Genome-Wide Association Scan of Serum Urea in European Populations Identifies Two Novel Loci

Lifelines Cohort Study group; Thio, Chris H L; Reznichenko, Anna; van der Most, Peter J; Kamali, Zoha; Vaez, Ahmad; Smit, Johannes H; Penninx, Brenda W J H; Haller, Toomas; Mihailov, Evelin

Published in:
American Journal of Nephrology

DOI:
10.1159/000496930

Original Report: Patient-Oriented, Translational Research

American Journal of Nephrology

Am J Nephrol 2019;49:193–202
DOI: 10.1159/000496930

Received: September 18, 2018
Accepted: January 11, 2019
Published online: February 26, 2019

Genome-Wide Association Scan of Serum Urea in European Populations Identifies Two Novel Loci

Chris H.L. Thio[a] Anna Reznichenko[b] Peter J. van der Most[a] Zoha Kamali[c] Ahmad Vaez[a, c] Johannes H. Smit[e, f] Brenda W.J.H. Penninx[e, f] Toomas Haller[g] Evelin Mihailov[g] Andres Metspalu[g] Jeffrey Damman[h, j] Martin H. de Borst[b, h] Pim van der Harst[i, k] Niek Verweij[i, k] Gerjan J. Navis[b, i] Ron T. Gansevoort[b, i] Ilja M. Nolte[a] Harold Snieder[a] Lifelines Cohort Study group[d]

[a]Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; [b]Department of Nephrology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; [c]Department of Bioinformatics, Isfahan University of Medical Sciences, Isfahan, Iran; [d]Lifelines Cohort Study and Biobank, Groningen, The Netherlands; [e]The Netherlands Study of Depression and Anxiety (NESDA), GGZ inGeest, Amsterdam, The Netherlands; [f]Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands; [g]Estonia Genome Center University of Tartu (EGCUT), Institute of Genomics, Tartu, Estonia; [h]TransplantLines, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; [i]Prevention of REnal and Vascular ENdstage Disease (PREVEND) Cohort Study, Groningen, The Netherlands; [j]Department of Pathology, Erasmus Medical Center, Rotterdam, The Netherlands; [k]Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Keywords
Genome-wide association studies · Serum urea · Kidney function

Abstract
Background: Serum urea level is a heritable trait, commonly used as a diagnostic marker for kidney function. Genome-wide association studies (GWAS) in East-Asian populations identified a number of genetic loci related to serum urea, however 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 7,379 participants of European ancestry. ***Results:*** We identified 6 genome-wide significant single nucleotide polymorphisms (SNPs) in or near 6 loci, of which 2 were novel (*POU2AF1* and *ADAMTS9-AS2*). Replication of East-Asian and Scottish data provided evidence for an additional 8 loci. SNPs tag regions previously associated with anthropometric traits, serum magnesium, and urinary albu-

C.H.L.T. and A.R. contributed equally to this work. R.T.G, I.M.N., and H.S. have joint senior authorship.

KARGER

E-Mail karger@karger.com
www.karger.com/ajn

Chris H.L. Thio
Unit of Genetic Epidemiology and Bioinformatics, Department of Epidemiology (HPC FA40), University Medical Center Groningen, University of Groningen
Hanzeplein 1, PO Box 30.001, NL–9700 RB Groningen (The Netherlands)
E-Mail c.h.l.thio@umcg.nl

min-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 into the genetic underpinnings of urea metabolism, with potential relevance to kidney function.

Background

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 2 markers are not equivalent in the 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 (GWASs) on BUN in East-Asians reported single nucleotide polymorphisms (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), 5 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 the 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 just 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 2-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.

Study Population

Stage I discovery analyses were performed in 13,312 subjects from the Lifelines Cohort Study. Stage II replication testing was performed in 7,379 subjects from the PREVEND (n = 3,387), NESDA (n = 2,523), EGCUT1 (n = 712), and EGCUT2 (n = 757) cohorts (online suppl. Note 1; for all online suppl. material, see www.karger.com/doi/10.1159/496930).

The Lifelines Cohort Study is a multidisciplinary prospective population-based cohort study with a unique 3-generation design that examines health and health-related behavior of 165,729 participants living in the north-eastern region of the Netherlands (https://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

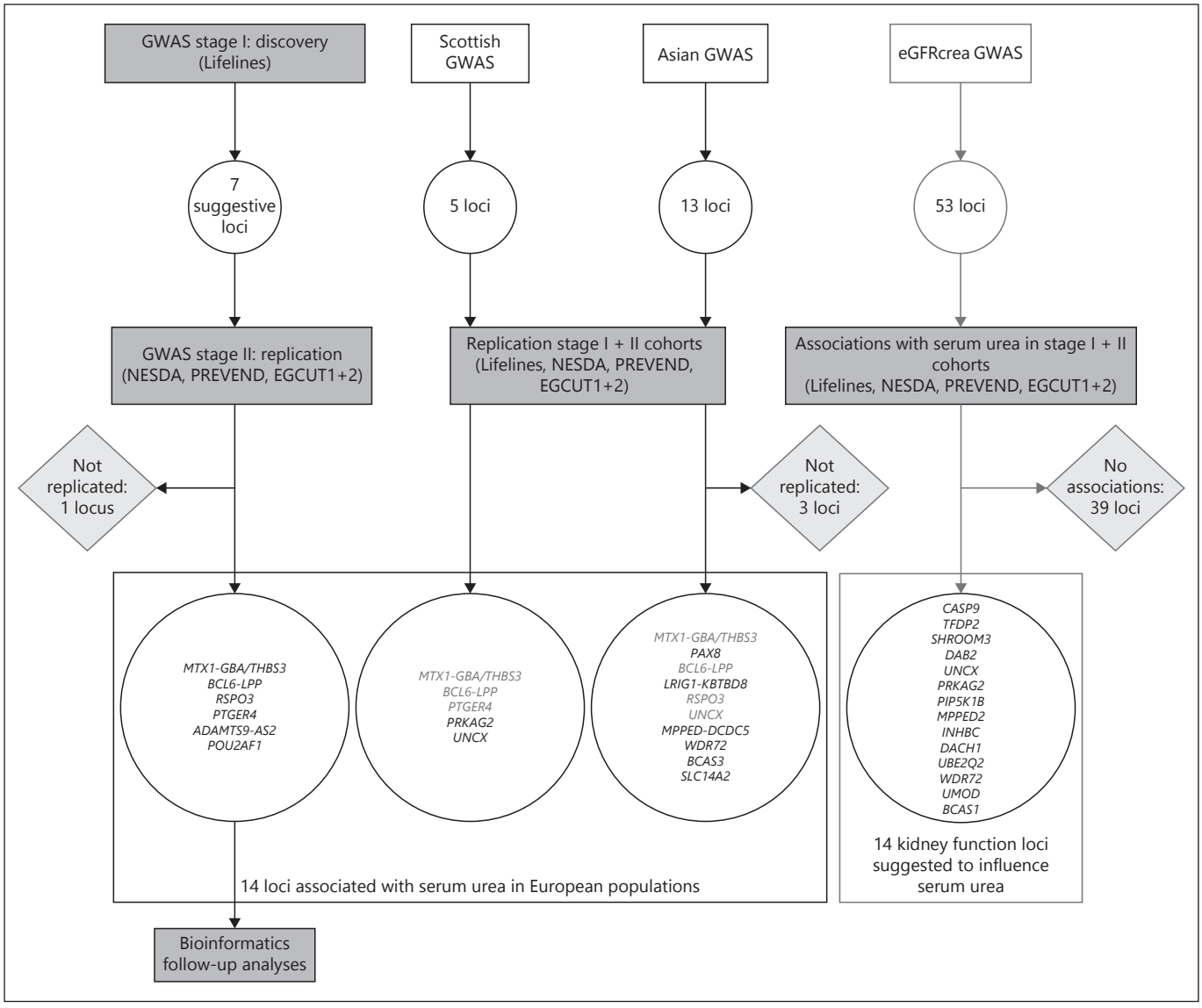

Fig. 1. Design and results of the present study. Genetic loci in GREY typefont indicate that these loci overlap between GWAS studies on serum urea/BUN.

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 Ethics Committee of the University Medical Center Groningen. Written informed consent was obtained from all participants during their visit to 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 SNPs were used in this study. SNPs were excluded when the call rate was <95%, when the minor allele frequency was <1%, or when the p value of the Hardy-Weinberg equilibrium 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 minor allele frequency <1% were excluded after imputation. The resulting number of SNPs available for analysis was 1.99×10^6. The procedure for genotyping, quality control, and imputation of the replication cohorts is described in online supplementary Note S1.

Phenotype Measurement in Lifelines

During the baseline examination, the study participants were asked to fill in a questionnaire before the visit. During the visit, a number of investigations were conducted and blood and 24 h-urine samples were taken. A total of 13,385 genotyped participants were included in 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 SDs from the mean. Serum urea measurements were performed with an ultraviolet kinetic assay on a Roche Modular. Serum creatinine was measured using 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 Study equation [24]. Body mass index (kg/m^2) was calculated by dividing the 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, and the first 10 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 3 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 4 replication cohorts separately. The replication results of these SNPs were meta-analyzed using an inverse variance weighted 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 the 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 5 SNPs at 5 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 5 cohorts (i.e., Lifelines, NESDA, PREVEND, EGCUT1 + 2) for these analyses. We considered a SNP replicated at a one-sided $p < 0.05$.

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 6 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 6 index SNPs, of the 6 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 (https://ldlink.nci.nih.gov/) using the effect sizes from the discovery sample (our 6 index SNPs) or from literature as weights.

Bioinformatics Characterization of the Replicated SNPs

We examined the 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 1,000 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 5a (February 20, 2015), including the 503 subjects of European ancestry. We used ANNOVAR (version July 16, 2017; http://annovar.openbioinformatics.org/) [31] for annotation of the index SNPs. We queried PolyPhen-2 (http://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 6 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 (https://www.gtexportal.org/) [35], and blood tissue (https://genenetwork.nl/bloodeqtlbrowser/) [36]. Furthermore, we examined eQTLs in donor kidney tissue in TransplantLines (detailed description of data and methods in online suppl. Note 11) [37, 38].

Results

Meta-Analysis Results

Manhattan plots of stage I for models 1 and 2 are shown in online supplementary Figure 2a. Regional association plots, showing location and significance of top hits for models 1 and 2 relative to known loci, are shown in online supplementary Figure S3. Risk of bias due to population stratification was assessed and considered acceptable ($\lambda = 1.05$; online suppl. Fig. S4). For models 1 and 2, 7 index SNPs were at least suggestive ($p < 1 \times 10^{-6}$) in stage I. Of these 7 SNPs, rs17586946 on chromosome 6 was only suggestive in the combined Stage I + II samples ($p = 1.4 \times 10^{-7}$) and hence not replicated. Table 1 shows results of the remaining 6 SNPs. For model 1, we replicated 3 SNPs (rs914615, rs4686914, rs2003313) at 3 genomic loci, significantly associated with serum urea at the genome-wide level ($p < 5 \times 10^{-8}$) in the combined

Stage I + II samples. In the second, *log*eGFRcrea-adjusted model, 2 SNPs from model 1 (rs4686914 and rs2003313) were again identified, while in addition 3 other SNPs (rs998394, rs11954639, rs2503107) were identified and replicated with genome-wide level significance. One SNP (rs914615) did not reach suggestive significance of $p < 1 \times 10^{-6}$ after *log*eGFRcrea adjustment ($p = 2.9 \times 10^{-6}$) and therefore was deemed non-significant for this model. A third, serum creatinine adjusted model, yielded essentially the same results as the *log*eGFRcrea-adjusted model (online suppl. Fig. 2a and Table S5).

Sex-stratified analysis yielded no additional loci: (1) we found no significant associations in female-only models, and (2) in male-only models, we identified 2 additional SNPs (rs9860469 and rs9820812) in high LD ($r^2 = 0.70$ and $r^2 = 1.0$, respectively) with a SNP already identified in models 1–2 (rs4686914; online suppl. Fig. S2b). Effects of rs4686914 and rs11954639 were stronger in men (online suppl. Table S6).

Replication of Previously Reported Urea Loci

We replicated 10 out of 13 East-Asian loci [8–11] at a one-sided $p < 0.05$ (online suppl. Table S7a). SNPs at 3 loci (*MECOM, C12orf51, GNAS*) were not replicated in the present study. All 5 Scottish loci [13] were replicated (online suppl. Table S7b). In total, 14 loci are now confirmed for Europeans (Fig. 2).

Associations with Kidney Function

One index SNP (rs2003313) was significantly associated with kidney function, though not in the expected direction (online suppl. Fig. S8a and Table S8b). 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 (online suppl. Fig. S9a and Tables S9b-c), more than could be expected through random chance alone (binomial distribution, 14/53, $\alpha = 0.05$, $p = 1.98 \times 10^{-7}$).

Proportion of Phenotypic Variance Explained in the NESDA Cohort

A GRS comprising all 6 index SNPs explained a small, but significant proportion of 0.43–0.45% of phenotypic variation in NESDA (online suppl. Table S10). This increased to 0.45–0.56% when 11 independent SNPs were added from the Scottish and East-Asian studies. A weight-

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				I^2, %
							B	SE	*p* value	*n*	B	SE	*p* value	*n*	B	SE	*p* value	*n*	
rs914615	1	153442516	Intronic	*THBS3*	A/G (0.476)	1	0.070	0.014	8.9E-07	13,312	0.065	0.020	1.3E-03	7,379	0.068	0.012	4.3E-09	20,689	0.0
						2*	0.064	0.014	2.9E-06	13,311	0.063	0.020	1.2E-03	7,335	0.064	0.011	1.3E-08	20,646	0.0
rs4686914	3	189200234	Intergenic	*LPP*	T/C (0.308)	1	–0.110	0.016	2.4E-12	13,312	–0.101	0.021	2.2E-06	7,378	–0.107	0.013	2.6E-17	20,690	0.0
						2	–0.106	0.015	2.3E-12	13,311	–0.098	0.021	2.1E-06	7,334	–0.103	0.012	2.3E-17	20,645	0.0
rs998394	3	64776227	ncRNA/intronic	*ADAMTS9-AS2*	A/G (0.458)	1*	–0.063	0.014	7.3E-06	13,312	–0.049	0.020	1.4E-02	7,379	–0.058	0.011	3.7E-07	20,691	0.0
						2	–0.067	0.014	7.5E-07	13,311	–0.058	0.019	2.2E-03	7,335	–0.064	0.011	7.1E-09	20,646	0.0
rs11954639	5	40710736	Intergenic	*PTGER4*	T/C (0.071)	1*	–0.165	0.037	5.8E-06	13,312	–0.170	0.040	2.4E-05	7,379	–0.168	0.027	6.1E-10	20,691	0.0
						2	–0.185	0.035	1.8E-07	13,311	–0.182	0.039	2.9E-06	7,335	–0.183	0.026	2.3E-12	20,646	0.0
rs2503107	6	127505069	Intronic	*RSPO3*	C/A (0.449)	1*	–0.075	0.017	8.6E-06	13,312	–0.051	0.020	1.2E-02	7,377	–0.065	0.013	4.9E-07	20,689	0.0
						2	–0.084	0.016	2.9E-07	13,311	–0.056	0.020	4.2E-03	7,333	–0.072	0.013	8.1E-09	20,644	18.0
rs2003313	11	110709203	intergenic	*POU2AF1*	T/A (0.448)	1	–0.088	0.015	6.0E-09	13,312	–0.048	0.020	1.7E-02	7,377	–0.073	0.012	1.3E-09	20,691	60.6
						2	–0.087	0.015	2.5E-09	13,311	–0.055	0.019	4.3E-03	7,333	–0.075	0.012	9.5E-11	20,644	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.
[a] position based on NCBI b36/hg18. [b] EAF in the complete sample (Stage I + II). * Not suggestive ($p \geq 1E\text{-}06$) in stage I for this model.
Model 1: adjusted for age, age^2, sex, body mass index, principal components 1–10.
Model 2: model 1 + log_{10} eGFRcrea.
B, unstandardized regression coefficient; Chr, chromosome; bp, basepair; EAF, effect allele frequency; I^2, heterogeneity statistics; SE, standard error; SNP, single nucleotide polymorphism.

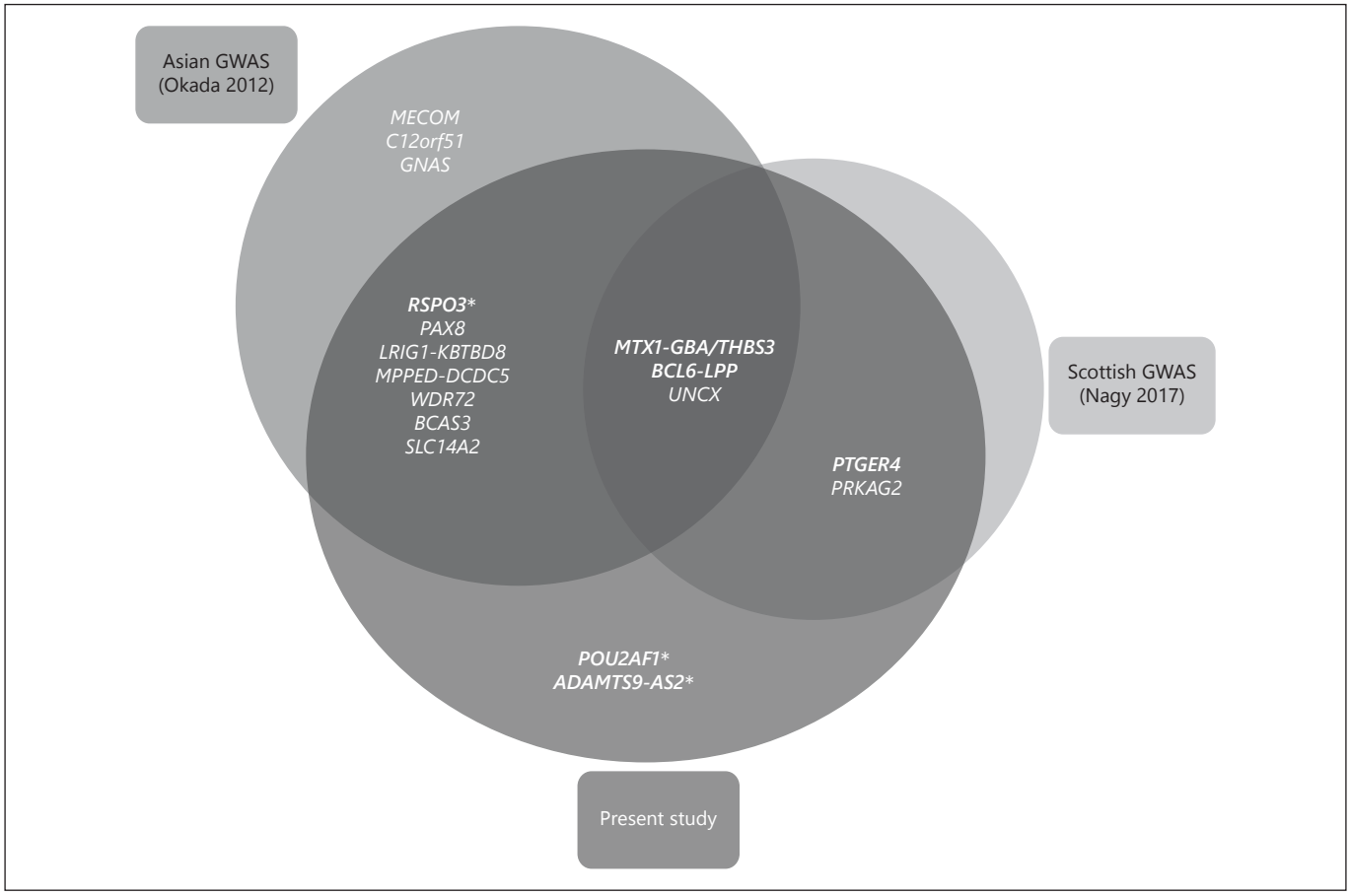

Fig. 2. Overview of all 17 currently identified loci in European and East-Asian populations. Overlap indicates replication in present study. The 6 BOLD loci are genome-wide significant ($p < 5 \times 10^{-8}$) in the present study; all other loci in overlapping areas were replicated in the present study at a one-sided $p < 0.05$. * Novel loci for European populations.

ed GRS comprised of all 17 SNPs showed a modest but significant linear trend ($p < 2.3 \times 10^{-4}$) in urea levels (Fig. 3). 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 2 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 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 to be associated with urinary albumin-to-creatinine ratio in diabetic subjects [39] (online suppl. Table S13). Using eQTL data publicly available from GTEx Portal, we found associations of 3 SNPs with gene expression in various tissues, and predominantly in gastro-intestinal tissues (online suppl. 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* (online suppl. Table S15). In kidney biopsy specimens, we found an association of the A allele of rs914615 with decreased mucin gene (*MUC1*) expression (online suppl. Table S16).

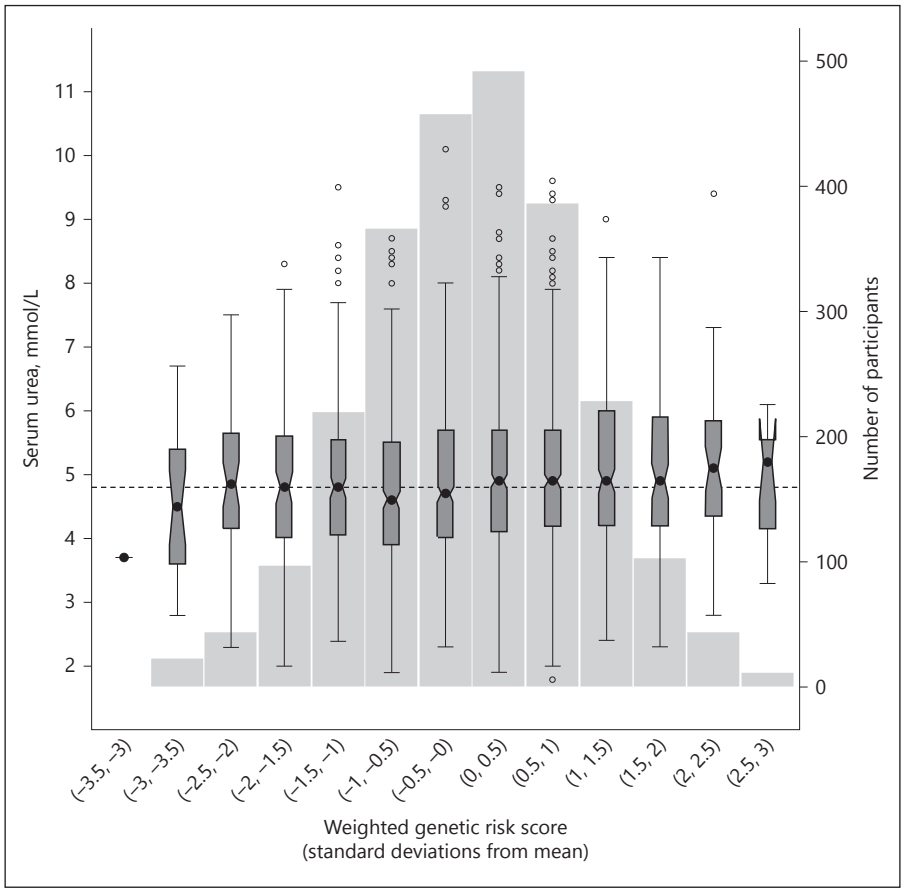

Fig. 3. Boxplots of serum urea levels (mmol/L) by categories of a weighted GRS comprised of all 17 currently identified serum urea SNPs in the NESDA cohort (n = 2,472). The black dots represent the medians, the grey boxes represent the observations between the 25th and the 75th percentile, the whiskers 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 grey histogram represents the population distribution of the GRS; 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). GRS, genetic risk score.

Discussion

In this meta-analysis of GWAS in European populations, we identified 6 index SNPs at 6 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 6 index SNPs, 2 (near *POU2AF1* and in *ADAMTS9-AS2*) are completely novel associations with urea, that is, 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 6 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, and 44%) according to the 1000G phase 3 East-Asian and European 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 2 SNPs (rs4686914, rs2003313) were unaffected by this adjustment, and are thus suggested to affect urea levels not

through kidney function but through other mechanisms. Associations of 3 SNPs (rs998394, rs11954639, rs2503107) were only significant in the *log*eGFRcrea-adjusted model, indicating positive confounding/suppression, that is, 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 2 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 [34]. 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. 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], that is, the effect of this SNP may be overestimated. Nonetheless, the strong significance of the combined meta-analysis of this locus indicates that 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 LD ($r^2 = 0.0$); we thus consider this SNP as 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 4 index SNPs are discussed in online 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 2 traits. Previously, Okada et al. [10] observed associations of *MPPED-DCDC5, BCAS3, WDR72,* and *UNCX* with both creatinine and BUN at the genome-wide level in East-Asians, indicating possible pleiotropy. In addition, the present study suggests pleiotropy for *PRKAG2, UNCX*, and *WDR72*, given that these known eGFRcrea loci are 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 GRS 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 6 genomic loci reproducibly associated with serum urea. We are the first to report 2 SNP associations

with urea near *POU2AF1* and in *ADAMTS9-AS2*. The identified regions have possible relevance to urea metabolism, as well as kidney function.

Acknowledgements

The authors wish to acknowledge the services of the Lifelines Cohort Study, the contributing research centres delivering data to Lifelines, and all the study participants. The Lifelines Biobank initiative has been made possible by funds from Fonds Economische Structuurversterking, Samenwerkingsverband Noord Nederland and REP (Ruimtelijk Economisch Programma). Funding and acknowledgements for the replication cohorts (NESDA, PREVEND, and EGCUT) are described in online supplementary Note S1.

LifeLines Cohort Study group members are: Behrooz Z. Alizadeh (Department of Epidemiology), H. Marieke Boezen (Department of Epidemiology), Lude Franke (Department of Genetics), Pim van der Harst (Department of Cardiology), Gerjan Navis (Department of Nephrology), Marianne Rots (Department of Pathology and Medical Biology), Harold Snieder (Department of Epidemiology), Morris Swertz (Department of Genetics), Bruce H.R. Wolffenbuttel (Department of Endocrinology), Cisca Wijmenga (Department of Genetics), all at University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.

Disclosure Statement

The authors declare no conflict of interests.

Funding Source

Lifelines Cohort Study and generation and management of GWAS genotype data for the Lifelines Cohort Study are supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (Fonds Economische Structuurversterking) of the Dutch Government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (Samenwerkingsverband Noord Nederland), the Province of Groningen, the University Medical Center Groningen, the University of Groningen, the Dutch Kidney Foundation, and the Dutch Diabetes Research Foundation.

References

1 Gotch FA: Urea is the best molecule to target adequacy of peritoneal dialysis. Perit Dial Int 2000;20(suppl 2):S58–S64.
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;116:466–473.
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 (Baltimore) 2010;89:197–203.
4 Baum N, Dichoso CC, Carlton CE: Blood urea nitrogen and serum creatinine. physiology and interpretations. Urology 1975;5:583–588.
5 Hosten AO: BUN and Creatinine; in Walker HK, Hall WD, Hurst JW (eds): Clinical Methods: The History, Physical, and Laboratory Examinations (3rd). 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 and Treatment 2019. New York, McGraw-Hill Education, 2019.
7 Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikäinen LP, et al: Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat Genet 2012;44:269–276.
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;42:210–215.
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;43:990–995.
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;44:904–909.
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: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. Sci Rep 2017;7:11008.
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;9:23.
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;42:373–375.
15 Kottgen A, Glazer NL, Dehghan A, Hwang SJ, Katz R, Li M, et al: Multiple loci associated with indices of renal function and chronic kidney disease. Nat Genet 2009;41:712–717.
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;42:376–384.
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;7:10023.
18 Gorski M, van der Most PJ, Teumer A, Chu AY, Li M, Mijatovic V, et al: 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. Sci Rep 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;44:1172–1180.
20 Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N: Representativeness of the LifeLines cohort study. PLoS One 2015;10:e0137203.
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;34:816–834.

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;44:955–959.
23 International HapMap Consortium, et al: A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851–861.
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;130:461–470.
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;81:559–575.
26 van der Most PJ, 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;30:1185–1186.
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;24:2938–2939.
28 Mägi R, Morris AP: GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 2010;11:288.
29 R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/2014.
30 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA: A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–1073.
31 Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38: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 2010;7:248–249.
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 2017;45: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;347:1260419.
35 Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al: The genotype-tissue expression (GTEx) project. Nat Genet 2013;45:580–585.
36 Westra HJ, 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;45:1238–1243.
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;99:1293–1300.
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;117:09438.
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;65:803–817.
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:D733–D745.
41 Wuttke M, Köttgen A: Insights into kidney diseases from genome-wide association studies 2016;12:549–562.
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;40:631–637.
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;33:177–182.
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;67:343–350.
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;42:949–960.
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 2013;9: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;40:638–645.
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;48:1279–1283.