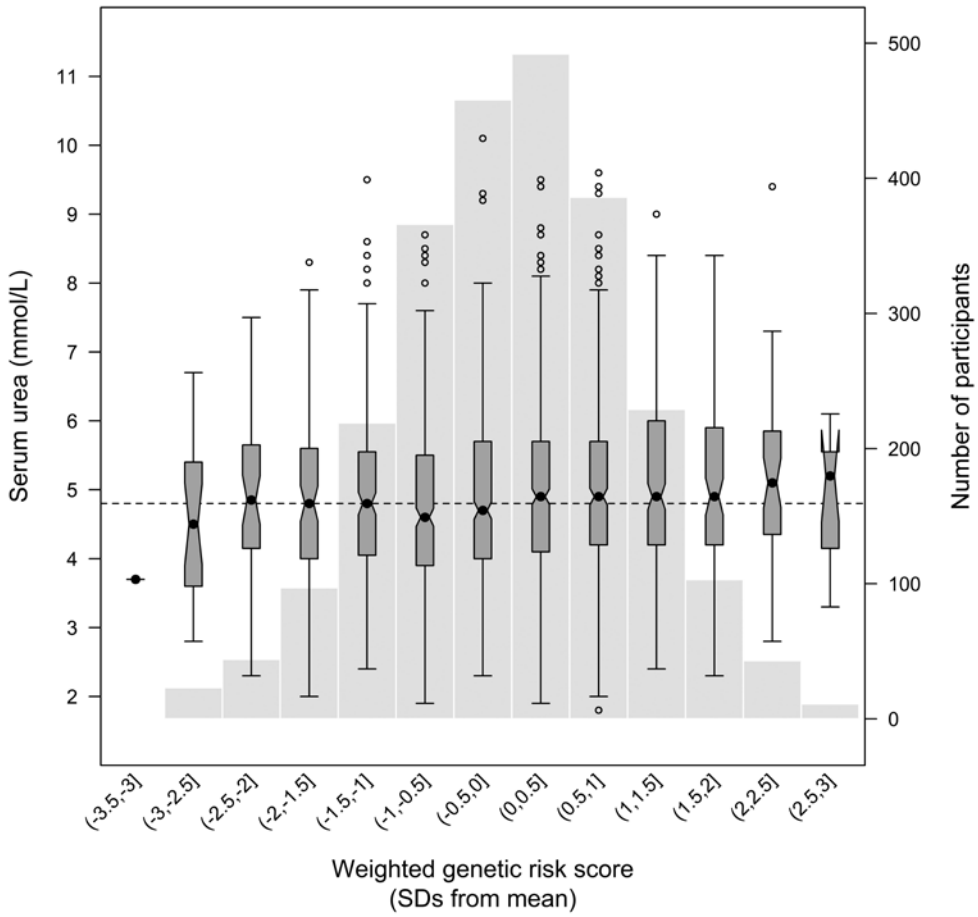
Associations with kidney function

One index SNP (rs2003313) was significantly associated with kidney function, though not in the expected direction (**Supplementary Figure and Table S8a-b**). rs914615 and rs2503107 were borderline significantly associated with kidney function (p=0.095 and p=0.085) in the expected direction. Conversely, 53 known eGFRcrea SNPs[17] were examined for potential associations with serum urea levels in all Stage I+II cohorts. After meta-analysis, 14 of the 53 SNPs were significantly associated with serum urea levels (**Supplementary Figure and Tables S9a-c**), more than could be expected through random chance alone (binomial distribution, 14/53, =0.05, $p=1.98 \times 10^{-7}$).

Proportion of phenotypic variance explained in the NESDA cohort

A GRS comprising all six index SNPs explained a small, but significant proportion of 0.43-0.45% of phenotypic variation in NESDA (**Supplementary Table S10**). This increased to 0.45-0.56% when 11 independent SNPs were added from the Scottish and East-Asian studies. A weighted GRS comprised of all 17 SNPs showed a modest but significant linear trend ($p<2.3x10^{-4}$) in urea levels (**Figure 4**).

Figure 4. Distribution of serum urea levels. Boxplots of serum urea levels (mmol/L) by categories of a weighted genetic risk score comprised of all 17 currently identified serum urea SNPs in the NESDA cohort (N=2472). The black dots represent the medians; the grey boxes represent the observations between the 25th and the 75th percentile; the whiskers represent (at maximum) 1.5 times the interquartile range; the notches represent the 95%CI of the median. In the rightmost boxplot, the notches extend to outside the box due to its wide 95%CI. The underlying light gray histogram represents the population distribution of the genetic risk score; its bell shape approximates a normal distribution. The dashed horizontal line depicts the median serum urea level in the NESDA cohort (4.8 mmol/L).



However, we observed no clinically relevant differences in serum urea between extremes of this GRS. The 53 SNPs identified to be associated with serum creatinine by the CKDGen consortium explained 0.18% of the variance in serum urea (p=0.02), but significance of this effect disappeared when correcting for *log*eGFRcrea or serum creatinine.

Bioinformatics characterization of the index SNPs

Our analyses returned 345 SNPs in at least moderate LD (r^2>0.50), of which 173 in at least high LD (r^2>0.80) and 49 in perfect LD (r^2=1). rs914615 is linked with two non-synonymous SNPs: rs760077 (*MTX1*), and rs4745 (*EFNA1*), both of which are predicted to be benign[32]. A number of proxy SNPs in high LD (r^2>0.8) with the index SNPs were reported in the literature as associated with other kidney-function or metabolically-relevant traits such as serum magnesium level and anthropomorphic traits. rs914615 was previously found associated with urinary albumin-to-creatinine ratio in diabetic subjects[39] (**Supplementary Table S13**). Using eQTL data publicly available from GTEx Portal, we found associations of three SNPs with gene expression in various tissues, and predominantly in gastro-intestinal tissues (Supplementary Table S14): rs914615 with expression of numerous genes, among others *EFNA1, MTX1, MUC1,* and *THBS3*; rs2003313 with *COLCA1* and *COLCA2*; and rs11954639 with *RPL37*. In whole blood, SNP rs914615 was associated with expression of *THBS3*, *ADAM15*, *KRTCAP2* (**Supplementary Table S15**). In kidney biopsy specimens, we found an association of the A allele of rs914615 with decreased mucin gene (*MUC1*) expression (**Supplementary Table S16**).

DISCUSSION

In this meta-analysis of GWAS in European populations, we identified six index SNPs at six genomic loci (in *THBS3, ADAMTS9-AS2, RSPO3,* or near *LPP, PTGER4, and POU2AF1*) that were associated with serum urea levels at a genome-wide significant level. Of these six index SNPs, two (near *POU2AF1* and in *ADAMTS9-AS2*) are completely novel associations with urea, i.e. not previously identified in either the East-Asian or Scottish studies. Three SNPs tag regions (*THBS3, LPP, and RSPO3*) previously identified in East-Asians. SNP rs11954639 near *PTGER4* is in high LD with a SNP previously identified in Scottish GWAS. Follow-up analysis of the six index SNPs yielded potential roles of a number of loci in urea metabolism.

In addition to our main meta-analysis, we examined 20 SNPs at 13 genetic loci previously associated with BUN in East-Asians[8-11]. Of these 20 SNPs, we replicated 15 at a one-sided p<0.05, confirming 10 previously identified loci (*MTX1-GBA*, *PAX8*, *BCL6-LPP*, *LRIG1-KBTBD8*, *RSPO3*, *UNCX*, *MPPED-DCDC5*, *WDR72*, *BCAS3*, and *SLC14A2*) but not *MECOM*, *C12orf51*, and *GNAS*. Of note, we replicated SNPs at the *SLC14A2* locus, a gene that encodes a renal tubular urea transporter (RefSeq release 89)[40]. Furthermore, we confirmed SNP associations at *MTX1, RP11-115 J16.1*, *PRKAG2, UNCX*, and an intergenic region near *PTGER4*, that were identified in a single-cohort GWAS in 19,293 Generation Scotland participants[13]. After replication, SNPs at 14 loci now have confirmed associations with serum urea in Europeans. SNPs tagging *PTGER4, PRKAG2, ADAMTS9-AS2,* and *POU2AF1* were specific to European studies, likely due to considerably lower minor allele frequencies in East-Asians (0%, 0%, 16%, and 12%, respectively) compared with Europeans (7%, 30%, 46%, 44%) according to the 1000G phase 3 East-Asian (EAS) and European (EUR) reference sets[30].

GWAS of biomarkers that are excreted through the kidney may be confounded by kidney function[41]. We therefore examined the effect of kidney function on SNP associations by running both unadjusted models and *log*eGFRcrea-adjusted models. Associations of two SNPs (rs4686914, rs2003313) were unaffected by this adjustment, thus are suggested to affect urea levels not through kidney function but through other mechanisms. Associations of three SNPs (rs998394, rs11954639, rs2503107) were only significant in the *log*eGFRcrea-adjusted model, indicating positive confounding/suppression, i.e. genetic effects were masked by kidney function. Associations of one SNP (rs914615) diminished after *log*eGFRcrea adjustment, suggesting that the effect of this SNP on serum urea is (partly) confounded or mediated through kidney function. In the following paragraphs, we discuss the two novel loci.

We report a novel association of urea with rs2003313, a SNP on chromosome 11 in an intergenic region near *POU2AF1*. We queried the GWAS catalog to find other phenotypes associated with this SNP, and SNPs in LD, r^2>0.50); however, we found none. eQTL analysis in GTEx[35] yielded significant associations of rs2003313 with expression of *COLCA2* and *COLCA1* (aliases *C11orf93* and *C11orf92*, respectively) in colon, esophagus, spleen, tibial artery and nerve, and adipose tissue. Protein function of *COLCA2* is currently unknown. *COLCA1* encodes a transmembrane protein of granular structures, such as crystalloid eosinophilic granules and other

granular organelles[40], with preferential expression in stomach, urinary bladder, and prostate. Both *COLCA2* and *COLCA1* have previously been associated to colorectal cancer[42]. Relevance of this locus to serum urea is unclear, and may be explored in future study. Against expectations, the T allele of rs2003313 was associated with lower serum urea in the present study, and with lower eGFRcrea in CKDGen data[17]. Whether this is due to unmeasured confounding or some unknown biological factor may be explored in future study. Potential biological mechanisms may be explored in future study. Of note, moderate heterogeneity was observed (I^2: 43-61%) with diminution of effect size in the replication phase, possibly indicative of Winner's curse[43], i.e. the effect of this SNP may be overestimated. Nonetheless, the strong significance of the combined meta-analysis of this locus indicates it is a non-spurious signal.

A second novel SNP is rs998394 on chromosome 3. Although in relative proximity (distance ~2Mb) to SNPs (near *LRIG1-KBTBD8*) previously identified in East-Asian GWAS on BUN, these are not in linkage disequilibrium (r^2=0.0); we thus consider this SNP independent and therefore a novel finding. rs998394 is located in *ADAMTS9-AS2*, a long non-coding RNA that is an antisense transcript of *ADAMTS9*. The protein encoded by *ADAMTS9* is a member of the ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) protein family. Members of this family have been implicated in the cleavage of proteoglycans, the control of organ shape during development, and the inhibition of proteoglycans[40]. *ADAMTS9* is localized to chromosome region 3p14.3-p14.2, an area known to be lost in hereditary renal tumors[44]. *ADAMTS9* has previously been associated with anthropomorphic traits[45, 46] and type 2 diabetes mellitus[47].

Loci tagged by the other four index SNPs are discussed in Supplementary Note S12. Briefly, we found potential roles of *MUC1* and *PTGER4* in urea metabolism and/or kidney function.

Sex-stratified analysis yielded no additional loci, although a marked difference in effect size was observed between men and women for rs4686914 and rs11954639. This is suggestive of gender-specific mechanisms of urea metabolism which may be investigated in future study.

Fourteen out of 53 (26%) known eGFRcrea loci were associated (one-sided p<0.05) with serum urea levels in our discovery cohort, more than could be expected through random chance alone. Furthermore, a GRS based on these loci was modestly but significantly associated with serum urea, supporting the notion of genetic overlap between the two traits. Previously, Okada et al observed associations of *MPPED-DCDC5, BCAS3, WDR72,* and *UNCX*, with both creatinine and BUN at the genome-wide level in East-Asians[10], indicating possible pleiotropy. In addition, the present study suggests pleiotropy for *PRKAG2, UNCX,* and *WDR72*, given that these known eGFRcrea loci also associated with serum urea in the present study.

To the best of our knowledge, the present study is the first meta-analysis of GWAS of serum urea in European populations. We were able to report new associations for European populations and confirm known associations from East-Asian studies. However, a genetic risk score combining all currently identified SNPs was only modestly associated with serum urea. Future study may involve imputation to the Haplotype Reference Consortium reference set[48], which due to its higher resolution may yield more precise results. Given the estimated explained variance of the identified SNPs (0.56%), and the estimated heritability of serum urea levels (44%), many of the genetic factors influencing serum urea are still to be found; larger samples are needed to detect these factors. Consequently, the immediate clinical relevance of our findings is limited.

In conclusion, we report the first meta-analysis of GWAS of serum urea levels in European populations. We identified six genomic loci reproducibly associated with serum urea. We are the first to report two SNP associations with urea near *POU2AF1* and in *ADAMTS9-AS2*. The identified regions have possible relevance to urea metabolism, as well as kidney function.

All Supplementary material can be accessed via the following link:
www.karger.com/Article/FullText/496930

REFERENCES

1. Gotch FA: Urea is the best molecule to target adequacy of peritoneal dialysis. *Perit Dial Int.* 2000;20 Suppl 2:S58-64.

2. Aronson D, Mittleman MA, Burger AJ: Elevated blood urea nitrogen level as a predictor of mortality in patients admitted for decompensated heart failure. *Am J Med.* 2004;116(7):466-473.

3. Gotsman I, Zwas D, Planer D, Admon D, Lotan C, Keren A: The significance of serum urea and renal function in patients with heart failure. *Medicine.* 2010;89(4):197-203.

4. Baum N, Dichoso CC, Carlton CE: Blood urea nitrogen and serum creatinine: physiology and interpretations. *Urology.* 1975;5(5):583-588.

5. Hosten AO: BUN and Creatinine; in 3rd, Walker HK, Hall WD, Hurst JW (eds): *Clinical Methods: The History, Physical, and Laboratory Examinations*. Boston, Butterworth Publishers, a division of Reed Publishing, 1990.

6. Dirkx TC, Woodell T: Kidney Disease; in Papadakis MA, McPhee SJ, Rabow MW (eds): *Current Medical Diagnosis & Treatment 2019*. New York, NY, McGraw-Hill Education, 2019.

7. Kettunen J, Tukiainen T, Sarin A, Ortega-Alonso A, Tikkanen E, Lyytikäinen L, et al: Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet.* 2012;44(3):269.

8. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, et al: Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet.* 2010;42(3):210.

9. 9. Kim YJ, Go MJ, Hu C, Hong CB, Kim YK, Lee JY, et al: Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat Genet.* 2011;43(10):990.

10. Okada Y, Sim X, Go MJ, Wu JY, Gu D, Takeuchi F, et al: Meta-analysis identifies multiple loci associated with kidney function-related traits in East Asian populations. *Nat Genet.* 2012;44(8):904-909

11. Lee J, Lee Y, Park B, Won S, Han JS, Heo NJ: Genome-wide association analysis identifies multiple loci associated with kidney disease-related traits in Korean populations. *PLoS One.* 2018;13(3):e0194044.

12. Prins BP, Kuchenbaecker KB, Bao Y, Smart M, Zabaneh D, Fatemifar G, et al: Genome-wide analysis of health-related biomarkers in the UK Household Longitudinal Study reveals novel associations. *Scientific Reports.* 2017;7(1):11008.

13. Nagy R, Boutin TS, Marten J, Huffman JE, Kerr SM, Campbell A, et al: Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. *Genome Med.* 2017;9(1):23.

14. Chambers JC, Zhang W, Lord GM, Van Der Harst P, Lawlor DA, Sehmi JS, et al: Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet.* 2010;42(5):373-375.

15. Kottgen A, Glazer NL, Dehghan A, Hwang S, Katz R, Li M, et al: Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet.* 2009;41(6):712-717.

16. Kottgen A, Pattaro C, Boger CA, Fuchsberger C, Olden M, Glazer NL, et al: New loci associated with kidney function and chronic kidney disease. *Nat Genet.* 2010;42(5):376-384.

17. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, et al: Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun.* 2016;7.

18. Gorski M, van dM, Teumer A, Chu AY, Li M, Mijatovic V, et al: 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. *Scientific Reports.* 2017;7:45040.

19. Scholtens S, Smidt N, Swertz MA, Bakker SJ, Dotinga A, Vonk JM, et al: Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol.* 2015 Aug;44(4):1172-80.

20. Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N: Representativeness of the LifeLines cohort study *PLoS One.* 2015 Sep 2;10(9)

21. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR: MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol.* 2010;34(8):816-834.

22. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR: Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012;44(8):955-959.

23. International HapMap Consortium: A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007;449(7164):851.

24. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D: A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Ann Intern Med.* 1999;130(6):461-470.

25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007 Sep;81(3):559-75.

26. van der Most, Peter J, Vaez A, Prins BP, Munoz ML, Snieder H, Alizadeh BZ, et al: QCGWAS: A flexible R package for automated quality control of genome-wide association results. *Bioinformatics.* 2014;30(8):1185-1186.

27. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'donnell CJ, De Bakker PI: SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* 2008;24(24):2938-2939.

28. Mägi R, Morris AP: GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics.* 2010;11(1):288.

29. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

30. 1000 Genomes Project Consortium: A map of human genome variation from population-scale sequencing. *Nature.* 2010;467(7319):1061-1073.

31. Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164-e164.

32. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al: A method and server for predicting damaging missense mutations. *Nat Methods.* 2010 Apr;7(4):248-9.

33. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al: The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 2016;45(D1):D896-D901.

34. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al: Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.

35. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al: The genotype-tissue expression (GTEx) project. *Nat Genet.* 2013;45(6):580.

36. Westra H, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al: Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013;45(10):1238.

37. Damman J, Bloks VW, Daha MR, van der Most PJ, Sanjabi B, van der Vlies P, et al: Hypoxia and Complement-and-Coagulation Pathways in the Deceased Organ Donor as the Major Target for Intervention to Improve Renal Allograft Outcome. *Transplantation.* 2015;99(6):1293-1300.

38. Wain LV, Vaez A, Jansen R, Joehanes R, van der Most PJ, Erzurumluoglu AM, et al: Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension.* 2017;70:e4-e19.

39. Teumer A, Tin A, Sorice R, Gorski M, Yeo NC, Chu AY, et al: Genome-wide association studies identify genetic loci associated with albuminuria in diabetes. *Diabetes.* 2016;65(3):803-81.

40. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al: Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2015;44(D1):D733-D745.

41. Wuttke M, Köttgen A: Insights into kidney diseases from genome-wide association studies. *Nat Rev Nephrol.* 2016;12(9):549.

42. Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, et al: Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet.* 2008;40(5):631.

43. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN: Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet.* 2003;33(2):177.

44. Clark ME, Kelner GS, Turbeville LA, Boyer A, Arden KC, Maki RA: ADAMTS9, a novel member of the ADAM-TS/metallospondin gene family. *Genomics.* 2000;67(3):343-350.

45. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al: Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet.* 2010;42(11):949.

46. Randall JC, Winkler TW, Kutalik Z, Berndt SI, Jackson AU, Monda KL, et al: Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* 2013;9(6):e1003500.

47. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al: Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008;40(5):638.

48. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al: A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48(10):1279.

PART III

Utilizing genetics to explain socieconomic disparities in chronic kidney disease