

# Multiple loci associated with indices of renal function and chronic kidney disease

Anna Köttgen<sup>1,25</sup>, Nicole L Glazer<sup>2,25</sup>, Abbas Dehghan<sup>3,24,25</sup>, Shih-Jen Hwang<sup>4,25</sup>, Ronit Katz<sup>5</sup>, Man Li<sup>1</sup>, Qiong Yang<sup>6</sup>, Vilmundur Gudnason<sup>7,8</sup>, Lenore J Launer<sup>9</sup>, Tamara B Harris<sup>9</sup>, Albert V Smith<sup>7</sup>, Dan E Arking<sup>10</sup>, Brad C Astor<sup>1</sup>, Eric Boerwinkle<sup>11</sup>, Georg B Ehret<sup>10,12</sup>, Ingo Ruczinski<sup>13</sup>, Robert B Scharpf<sup>13</sup>, Yii-Der Ida Chen<sup>14</sup>, Ian H de Boer<sup>15</sup>, Talin Haritunians<sup>14</sup>, Thomas Lumley<sup>5</sup>, Mark Sarnak<sup>16</sup>, David Siscovick<sup>17</sup>, Emelia J Benjamin<sup>18</sup>, Daniel Levy<sup>4</sup>, Ashish Upadhyay<sup>19</sup>, Yurii S Aulchenko<sup>3</sup>, Albert Hofman<sup>3</sup>, Fernando Rivadeneira<sup>20</sup>, André G Uitterlinden<sup>20</sup>, Cornelia M van Duijn<sup>3</sup>, Daniel I Chasman<sup>21</sup>, Guillaume Paré<sup>21</sup>, Paul M Ridker<sup>21</sup>, W H Linda Kao<sup>1</sup>, Jacqueline C Witteman<sup>3,24,26</sup>, Josef Coresh<sup>1,13,26</sup>, Michael G Shlipak<sup>22,26</sup> & Caroline S Fox<sup>4,23,26</sup>

**Chronic kidney disease (CKD) has a heritable component and is an important global public health problem because of its high prevalence and morbidity<sup>1</sup>. We conducted genome-wide association studies (GWAS) to identify susceptibility loci for glomerular filtration rate, estimated by serum creatinine (eGFR<sub>crea</sub>) and cystatin C (eGFR<sub>cys</sub>), and CKD (eGFR<sub>crea</sub> < 60 ml/min/1.73 m<sup>2</sup>) in European-ancestry participants of four population-based cohorts (ARIC, CHS, FHS, RS; *n* = 19,877; 2,388 CKD cases), and tested for replication in 21,466 participants (1,932 CKD cases). We identified significant SNP associations (*P* < 5 × 10<sup>-8</sup>) with CKD at the *UMOD* locus, with eGFR<sub>crea</sub> at *UMOD*, *SHROOM3* and *GATM-SPATA5L1*, and with eGFR<sub>cys</sub> at *CST* and *STC1*. *UMOD* encodes the most common protein in human urine, Tamm-Horsfall protein<sup>2</sup>, and rare mutations in *UMOD* cause mendelian forms of kidney disease<sup>3</sup>. Our findings provide new insights into CKD pathogenesis and underscore the importance of common genetic variants influencing renal function and disease.**

CKD affects 10–13% of US adults<sup>4</sup>. Estimates from Europe are similar<sup>5</sup>, and incidence and prevalence are increasing worldwide. Its most severe form, end-stage renal disease requires dialysis and currently affects over 500,000 US adults<sup>6</sup>. In addition to conferring risk for end-stage renal disease, CKD increases the risk of cardiovascular disease<sup>7</sup> and all-cause mortality<sup>8</sup>.

Multiple studies such as familial aggregation studies have provided evidence for a genetic component to kidney disease. Heritability estimates of eGFR<sub>crea</sub>, the most commonly used measure of kidney function, are reported between 0.41 and 0.75 in individuals with the major CKD risk factors hypertension or diabetes<sup>9,10</sup>, and as 0.33 in a population-based sample<sup>11</sup>. Heritability estimates of GFR<sub>cys</sub> are similar. Although rare genetic variants causing different forms of monogenetic kidney disease have been identified, common CKD susceptibility variants have been difficult to detect reproducibly by linkage or candidate gene studies<sup>12</sup>.

To discover such variants, we conducted meta-analyses of study-specific GWAS for indices of renal function (eGFR<sub>crea</sub> and eGFR<sub>cys</sub>) and for CKD from four population-based, unselected cohorts of the CHARGE Consortium<sup>13</sup>: Atherosclerosis Risk in Communities

<sup>1</sup>Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, USA. <sup>2</sup>Cardiovascular Health Research Unit and Department of Medicine, University of Washington, Seattle, Washington, USA. <sup>3</sup>Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. <sup>4</sup>National Heart, Lung, and Blood Institute Framingham Heart Study and the Center for Population Studies, Framingham, Massachusetts, USA. <sup>5</sup>Department of Biostatistics, University of Washington, Seattle, Washington, USA. <sup>6</sup>Department of Biostatistics, Boston University, Boston, Massachusetts, USA. <sup>7</sup>Icelandic Heart Association Research Institute, Kopavogur, Iceland. <sup>8</sup>University of Iceland, Reykjavik, Iceland. <sup>9</sup>Laboratory of Epidemiology, Demography, and Biometry, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, USA. <sup>10</sup>McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA. <sup>11</sup>Human Genetics Center and Division of Epidemiology, University of Texas Health Science Center at Houston, Houston, Texas, USA. <sup>12</sup>Division of Cardiology, Geneva University Hospital, Switzerland. <sup>13</sup>Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA. <sup>14</sup>Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. <sup>15</sup>Division of Nephrology, Department of Medicine, University of Washington, Seattle, Washington, USA. <sup>16</sup>Tufts-New England Medical Center, Boston, Massachusetts, USA. <sup>17</sup>Departments of Medicine and Epidemiology, Cardiovascular Health Research Unit, University of Washington, Seattle, Washington, USA. <sup>18</sup>Boston University Schools of Medicine and Public Health, Boston, Massachusetts, USA. <sup>19</sup>Renal Section, Boston Medical Center and Boston University School of Medicine, Boston, Massachusetts, USA. <sup>20</sup>Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands. <sup>21</sup>Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. <sup>22</sup>General Internal Medicine Division, San Francisco VA Medical Center, University of California, San Francisco, California, USA. <sup>23</sup>Brigham and Women's Hospital Division of Endocrinology and Harvard Medical School, Boston, Massachusetts, USA. <sup>24</sup>Member of the Netherlands Consortium on Healthy Aging (NCHA). <sup>25</sup>These authors contributed equally to this work. <sup>26</sup>These authors jointly directed the project. Correspondence should be addressed to C.S.F. (foxca@nhlbi.nih.gov), J.C.W. (j.witteman@erasmusmc.nl), J.C. (coresh@jhu.edu) or M.G.S. (michael.shlipak@ucsf.edu).

Received 27 January; accepted 3 April; published online 10 May 2009; doi:10.1038/ng.377

**Table 1** Characteristics of the study samples

	ARIC	CHS	FHS	RS	AGES	WGHS
<b>Cohort information</b>			Discovery			Replication
Study design	Population-based prospective multicenter	Population-based prospective multicenter	Community-based family multigeneration	Population-based prospective	Population-based prospective	Trial
Analyzed sample size, CKD/eGFR <sub>crea</sub> /eGFR <sub>cys</sub>	8,069 / 6,525 / 6,430	3,278 / 3,278 / 2,844	4,140 / 3,934 / 2,992	4,390 / 4,390 / NA	3,219 / 3,219 / NA	18,247 / 18,247 / NA
<b>Sample characteristics<sup>a</sup></b>	Mean (s.d.) or % ( <i>n</i> )	Mean (s.d.) or % ( <i>n</i> )	Mean (s.d.) or % ( <i>n</i> )	Mean (s.d.) or % ( <i>n</i> )	Mean (s.d.) or % ( <i>n</i> )	Mean (s.d.) or % ( <i>n</i> )
Age, years	63.1 (5.6)	72.4 (5.4)	61.9 (8.9)	70.0 (9.0)	76.4 (5.5)	54.7 (7.1)
Male	46.7 (3,047)	39.2 (1,286)	44.4 (1,747)	38.6 (1,694)	42 (1,352)	0
Hypertension prevalence	42.8 (2,793)	52.4 (1,719)	41.8 (1,644)	34.1 (1,497)	80.6 (2,596)	24.4 (4,460)
Diabetes mellitus	13.7 (893)	11.8 (385)	11.3 (445)	10.7 (470)	11.4 (368)	2.6 (474)
eGFR cystatin, ml/min/1.73 m <sup>2</sup>	84.1 (19.7)	79.9 (18.3)	77.9 (16.9)	NA	NA	NA
eGFR creatinine, ml/min/1.73 m <sup>2</sup>	80.6 (17.2)	80.0 (22.6)	85.2 (23.5)	77.1 (17.2)	73.0 (20)	90.4 (22.8)
CKD, eGFR < 60 ml/min/1.73 m <sup>2</sup>	9.1 (731)	18.7 (612)	10.7 (445)	13.7 (600)	24.3 (781)	6.3 (1,151)

<sup>a</sup>Information on demographics, hypertension and diabetes were ascertained at the visit eGFR<sub>crea</sub> was measured. Sample sizes for CKD and eGFR<sub>crea</sub> in ARIC and FHS differ, as these studies use a cumulative definition of CKD (see Methods).

ARIC: Atherosclerosis Risk in Communities Study, CHS: Cardiovascular Health Study, FHS: Framingham Heart Study, RS: Rotterdam Study, AGES: Age Gene/Environment Susceptibility-Reykjavik Study, WGHS: Women's Genome Health Study, CKD: chronic kidney disease, eGFR: estimated glomerular filtration rate. NA, eGFR<sub>cys</sub> was not available in the replication samples.

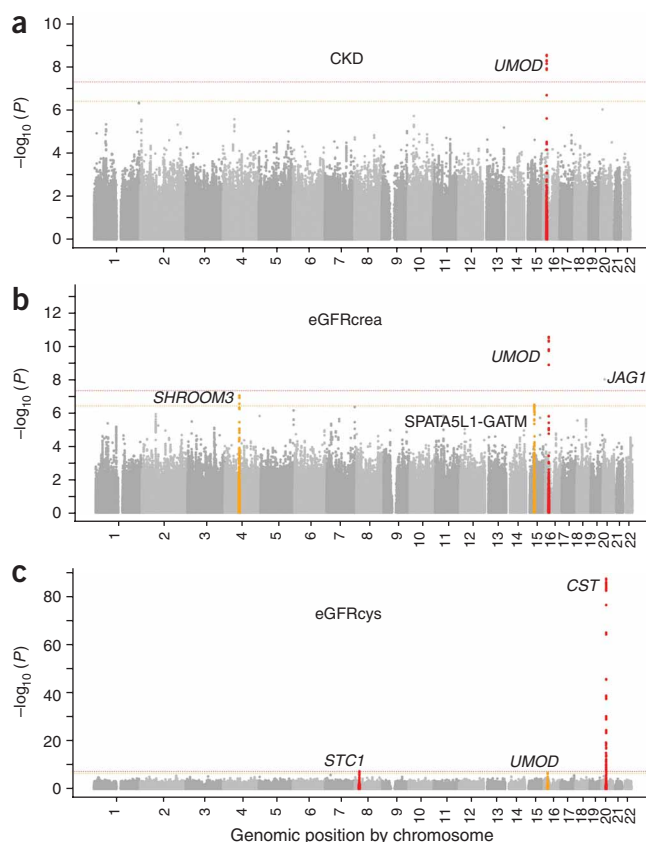
Study (ARIC), Cardiovascular Health Study (CHS), Framingham Heart Study (FHS) and Rotterdam Study (RS). As a direct measurement of kidney function is not feasible in population-based studies, we applied commonly used estimating equations to determine eGFR<sub>crea</sub><sup>14</sup> and eGFR<sub>cys</sub><sup>15</sup>. Population-based measures of GFR are imperfect<sup>16</sup>, and using two different biomarkers to estimate GFR can therefore help to uncover true signals. CKD was defined as eGFR<sub>crea</sub> < 60 ml/min/1.73 m<sup>2</sup> according to US national guidelines<sup>17</sup>, as detailed in the Methods. Genotypes for >2.5 million SNPs were imputed within each study using reference genotype data from the HapMap CEU population. Study-specific details on genotyping and imputation are provided in **Supplementary Table 1** online. SNPs showing evidence of suggestive ( $P < 4 \times 10^{-7}$ ) or significant ( $P < 5 \times 10^{-8}$ ) genome-wide association were tested for *in silico* replication in independent study samples: the Age Gene/Environment Susceptibility-Reykjavik Study (AGES) and the Women's Genome Health Study (WGHS). Detailed information on the study samples is provided in **Supplementary Methods** online.

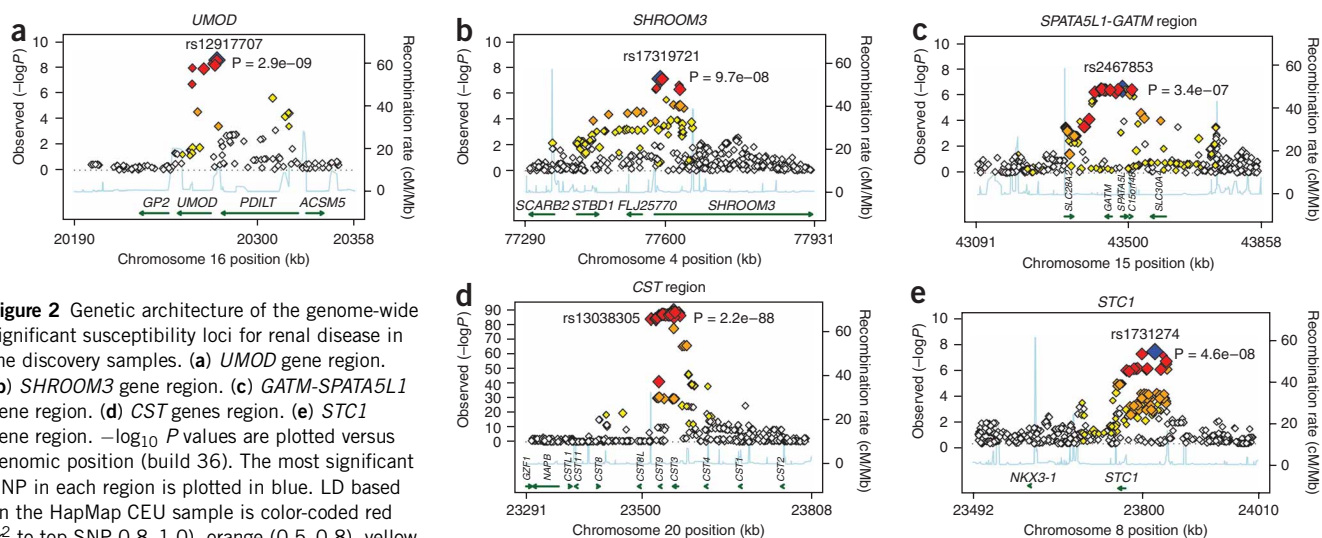
Characteristics of the four discovery and two validation study samples are shown in **Table 1**; 19,877 participants with 2,388 CKD cases and 21,466 participants with 1,932 CKD cases contributed information, respectively. CKD prevalence was higher in cohorts with older participants, ranging from 6.3% (WGHS) to 24.3% (AGES). Characteristics among CKD cases are provided in **Supplementary Table 2** online. **Figure 1** summarizes meta-analysis results for CKD, eGFR<sub>crea</sub> and eGFR<sub>cys</sub> across the discovery samples. The observed versus expected *P*-value distributions (quantile-quantile plots) are shown in **Supplementary Figure 1** online: study-specific genomic inflation factors did not indicate substantial inflation of the test statistics for any of the traits.

**Table 2** lists the most significant SNP at each genomic locus associated with CKD, eGFR<sub>crea</sub> and eGFR<sub>cys</sub>, and replication results. Study-specific results are presented in **Supplementary Table 3** online.

**Figure 1** Overview of GWAS results. (a–c) Meta-analysis  $-\log_{10}(P)$  value versus genomic position plots for CKD (a), eGFR<sub>crea</sub> (b) and eGFR<sub>cys</sub> (c) in the discovery samples. Genomic loci with evidence of suggestive association ( $P < 4 \times 10^{-7}$ ) are plotted in orange and those with significant association ( $P < 5 \times 10^{-8}$ ) are plotted in red, with the exception of the SNP at the *JAG1* locus on chromosome 20 (b, gray), which did not replicate.

For CKD, we identified SNP rs12917707 in a highly evolutionary conserved region 3.6 kb upstream of the *UMOD* (uromodulin) gene on chromosome 16 ( $P = 5 \times 10^{-16}$  across discovery and replication samples; **Fig. 2** and **Table 2**). Seven SNPs in or upstream of *UMOD* in high LD ( $r^2 > 0.8$ ) with rs12917707 were also associated with CKD at a genome-wide significant level. The minor T allele at rs12917707 was associated with 20% reduced risk of CKD (meta-analysis OR = 0.80,  $P = 2 \times 10^{-12}$ ; **Table 2**). The association of rs12917707 with CKD, which was not significant in the FHS Study, showed some





**Figure 2** Genetic architecture of the genome-wide significant susceptibility loci for renal disease in the discovery samples. (a) *UMOD* gene region. (b) *SHROOM3* gene region. (c) *GATM-SPATA5L1* gene region. (d) *CST* genes region. (e) *STC1* gene region.  $-\log_{10} P$  values are plotted versus genomic position (build 36). The most significant SNP in each region is plotted in blue. LD based on the HapMap CEU sample is color-coded red ( $r^2$  to top SNP 0.8–1.0), orange (0.5–0.8), yellow (0.2–0.5) and white ( $<0.2$ ). Gene annotations are based on UCSC Genome Browser (RefSeq Genes, b36) and arrows present direction of transcription.  $P$  values are obtained from the discovery traits: CKD (*UMOD*), eGFRcrea (*SHROOM3*, *GATM*), eGFRcys (*CST*, *STC1*).

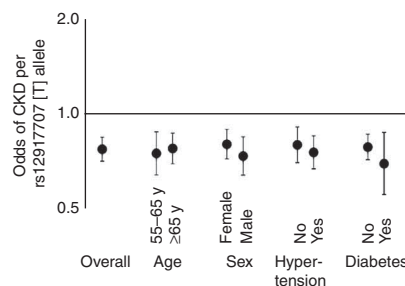
heterogeneity across studies ( $P$  heterogeneity = 0.02). Findings were consistent in models adjusting for major CKD risk factors including systolic blood pressure, hypertension medication intake and diabetes mellitus, as well as stratified for age, sex, hypertension and diabetes status (Fig. 3). Prospective information from the ARIC Study demonstrated that the T allele of rs12917707 was associated with a lower relative hazard of incident CKD (HR 0.81, 95% CI = 0.72–0.92,  $P = 0.001$ ) over 14.7 years of follow-up ( $n = 952$  cases; see Methods).

Rare *UMOD* mutations cause autosomal dominant forms of kidney disease, medullary cystic kidney disease type 2 (MCKD2), familial juvenile hyperuricemic nephropathy (FJHN), and glomerulocystic kidney disease (GCKD) (MIM603860, MIM162000, MIM609886)<sup>3,18,19</sup>. As the syndromes caused by rare *UMOD* mutations are often accompanied by hyperuricemia and gout, we explored the association of rs12917707 with these traits; no significant associations were observed. Although this does not exclude the presence of rare *UMOD* variants among our study participants, our study identifies another example of a genomic risk locus containing susceptibility variants across the spectrum of risk allele frequencies.

One study of *Umod* knockout mice reported 63% lower creatinine clearance in knockout compared to wild-type mice<sup>20</sup>. *UMOD* encodes the most abundant protein in the urine of healthy individuals, Tamm-Horsfall protein. The physiological functions of Tamm-Horsfall protein are not well understood but may include protection against inflammation and infection<sup>2</sup>. A possible role for *UMOD* in renal development was also recently reported<sup>21</sup>. *UMOD* is transcribed exclusively in renal tubular cells of the thick ascending limb of the loop of Henle. Our findings therefore suggest a common mechanism for CKD pathogenesis localized to the nephron's loop of Henle, which has previously received little attention. The major risk factors for kidney disease, hypertension and diabetes, are thought to affect the glomerulus primarily, and glomerular damage is typically characterized by albuminuria. Our findings, however, indicate that the association is consistent across strata of hypertension and diabetes, and we observed no association with albuminuria. Thus, our findings provide new insights into CKD pathogenesis and highlight the need to understand the production and functions of Tamm-Horsfall protein within the kidney. The broad CKD definition we chose, including a variety of causes of CKD such as hypertension and diabetes, indicates that the

search for susceptibility variants for complex diseases may not only benefit from a narrower definition of the phenotype, but also from evaluation of a broad phenotype definition in order to identify common disease mechanisms.

Four loci were identified in association with eGFRcrea: the strongest association was for SNP rs12917707 at the *UMOD* locus ( $P$  overall =  $5 \times 10^{-16}$ ; Table 2). The most significant SNP at the second significant eGFRcrea locus was the intronic SNP rs17319721 located in a highly evolutionary conserved region in *SHROOM3* (shroom family member 3) on chromosome 4 ( $P$  overall =  $1 \times 10^{-12}$ ; Fig. 2b and Table 2). The *SHROOM3* gene product is expressed in human kidney and is reported to have a role in epithelial cell shape regulation<sup>22</sup>. The association at the third eGFRcrea locus, the intronic SNP rs6040055 in *JAG1* (jagged 1) on chromosome 20 ( $P = 1 \times 10^{-8}$ ; Table 2 and Supplementary Fig. 2 online), did not replicate ( $P$  overall = 0.006). The finding may therefore be a false positive, although a biological role of *JAG1* in kidney disease is supported by rare *JAG1* mutations causing Alagille syndrome (MIM118450)<sup>23</sup>. Lastly, the intronic SNP rs2467853 in *SPATA5L1* (spermatogenesis associated 5-like 1) at the *GATM-SPATA5L1* locus on chromosome 15 was significantly associated with eGFRcrea ( $P$  overall =  $6 \times 10^{-14}$ ; Fig. 2c and Table 2). *GATM* encodes glycine amidinotransferase, an enzyme involved in creatine biosynthesis. SNPs at this locus are therefore likely related to serum levels of creatinine without influencing susceptibility to kidney disease



**Figure 3** Meta-analysis of the odds of CKD per each additional copy of the minor T allele at *UMOD* rs12917707 across strata of major kidney disease risk factors. Error bars correspond to 95% confidence intervals. Meta-analysis values obtained from the discovery samples.

**Table 2 Results from meta-analyses of top GWAS signals at each locus ( $P < 4 \times 10^{-7}$ ) for CKD, eGFRcrea and eGFRcys in discovery samples and after replication**

Locus				GWAS meta-analysis					GWAS discovery and replication meta-analysis			
SNP	Chr	Position	In (near) gene	Minor / major allele	OE var ratio	MAF	beta / OR	s.e. / 95% CI	$P$	beta / OR	s.e. / 95% CI	$P$
<b>CKD (<math>n = 19,877</math>)</b>												
rs12917707	16	20275191	<i>UMOD</i>	T/G	0.96	0.18	0.76	0.70–0.83	2.8E–09	0.80	0.75–0.85	2.3E–12
<b>eGFRcrea (<math>n = 18,127</math>)</b>												
rs17319721	4	77587871	<i>SHROOM3</i>	A/G	0.96	0.44	–0.014	0.003	9.7E–08	–0.012	0.002	1.2E–12
rs2467853	15	43486085	<i>SPATA5L1</i> , ( <i>GATM</i> )	G/T	0.97	0.38	–0.013	0.003	3.4E–07	–0.013	0.002	6.2E–14
rs12917707	16	20275191	<i>UMOD</i>	T/G	0.96	0.18	0.022	0.003	3.0E–11	0.018	0.002	5.2E–16
rs6040055	20	10581313	<i>JAG1</i>	T/C	0.73	0.39	–0.017	0.003	1.0E–08	–0.005	0.002	5.9E–03
<b>eGFRcys (<math>n = 12,266</math>)</b>												
rs1731274	8	23822264	<i>STC1</i>	G/A	0.96	0.43	–0.017	0.003	4.6E–08	NA	NA	NA
rs12917707	16	20275191	<i>UMOD</i>	T/G	0.96	0.18	0.021	0.004	2.0E–07	NA	NA	NA
rs13038305	20	23558262	<i>CST3-CST9</i>	T/C	0.95	0.21	0.076	0.004	2.2E–88	NA	NA	NA

Meta-analysis  $P$  values are adjusted for the trait- and study-specific genomic-control parameters (see **Supplementary Fig. 1**). For eGFR, betas indicate the change in eGFR per minor allele for the natural logarithmic transformation of eGFR. NA: eGFRcys was not available in the replication samples, but rs1731274 was associated with eGFRcrea after replication at  $P = 2 \times 10^{-7}$ , beta = –0.009, s.e. = 0.002. OE var ratio: sample-size weighted mean of the observed by expected variance ratio for each SNP across the discovery samples; MAF: minor allele frequency, sample-size weighted mean of MAF across the discovery samples. SNP position based on dbSNP129.

(**Table 3**). Although rs2467853 is located in *SPATA5L1*, strong LD extends into the region of the *GATM* gene.

We identified three loci in association with eGFRcys: the strongest association was for the intergenic SNP rs13038305 between *CST3* (cystatin C) and *CST9* (cystatin 9) ( $P = 2.2 \times 10^{-88}$ ; **Fig. 2d** and **Table 2**). SNPs within the cystatin (*CST*) superfamily gene cluster on chromosome 20 have been previously reported as associated with serum cystatin C levels<sup>24</sup>. The genes in the *CST* superfamily encode cystatin proteins. SNPs in these genes likely influence serum levels of cystatin C and therefore estimated eGFRcys, but not true GFR or CKD susceptibility (**Table 3**). Second, we identified the intergenic SNP rs1731274, located 54 kb from the *STC1* (stanniocalcin 1) gene on chromosome 8 ( $P = 4.6 \times 10^{-8}$ ; **Fig. 2e** and **Table 2**). *STC1* encodes stanniocalcin 1, a hormone regulating calcium homeostasis in fish. In mammals, it is highly expressed in the renal nephron and may influence local calcium and phosphate homeostasis through a paracrine mechanism<sup>25</sup>. A recent study in *STC1* transgenic mice reported *STC1* as a renal protective protein with a potent anti-inflammatory role<sup>26</sup>. As the replication samples did not have cystatin C measurements available, we explored the association of rs1731274 with eGFRcrea across the discovery and replication samples ( $P = 2 \times 10^{-7}$ , **Table 2**). Finally, rs12917707 at the *UMOD* locus was associated with eGFRcys at  $P = 2 \times 10^{-7}$ .

**Table 3** presents the association of all genome-wide significant SNPs across the three renal traits. SNPs in *UMOD*, *SHROOM3* and *STC1* showed direction-consistent association across traits. For example, rs12917707 at *UMOD* was associated with both higher eGFRcrea and eGFRcys representing better kidney function, and with lower odds of CKD, conferring disease protection. SNPs at the *GATM-SPATA5L1* and *CST* regions were only associated with the respective discovery trait; the association of rs2467853 at the

*GATM-SPATA5L1* locus with CKD likely results from the eGFRcrea-based definition of CKD. All SNPs associated with CKD, eGFRcrea and eGFRcys at  $P < 4 \times 10^{-7}$  are listed in **Supplementary Tables 4, 5 and 6** online, respectively.

Together, loci for eGFRcrea explain 0.7% of the eGFRcrea variance (0.43% without the *GATM* locus), and loci for eGFRcys explain 3.2% of the eGFRcys variance (0.24% without the *CST* locus), suggesting that additional yet-undiscovered genetic variants influence variability in renal function. In accordance with small absolute differences observed in other GWAS for continuous human traits, the multi-variable adjusted eGFR difference across genotypes for any one locus was small. As risk alleles may act in an additive fashion, we created a risk score for each individual as the sum of risk alleles at *UMOD*, *SHROOM3* and *STC1*. These analyses were conducted in the ARIC study, the largest individual study contributing data and with available prospective information. The mean eGFRcrea was 10 ml/min/1.73 m<sup>2</sup> lower in individuals with all six risk alleles across the three loci

**Table 3 Association of significant SNPs ( $P < 5 \times 10^{-8}$ ) across indices of renal function and CKD in up to 19,877 participants of the ARIC, CHS, FHS and RS studies**

Gene		<i>UMOD</i>	<i>SHROOM3</i>	<i>SPATA5L1-GATM</i>	<i>STC1</i>	<i>CST3-CST9</i>
		rs12917707	rs17319721	rs2467853	rs1731274	rs13038305
CKD	OR	0.76	1.07	1.13	1.06	0.94
	95% CI	(0.70–0.83)	(1.00–1.15)	(1.06–1.21)	(0.99–1.13)	(0.87–1.03)
	$P$	2.8E–09	0.04	3.9E–04	0.09	0.18
eGFRcrea	Beta	0.022	–0.014	–0.013	–0.009	0.004
	s.e.	0.003	0.003	0.003	0.003	0.003
	$P$	3.0E–11	9.7E–08	3.4E–07	1.8E–04	0.22
eGFRcys	Beta	0.021	–0.013	–0.0001	–0.017	0.076
	s.e.	0.004	0.003	0.003	0.003	0.004
	$P$	2.0E–07	4.2E–05	0.95	4.6E–08	2.2E–88

Values for the discovery trait for each SNP are shaded in light right gray; values correspond to **Table 2**.

compared to those with no risk alleles ( $P = 2 \times 10^{-8}$  per unit score increase). CKD prevalence ranged from 0% in those without any risk alleles to 12.1% in individuals carrying all six risk alleles.

In summary, we have identified and validated common variants at several previously unknown loci conferring susceptibility for kidney dysfunction and CKD in large, unselected population-based studies. Studies to understand the production and functions of Tamm-Horsfall protein are warranted and may eventually lead to novel prevention and intervention options to reduce CKD risk.

## METHODS

**Study samples.** Four large, population-based cohorts of the CHARGE consortium had GWAS data available and formed the discovery sample: ARIC, CHS, FHS and RS. Detailed information about these cohorts, including the design papers, is provided in the **Supplementary Methods**. Briefly, the studies were initiated to study cardiovascular disease and its risk factors and diseases related to aging. The population-based ARIC cohort recruited 15,792 middle-aged participants from 1987–1989 in four US communities. The population-based CHS cohort recruited 5,888 participants aged 65 years and older from 1989–1990 and 1992–1993 in four US communities. The FHS is a community-based study with a family component, including the Original ( $n = 5,209$ , recruited 1948) and Offspring ( $n = 5,214$ , recruited 1971) components. The community-based RS recruited 7,983 participants aged 55 years or older from 1990–1993. Two independent study samples were used to replicate results. In AGES, 5,764 survivors of the Reykjavik Study were examined from 2002 to 2006 and contributed to information. The WGHS is a sample drawn in 2006 from the original Women's Health Study. Each participant provided written informed consent, and institutional review boards of the participating institutions approved the study protocols. African-American participants from ARIC and CHS did not contribute information to the present study.

**Genotyping and imputation.** Detailed information about genotyping and imputation methods is provided in **Supplementary Table 1**, and details about data cleaning are provided in the **Supplementary Methods**. Briefly, all studies directly genotyped between 300,000 and 900,000 SNPs using whole-genome genotyping arrays by either Affymetrix (6.0, ARIC; 500K and 50K gene-centric, FHS) or Illumina (Human CNV370, AGES and CHS; 550K, RS; Human-Hap300 Duo-Plus or a combination of HumanHap300 and iSelect, WGHS). All genotyping was carried out according to the manufacturer's instructions between 2006–2008. Using the Phase II CEU HapMap individuals as a reference panel, genotypes were imputed to a common set of  $\sim 2.5$  million high-quality HapMap SNPs. Software used for imputation were BimBam v0.99 (ref. 27, CHS) and MACH v1.0.15/16 (all others, see MACH URL below); FHS accounted for relatedness of participants. Imputed genotypes were expressed as an allelic dosage, a fractional value between 0 and 2. The WGHS did not impute genotypes.

**eGFRcrea, eGFRcys and CKD.** Serum creatinine was measured using a modified kinetic Jaffe reaction in all studies but AGES, where an enzymatic method was used. eGFRcrea was calculated using the Modification of Diet in Renal Disease (MDRD) Study equation<sup>14</sup>:  $eGFR_{crea} \text{ (ml/min/1.73 m}^2\text{)} = 186.3 \times \text{serum creatinine (mg/dl)}^{-1.154} \times \text{age}^{-0.203} \times 0.742$  (if female). To be comparable across studies, creatinine values in all studies were calibrated using regression to age- and sex-adjusted mean values from a nationally representative US survey as described previously<sup>28</sup>. Cystatin C was measured by a particle-enhanced immunonephelometric assay (N Latex Cystatin C, Dade Behring) at ARIC visit 4, CHS baseline exam, and FHS offspring exam 7 with a nephelometer (BNII, Dade-Behring). eGFRcys was then calculated using the formula  $eGFR_{cys} = 76.7 \times (\text{serum cystatin C})^{-1.19}$  (ref. 15). CKD was defined as  $eGFR_{crea} < 60 \text{ ml/min/1.73 m}^2$  according to National Kidney Foundation guidelines<sup>17</sup>.

CKD in CHS, RS, WGHS and AGES was defined on the basis of a single measurement of serum creatinine at the baseline visit. FHS and ARIC used a cumulative definition of CKD based on serum creatinine measurements at several study visits as detailed in the **Supplementary Methods**. Incident

CKD in ARIC was defined as  $eGFR_{crea} < 60 \text{ ml/min/1.73 m}^2$  at study visits 2 or 4 in individuals with  $eGFR_{crea} \geq 60 \text{ ml/min/1.73 m}^2$  at study visit 1, or a kidney-disease specific ICD code on a hospital discharge record or death certificate from study inception in 1987 through January 1, 2005 (ref. 29).

Information on age and sex was collected at each study visit, and ancestry was self-reported. Potential population stratification was assessed as detailed in the **Supplementary Methods**.

**Statistical analysis.** GWAS was conducted within each cohort for eGFRcrea, eGFRcys and CKD, followed by meta-analysis of the study-specific associations for each trait. SNPs showing genome-wide significant association with any of the three traits in meta-analyses were then explored for their association with the other two traits.

The phenotype for the eGFR analyses in all studies was created by calculating a natural logarithmic transformation of eGFR obtained from the respective equations for eGFRcrea and eGFRcys. All studies but CHS then created sex-specific age- and study-site (ARIC) or cohort (FHS) adjusted residuals. CHS adjusted for age, sex and study site in multivariable regression models. Incident CKD in ARIC was analyzed using multivariable-adjusted Cox proportional hazards regression. Software packages used by the individual studies to conduct linear and logistic regression are listed in **Supplementary Table 1**. FHS accounted for the relatedness of individuals in the analyses as detailed in the **Supplementary Methods**. Pedigree correlations were adjusted using the robust variance option. All studies used an additive genetic model.

Meta-analysis was conducted using inverse-variance weighting as implemented in METAL (see URLs section below). Prior to meta-analysis, the genomic control parameter was calculated within each study for each trait to assess potential inflation of the test statistics. If the parameter was larger than 1, an adjustment was carried out. Only SNPs with minor allele frequency (MAF)  $\geq 2\%$  were analyzed based on the number of CKD cases, corresponding to approximately 50 carriers of the minor allele with CKD. Statistical heterogeneity was evaluated using Cochran's  $\chi^2$  test (Q-test).

The most significant SNPs at genomic loci with evidence of suggestive association ( $P < 4 \times 10^{-7}$ ) for any of the traits were tested for replication in the independent samples. This threshold corresponds to 1/2.5 million tests conducted and corresponds to one or less expected false-positive findings<sup>30</sup>. A threshold of  $P < 5 \times 10^{-8}$  was used to indicate genome-wide significance, corresponding to a Bonferroni correction for the estimated 1 million common independent SNPs across the genome (0.05/1 million). The SNAP program with the HapMap CEU sample as a reference was used to identify the best proxy in the WGHS dataset, to evaluate LD to nearby coding SNPs, and to evaluate LD between imputed SNPs and proxy SNPs that were directly genotyped.

**URLs.** MACH, <http://www.sph.umich.edu/csg/abecasis/MACH/>; METAL, <http://www.sph.umich.edu/csg/abecasis/Metal/index.html>; SNAP, <http://www.broad.mit.edu/mpg/snap/ldsearch.php>.

*Note: Supplementary information is available on the Nature Genetics website.*

## ACKNOWLEDGMENTS

We are indebted to the staff and participants of the AGES Reykjavik Study, the ARIC Study, the CHS Study, the FHS Study, the Rotterdam Study and the WGHS Study for their contributions.

**AGES:** The Age, Gene/Environment Susceptibility Reykjavik Study has been funded by US National Institutes of Health contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association) and the Althingi (the Icelandic Parliament). We thank T. Aspelund and G. Eiriksdottir for their contribution to collecting, analyzing and preparing the AGES Reykjavik Study data.

**ARIC:** The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number

UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. A.K. was supported by a German Research Foundation Fellowship. W.H.L.K. was supported by K01DK067207.

**CHS:** The CHS research reported in this article was supported by contract numbers N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01-HC-15103, N01-HC-55222, N01-HC-75150, N01-HC-45133, grant numbers U01 HL080295 and R01 HL087652, and R01 AG027002 from the National Heart, Lung, and Blood Institute, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal CHS investigators and institutions can be found at <http://www.chs-nhlbi.org/pi.htm>. DNA handling and genotyping was supported in part by National Center for Research Resources grant M01RR00069 to the Cedars-Sinai General Clinical Research Center Genotyping core and National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

**FHS:** This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart, Lung, and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study (contract no. N01-HC-25195) and its contract with Affymetrix for genotyping services (contract no. N02-HL-6-4278). A portion of this research used the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

**RS:** The Rotterdam Study is supported by the Erasmus Medical Center and Erasmus University Rotterdam; the Netherlands Organization for Scientific Research; the Netherlands Organization for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly; The Netherlands Heart Foundation; the Ministry of Education, Culture and Science; the Ministry of Health Welfare and Sports; the European Commission; and the Municipality of Rotterdam. Support for genotyping was provided by the Netherlands Organization for Scientific Research (NWO) (175.010.2005.011, 911.03.012) and Research Institute for Diseases in the Elderly (RIDE). This study was further supported by the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. We thank P. Arp, M. Jhamai, M. Moorhouse, M. Verkerk and S. Bervoets for their help in creating the Rotterdam database and M. Struchalin for his contributions to the imputations of the Rotterdam data.

**WGHS:** The WGHS was supported by the National Heart, Lung, and Blood Institute (HL 043851) and the National Cancer Institute (CA 047988). Collaborative scientific and genotyping support was provided by Amgen.

#### AUTHOR CONTRIBUTIONS

A.K., N.L.G., A.D., S.-J.H., Q.Y., I.H.d.B., T.L., D.S., D.L., V.G., J.C., M.G.S. and C.S.F. contributed to the design of this analysis. A.K., N.L.G., A.D., S.-J.H., Q.Y., E.B., I.R., I.H.d.B., T.L., D.S., D.L., A.V.S., V.G., W.H.L.K., J.C.W., J.C., M.G.S. and C.S.F. contributed to the interpretation of the results. A.K., N.L.G., A.D., S.-J.H., J.C., M.G.S. and C.S.F. drafted the manuscript; all others reviewed and commented on the manuscript. A.K., N.L.G., A.D., S.-J.H., R.K., M.L., Q.Y., D.E.A., G.B.E., I.R., R.B.S., T.L., F.R., C.M.v.D., A.V.S., D.I.C., G.P. and J.C. contributed to statistical methods and analysis. D.S., E.J.B., D.L., V.G., J.C. and C.S.F. contributed to recruitment and follow up of subjects. E.B., Y.-D.I.C., T.H., F.R. and A.G.U. contributed to genotyping; M.L., Q.Y., G.B.E., Y.S.A., F.R., A.G.U., A.V.S. and G.P. contributed to bioinformatics. V.G., L.J.L., T.B.H. and A.V.S. participated in the AGES Study. A.K., M.L., W.H.L.K., J.C., D.E.A., B.C.A., E.B., G.B.E., I.R. and R.B.S. participated in the ARIC Study. N.L.G., M.G.S., Y.-D.I.C., I.H.d.B., T.H., T.L., M.S. and D.S. participated in the CHS Study. E.J.B., D.L., A.U., S.-J.H., Q.Y. and C.S.F. participated in the FHS Study. Y.S.A., A.H., F.R., A.G.U., C.M.v.D., A.D. and J.C.W. participated in the RS Study. D.I.C., G.P. and P.M.R. participated in the WGHS Study.

Published online at <http://www.nature.com/naturegenetics/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. Levey, A.S. *et al.* Chronic kidney disease as a global public health problem: approaches and initiatives — a position statement from Kidney Disease Improving Global Outcomes. *Kidney Int.* **72**, 247–259 (2007).

2. Serafini-Cessi, F., Malagolini, N. & Cavallone, D. Tamm-Horsfall glycoprotein: biology and clinical relevance. *Am. J. Kidney Dis.* **42**, 658–676 (2003).
3. Hart, T.C. *et al.* Mutations of the *UMOD* gene are responsible for medullary cystic kidney disease 2 and familial juvenile hyperuricaemic nephropathy. *J. Med. Genet.* **39**, 882–892 (2002).
4. Coresh, J. *et al.* Prevalence of chronic kidney disease in the United States. *J. Am. Med. Assoc.* **298**, 2038–2047 (2007).
5. Zhang, Q.L. & Rothenbacher, D. Prevalence of chronic kidney disease in population-based studies: systematic review. *BMC Public Health* **8**, 117 (2008).
6. National Institutes of Health. *USRDS 2008 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States* (National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland, 2008).
7. Sarnak, M.J. *et al.* Kidney disease as a risk factor for development of cardiovascular disease: a statement from the American Heart Association councils on kidney in cardiovascular disease, high blood pressure research, clinical cardiology, and epidemiology and prevention. *Circulation* **108**, 2154–2169 (2003).
8. Weiner, D.E. *et al.* Chronic kidney disease as a risk factor for cardiovascular disease and all-cause mortality: a pooled analysis of community-based studies. *J. Am. Soc. Nephrol.* **15**, 1307–1315 (2004).
9. Langefeld, C.D. *et al.* Heritability of GFR and albuminuria in Caucasians with type 2 diabetes mellitus. *Am. J. Kidney Dis.* **43**, 796–800 (2004).
10. Bochud, M. *et al.* Heritability of renal function in hypertensive families of African descent in the Seychelles (Indian Ocean). *Kidney Int.* **67**, 61–69 (2005).
11. Fox, C.S. *et al.* Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: the Framingham Heart Study. *J. Am. Soc. Nephrol.* **15**, 2457–2461 (2004).
12. Freedman, B.I. & Satko, S.G. Genes and renal disease. *Curr. Opin. Nephrol. Hypertens.* **9**, 273–277 (2000).
13. Psaty, B.M. *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from five cohorts. *Circulation: Cardiovascular Genetics* **2**, 73–80 (2009).
14. Levey, A.S. *et al.* A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann. Intern. Med.* **130**, 461–470 (1999).
15. Stevens, L.A. *et al.* Estimating GFR using serum cystatin C alone and in combination with serum creatinine: a pooled analysis of 3,418 individuals with CKD. *Am. J. Kidney Dis.* **51**, 395–406 (2008).
16. Stevens, L.A., Coresh, J., Greene, T. & Levey, A.S. Assessing kidney function—measured and estimated glomerular filtration rate. *N. Engl. J. Med.* **354**, 2473–2483 (2006).
17. National Kidney Foundation. *K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification.* *Am. J. Kidney Dis.* **39**, S1–S266 (2002).
18. Vylet' al, P. *et al.* Alterations of uromodulin biology: a common denominator of the genetically heterogeneous FJHN/MCKD syndrome. *Kidney Int.* **70**, 1155–1169 (2006).
19. Rampoldi, L. *et al.* Allelism of MCKD, FJHN and GCKD caused by impairment of uromodulin export dynamics. *Hum. Mol. Genet.* **12**, 3369–3384 (2003).
20. Bachmann, S. *et al.* Renal effects of Tamm-Horsfall protein (uromodulin) deficiency in mice. *Am. J. Physiol. Renal Physiol.* **288**, F559–F567 (2005).
21. Benetti, E. *et al.* Immature renal structures associated with a novel *UMOD* sequence variant. *Am. J. Kidney Dis.* **53**, 327–331 (2009).
22. Nishimura, T. & Takeichi, M. Shroom3-mediated recruitment of Rho kinases to the apical cell junctions regulates epithelial and neuroepithelial planar remodeling. *Development* **135**, 1493–1502 (2008).
23. Anad, F. *et al.* Alagille syndrome and deletion of 20p. *J. Med. Genet.* **27**, 729–737 (1990).
24. Hwang, S.J., Yang, Q., Meigs, J.B., Pearce, E.N. & Fox, C.S. A genome-wide association for kidney function and endocrine-related traits in the NHLBI's Framingham Heart Study. *BMC Med. Genet.* **8**(Suppl. 1), S10 (2007).
25. Yoshiko, Y. & Aubin, J.E. Stanniocalcin 1 as a pleiotropic factor in mammals. *Peptides* **25**, 1663–1669 (2004).
26. Huang, L. *et al.* Anti-inflammatory and renal protective actions of stanniocalcin-1 in a model of anti-glomerular basement membrane glomerulonephritis. *Am. J. Pathol.* **174**, 1368–1378 (2009).
27. Servin, B. & Stephens, M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet.* **3**, e114 (2007).
28. Coresh, J. *et al.* Calibration and random variation of the serum creatinine assay as critical elements of using equations to estimate glomerular filtration rate. *Am. J. Kidney Dis.* **39**, 920–929 (2002).
29. Hsu, C.C. *et al.* Apolipoprotein E and progression of chronic kidney disease. *J. Am. Med. Assoc.* **293**, 2892–2899 (2005).
30. Gordon, A., Glazko, G., Qui, X. & Yakovlev, A. Control of the mean number of false discoveries, Bonferroni and stability of multiple testing. *The Annals of Applied Statistics* **1**, 179–190 (2007).