

A catalog of genetic loci associated with kidney function from analyses of a million individuals

Chronic kidney disease (CKD) is responsible for a public health burden with multi-systemic complications. Through trans-ancestry meta-analysis of genome-wide association studies of estimated glomerular filtration rate (eGFR) and independent replication ($n = 1,046,070$), we identified 264 associated loci (166 new). Of these, 147 were likely to be relevant for kidney function on the basis of associations with the alternative kidney function marker blood urea nitrogen ($n = 416,178$). Pathway and enrichment analyses, including mouse models with renal phenotypes, support the kidney as the main target organ. A genetic risk score for lower eGFR was associated with clinically diagnosed CKD in 452,264 independent individuals. Colocalization analyses of associations with eGFR among 783,978 European-ancestry individuals and gene expression across 46 human tissues, including tubulo-interstitial and glomerular kidney compartments, identified 17 genes differentially expressed in kidney. Fine-mapping highlighted missense driver variants in 11 genes and kidney-specific regulatory variants. These results provide a comprehensive priority list of molecular targets for translational research.

CKD is a major public health issue, with increasing incidence and prevalence worldwide¹. Its associated burden of disease encompasses metabolic disturbances, end-stage kidney disease and multi-systemic complications such as cardiovascular disease^{1–4}. CKD is a leading cause of death⁵ and has shown one of the highest increases in disease-attributable mortality over the last decade². Nevertheless, public and clinical awareness remain low³. Moreover, clinical trials in nephrology are still under-represented⁶, which has resulted in a scarcity of therapeutic options to alter disease progression and high costs for health systems⁷. A major barrier to developing new therapeutics is the limited understanding of the mechanisms underlying kidney function in health and disease, with the consequent lack of therapeutic targets.

Genome-wide association studies (GWAS) and exome-chip studies of the glomerular filtration rate estimated from serum creatinine (eGFR), the main biomarker to quantify kidney function and define CKD, have identified nearly 100 eGFR-associated genetic loci⁸ in samples of European^{9–15}, Asian^{16–19} and multiple²⁰ ancestry. However, similarly to other complex traits and diseases, identifying causal genes and molecular mechanisms implicated by genetic associations is challenging and has only been successful for a few kidney-function-associated loci^{21,22}. Advanced statistical fine-mapping approaches and newly emerging multi-tissue gene expression data provide new opportunities for prioritizing putative causal variants, effector genes and target tissues from the results of large-scale GWAS meta-analyses.

We therefore conducted a trans-ancestry GWAS meta-analysis in the CKD Genetics (CKDGen) Consortium ($n = 765,348$) and replicated findings in the Million Veteran Program (MVP; $n = 280,722$)²³, for a combined sample size of greater than 1 million participants. The first aim of this study was to identify new globally important loci for kidney function through maximizing statistical power (Supplementary Fig. 1). Results from GWAS of the complementary kidney function marker blood urea nitrogen (BUN; $n = 416,178$) were used to prioritize the eGFR-associated loci on the basis of those most likely to be relevant for kidney function. A genetic risk score (GRS) for low eGFR was used to test relevance for clinically diagnosed CKD among 452,264 independent individuals. The second aim was to characterize replicated eGFR-associated

loci through complementary computational approaches, including various enrichment and network analyses, fine-mapping, and colocalization with gene expression in 46 tissues and protein levels (Supplementary Fig. 1). We focused this aim on European-ancestry individuals, as fine-mapping based on summary statistics requires linkage disequilibrium (LD) reference panels whose sample size scales with that of the GWAS²⁴. The resulting list of prioritized variants and genes provides a rich resource of potential therapeutic targets to improve CKD treatment and prevention.

Results

Discovery trans-ancestry meta-analysis. We performed 121 GWAS encompassing 765,348 individuals of European ($n = 567,460$), East Asian ($n = 165,726$), African-American ($n = 13,842$), South Asian ($n = 13,359$) and Hispanic ($n = 4,961$) ancestry (median age, 54 years; 50% female; Supplementary Table 1). The median of the study-specific mean eGFR values was $89 \text{ ml min}^{-1} \text{ per } 1.73 \text{ m}^2$ (interquartile range, IQR: 81, 94). GWAS were based on genotypes imputed from Haplotype Reference Consortium²⁵ or 1000 Genomes Project²⁶ reference panels (Methods and Supplementary Table 2). Following study-specific variant filtering and quality-control procedures, we performed a fixed-effects inverse-variance-weighted meta-analysis, finding no evidence of unmodeled population structure (LD score regression intercept = 1.04; genomic control factor $\lambda_{\text{GC}} = 1.05$). After variant filtering, 8,221,591 SNPs were used for downstream analysis (Methods).

We discovered 308 loci containing at least one eGFR-associated SNP at genome-wide significance (Methods), of which 200 were new and 108 contained an index SNP reported by previous GWAS of eGFR (Fig. 1 and Supplementary Table 3). Regional association plots are shown in Supplementary Fig. 2. The minor alleles across index SNPs showed both decreasing and increasing effects on eGFR, with larger effects observed for lower-frequency SNPs (Fig. 1, inset). The 308 index SNPs explained 7.1% of the eGFR variance, nearly doubling recent GWAS-based estimates⁹, and 19.6% of eGFR genetic heritability ($h^2 = 39\%$, 95% credible interval = 32%, 47%), estimated in a participating general-population-based pedigree study (Methods and Supplementary Fig. 3). The effects of index SNPs were largely homogeneous across studies (Fig. 2a and

Supplementary Table 3) and ancestry groups (Supplementary Table 4 and Supplementary Note 1).

Replication and meta-analysis of more than 1 million individuals. We assessed replication in an independent trans-ancestry GWAS meta-analysis of eGFR performed among 280,722 MVP participants²³. Effect estimates, available for 305 of the 308 SNPs, showed almost perfect directional consistency (302/305 SNPs, 99%) and very strong correlation with the discovery results (Fig. 2b). For these 305 SNPs, we performed a meta-analysis of the 1,046,070 discovery and replication samples. Replication was met by 262 SNPs (Fig. 1, Methods and Supplementary Table 3). Of the three SNPs not available in MVP, the index SNPs at *SHROOM3* ($P=3.5 \times 10^{-120}$) and *SH3YL1* ($P=1.2 \times 10^{-11}$) were also considered to be replicated on the basis of previous evidence^{15,27}, resulting in a total of 264 replicated SNPs (166 new). Of these, 74 SNPs were genome-wide significant in MVP alone (Supplementary Table 3).

Association of eGFR loci with BUN and CKD. To evaluate whether associations with creatinine-based eGFR were probably related to kidney function or potentially to creatinine metabolism, we assessed the association of the 264 eGFR-associated index SNPs with BUN, an alternative marker of kidney function that is inversely correlated with eGFR. Trans-ancestry meta-analysis of 65 GWAS for BUN ($n=416,178$; Supplementary Table 1) showed no evidence of unmodeled population structure ($\lambda_{GC}=1.03$; LD score regression intercept=0.98) and yielded 111 genome-wide-significant loci (15 known, 96 new; Supplementary Fig. 4 and Supplementary Table 5).

Of the 264 replicated eGFR index SNPs, 34 and 146 showed genome-wide-significant and nominally significant ($P<0.05$) association with BUN, respectively (Supplementary Table 6). SNP effects were inversely correlated ($r=-0.65$; Fig. 2c). Relevance to kidney function was classified as 'likely' for 147 eGFR index SNPs with inverse, significant associations with BUN (one-sided $P<0.05$); 'inconclusive' for 102 eGFR index SNPs not associated with BUN ($P \geq 0.05$); and 'unlikely' for 15 eGFR index SNPs showing concordant, significant association with BUN (one-sided $P<0.05$; Supplementary Table 6). This comparative analysis of complementary biomarkers supports the idea that signals at the majority of eGFR-associated loci probably reflect kidney function.

Next, we investigated the effects of the eGFR index SNPs on CKD in CKDGen studies ($n=625,219$, including 64,164 CKD cases; Methods). GWAS meta-analysis of CKD identified 23 genome-wide-significant loci, including 17 likely relevant for kidney function (*SDCCAG8*, *LARP4B*, *DCDC1*, *WDR72*, *UMOD-*PDILT**, *MYO19*, *AQP4*, *NFATC1*, *PSD4*, *HOXD8*, *NRIP1*, *SHROOM3*, *FGF5*, *SLC34A1*, *DAB2*, *UNCX* and *PRKAG2*; Supplementary Table 6). The majority of replicated eGFR index SNPs (224 of 264) were associated with CKD (one-sided $P<0.05$; Fig. 1, inset), including 130 likely relevant for kidney function (Supplementary Table 6).

Finally, we tested whether a GRS based on the combined effect of the 147 eGFR index SNPs likely relevant for kidney function was associated with clinically diagnosed CKD and CKD-related outcomes in the UK Biobank ($n=452,264$; Methods). A lower GRS, reflecting genetically lower eGFR, was associated with higher odds ratios (ORs) of chronic renal failure, glomerular diseases, acute renal failure and hypertensive diseases (Fig. 2d and Supplementary Fig. 5). The OR of chronic renal failure per 10% lower GRS-predicted eGFR was 2.13 (95% CI=1.90, 2.39; $P=8.1 \times 10^{-38}$). A significant protective association with urolithiasis may reflect a reduced ability to concentrate urine at lower eGFR.

Genetic correlations of eGFR and BUN with other phenotypes.

We assessed genome-wide genetic correlations (r_g) of eGFR associations with each of 748 complex traits and diseases (Methods)²⁸. We observed 37 significant correlations ($P<6.7 \times 10^{-5}=0.05/748$;

Supplementary Fig. 6 and Supplementary Table 7). After serum creatinine, the largest negative correlations were observed between eGFR and serum citrate ($r_g=-0.27$) and urate ($r_g=-0.23$), followed by anthropometric traits including lean mass and physical fitness (for example, $r_g=-0.20$ for left hand grip strength). While the inverse correlation with muscle-mass-related traits probably reflects higher creatinine generation leading to lower creatinine-based eGFR, the correlations with citrate and urate levels probably reflect reduced filtration function, as does the positive correlation with GFR estimated from cystatin C ($r_g=0.53$).

A very similar pattern of genetic correlations was observed for BUN (Supplementary Table 7), but the genetic correlations with muscle-mass-related traits were generally lower than for eGFR. The largest genetic correlation for BUN was observed with CKD ($r_g=0.47$), as compared to creatinine-based ($r_g=-0.29$) and cystatin C-based ($r_g=-0.26$) eGFR.

In summary, significant genetic correlations with eGFR reflect the two biological components that govern serum creatinine concentrations: its excretion via the kidney and its generation in muscle. The fact that genetic correlations between BUN and muscle-mass-related traits are generally lower than was observed for eGFR underscores the value of using genetic associations with BUN to help prioritize eGFR-associated loci most likely to be relevant for kidney function.

Functional enrichment and pathway analyses. To identify molecular mechanisms and tissues of importance for kidney function, we assessed the enrichment of the eGFR and BUN genetic associations by using tissue-specific gene expression, regulatory annotations, and gene sets and pathways (Methods). First, we used eGFR-associated SNPs ($P<5 \times 10^{-8}$) to explore enriched pathways, tissues and cell types on the basis of gene expression data with DEPICT²⁹. We identified 16 significantly enriched physiological systems, cell types and tissues highlighting several aspects of kidney function, physiology and disease. The strongest enrichment was observed for urogenital and renal physiological systems and tissues (kidney, kidney cortex and urinary tract; false-discovery rate (FDR) <0.05 ; Supplementary Fig. 7a,b). Pathway and gene set enrichment analysis identified three highly correlated and strongly associated meta gene sets ($P<1 \times 10^{-6}$, FDR <0.05), including some relevant to the kidney such as polyuria, dilated renal tubules and expanded mesangial matrix, as well as signaling and transcription, and energy metabolism (Supplementary Fig. 7c). Tissue and cell-type enrichment analysis of BUN-associated SNPs associated at $P<5 \times 10^{-8}$ highlighted a very similar pattern (Supplementary Fig. 8) but without enrichment for muscle tissues, further supporting the use of BUN to prioritize the loci most likely to be related to kidney function.

Second, we used stratified LD score regression³⁰ on the genome-wide eGFR and BUN summary statistics to identify cell-type groups with enriched heritability on the basis of data from diverse cell-type-specific functional genomic elements. The strongest enrichment for eGFR was observed for the kidney (13.2-fold), followed by the liver (7.3-fold) and adrenal/pancreas (5.7-fold enrichment; Supplementary Table 8). The kidney was also the most enriched cell-type group for BUN (11.5-fold enrichment; Supplementary Table 8).

Finally, by using a complementary approach, we assessed enrichment of eGFR-associated variants in genes in which disruption results in kidney phenotypes in genetically manipulated mice³¹. From the Mouse Genome Informatics (MGI) database, we selected all genes for which disruption causes abnormal GFR ($n=24$), abnormal kidney physiology ($n=453$) or abnormal kidney morphology ($n=764$) and interrogated their human orthologs in the eGFR summary statistics (Methods). We identified significant associations in ten genes linked to abnormal GFR in mice (enrichment $P=8.9 \times 10^{-4}$), 55 linked to abnormal kidney physiology

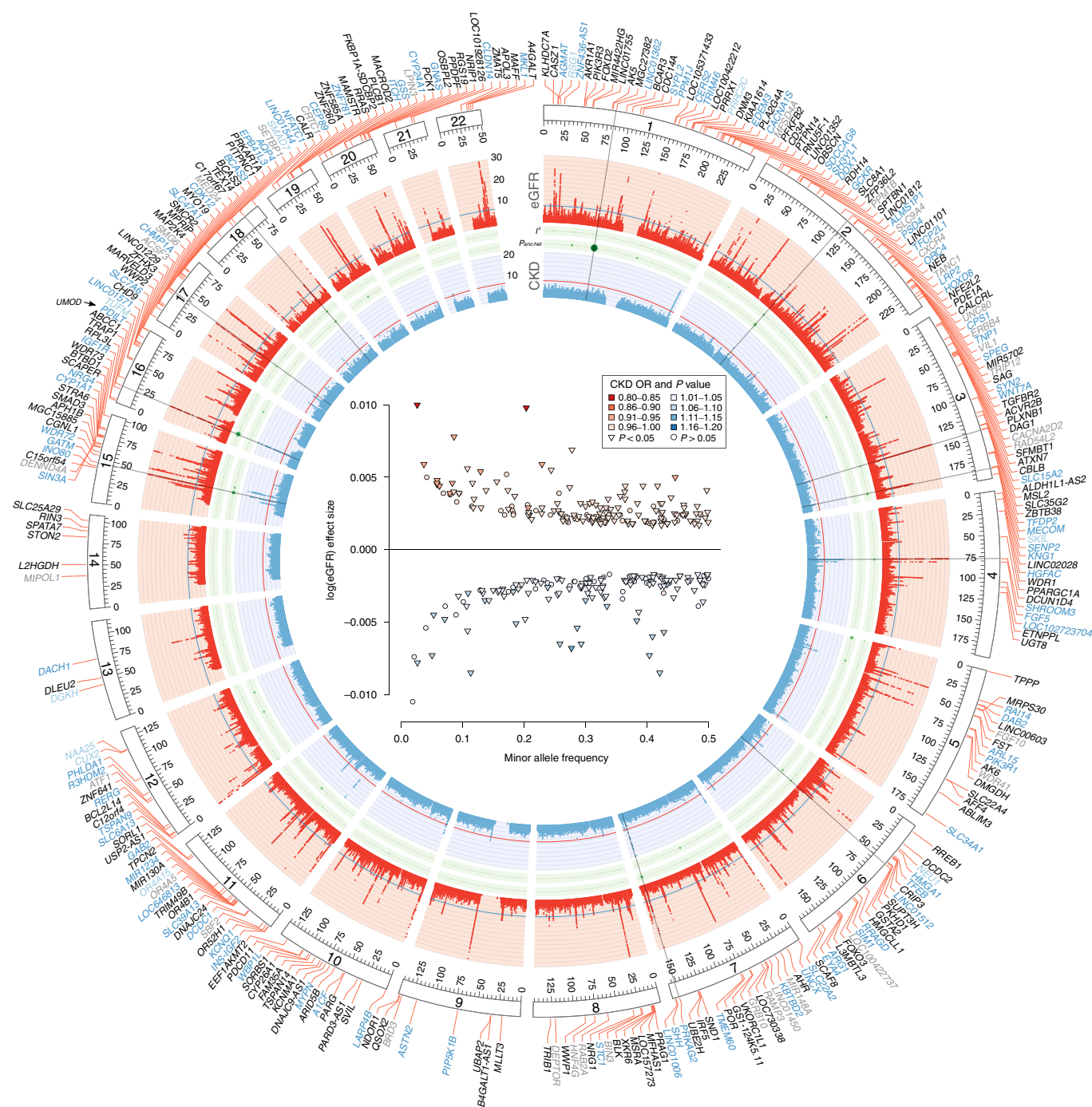


Fig. 1 | Trans-ancestry GWAS meta-analysis identifies 308 loci associated with eGFR. Circos plot. The red band corresponds to $-\log_{10}(P)$ for association with eGFR (y axis truncated at 30), by chromosomal position. The blue line indicates genome-wide significance ($P=5 \times 10^{-8}$). Black gene labels indicate new loci, while blue labels indicate known loci. Non-replicating loci are colored in gray (new) or light blue (known). The green band corresponds to measures of heterogeneity related to the index SNPs associated with eGFR. Dot sizes are proportional to I^2 or ancestry-related heterogeneity ($P_{\text{anc-het}}$). The blue band corresponds to $-\log_{10}(P)$ for association with CKD (y axis truncated at 20), by chromosomal position. The red line indicates genome-wide significance ($P=5 \times 10^{-8}$). Radial lines mark regions with $P_{\text{anc-het}} < 1.6 \times 10^{-4} = 0.05/308$ or $I^2 > 25\%$. Inset, effects of all 308 index SNPs on log(eGFR) by minor allele frequency, colored by the associated OR for CKD (red scale for OR ≤ 1 , blue scale for OR > 1). The largest effects on CKD were observed for rs77924615 at *UMOD-PCITL* (OR=0.81, 95% confidence interval (CI)=0.80, 0.83), rs187355703 at *HOXD8* (OR=0.82, 95% CI=0.77, 0.87) and rs10254101 at *PRKAG2* (OR=1.11, 95% CI=1.09, 1.11). Triangles highlight SNPs that were associated with CKD (one-sided $P < 0.05$).

(enrichment $P=1.1 \times 10^{-4}$) and 96 linked to abnormal kidney morphology (enrichment $P=1.8 \times 10^{-5}$; Fig. 3 and Methods). Of these, 25 genes represent new eGFR candidate genes in humans; that is, they have not previously been reported to contain genome-wide-significant eGFR-associated SNPs or map near known loci

(Supplementary Table 9). The existing mouse models may pave the way for experimental confirmation of these findings.

Fine-mapping and secondary signal analysis in European-ancestry individuals. Conditional and fine-mapping analyses were

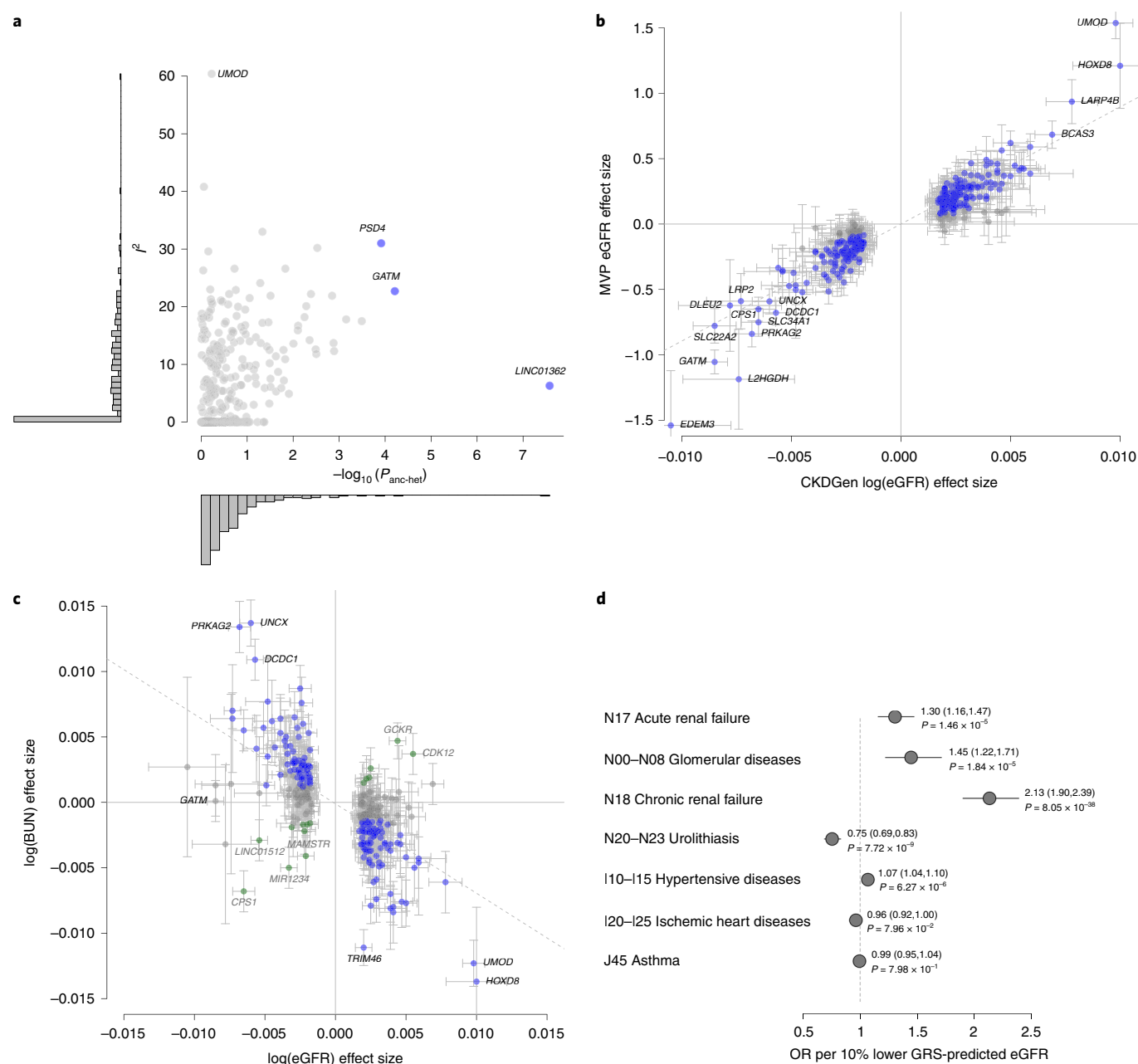


Fig. 2 | Generalizability with respect to other populations and other kidney function markers. a, Measures of heterogeneity for the 308 eGFR-associated index SNPs. Each variant's heterogeneity quantified as I^2 from the trans-ancestry meta-analysis (y axis) is compared to the ancestry-related heterogeneity from meta-regression ($-\log_{10}(P_{\text{anc-het}})$; x axis). Histograms summarize the distribution of the heterogeneity measures on both axes. SNPs with ancestry-related heterogeneity ($P_{\text{anc-het}} < 1.6 \times 10^{-4} = 0.05/308$) are marked in blue and labeled; SNPs with $I^2 > 50\%$ are labeled. **b**, Comparison of genetic effect estimates between CKDGen Consortium discovery (x axis) and MVP replication (y axis). Blue font indicates one-sided $P < 0.05$ in the MVP. Error bars correspond to 95% CIs. The dashed line corresponds to the line of best fit. Pearson's correlation coefficient $r = 0.92$ (95% CI = 0.90, 0.94). **c**, The magnitude of genetic effects on eGFR (x axis) as compared to BUN (y axis) for the 264 replicated eGFR-associated index SNPs. Color coding reflects evidence of kidney function relevance (Methods), which is coded as 'likely' (blue), 'inconclusive' (gray) or 'unlikely' (green). Error bars correspond to 95% CIs. The dashed line corresponds to the line of best fit. Pearson's correlation coefficient $r = -0.65$ (95% CI = -0.72 , -0.58). **d**, Association of lower genetically predicted eGFR based on a GRS of 147 SNPs likely to be most relevant for kidney function with ICD-10-based clinical diagnoses for 452,264 individuals from the UK Biobank. Asthma was included as a negative control. Results are displayed as the OR and 95% CI per 10% lower GRS-predicted eGFR (Methods).

restricted to European-ancestry participants, for whom data to construct a large enough LD reference panel were publicly available (Methods). Meta-analysis of 85 European-ancestry CKDGen GWAS identified 256 genome-wide-significant loci (Supplementary Table 10). Replication among 216,518 European-ancestry MVP

participants confirmed 228 SNPs, including 227 index SNPs that met replication criteria and the *SHROOM3* index SNP (Methods and Supplementary Table 10). Of these 228 SNPs, 221 mapped to one of the 264 replicated loci from the trans-ancestry analysis (≤ 500 kb up- or downstream of the trans-ancestry index SNP), and

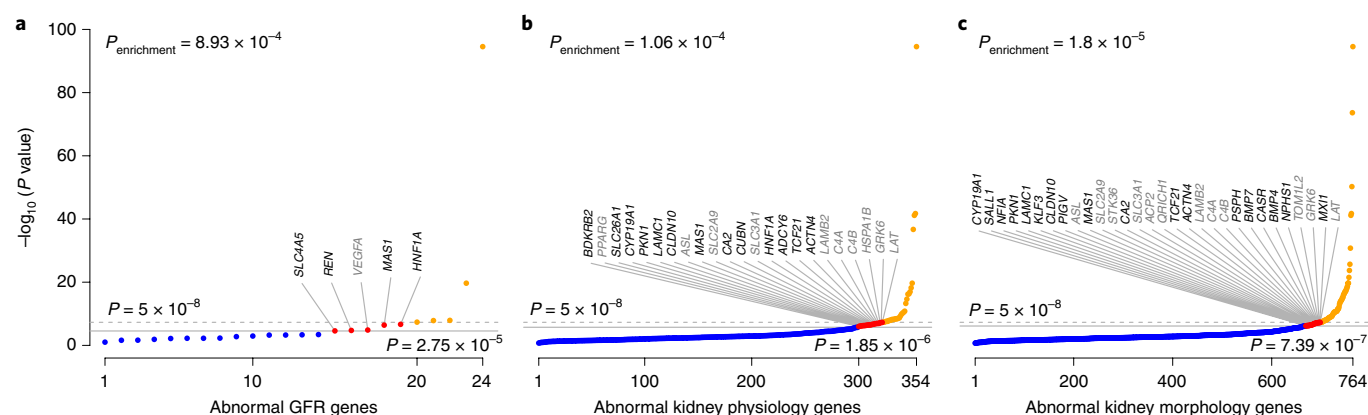


Fig. 3 | Human orthologs of genes with renal phenotypes in genetically manipulated mice are enriched for association signals with eGFR. a–c, Signals in candidate genes identified on the basis of the mouse phenotypes of abnormal GFR (**a**), abnormal kidney physiology (**b**) and abnormal kidney morphology (**c**). The y axis shows $-\log_{10}(P)$ for association with eGFR in the trans-ancestry meta-analysis for the variant with the lowest P value in each candidate gene. The dashed line corresponds to genome-wide significance ($P=5 \times 10^{-8}$), and the solid gray line corresponds to the experiment-wide significance threshold for each nested candidate gene analysis. Orange, genome-wide significance; red, experiment-wide but not genome-wide significance; blue, no significantly associated SNPs. Genes are labeled if they reached experiment- but not genome-wide significance; black font indicates genes not mapping to loci reported in the main analysis. Enrichment P values correspond to the observed number of genes with association signals below the experiment-wide threshold against the number expected on the basis of the complementary cumulative binomial distribution (Methods).

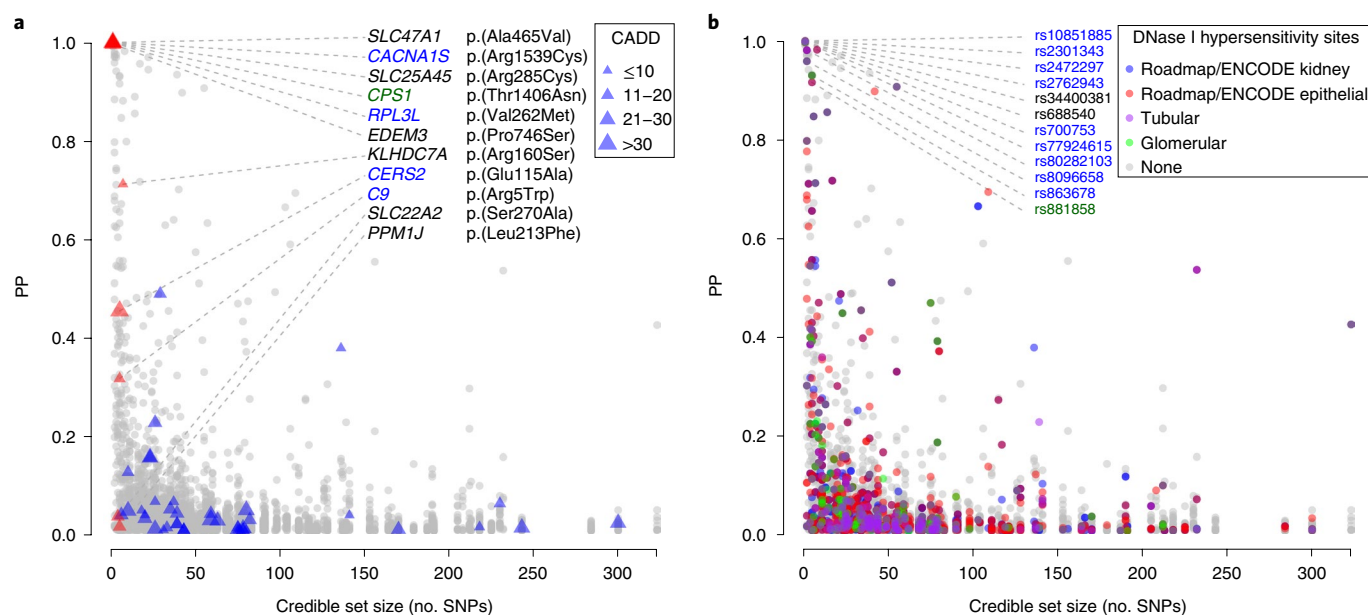


Fig. 4 | Credible set size plotted against variant posterior probability for 3,655 SNPs in 253 99% credible sets according to variant annotation. a, Exonic variants. SNPs are marked by triangles, with triangle size proportional to CADD score. Red triangles indicate missense SNPs mapping to small credible sets (≤ 5 SNPs) or to sets containing SNPs with high individual PP of driving the association signal ($> 50\%$). **b,** SNPs with regulatory potential. Symbol color corresponds to regulatory potential as derived from DNase I hypersensitivity analysis in target tissues (Methods). Annotation was restricted to variants with $PP > 1\%$; SNPs with $PP \geq 90\%$ contained in credible sets with ≤ 10 SNPs are labeled. Data are plotted as credible set size (x axis) against variant PP (y axis). Blue and green color coding for gene and SNP labels refers to kidney-function relevance and has the same meaning as in Fig. 2.

the remaining 7 showed $P \leq 3.3 \times 10^{-6}$ in the trans-ancestry discovery analysis. BUN GWAS meta-analysis of CKDGen European-ancestry studies ($n=243,029$) allowed us to classify 122 SNPs as likely relevant for kidney function, 90 as inconclusive and 16 as unlikely (Supplementary Table 10).

To conduct statistical fine-mapping of the 228 eGFR loci, we first performed summary-statistics-based conditional analysis and identified 253 independent genome-wide-significant SNPs (Supplementary Table 11) mapping to 189 regions (Methods). For each independent

variant, we computed a 99% credible set³², with a median set size of 26 SNPs (IQR: 6, 60). We observed 58 small credible sets (≤ 5 SNPs), including 20 single-SNP sets: *EDEM3*, *CACNA1S*, *HOXD11*, *CPS1*, *DAB2*, *SLC34A1*, *LINC01512*, *LARP4B*, *DCDC1*, *SLC25A45*, *SLC6A13*, *GATM*, *CGNL1*, *CYP1A1*, *NRG4*, *RPL3L*, *UMOD-PDILT*, *SLC47A1* and two independent sets at *BCL2L14* (Fig. 4 and Supplementary Table 11). Of the 58 small credible sets, 33 were likely relevant for kidney function and contain genes and SNPs that can now be prioritized for further study (Supplementary Table 11).

Table 1 | Genes implicated as causal via identification of missense SNPs with high probability of driving the eGFR association signal

Gene	SNP ^a	Credible set size	SNP PP	Functional consequence	CADD score	DHSs, tissue	Brief summary of the gene's function and relevant literature
<i>CACNA1S</i>	rs3850625	1	1.00	p.(Arg1539Cys) (NP_000060.2)	34.0	–	Encodes a subunit of the slowly inactivating L-type voltage-dependent calcium channel in skeletal muscle. Reports of altered expression in kidney cancer ⁴⁸ and after indoxyl sulfate treatment ⁴⁹ . Rare variants can cause autosomal dominant hypokalemic periodic paralysis, type 1 (MIM 170400) or malignant hyperthermia susceptibility (MIM 601887). Common variation at this locus has been reported as associated with eGFR in previous GWAS ^{10,50} .
<i>CPS1</i>	rs1047891	1	1.00	p.(Thr1406Asn) (NP_001866.2)	22.1	–	Encodes a key mitochondrial enzyme of the urea cycle that catalyzes the synthesis of carbamoyl phosphate from ammonia and bicarbonate to remove excess urea. Rare mutations cause autosomal recessive carbamoyl phosphate synthetase I deficiency (MIM 237300). GWAS locus for eGFR ¹³ , serum metabolites ⁵¹ and urinary glycine ⁵² , as well as for many other quantitative biomarkers. This variant has been reported to associate with hyperammonemia after valproate therapy ⁵³ .
<i>EDEM3</i>	rs78444298	1	1.00	p.(Pro746Ser) (NP_079467.3)	24.6	–	The gene product accelerates proteasome-mediated ER-associated degradation of glycoproteins by catalyzing mannose trimming from Man ₈ GlcNAc ₂ to yield Man ₇ GlcNAc ₂ on N-glycans. This variant has been identified by a previous exome chip association study with eGFR ²⁷ .
<i>KLHDC7A</i>	rs11261022	7	0.71	p.(Arg160Ser) (NP_689588.2)	1.1	Roadmap + ENCODE, kidney	This gene encodes the Kelch-domain-containing 7A protein and is a paralog of <i>KBTD11</i> . No specific entry in relation to kidney disease in PubMed.
<i>RPL3L</i>	rs113956264	1	1.00	p.(Val262Met) (NP_005052.1)	27.2	–	The gene product has sequence similarity with ribosomal protein L3. It has a tissue-specific expression pattern, with the highest levels in skeletal muscle and heart.
<i>SLC25A45</i>	rs34400381	1	1.00	p.(Arg285Cys) (NP_001070709.2)	26.0	ENCODE, kidney	The encoded protein belongs to the SLC25 family of mitochondrial carrier proteins and is an orphan transporter. This variant has already been identified in a GWAS of symmetric dimethylarginine levels ⁵⁴ and in a whole-genome-sequencing analysis of serum creatinine ⁵⁵ . <i>SLC25A45</i> may have a role in biosynthesis of arginine, which is involved in the synthesis of creatine.
<i>SLC47A1</i>	rs111653425	1	1.00	p.(Ala465Val) (NP_060712.2)	24.6	–	Encodes a multidrug and toxin extrusion protein (MATE1), a transport protein responsible for the secretion of cationic drugs and creatinine across brush border membranes. This variant has already been identified in a whole-genome-sequencing analysis of serum creatinine from Iceland ⁵⁵ . Rare and common variants in the locus have been identified in exome chip studies ²⁷ and GWAS ¹³ of eGFR, respectively. <i>Slc47a1</i> -knockout mice show higher levels of serum creatinine and BUN ³⁴ , arguing against a sole effect on creatinine transport and supporting an effect on kidney function.
<i>PPM1J</i>	rs34611728	5	0.02	p.(Leu213Phe) (NP_005158.5)	13.1	ENCODE, kidney	This gene encodes a serine/threonine protein phosphatase. The variant has been reported in association with eGFR in an exome chip association study ²⁷ .

Continued

Table 1 | Genes implicated as causal via identification of missense SNPs with high probability of driving the eGFR association signal (Continued)

Gene	SNP ^a	Credible set size	SNP PP	Functional consequence	CADD score	DHSs, tissue	Brief summary of the gene's function and relevant literature
<i>CERS2</i>	rs267738	5	0.46	p.(Glu115Ala) (NP_071358.1)	32.0/28.2	-	Encodes ceramide synthase 2, which may be involved in sphingolipid synthesis. Changes in ceramide levels were reported as essential in renal Madin-Darby canine kidney (MDCK) cell differentiation ⁵⁶ . <i>Cers2</i> -knockout mice show strongly reduced ceramide levels in the kidney and develop renal parenchyma abnormalities ⁵⁷ . This variant has been reported as associated with the rate of albuminuria increase in individuals with diabetes ⁵⁸ .
<i>C9</i>	rs700233	5	0.32	p.(Arg5Trp) (NP_001728.1)	6.6	-	Encodes a constituent of the membrane attack complex that has a key role in the innate and adaptive immune responses. Rare mutations can cause C9 deficiency (MIM 613825). C9 is mentioned in several kidney disease case reports, including for patients with congenital factor 9 deficiency showing IgA nephropathy ⁵⁹ .
<i>SLC22A2</i>	rs316019	4	0.04	p.(Ser270Ala) (NP_003049.2)	12.7	-	Encodes the polyspecific organic cation transporter (OCT2) that is primarily expressed in the kidney, where it mediates tubular uptake of organic compounds including creatinine from the circulation. Many publications relate <i>SLC22A2</i> to kidney function. rs316019 is a known pharmacogenomics variant associated with response to metformin and other drugs such as cisplatin. Carriers of the risk allele have a higher risk of cisplatin-induced nephrotoxicity ⁴³ , indicating that this transporter is essential in excreting toxins. The locus has been reported in previous GWAS of eGFR ¹³ .

Genes are included if they contain a missense SNP with a PP of association of >50% or map to a small credible set (≤5 SNPs). PP, posterior probability; CADD score, combined annotation-dependent depletion (CADD) Phred-like score (Methods); DHSs, DNase I-hypersensitive sites. ^aBoldface indicates the SNPs most likely to be relevant for kidney function on the basis of combined effects on eGFR and BUN.

Credible set SNPs were annotated with respect to their functional consequence and regulatory potential. Missense SNPs with >50% posterior probability (PP) of driving the association and/or mapping to a small credible set are of particular interest because they directly implicate the affected gene. Such missense SNPs were identified in 11 genes (*SLC47A1*, *RPL3L*, *SLC25A45*, *CACNA1S*, *EDEM3*, *CPS1*, *KLHDC7A*, *PPM1J*, *CERS2*, *C9* and *SLC22A2*; Supplementary Table 12), of which *CACNA1S*, *RPL3L*, *CERS2* and *C9* were likely relevant for kidney function (Fig. 4a). The majority of the 11 variants had a combined annotation-dependent depletion (CADD) score greater than 15, indicating potential deleteriousness³³. Several identified genes are plausible biological candidates for driving the association signal (Table 1). For example, the missense p.(Ala465Val) SNP in *SLC47A1* (PP > 99%) alters the encoded multidrug and toxin extrusion protein (MATE1), a transport protein responsible for the secretion of cationic drugs, toxins and internal metabolites including creatinine across brush border membranes, including kidney-proximal tubules. The fact that *Slc47a1*-knockout mice have higher blood levels of both creatinine and BUN³⁴ argues against a sole effect on creatinine transport.

To evaluate the regulatory potential of SNPs from small credible sets in the kidney, we annotated them to open chromatin regions identified from primary human tubular and glomerular cell cultures³⁵, as well as from publicly available kidney cell types (Methods). We identified 72 SNPs mapping to one of these annotations, which may thus represent causal regulatory variants (Supplementary Table 12). A particularly interesting finding was the

intronic rs77924615 SNP in *PDILT*, which showed PP > 99% of driving the association at the *UMOD* locus and mapped to open chromatin in all evaluated resources (native kidney cells, ENCODE and Roadmap kidney cell types; Fig. 4b).

Gene prioritization: colocalization with gene expression. We performed colocalization analyses for each eGFR-associated locus with gene expression in *cis* across 46 tissues, including kidney glomerular and tubulo-interstitial compartments (Methods). PP > 80% of colocalization in at least one kidney tissue was observed for 17 transcripts mapping to 16 of the 228 replicated loci (Fig. 5), pointing toward a shared underlying SNP associated with both eGFR and gene expression and implicating the gene encoding the colocalized transcript as the effector gene for the locus.

New insights emerged on several levels: first, *UMOD* is a well-established causal gene for CKD and can therefore be used to evaluate our workflow. In the tubulo-interstitial compartment, we observed a shared underlying variant associated with higher *UMOD* gene expression and lower eGFR (Fig. 5), in agreement with previous GWAS of urinary uromodulin concentration, in which alleles associated with lower eGFR at *UMOD*¹⁵ were associated with higher urinary uromodulin concentrations³⁶. The lead SNP at this locus was rs77924615, highlighted above as the candidate causal regulatory variant mapping to an intron of *PDILT* (upstream of *UMOD*). The association with differential *UMOD* but not *PDILT* gene expression supports *UMOD* as the causal gene and rs77924615 as a regulatory SNP.

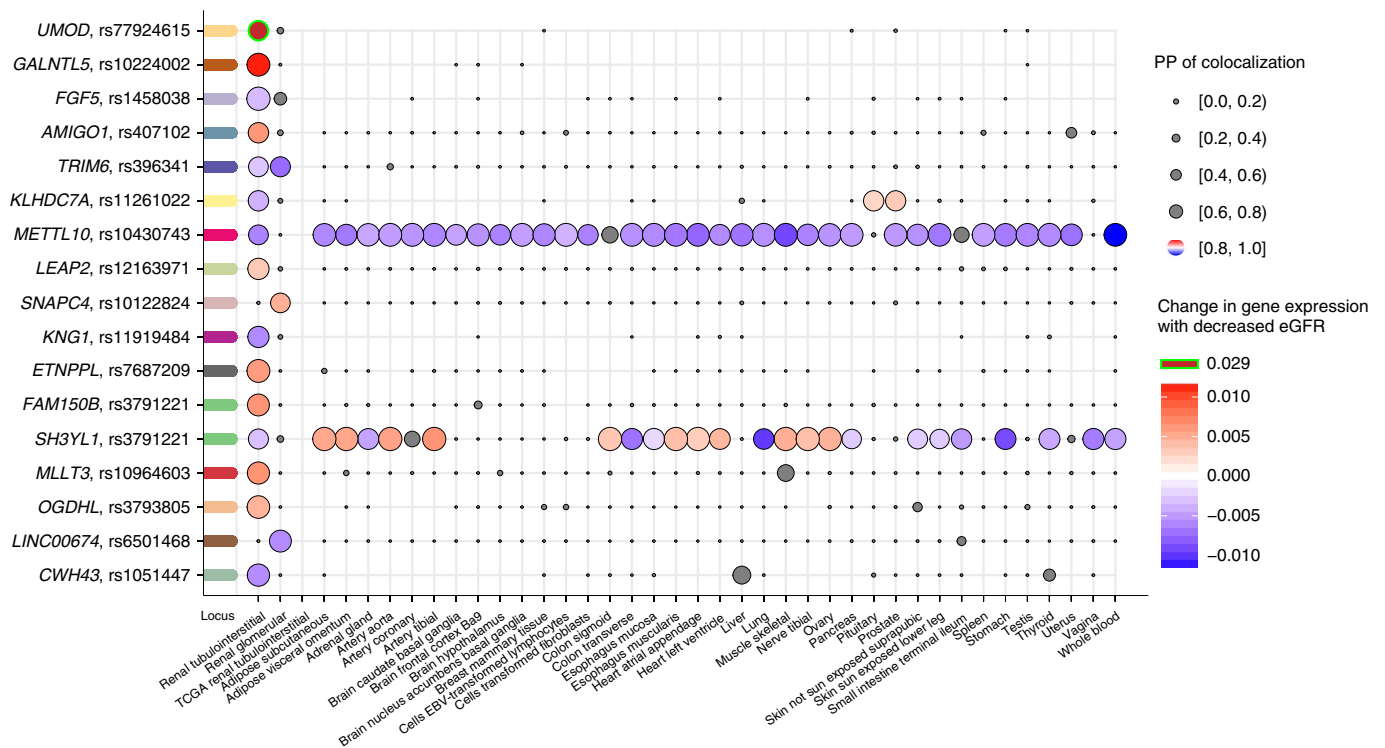


Fig. 5 | Colocalization of eGFR association signals with gene expression in kidney tissues. All eGFR loci were tested for colocalization with all eQTLs where the eQTL *cis* window overlapped (± 100 kb) the sentinel genetic variant. Genes with at least one positive colocalization (PP of one common causal variant (H_4) $\geq 80\%$) in a kidney tissue are shown with the respective sentinel SNP (y axis). Colocalizations across all tissues (x axis) are illustrated as dots, where dot size corresponds to the PP of colocalization. Negative colocalizations (PP for $H_4 < 80\%$) are gray, while positive colocalizations are colored according to the predicted change in expression relative to the allele associated with lower eGFR.

Second, new biologically plausible candidates emerged. For example, our results suggest *KNG1* and *FGF5* as effector genes in the respective eGFR-associated loci (Fig. 5 and Supplementary Table 13). *KNG1* encodes the high-molecular-weight kininogen, which is cleaved to bradykinin. Bradykinin influences blood pressure, natriuresis and diuresis and can be linked to kidney function via the renin-angiotensin-aldosterone system³⁷. *FGF5* encodes fibroblast growth factor 5, and the index SNPs for eGFR or highly correlated SNPs ($r^2 > 0.9$) have been identified in multiple GWAS of blood pressure, atrial fibrillation, coronary artery disease, hematocrit and multiple kidney-function-related traits (Supplementary Table 13). The eGFR index SNP rs1458038 (PP $> 50\%$, CADD score = 14.8; Supplementary Table 13) colocalized with the eGFR signal only in the tubulo-interstitial kidney portion (Fig. 5), supporting its regulatory potential in controlling the expression levels of *FGF5* in this compartment. Both *KNG1* and *FGF5* index SNPs were associated with BUN and CKD and are thus probably related to kidney function.

Third, for loci that showed colocalization of eGFR signals with gene expression in kidney and multiple other tissues, in some cases the allelic effect direction on gene expression was concordant across all tissues (for example, *METTL10*), whereas in other cases it differed by tissue (for example, *SH3YL1*; Fig. 5). These observations were also reflected broadly across all transcripts with evidence of colocalization in any tissue (Supplementary Fig. 9) and highlight tissue-shared and tissue-specific signals^{38,39}.

Finally, *trans* expression quantitative trait locus (*trans*-eQTL) annotation of the index SNPs in whole and peripheral blood identified a reproducible link of rs10774625 (12q24.11) with several transcripts (Methods, Supplementary Tables 14 and 15, and Supplementary Note 2).

Colocalization with uromodulin protein levels in urine. The *UMOD* locus is of particular clinical interest for CKD research²¹: rare *UMOD* mutations cause autosomal dominant tubulo-interstitial kidney disease⁴⁰, and common variants at *UMOD* give rise to the strongest eGFR and CKD GWAS signals¹⁵. We therefore performed conditional analyses based on European-ancestry-specific summary statistics and found two independent variants: rs77924615, mapping upstream of *PDILT*, and rs34882080, mapping to an intron of *UMOD* (Fig. 6a). SNP association with the urinary uromodulin-to-creatinine ratio (UUCR) in one participating cohort (Fig. 6b) matched the eGFR association pattern. Colocalization of the conditional eGFR and UUCR associations was evaluated separately for rs34882080 (Fig. 6c) and rs77924615 (Fig. 6d). Both regions showed high probability of a shared underlying variant driving the respective associations with eGFR and UUCR levels (PP = 97% and 96%, respectively), further supporting rs77924615 as a causal regulatory variant and *UMOD* as its effector gene.

A summary of the various gene characterization results for replicated loci from the European-ancestry analysis is shown in Supplementary Table 16, to facilitate selection of the most promising candidates for further experimental studies.

Discussion

This trans-ancestry study is fivefold larger than previous GWAS meta-analyses for eGFR and identified 264 replicated loci, 166 of which are reported here for the first time. By also analyzing BUN, an established complementary marker of kidney function, we highlight eGFR-associated loci that are likely to be important for kidney function as opposed to creatinine metabolism and provide a comprehensive annotation resource. Clinical relevance is supported by associations of a GRS for low eGFR with higher odds of

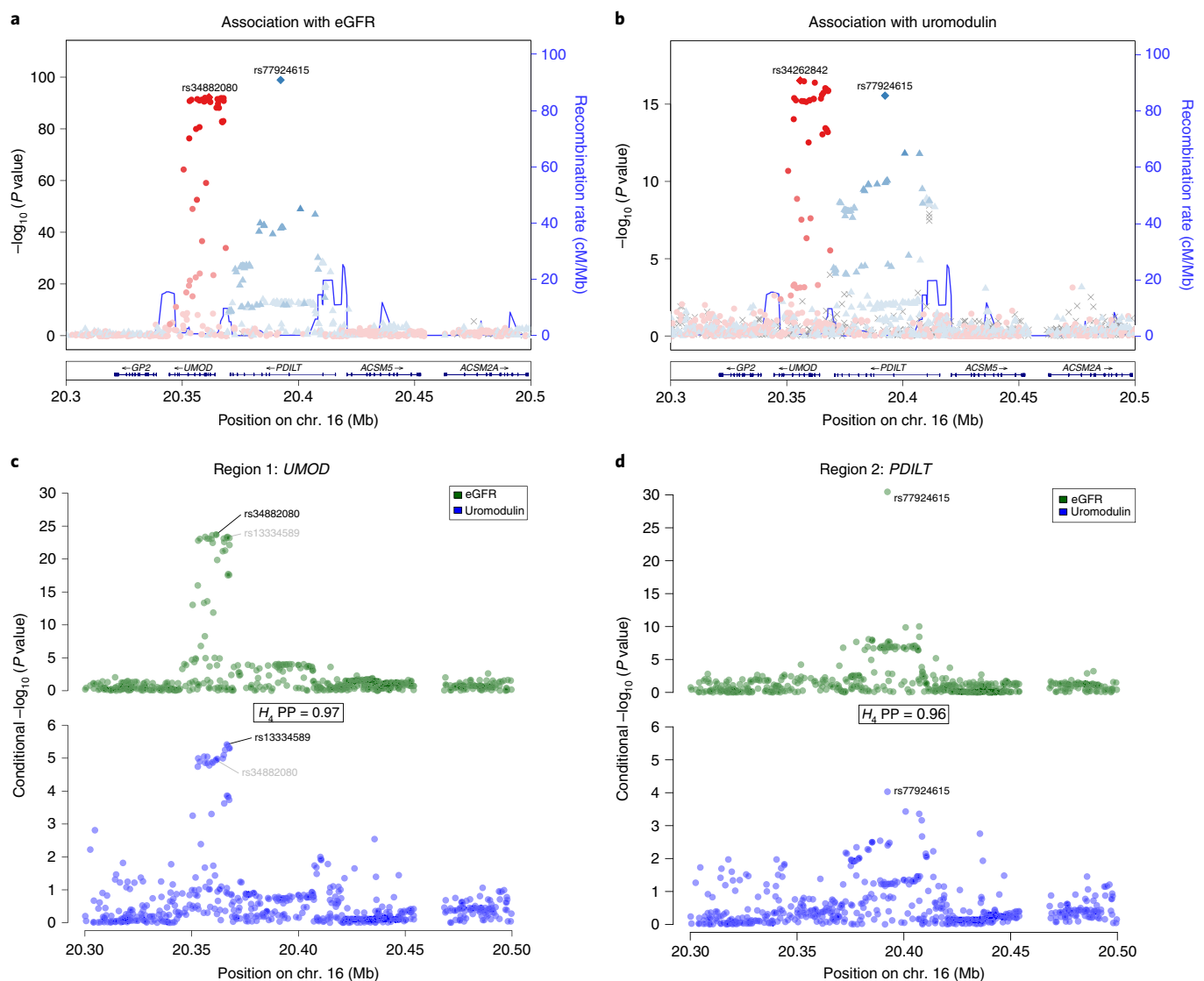


Fig. 6 | Colocalization of independent eGFR association signals at the *UMOD*-*PDILT* locus with urinary uromodulin concentrations (UUCR) supports *UMOD* as the effector gene. Association plots show association $-\log_{10}(P \text{ value})$ (y axis) plotted against chromosomal position (x axis). **a**, Approximate conditional analyses among European-ancestry individuals support the presence of two independent eGFR-associated signals. **b**, The association signal for uromodulin (UUCR) levels is similar; $r^2 = 0.93$ between rs34882080 and rs34262842. **c,d**, Colocalization of association with eGFR (top) and uromodulin (UUCR) levels (bottom) for the independent regions centered on *UMOD* (**c**) and *PDILT* (**d**) supports a shared underlying variant in both regions with high PP.

clinically diagnosed CKD, CKD-related phenotypes and hypertension. Enrichment analyses confirm the kidney as the main target organ. Colocalization of associations with eGFR and gene expression in the kidney implicates specific target genes for follow-up. Conditional analyses, fine-mapping and functional annotation at 228 replicated eGFR-associated loci among European-ancestry participants implicate single potentially causal variants at 20 loci.

Most previous GWAS meta-analyses for eGFR have been limited to a single ancestry group⁸ and did not prioritize causal variants or effector genes in associated loci. Although underpowered to uncover new loci, one previous trans-ancestry study used fine-mapping, resolving one signal to a single variant²⁰, rs77924615 at *UMOD*-*PDILT*, which is also identified in our study. At this locus, we further characterized the relationship between the causal variant, *UMOD* expression in the target tissue and uromodulin protein levels. This increase in resolution—from a locus to a single potentially causal variant with its effector gene, protein and target tissue—represents a critical advance over 10 years of eGFR GWAS¹⁵ and is a prerequisite for translational research.

The complementary multi-tissue approaches, including enrichment analyses based on gene expression, regulatory annotations, and gene sets and pathways, highlight the kidney as the most important target organ. However, relatively few kidney-specific experimental datasets are publicly available. For example, the kidney is not well represented in the Genotype-Tissue Expression (GTEx) Project and is not included in its tissue-specific eQTL datasets³⁸, emphasizing the value of open-access resources and in-depth characterization of uncommon tissues and cell types. We were able to specifically investigate the kidney by using a recently published eQTL dataset from glomerular and tubulo-interstitial portions of microdissected human kidney biopsies⁴¹, kidney-specific regulatory information from the ENCODE and Roadmap Epigenomics resources, and by obtaining regulatory information from primary cultures of human glomerular and tubulo-interstitial cells³⁵.

Functional follow-up studies of potentially causal variants should benefit from prioritized loci that show clear evidence supporting one or a few SNPs driving the association signal. The fine-mapping workflow allowed us to prioritize several SNPs at single-SNP

resolution or at a resolution of ≤ 5 SNPs, some of which may have broader clinical relevance. For example, the OCT2 protein encoded by *SLC22A2* transports several cationic drugs such as metoprolol, cisplatin, metformin and cimetidine across the basolateral membrane of renal tubular cells⁴². The prioritized missense SNP encodes p.(Ser270Ala), a known pharmacogenomic variant that alters the transport of these drugs and their side effects, such as cisplatin-induced nephrotoxicity⁴³. Along the same lines, the prioritized SNP encoding the p.(Ala465Val) substitution in the transporter MATE1 encoded by *SLC47A1* may affect the ability to secrete drugs and other toxins from proximal tubular cells into the urine⁴⁴ and hence alter CKD risk.

Strengths of this project include the large sample size with dense genotype imputation, standardized and automated phenotype generation and quality control, and independent replication, as well as the advanced and comprehensive downstream bioinformatics analyses. Further strengths are the use of BUN to prioritize eGFR-associated loci likely relevant for kidney function and to provide genome-wide BUN summary statistics as an annotation resource for other studies of eGFR. Moreover, we evaluated a GRS for eGFR for association with clinically diagnosed CKD in a large independent study. Among the limitations, non-European populations are still under-represented in our study, as in many other genomic efforts⁴⁵. Statistical fine-mapping using trans-ancestry data with different LD structures can potentially narrow association signals. However, a sufficiently large reference dataset to compute ancestry-matched LD structure for summary-statistics-based fine-mapping was only available for European ancestry, highlighting the potential of future large-scale efforts with trans-ancestry fine-mapping and the need to generate data from non-European-ancestry populations, thereby enabling such endeavors. Finally, several SNPs had small effective sample sizes in some subpopulations, which might have affected the ability to assess between-ancestry heterogeneity and potentially underestimated true heterogeneity.

We estimated GFR from serum creatinine, as done in clinical practice and observational studies, because direct measurement of kidney function is invasive, time-consuming and burdensome. Under the assumption that genetic associations supported by multiple markers are less likely to reflect marker metabolism, we used BUN to prioritize eGFR-associated loci likely to be relevant to kidney function. Blood creatinine, urea and cystatin C concentrations are influenced not only by glomerular filtration but also by the synthesis, active secretion and reabsorption of these molecules, as illustrated by loci detected in our study: for example, the *GATM* locus was associated with eGFR but not with BUN, in agreement with the function of the encoded protein as a rate-limiting enzyme in creatine synthesis⁴⁶. Conversely, the *SLC14A2* locus was associated with BUN but not with eGFR, in line with the function of the encoded protein as a urea transporter⁴⁷. Even so, lack of association for a SNP with one kidney function marker based on a combination of *P* value and effect direction may not necessarily mean that the locus is not relevant to kidney function. Our categorization of the eGFR loci into three classes on the basis of direction of effect and significance of BUN association should be interpreted with caution, with 'likely' and 'unlikely' reflecting uncertainty of the assignment. Factors complicating the comparison of eGFR and BUN associations at the locus level are differential statistical power, differential ancestry distribution and potential allelic heterogeneity. Further large-scale studies with multiple kidney function markers measured in the same individuals are therefore warranted.

To identify broadly representative and generalizable association signals, we focused on SNPs that were present in the majority of the participating studies. This choice might have limited our ability to uncover new variants or to fine-map low-frequency or population-specific variants, which represents a complementary avenue of research. Moreover, even with well-powered

fine-mapping approaches, potentially causal SNPs need to be confirmed as functional variants in experimental studies. Although colocalization with gene expression can help prioritize effector genes, these associations are based on measures from a single time point and hence cannot answer whether changes in gene expression precede or follow changes in kidney function.

In summary, we have identified and characterized a large number of loci associated with eGFR and prioritized potential effector genes, driver variants and target tissues. These findings will help direct functional studies and advance the understanding of kidney function biology, a prerequisite to develop novel therapies to reduce the burden of CKD.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at <https://doi.org/10.1038/s41588-019-0407-x>.

Received: 16 September 2018; Accepted: 29 March 2019;

Published online: 31 May 2019

References

- Eckardt, K. U. et al. Evolving importance of kidney disease: from subspecialty to global health burden. *Lancet* **382**, 158–169 (2013).
- Jha, V. et al. Chronic kidney disease: global dimension and perspectives. *Lancet* **382**, 260–272 (2013).
- Ene-Iordache, B. et al. Chronic kidney disease and cardiovascular risk in six regions of the world (ISN-KDDC): a cross-sectional study. *Lancet Glob. Health* **4**, e307–e319 (2016).
- Go, A. S., Chertow, G. M., Fan, D., McCulloch, C. E. & Hsu, C. Y. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N. Engl. J. Med.* **351**, 1296–1305 (2004).
- GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1151–1210 (2017).
- Inrig, J. K. et al. The landscape of clinical trials in nephrology: a systematic review of ClinicalTrials.gov. *Am. J. Kidney Dis.* **63**, 771–780 (2014).
- Levin, A. et al. Global kidney health 2017 and beyond: a roadmap for closing gaps in care, research, and policy. *Lancet* **390**, 1888–1917 (2017).
- Wuttke, M. & Kottgen, A. Insights into kidney diseases from genome-wide association studies. *Nat. Rev. Nephrol.* **12**, 549–562 (2016).
- Gorski, M. et al. 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. *Sci. Rep.* **7**, 45040 (2017).
- Pattaro, C. et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.* **7**, 10023 (2016).
- Chasman, D. I. et al. Integration of genome-wide association studies with biological knowledge identifies six novel genes related to kidney function. *Hum. Mol. Genet.* **21**, 5329–5343 (2012).
- Pattaro, C. et al. Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet.* **8**, e1002584 (2012).
- Kottgen, A. et al. New loci associated with kidney function and chronic kidney disease. *Nat. Genet.* **42**, 376–384 (2010).
- Chambers, J. C. et al. Genetic loci influencing kidney function and chronic kidney disease. *Nat. Genet.* **42**, 373–375 (2010).
- Kottgen, A. et al. Multiple loci associated with indices of renal function and chronic kidney disease. *Nat. Genet.* **41**, 712–717 (2009).
- Kanai, M. et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
- Okada, Y. et al. Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat. Genet.* **44**, 904–909 (2012).
- Hishida, A. et al. Genome-wide association study of renal function traits: results from the Japan Multi-institutional Collaborative Cohort study. *Am. J. Nephrol.* **47**, 304–316 (2018).
- Lee, J. et al. Genome-wide association analysis identifies multiple loci associated with kidney disease-related traits in Korean populations. *PLoS One* **13**, e0194044 (2018).
- Mahajan, A. et al. Trans-ethnic fine mapping highlights kidney-function genes linked to salt sensitivity. *Am. J. Hum. Genet.* **99**, 636–646 (2016).
- Devuyst, O. & Pattaro, C. The *UMOD* locus: insights into the pathogenesis and prognosis of kidney disease. *J. Am. Soc. Nephrol.* **29**, 713–726 (2018).

22. Yeo, N. C. et al. Shroom3 contributes to the maintenance of the glomerular filtration barrier integrity. *Genome Res.* **25**, 57–65 (2015).
23. Gaziano, J. M. et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).
24. Benner, C. et al. Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *Am. J. Hum. Genet.* **101**, 539–551 (2017).
25. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
26. Abecasis, G. R. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
27. Li, M. et al. *SOS2* and *ACPI* loci identified through large-scale exome chip analysis regulate kidney development and function. *J. Am. Soc. Nephrol.* **28**, 981–994 (2017).
28. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
29. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
30. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
31. Jing, J. et al. Combination of mouse models and genomewide association studies highlights novel genes associated with human kidney function. *Kidney Int.* **90**, 764–773 (2016).
32. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).
33. Dong, C. et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).
34. Tsuda, M. et al. Targeted disruption of the multidrug and toxin extrusion 1 (*Mate1*) gene in mice reduces renal secretion of metformin. *Mol. Pharm.* **75**, 1280–1286 (2009).
35. Sieber, K. B. et al. Integrated functional genomic analysis enables annotation of kidney genome-wide association study loci. *J. Am. Soc. Nephrol.* **30**, 421–441 (2019).
36. Olden, M. et al. Common variants in *UMOD* associate with urinary uromodulin levels: a meta-analysis. *J. Am. Soc. Nephrol.* **25**, 1869–1882 (2014).
37. Moreau, M. E. et al. The kallikrein-kinin system: current and future pharmacological targets. *J. Pharm. Sci.* **99**, 6–38 (2005).
38. Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
39. Gamazon, E. R. et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat. Genet.* **50**, 956–967 (2018).
40. Eckardt, K. U. et al. Autosomal dominant tubulointerstitial kidney disease: diagnosis, classification, and management—a KDIGO consensus report. *Kidney Int.* **88**, 676–683 (2015).
41. Gillies, C. E. et al. An eQTL landscape of kidney tissue in human nephrotic syndrome. *Am. J. Hum. Genet.* **103**, 232–244 (2018).
42. Dudley, A. J., Bleasby, K. & Brown, C. D. The organic cation transporter OCT2 mediates the uptake of β -adrenoceptor antagonists across the apical membrane of renal LLC-PK₁ cell monolayers. *Br. J. Pharm.* **131**, 71–79 (2000).
43. Filipski, K. K., Mathijssen, R. H., Mikkelsen, T. S., Schinkel, A. H. & Sparreboom, A. Contribution of organic cation transporter 2 (OCT2) to cisplatin-induced nephrotoxicity. *Clin. Pharm. Ther.* **86**, 396–402 (2009).
44. Motohashi, H. & Inui, K. Organic cation transporter OCTs (SLC22) and MATes (SLC47) in the human kidney. *AAPS J.* **15**, 581–588 (2013).
45. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).
46. Humm, A., Huber, R. & Mann, K. The amino acid sequences of human and pig L-arginine:glycine amidinotransferase. *FEBS Lett.* **339**, 101–107 (1994).
47. Olives, B. et al. Molecular characterization of a new urea transporter in the human kidney. *FEBS Lett.* **386**, 156–160 (1996).
48. Phan, N. N. et al. Voltage-gated calcium channels: novel targets for cancer therapy. *Oncol. Lett.* **14**, 2059–2074 (2017).
49. Thi Do, D., Phan, N. N., Wang, C. Y., Sun, Z. & Lin, Y. C. Novel regulations of MEF2-A, MEF2-D, and CACNA1S in the functional incompetence of adipose-derived mesenchymal stem cells by induced indoxyl sulfate in chronic kidney disease. *Cytotechnology* **68**, 2589–2604 (2016).
50. Parsa, A. et al. Common variants in Mendelian kidney disease genes and their association with renal function. *J. Am. Soc. Nephrol.* **24**, 2105–2117 (2013).
51. Xie, W. et al. Genetic variants associated with glycine metabolism and their role in insulin sensitivity and type 2 diabetes. *Diabetes* **62**, 2141–2150 (2013).
52. Raffler, J. et al. Genome-wide association study with targeted and non-targeted NMR metabolomics identifies 15 novel loci of urinary human metabolic individuality. *PLoS Genet.* **11**, e1005487 (2015).
53. Janicki, N. et al. Increased occurrence of valproic acid-induced hyperammonemia in carriers of T1405N polymorphism in carbamoyl phosphate synthetase 1 gene. *ISRN Neurol.* **2013**, 261497 (2013).
54. Seppala, A. et al. Genome-wide association study on dimethylarginines reveals novel AGXT2 variants associated with heart rate variability but not with overall mortality. *Eur. Heart J.* **35**, 524–531 (2014).
55. Sveinbjornsson, G. et al. Rare mutations associating with serum creatinine and chronic kidney disease. *Hum. Mol. Genet.* **23**, 6935–6943 (2014).
56. Pescio, L. G. et al. Changes in ceramide metabolism are essential in Madin-Darby canine kidney cell differentiation. *J. Lipid Res.* **58**, 1428–1438 (2017).
57. Imgrund, S. et al. Adult ceramide synthase 2 (CERS2)-deficient mice exhibit myelin sheath defects, cerebellar degeneration, and hepatocarcinomas. *J. Biol. Chem.* **284**, 33549–33560 (2009).
58. Shiffman, D. et al. A gene variant in *CERS2* is associated with rate of increase in albuminuria in patients with diabetes from ONTARGET and TRANSCEND. *PLoS One* **9**, e106631 (2014).
59. Yoshioka, K. et al. IgA nephropathy in patients with congenital C9 deficiency. *Kidney Int.* **42**, 1253–1258 (1992).

Acknowledgements

We thank D. Di Domizio (Eurac Research) and J. Knaus (University of Freiburg) for IT assistance and T. Johnson (GlaxoSmithKline) for sharing his code and discussion on credible set fine-mapping and colocalization analysis. This research has been conducted using the UK Biobank resource under application number 20272. Study-specific acknowledgements and funding sources are listed in the Supplementary Information.

Author contributions

Manuscript writing group: M. Wuttke, Y.L., M. Li, K.B.S., M.F., M. Gorski, A. Tin, L. Wang, H. Kirsten, T.A., K. Ho, I.H., M. Scholz, A. Teumer, A. Köttgen, C.P. **Design of the study:** C.A.B., C.F., M. Gorski, A. Köttgen, A.P.M., C.P., A. Teumer, A. Tin, M. Wuttke. **Management of an individual contributing study:** T.S.A., E.d.A., S. Akilesh, S.J.B., G.B., M. Bochud, M. Boehnke, E.B., M.H.d.B., H.B., A.S.B., C.A.B., A.C., R.J.C., J.C.C., D.I.C., C.-Y.C., K.C., R.C., M. Ciullo, J.C., D.C., R.M.v.D., J. Danesh, O.D., C.M.v.D., K.-U.E., G.E., P.E., M.K.E., J.F.F., O.H.F., B.I.F., Y.F., R.T.G., H.G., P.G., J.M.G., V. Giedraitis, C.G., F.G., A.D.G., V. Gudnason, T.B.H., P.v.d.H., C.A.H., C.H., C.-K.H., A.A.H., K. Ho, A.M.H., M.A.I., O.S.I., E.I., V.W.J., J.B.J., B.J., C.M.K., C.-C.K., W. Kiess, M.E.K., W. Koenig, J.S.K., H. Kramer, F.K., B.K.K., M. Kubo, J.K., M. Kähönen, A. Körner, A. Köttgen, T.L., Y.L., S.-C.L., M. Loeffler, R.J.L., S.L., M.A.L., P.K.M., N.G.M., D.M., K. Matsuda, O.M., A. Metspalu, E.K.M., Y.M., K.L.M., G.W.M., A.P.M., R.d.M., W.M., G.N.N., J.O.C., M.L.O'D., A.J.O., M.O.-M., W.H.O., A.P., C.P., S.A.P., B.W.P., T. Perls, M. Perola, M. Pirastu, O.P., B.P., P.P.P., M.A.P., B.M.P., T.J.R., O.T.R., D.F.R., R. Rettig, M.R., P.M.R., D.J.R., P.R., I.R., C.S., V.S., K.-U.S., H. Schmidt, R.S., M. Scholz, B.S., X.S., H. Snieder, N. Soranzo, C.N.S., K. Stefansson, C. Strauch, M. Sturm, G.S., P.O.S., E.-S.T., B.O.T., Y.-C.T., J. Thiery, A. Tin, D.T., J. Tremblay, I.T., A. Tönjes, P.V., A.P.d.V., U.V., G.W., L. Wallentin, Y.X.W., D.M.W., W.B.W., H.W., J.B.W., S.H.W., J.G.W., C. Wong, T.-Y.W., M. Wuttke, L.X., Q.Y., M.Y., W.Z., A.B.Z. **Statistical methods and analysis:** T.S.A., M.A., P.A., M.L.B., G.B., M. Boissel, T.S.B., M. Brumat, C.A.B., M. Canouil, R.J.C., J.-F.C., D.I.C., Miao-Li Chee, X.C., Y.C., A.Y.C., M. Cocca, M.P.C., J.P.C., T.C., A. Dehghan, G.D., A. Demirkan, J. Divers, R.D., D.R.V.E., T.L.E., M.F.F., J.F.F., B.I.F., S.F.-W., C.F., S.G., A.G., M. Gorski, D.F.G., M. Gögele, T.H., P.H., P.v.d.H., I.M.H., J.N.H., E.H., A.H., K. Horn, S.-J.H., J.J., P.K.J., N.S.J., B.J., Y.K., M. Kanai, C.-C.K., H. Kirsten, M.E.K., A. Krajcoviechova, H. Kramer, M. Kuokkanen, A. Köttgen, B.K., L.A.L., C.D.L., M. Li, Y.L., Jianjun Liu, Jun Liu, L.-P.L., A. Mahajan, J. Marten, J. Martins, K. Miliku, P.P.M., N.M., A.P.M., P.J.v.d.M., W.M., M.A.N., M.N., B.N., D.N., I.M.N., R.N., T.N., Y.O., C.P., S.A.P., N.P., M.H.P., B.P.P., L.M.R., M.R., K.M.R., F. Rivadeneira, F. Rizzi, R. Rueedi, K.A.R., Y. Saba, E.S., M. Scholz, C.-A.S., S. Sedaghat, Y. Shi, K.B.S., X.S., A.V.S., C.N.S., H.M.S., G.S., S. Szymczak, S.M.T., B.O.T., A. Teumer, C.H.T., H.T., G.T., J. Tremblay, N.V., V.V., S. Vögelezang, C. Wang, L. Wang, J.F.W., M.K.W., M. Wuttke, Y.X., Q.Y., L.M.Y.-A., W.Z. **Bioinformatics:** T.S.A., S. Akilesh, P.A., D.B., S.B., A.S.B., C.A.B., E.C., R.J.C., X.C., A.Y.C., M. Cocca, M.P.C., T.C., E.W.D., F.D., A. Dehghan, J. Divers, R.D., G.E., A.F., H.G., S.G., A.G., S.D.G., M. Gorski, P.H., I.M.H., E.H., A.H., K. Horn, J.J., N.S.J., C.-C.K., H. Kirsten, M.E.K., A. Krajcoviechova, A. Köttgen, C.D.L., B.L., M. Li, Y.L., Jianjun Liu, L.-P.L., J. Marten, J. Martins, Y.M., P.P.M., K.L.M., D.O.M.-K., P.J.v.d.M., R.M., W.M., R.N., T.N., S.A.P., N.P., G.P., A.I.P., M.H.P., B.P.P., F. Rizzi, R. Rueedi, Y. Saba, E.S., M. Scholz, C.-A.S., S. Sedaghat, C.M.S., K.B.S., A.V.S., C.N.S., S. Szymczak, H.T., J. Tremblay, C. Wang, J.F.W., M. Wuttke, Y.X., L.M.Y.-A., Z.Y., W.Z. **Interpretation of results:** T.S.A., E.d.A., C.A.B., C.-Y.C., K.D., J. Divers, R.D., K.E., M.F.F., J.F.F., B.I.F., S.G., C.G., A.G., M. Gorski, P.H., P.v.d.H., H.T., I.M.H., K. Ho, K. Horn, W.H., S.-J.H., B.J., H. Kirsten, W. Koenig, A. Krajcoviechova, A. Köttgen, M. Laakso, C.D.L., M. Li, Y.L., P.K.M., J. Marten, K. Miliku, K.L.M., A.P.M., N.D.P., C.P., S.A.P., B.P.P., D.F.R., M.R., P.M.R., M. Scholz, S. Sedaghat, K.B.S., C.N.S., P.O.S., B.O.T., A. Teumer, A. Tin, J. Tremblay, I.T., A.G.U., N.V., V.V., S. Vögelezang, L. Wallentin, H.W., M. Wuttke, Y.X., M.Y., L.M.Y.-A. **Genotyping:** N.A., D.B., R.B., A.S.B., C.A.B., A.C., H.C., D.I.C., C.-Y.C., E.W.D., A. Demirkan, R.D., C.M.v.D., G.E., M.K.E., M.F.F., A.E., Y.F., C.F., R.T.G., H.G., S.D.G., P.H., P.v.d.H., H.T., C.H., C.-K.H., W.H., E.I., C.-C.K., M.E.K., W. Koenig, J.S.K., P.K., A.T.K., A. Krajcoviechova, F.K., M. Kubo, M. Kähönen, A. Körner, L.A.L., T.L., L.-P.L., P.K.M., T.M., O.M., Y.M., K.L.M., N.M., G.W.M., D.O.M.-K., A.P.M., J.C.M., W.M., M.A.N., M.O.-M.,

S.P., N.D.P., B.W.P., M. Perola, D.J.P., M.H.P., O.T.R., D.F.R., F. Rivadeneira, F. Rizzi, J.I.R., D.R., V.S., E.S., B.H.S., C.N.S., S.M.T., K.T., A. Teumer, D.T., J. Tremblay, A.G.U., S. Vaccargiu, U.V., M. Waldenberger, C. Wang, L. Wang, Y.X.W., J.G.W., M.K.W., A.B.Z., J.Ä. **Critical review of manuscript:** T.S.A., S. Akilesh, P.A., E.d.A., S.J.B., N.B., M.L.B., G.B., M.H.d.B., E.P.B., T.S.B., H.B., A.S.B., C.A.B., H.C., D.I.C., X.C., Y.C., A.Y.C., M. Ciullo, J.C., R.M.v.D., G.D., O.D., J. Divers, R.D., K.-U.E., D.R.V.E., T.L.E., P.E., K.E., M.K.E., M.F.F., J.F.F., O.H.F., A.F., B.I.F., Y.F., C.F., H.G., S.G., C.G., A.G., S.D.G., M. Gorski, D.F.G., P.H., T.B.H., P.v.d.H., C.A.H., C.H., I.M.H., J.N.H., C.-K.H., K. Ho, A.H., W.H., N.H.-K., S.-J.H., O.S.I., E.L., V.W.J., J.J., J.B.J., P.K.J., B.J., M. Kastarinen, S.M.K., M.E.K., W. Koenig, A.T.K., H. Kramer, F.K., B.K.K., M. Kuokkanen, M. Kähönen, A. Körner, A. Köttgen, B.K., M. Laakso, L.A.L., C.D.L., J.J.-M.L., T.L., M. Li, Y.L., W.L., L.L., C.M.L., M. Loeffler, R.J.L., L.-P.L., P.K.M., A. Mahajan, J. Marten, N.G.M., D.M., C.M., T.M., O.M., E.K.M., K. Miliku, K.L.M., G.W.M., D.O.M.-K., R.d.M., W.M., G.N.N., M.A.N., M.N., K.N., B.N., I.M.N., R.N., T.N., M.L.O'D., A.J.O., M.O.-M., N.D.P., R.P., A.P., C.P., S.A.P., B.W.P., M. Perola, O.P., M.H.P., B.P.P., B.M.P., T.J.R., L.M.R., O.T.R., R. Rettig, M.R., K.M.R., P.M.R., F. Rivadeneira, D.J.R., P.R., I.R., D.R., C.S., V.S., K.-U.S., M. Scholz, C.-A.S., N. Schupf, B.S., S. Sedaghat, K.B.S., X.S., A.V.S., H. Snieder, C.N.S., K. Strauch, G.S., P.O.S., S.M.T., N.Y.Q.T., B.O.T., A. Teumer, H.T., A. Tin, J. Tremblay, I.T., A. Tönjes, A.G.U., N.V., V.V., S. Voegelzang, A.P.d.V., U.V., M. Waldenberger, L. Wallentin, D.M.W., H.W., J.B.W., S.H.W., J.G.W., M. Wuttke, Q.Y., Z.Y., A.B.Z. **Subject recruitment:** S. Afaq, E.P.B., H.B., C.A.B., A.C., H.C., J.C.C., Miao-Ling Chee, K.C., R.C., M. Ciullo, D.C., K.D., M.K.E., V.H.X.F., B.I.F., R.T.G., V. Gudnason, C.A.H., W.H., N.H.-K., O.S.I., M.L., V.W.J., J.B.J., B.J., C.M.K., M. Kastarinen, J.S.K., A. Krajcoviechova, F.K., M. Kubo, M. Kähönen, A. Köttgen, M. Laakso, J.J.-M.L., T.L., W.L., L.L., N.G.M., K. Matsuda, C.M., A. Metspalu, R.d.M., W.M., K.N., M.L.O'D., I.O., A.J.O., S.P., C.P., S.A.P., B.W.P., M. Perola, O.P., B.P., D.J.P., T. Poulain, M.A.P., T.J.R., O.T.R., M.R., P.M.R., P.R., I.R., D.R., V.S., R.S., B.H.S., P.O.S., N.Y.Q.T., A. Teren, Y.-C.T., J. Tremblay, I.T., A. Tönjes, S. Vaccargiu, S. Voegelzang, P.V., A.P.d.V., G.W., L. Wallentin, H.W., J.B.W., S.H.W., J.G.W., A.B.Z., J.Ä.

Competing interests

W. Koenig reports modest consultation fees for advisory board meetings from Amgen, DalCor, Kowa, Novartis, Pfizer and Sanofi and modest personal fees for lectures from Amgen, AstraZeneca, Novartis, Pfizer and Sanofi, all outside the scope of the submitted work. W.M. is employed with Synlab Services and holds shares of Synlab Holding Deutschland. D.O.M.-K. is a part-time research consultant at Metabolon. M.A.N.

is supported by a consulting contract between Data Tecnica International and the National Institute on Aging (NIA), National Institutes of Health (NIH) and consults for Illumina, the Michael J. Fox Foundation and University of California Healthcare. O.H.F. works in ErasmusAGE, a center for aging research across the life course funded by Nestlé Nutrition (Nestec); Metagenics; and AXA. K.B.S., L.Y.-A., D.M.W. and M.A.L. are full-time employees of GlaxoSmithKline. M.L.O'D. received grant support from GlaxoSmithKline, MSD, Eisai, AstraZeneca, MedCo and Janssen. H.W. received grants and non-financial support from GlaxoSmithKline, during the conduct of the study; grants from Sanofi-Aventis, Eli Lilly, the National Institute of Health, Omthera Pharmaceuticals, Pfizer New Zealand, Eisai Inc. and Dalcro Pharma UK; honoraria and non-financial support from AstraZeneca; and is on advisory boards for Sirtex and Acetillon and received personal fees from CSL Behring and American Regent outside the scope of the submitted work. L. Wallentin received institutional grants from GlaxoSmithKline, AstraZeneca, BMS, Boehringer-Ingelheim, Pfizer, MSD and Roche Diagnostics. D.F.R. and A.I.P. are employees of MSD. M. Scholz received consultancy of and grant support from Merck Serono not related to this project. B.M.P. serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the steering committee of the Yale Open Data Access Project funded by Johnson & Johnson. J. Danesh is a member of the Novartis Cardiovascular and Metabolic Advisory Board and received grant support from Novartis. A.S.B. received grants from MSD, Pfizer, Novartis, Biogen and Bioerativ and personal fees from Novartis. V.S. has participated in a conference trip sponsored by Novo Nordisk and received a honorarium from the same source for participating in an advisory board meeting. A. Köttgen received grant support from Gruenenthal. All other authors declare no conflicts of interest.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41588-019-0407-x>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.K. or C.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Matthias Wuttke^{1,2,267}, **Yong Li**^{1,267}, **Man Li**^{3,267}, **Karsten B. Sieber**^{4,267}, **Mary F. Feitosa**^{5,267}, **Mathias Gorski**^{6,7,267}, **Adrienne Tin**^{8,9}, **Lihua Wang**⁵, **Audrey Y. Chu**¹⁰, **Anselm Hoppmann**¹, **Holger Kirsten**^{11,12}, **Ayush Giri**^{13,14}, **Jin-Fang Chai**¹⁵, **Gardar Sveinbjornsson**¹⁶, **Bamidele O. Tayo**¹⁷, **Teresa Nutile**¹⁸, **Christian Fuchsberger**¹⁹, **Jonathan Marten**²⁰, **Massimiliano Cocca**²¹, **Sahar Ghasemi**^{22,23}, **Yizhe Xu**³, **Katrin Horn**^{11,12}, **Damia Noce**¹⁹, **Peter J. van der Most**²⁴, **Sanaz Sedaghat**²⁵, **Zhi Yu**^{8,26}, **Masato Akiyama**^{27,28}, **Saima Afaq**^{29,30}, **Tarunveer S. Ahluwalia**³¹, **Peter Almgren**³², **Najaf Amin**²⁵, **Johan Ärnlöv**^{33,34}, **Stephan J. L. Bakker**³⁵, **Nisha Bansal**^{36,37}, **Daniela Baptista**³⁸, **Sven Bergmann**^{39,40,41}, **Mary L. Biggs**^{42,43}, **Ginevra Biino**⁴⁴, **Michael Boehnke**⁴⁵, **Eric Boerwinkle**⁴⁶, **Mathilde Boissel**⁴⁷, **Erwin P. Bottinger**^{48,49}, **Thibaud S. Boutin**²⁰, **Hermann Brenner**^{50,51}, **Marco Brumat**⁵², **Ralph Burkhardt**^{12,53,54}, **Adam S. Butterworth**^{55,56}, **Eric Campana**⁵², **Archie Campbell**⁵⁷, **Harry Campbell**⁵⁸, **Mickaël Canouil**⁴⁷, **Robert J. Carroll**⁵⁹, **Eulalia Catamo**²¹, **John C. Chambers**^{29,60,61,62,63}, **Miao-Ling Chee**⁶⁴, **Miao-Li Chee**⁶⁴, **Xu Chen**⁶⁵, **Ching-Yu Cheng**^{64,66,67}, **Yurong Cheng**¹, **Kaare Christensen**⁶⁸, **Renata Cifkova**^{69,70}, **Marina Ciullo**^{18,71}, **Maria Pina Concas**²¹, **James P. Cook**⁷², **Josef Coresh**⁸, **Tanguy Corre**^{39,40,73}, **Cinzia Felicita Sala**⁷⁴, **Daniele Cusi**^{75,76}, **John Danesh**⁷⁷, **E. Warwick Daw**⁵, **Martin H. de Borst**³⁵, **Alessandro De Grandi**¹⁹, **Renée de Mutsert**⁷⁸, **Aiko P. J. de Vries**⁷⁹, **Frauke Degenhardt**⁸⁰, **Graciela Delgado**⁸¹, **Ayse Demirkan**²⁵, **Emanuele Di Angelantonio**^{82,83}, **Katalin Dittrich**^{84,85}, **Jasmin Divers**⁸⁶, **Rajkumar Dorajoo**⁸⁷, **Kai-Uwe Eckardt**^{88,89}, **Georg Ehret**³⁸, **Paul Elliott**^{90,91,92,93}, **Karlhans Endlich**^{23,94}, **Michele K. Evans**⁹⁵, **Janine F. Felix**^{25,96,97}, **Valencia Hui Xian Foo**⁶⁴, **Oscar H. Franco**^{25,98}, **Andre Franke**⁸⁰, **Barry I. Freedman**⁹⁹, **Sandra Freitag-Wolf**¹⁰⁰, **Yechiel Friedlander**¹⁰¹, **Philippe Froguel**^{47,102}, **Ron T. Gansevoort**³⁵, **He Gao**⁹⁰, **Paolo Gasparini**^{21,52}, **J. Michael Gaziano**¹⁰³, **Vilmantas Giedraitis**¹⁰⁴,

Christian Gieger^{105,106,107}, Giorgia Grotto^{21,52}, Franco Giulianini¹⁰⁸, Martin Gögele¹⁹, Scott D. Gordon¹⁰⁹, Daniel F. Gudbjartsson¹⁶, Vilmundur Gudnason^{110,111}, Toomas Haller¹¹², Pavel Hamet^{113,114}, Tamara B. Harris¹¹⁵, Catharina A. Hartman¹¹⁶, Caroline Hayward²⁰, Jacklyn N. Hellwege^{117,118,119}, Chew-Kiat Heng^{120,121}, Andrew A. Hicks¹⁹, Edith Hofer^{122,123}, Wei Huang^{124,125}, Nina Hutri-Kähönen^{126,127}, Shih-Jen Hwang^{128,129}, M. Arfan Ikram²⁵, Olafur S. Indridason¹³⁰, Erik Ingelsson^{131,132,133,134}, Marcus Ising¹³⁵, Vincent W. V. Jaddoe^{25,96,97}, Johanna Jakobsdottir¹³⁶, Jost B. Jonas^{137,138}, Peter K. Joshi⁵⁸, Navya Shilpa Josyula¹³⁹, Bettina Jung⁶, Mika Kähönen^{140,141}, Yoichiro Kamatani^{27,142}, Candace M. Kammerer¹⁴³, Masahiro Kanai^{27,144}, Mika Kastarinen¹⁴⁵, Shona M. Kerr²⁰, Chiea-Chuen Khor^{64,87}, Wieland Kiess^{12,84,85}, Marcus E. Kleber⁸¹, Wolfgang Koenig^{146,147,148}, Jaspal S. Kooner^{61,62,63,149}, Antje Körner^{12,84,85}, Peter Kovacs¹⁵⁰, Aldi T. Kraja⁵, Alena Krajcoviechova^{69,70}, Holly Kramer^{17,151}, Bernhard K. Krämer⁸¹, Florian Kronenberg¹⁵², Michiaki Kubo¹⁵³, Brigitte Kühnel¹⁰⁵, Mikko Kuokkanen^{154,155}, Johanna Kuusisto^{145,156}, Martina La Bianca²¹, Markku Laakso^{145,156}, Leslie A. Lange¹⁵⁷, Carl D. Langefeld⁸⁶, Jeannette Jen-Mai Lee¹⁵, Benjamin Lehne²⁹, Terho Lehtimäki^{158,159}, Wolfgang Lieb¹⁶⁰, Lifelines Cohort Study¹⁶¹, Su-Chi Lim^{15,162}, Lars Lind¹⁶³, Cecilia M. Lindgren^{164,165}, Jun Liu²⁵, Jianjun Liu^{87,166}, Markus Loeffler^{11,12}, Ruth J. F. Loos^{48,167}, Susanne Lucae¹³⁵, Mary Ann Lukas¹⁶⁸, Leo-Pekka Lyytikäinen^{158,159}, Reedik Mägi¹¹², Patrik K. E. Magnusson⁶⁵, Anubha Mahajan^{169,170}, Nicholas G. Martin¹⁰⁹, Jade Martins¹⁷¹, Winfried März^{172,173,174}, Deborah Mascalzoni¹⁹, Koichi Matsuda¹⁷⁵, Christa Meisinger^{176,177}, Thomas Meitinger^{147,178,179}, Olle Melander¹⁸⁰, Andres Metspalu¹¹², Evgenia K. Mikaelsdottir¹⁶, Yuri Milaneschi¹⁸¹, Kozeta Miliku^{25,96,97}, Pashupati P. Mishra^{158,159}, V. A. Million Veteran Program¹⁶¹, Karen L. Mohlke¹⁸², Nina Mononen^{158,159}, Grant W. Montgomery¹⁸³, Dennis O. Mook-Kanamori^{78,184}, Josyf C. Mychaleckyj¹⁸⁵, Girish N. Nadkarni^{48,186}, Mike A. Nalls^{187,188}, Matthias Nauck^{23,189}, Kjell Nikus^{190,191}, Boting Ning¹⁹², Ilja M. Nolte²⁴, Raymond Noordam¹⁹³, Jeffrey O'Connell¹⁹⁴, Michelle L. O'Donoghue^{195,196}, Isleifur Olafsson¹⁹⁷, Albertine J. Oldehinkel¹¹⁶, Marju Orho-Melander³², Willem H. Ouwehand⁷⁷, Sandosh Padmanabhan¹⁹⁸, Nicholette D. Palmer¹⁹⁹, Runolfur Palsson^{111,130}, Brenda W. J. H. Penninx¹⁸¹, Thomas Perls²⁰⁰, Markus Perola²⁰¹, Mario Pirastu²⁰², Nicola Pirastu⁵⁸, Giorgio Pistis²⁰³, Anna I. Podgornaia¹⁰, Ozren Polasek^{204,205}, Belen Ponte²⁰⁶, David J. Porteous^{57,207}, Tanja Poulain¹², Peter P. Pramstaller¹⁹, Michael H. Preuss⁴⁸, Bram P. Prins⁵⁵, Michael A. Province⁵, Ton J. Rabelink^{79,208}, Laura M. Raffield¹⁸², Olli T. Raitakari^{209,210}, Dermot F. Reilly¹⁰, Rainer Rettig²¹¹, Myriam Rheinberger⁶, Kenneth M. Rice⁴³, Paul M. Ridker^{108,212}, Fernando Rivadeneira^{25,213}, Federica Rizzi^{214,215}, David J. Roberts²¹⁶, Antonietta Robino²¹, Peter Rossing³¹, Igor Rudan⁵⁸, Rico Rueedi^{39,40}, Daniela Ruggiero^{18,71}, Kathleen A. Ryan²¹⁷, Yasaman Saba²¹⁸, Charumathi Sabanayagam⁶⁴, Veikko Salomaa²⁰¹, Erika Salvi^{214,219}, Kai-Uwe Saum⁵⁰, Helena Schmidt²²⁰, Reinhold Schmidt¹²², Ben Schöttker^{50,51}, Christina-Alexandra Schulz³², Nicole Schupf^{221,222,223}, Christian M. Shaffer⁵⁹, Yuan Shi⁶⁴, Albert V. Smith¹¹¹, Blair H. Smith²²⁴, Nicole Soranzo²²⁵, Cassandra N. Spracklen¹⁸², Konstantin Strauch^{226,227}, Heather M. Stringham⁴⁵, Michael Stumvoll²²⁸, Per O. Svensson^{229,230}, Silke Szymczak¹⁰⁰, E-Shyong Tai^{15,166,231}, Salman M. Tajuddin⁹⁵, Nicholas Y. Q. Tan⁶⁴, Kent D. Taylor²³², Andrej Teren^{12,233}, Yih-Chung Tham⁶⁴, Joachim Thiery^{12,53}, Chris H. L. Thio²⁴, Hauke Thomsen²³⁴, Gudmar Thorleifsson¹⁶, Daniela Toniolo⁷⁴, Anke Tönjes²²⁸, Johanne Tremblay^{113,235}, Ioanna Tzoulaki^{90,236}, André G. Uitterlinden²¹³, Simona Vaccargiu²⁰², Rob M. van Dam^{15,166}, Pim van der Harst^{237,238,239}, Cornelia M. van Duijn²⁵, Digna R. Velez Edward^{119,240}, Niek Verweij²³⁷,

Suzanne Vorgelezang^{25,96,97}, Uwe Völker^{23,241}, Peter Vollenweider²⁴², Gerard Waeber²⁴²,
 Melanie Waldenberger^{105,106,147}, Lars Wallentin^{243,244}, Ya Xing Wang¹³⁸, Chaolong Wang^{87,245},
 Dawn M. Waterworth⁴, Wen Bin Wei²⁴⁶, Harvey White²⁴⁷, John B. Whitfield¹⁰⁹, Sarah H. Wild²⁴⁸,
 James F. Wilson^{20,58}, Mary K. Wojczynski⁵, Charlene Wong⁶⁷, Tien-Yin Wong^{64,67}, Liang Xu¹³⁸,
 Qiong Yang¹⁹², Masayuki Yasuda^{64,249}, Laura M. Yerges-Armstrong⁴, Weihua Zhang^{61,90},
 Alan B. Zonderman⁹⁵, Jerome I. Rotter^{232,250,251}, Murielle Bochud⁷³, Bruce M. Psaty^{252,253},
 Veronique Vitart²⁰, James G. Wilson²⁵⁴, Abbas Dehghan^{29,90}, Afshin Parsa^{255,256},
 Daniel I. Chasman^{108,212}, Kevin Ho^{257,258}, Andrew P. Morris^{72,169}, Olivier Devuyst²⁵⁹,
 Shreeram Akilesh^{37,260}, Sarah A. Pendergrass²⁶¹, Xueling Sim¹⁵, Carsten A. Böger^{6,262},
 Yukinori Okada^{263,264}, Todd L. Edwards^{119,265}, Harold Snieder²⁴, Kari Stefansson¹⁶,
 Adriana M. Hung^{119,266}, Iris M. Heid^{7,268}, Markus Scholz^{11,12,268}, Alexander Teumer^{22,23,268},
 Anna Köttgen^{1,8,268*} and Cristian Pattaro^{19,268*}

¹Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Center-University of Freiburg, Freiburg, Germany. ²Renal Division, Department of Medicine IV, Faculty of Medicine and Medical Center-University of Freiburg, Freiburg, Germany. ³Division of Nephrology and Hypertension, Department of Medicine, University of Utah, Salt Lake City, USA. ⁴Target Sciences-Genetics, GlaxoSmithKline, Collegeville, PA, USA. ⁵Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA. ⁶Department of Nephrology, University Hospital Regensburg, Regensburg, Germany. ⁷Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany. ⁸Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. ⁹Welch Center for Prevention, Epidemiology and Clinical Research, Baltimore, MD, USA. ¹⁰Genetics, Merck & Co., Inc, Kenilworth, NJ, USA. ¹¹Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. ¹²LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany. ¹³Division of Quantitative Sciences, Department of Obstetrics & Gynecology, Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, USA. ¹⁴Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA. ¹⁵Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore. ¹⁶deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. ¹⁷Department of Public Health Sciences, Loyola University Chicago, Maywood, IL, USA. ¹⁸Institute of Genetics and Biophysics 'Adriano Buzzati-Traverso'-CNR, Naples, Italy. ¹⁹Eurac Research, Institute for Biomedicine (affiliated with the University of Lübeck), Bolzano, Italy. ²⁰Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. ²¹Institute for Maternal and Child Health, IRCCS 'Burlo Garofolo', Trieste, Italy. ²²Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. ²³DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany. ²⁴Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ²⁵Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. ²⁶Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. ²⁷Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences (IMS), Yokohama, Japan. ²⁸Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. ²⁹Department of Epidemiology and Biostatistics, Faculty of Medicine, School of Public Health, Imperial College London, London, UK. ³⁰Institute of Public Health & Social Sciences, Khyber Medical University, Peshawar, Pakistan. ³¹Steno Diabetes Center Copenhagen, Gentofte, Denmark. ³²Diabetes and Cardiovascular Disease-Genetic Epidemiology, Department of Clinical Sciences in Malmö, Lund University, Malmö, Sweden. ³³Division of Family Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden. ³⁴School of Health and Social Studies, Dalarna University, Stockholm, Sweden. ³⁵Division of Nephrology, Department of Internal Medicine, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ³⁶Division of Nephrology, University of Washington, Seattle, WA, USA. ³⁷Kidney Research Institute, University of Washington, Seattle, WA, USA. ³⁸Cardiology, Geneva University Hospitals, Geneva, Switzerland. ³⁹Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. ⁴⁰Swiss Institute of Bioinformatics, Lausanne, Switzerland. ⁴¹Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa. ⁴²Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. ⁴³Department of Biostatistics, University of Washington, Seattle, WA, USA. ⁴⁴Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy. ⁴⁵Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. ⁴⁶Human Genetics Center, University of Texas Health Science Center, Houston, TX, USA. ⁴⁷CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, University of Lille, Lille, France. ⁴⁸Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁴⁹Digital Health Center, Hasso Plattner Institute and University of Potsdam, Potsdam, Germany. ⁵⁰Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁵¹Network Aging Research, University of Heidelberg, Heidelberg, Germany. ⁵²Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy. ⁵³Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig, Leipzig, Germany. ⁵⁴Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany. ⁵⁵MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ⁵⁶National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK. ⁵⁷Center for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. ⁵⁸Center for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. ⁵⁹Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. ⁶⁰Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. ⁶¹Department of Cardiology, Ealing Hospital, Middlesex, UK. ⁶²Imperial College Healthcare NHS Trust, Imperial College London, London, UK. ⁶³MRC-PHE Center for Environment and Health, School of Public Health, Imperial College London, London, UK. ⁶⁴Singapore Eye Research Institute, Singapore National Eye Center, Singapore, Singapore. ⁶⁵Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ⁶⁶Ophthalmology and Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore.

⁶⁷Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. ⁶⁸Unit of Epidemiology, Biostatistics and Biodemography, Department of Public Health, Southern Denmark University, Odense, Denmark. ⁶⁹Center for Cardiovascular Prevention, Charles University in Prague, First Faculty of Medicine and Thomayer Hospital, Prague, Czech Republic. ⁷⁰Department of Medicine II, Charles University in Prague, First Faculty of Medicine, Prague, Czech Republic. ⁷¹IRCCS Neuromed, Pozzilli, Italy. ⁷²Department of Biostatistics, University of Liverpool, Liverpool, UK. ⁷³Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland. ⁷⁴San Raffaele Research Institute, Milan, Italy. ⁷⁵Institute of Biomedical Technologies, National Research Council of Italy, Milan, Italy. ⁷⁶Bio4Dreams–Business Nursery for Life Sciences, Milan, Italy. ⁷⁷Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge, UK. ⁷⁸Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. ⁷⁹Section of Nephrology, Department of Internal Medicine, Leiden University Medical Center, Leiden, the Netherlands. ⁸⁰Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany. ⁸¹Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany. ⁸²Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ⁸³NHS Blood and Transplant, Cambridge, UK. ⁸⁴Department of Women and Child Health, Hospital for Children and Adolescents, University of Leipzig, Leipzig, Germany. ⁸⁵Center for Pediatric Research, University of Leipzig, Leipzig, Germany. ⁸⁶Public Health Sciences–Biostatistics, Wake Forest School of Medicine, Winston-Salem, NC, USA. ⁸⁷Genome Institute of Singapore, Agency for Science Technology and Research, Singapore, Singapore. ⁸⁸Intensive Care Medicine, Charité, Berlin, Germany. ⁸⁹Department of Nephrology and Hypertension, Friedrich Alexander University Erlangen-Nürnberg (FAU), Erlangen, Germany. ⁹⁰Department of Epidemiology and Biostatistics, MRC-PHE Center for Environment and Health, School of Public Health, Imperial College London, London, UK. ⁹¹Imperial College NIHR Biomedical Research Center, Imperial College London, London, UK. ⁹²Dementia Research Institute, Imperial College London, London, UK. ⁹³Health Data Research UK–London, London, UK. ⁹⁴Department of Anatomy and Cell Biology, University Medicine Greifswald, Greifswald, Germany. ⁹⁵Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, US National Institutes of Health, Baltimore, MD, USA. ⁹⁶Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. ⁹⁷Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. ⁹⁸Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland. ⁹⁹Section on Nephrology, Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA. ¹⁰⁰Institute of Medical Informatics and Statistics, Kiel University, University Hospital Schleswig-Holstein, Kiel, Germany. ¹⁰¹School of Public Health and Community Medicine, Hebrew University of Jerusalem, Jerusalem, Israel. ¹⁰²Department of Genomics of Common Disease, Imperial College London, London, UK. ¹⁰³Massachusetts Veterans Epidemiology Research and Information Center, VA Cooperative Studies Program, VA Boston Healthcare System, Boston, MA, USA. ¹⁰⁴Molecular Geriatrics, Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden. ¹⁰⁵Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. ¹⁰⁶Institute of Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. ¹⁰⁷German Center for Diabetes Research (DZD), Neuherberg, Germany. ¹⁰⁸Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA. ¹⁰⁹QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. ¹¹⁰Icelandic Heart Association, Kopavogur, Iceland. ¹¹¹Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. ¹¹²Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia. ¹¹³Montreal University Hospital Research Center, CHUM, Montreal, Quebec, Canada. ¹¹⁴Medpharmgene, Montreal, Quebec, Canada. ¹¹⁵Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, US National Institutes of Health, Bethesda, MD, USA. ¹¹⁶Interdisciplinary Center of Psychopathology and Emotion Regulation (ICPE), University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ¹¹⁷Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. ¹¹⁸Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. ¹¹⁹Department of Veteran's Affairs, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA. ¹²⁰Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ¹²¹Khoo Teck Puat–National University Children's Medical Institute, National University Health System, Singapore, Singapore. ¹²²Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria. ¹²³Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria. ¹²⁴Department of Genetics, Shanghai–MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center, Shanghai, China. ¹²⁵Shanghai Industrial Technology Institute, Shanghai, China. ¹²⁶Department of Pediatrics, Tampere University Hospital, Tampere, Finland. ¹²⁷Department of Pediatrics, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. ¹²⁸NHLBI's Framingham Heart Study, Framingham, MA, USA. ¹²⁹The Center for Population Studies, NHLBI, Framingham, MA, USA. ¹³⁰Division of Nephrology, Internal Medicine Services, Landspítali–The National University Hospital of Iceland, Reykjavik, Iceland. ¹³¹Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. ¹³²Stanford Cardiovascular Institute, Stanford University, Stanford, CA, USA. ¹³³Molecular Epidemiology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden. ¹³⁴Stanford Diabetes Research Center, Stanford University, Stanford, CA, USA. ¹³⁵Max Planck Institute of Psychiatry, Munich, Germany. ¹³⁶The Center of Public Health Sciences, University of Iceland, Reykjavik, Iceland. ¹³⁷Department of Ophthalmology, Medical Faculty Mannheim, University Heidelberg, Mannheim, Germany. ¹³⁸Beijing Institute of Ophthalmology, Beijing Key Laboratory of Ophthalmology and Visual Sciences, Beijing Tongren Hospital, Capital Medical University, Beijing, China. ¹³⁹Geisinger Research, Biomedical and Translational Informatics Institute, Rockville, MD, USA. ¹⁴⁰Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland. ¹⁴¹Department of Clinical Physiology, Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. ¹⁴²Kyoto–McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. ¹⁴³Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. ¹⁴⁴Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. ¹⁴⁵Department of Medicine, Kuopio University Hospital, Kuopio, Finland. ¹⁴⁶Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. ¹⁴⁷DZHK (German Center for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany. ¹⁴⁸Institute of Epidemiology and Biostatistics, University of Ulm, Ulm, Germany. ¹⁴⁹National Heart and Lung Institute, Imperial College London, London, UK. ¹⁵⁰Integrated Research and Treatment Center Adiposity Diseases, University of Leipzig, Leipzig, Germany. ¹⁵¹Division of Nephrology and Hypertension, Loyola University Chicago, Chicago, IL, USA. ¹⁵²Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria. ¹⁵³RIKEN Center for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan. ¹⁵⁴The Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland. ¹⁵⁵Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland. ¹⁵⁶Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland. ¹⁵⁷Division of Biomedical Informatics and Personalized Medicine, School of Medicine, University of Colorado Denver–Anschutz Medical Campus, Aurora, CO, USA. ¹⁵⁸Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. ¹⁵⁹Department of Clinical Chemistry, Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. ¹⁶⁰Institute of Epidemiology and Biobank Popgen, Kiel University, Kiel, Germany. ¹⁶¹A list of members and affiliations appears in the Supplementary Note. ¹⁶²Diabetes Center, Khoo Teck Puat Hospital, Singapore, Singapore. ¹⁶³Cardiovascular Epidemiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden. ¹⁶⁴Nuffield Department of Medicine, University of Oxford, Oxford, UK. ¹⁶⁵Broad Institute of Harvard and MIT, Cambridge, MA, USA. ¹⁶⁶Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. ¹⁶⁷The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

¹⁶⁸Target Sciences—Genetics, GlaxoSmithKline, Albuquerque, NM, USA. ¹⁶⁹Wellcome Trust Center for Human Genetics, University of Oxford, Oxford, UK. ¹⁷⁰Oxford Center for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. ¹⁷¹Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. ¹⁷²Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany. ¹⁷³Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria. ¹⁷⁴Medical Clinic V, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany. ¹⁷⁵Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. ¹⁷⁶Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. ¹⁷⁷Chair of Epidemiology, Ludwig-Maximilians-Universität München at UNIKA-T Augsburg, Augsburg, Germany. ¹⁷⁸Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany. ¹⁷⁹Institute of Human Genetics, Technische Universität München, Munich, Germany. ¹⁸⁰Hypertension and Cardiovascular Disease, Department of Clinical Sciences Malmö, Lund University, Malmö, Sweden. ¹⁸¹Department of Psychiatry, VU University Medical Center, Amsterdam, the Netherlands. ¹⁸²Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. ¹⁸³Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland, Australia. ¹⁸⁴Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands. ¹⁸⁵Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. ¹⁸⁶Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ¹⁸⁷Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. ¹⁸⁸Data Tecnica International, Glen Echo, MD, USA. ¹⁸⁹Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. ¹⁹⁰Department of Cardiology, Heart Center, Tampere University Hospital, Tampere, Finland. ¹⁹¹Department of Cardiology, Finnish Cardiovascular Research Center—Tampere, Faculty of Medicine and Life Sciences, Tampere University, Tampere, Finland. ¹⁹²Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. ¹⁹³Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Center, Leiden, the Netherlands. ¹⁹⁴University of Maryland School of Medicine, Baltimore, MD, USA. ¹⁹⁵Cardiovascular Division, Brigham and Women's Hospital, Boston, MA, USA. ¹⁹⁶TIMI Study Group, Boston, MA, USA. ¹⁹⁷Department of Clinical Biochemistry, Landspítali University Hospital, Reykjavik, Iceland. ¹⁹⁸Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. ¹⁹⁹Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. ²⁰⁰Department of Medicine, Geriatrics Section, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA. ²⁰¹National Institute for Health and Welfare, Helsinki, Finland. ²⁰²Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Li Punti, Sassari, Italy. ²⁰³Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland. ²⁰⁴Faculty of Medicine, University of Split, Split, Croatia. ²⁰⁵Gen-info Ltd, Zagreb, Croatia. ²⁰⁶Service de Néphrologie, Geneva University Hospitals, Geneva, Switzerland. ²⁰⁷Center for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. ²⁰⁸Einthoven Laboratory of Experimental Vascular Research, Leiden University Medical Center, Leiden, the Netherlands. ²⁰⁹Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. ²¹⁰Research Center of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. ²¹¹Institute of Physiology, University Medicine Greifswald, Karlsburg, Germany. ²¹²Harvard Medical School, Boston, MA, USA. ²¹³Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. ²¹⁴Department of Health Sciences, University of Milan, Milano, Italy. ²¹⁵ePhood Scientific Unit, ePhood SRL, Milano, Italy. ²¹⁶NHS Blood and Transplant, BRC Oxford Haematology Theme; Nuffield Division of Clinical Laboratory Sciences; University of Oxford, Oxford, UK. ²¹⁷Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA. ²¹⁸Molecular Biology and Biochemistry, Gottfried Schatz Research Center for Cell Signaling, Metabolism and Aging, Medical University of Graz, Graz, Austria. ²¹⁹Neurology Unit, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan, Italy. ²²⁰Institute of Molecular Biology and Biochemistry, Center for Molecular Medicine, Medical University of Graz, Graz, Austria. ²²¹Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, NY, USA. ²²²Gertrude H. Sergievsky Center, Columbia University Medical Center, New York, NY, USA. ²²³Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, NY, USA. ²²⁴Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. ²²⁵Human Genetics, Wellcome Sanger Institute, Hinxton, UK. ²²⁶Institute of Genetic Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. ²²⁷Chair of Genetic Epidemiology, IBE, Faculty of Medicine, Ludwig-Maximilians-Universität München, München, Germany. ²²⁸Department of Endocrinology and Nephrology, University of Leipzig, Leipzig, Germany. ²²⁹Department of Clinical Science and Education, Karolinska Institutet, Södersjukhuset, Stockholm, Sweden. ²³⁰Department of Cardiology, Södersjukhuset, Stockholm, Sweden. ²³¹Duke-NUS Medical School, Singapore, Singapore. ²³²The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA. ²³³Heart Center Leipzig, Leipzig, Germany. ²³⁴Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ²³⁵CRCHUM, Montreal, Canada. ²³⁶Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece. ²³⁷Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ²³⁸Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ²³⁹Durrer Center for Cardiovascular Research, The Netherlands Heart Institute, Utrecht, the Netherlands. ²⁴⁰Department of Obstetrics and Gynecology, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, USA. ²⁴¹Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. ²⁴²Internal Medicine, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland. ²⁴³Cardiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden. ²⁴⁴Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden. ²⁴⁵School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. ²⁴⁶Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China. ²⁴⁷Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, New Zealand. ²⁴⁸Center for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. ²⁴⁹Department of Ophthalmology, Tohoku University Graduate School of Medicine, Miyagi, Japan. ²⁵⁰Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA, USA. ²⁵¹Department of Medicine, Harbor-UCLA Medical Center, Torrance, CA, USA. ²⁵²Cardiovascular Health Research Unit, Department of Medicine, Department of Epidemiology, Department of Health Service, University of Washington, Seattle, WA, USA. ²⁵³Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. ²⁵⁴Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA. ²⁵⁵Division of Kidney, Urologic and Hematologic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. ²⁵⁶Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. ²⁵⁷Kidney Health Research Institute (KHRI), Geisinger, Danville, PA, USA. ²⁵⁸Department of Nephrology, Geisinger, Danville, PA, USA. ²⁵⁹Institute of Physiology, University of Zurich, Zurich, Switzerland. ²⁶⁰Anatomic Pathology, University of Washington Medical Center, Seattle, WA, USA. ²⁶¹Geisinger Research, Biomedical and Translational Informatics Institute, Danville, PA, USA. ²⁶²Department of Nephrology and Rheumatology, Kliniken Südostbayern, Regensburg, Germany. ²⁶³Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences (IMS), Osaka, Japan. ²⁶⁴Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan. ²⁶⁵Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. ²⁶⁶Vanderbilt University Medical Center, Division of Nephrology & Hypertension, Nashville, TN, USA. ²⁶⁷These authors contributed equally: Matthias Wuttke, Yong Li, Man Li, Karsten B. Sieber, Mary F. Feitosa, Mathias Gorski. ²⁶⁸These authors jointly supervised this work: Iris M. Heid, Markus Scholz, Alexander Teumer, Anna Köttgen, Cristian Pattaro.

*e-mail: anna.koettgen@uniklinik-freiburg.de; cristian.pattaro@eurac.edu

Methods

Overview. We set up a collaborative meta-analysis based on a distributive data model and quality-control procedures. To maximize phenotype standardization across studies, an analysis plan and a command line script (<https://github.com/genepi-freiburg/ckdgen-pheno>) were created centrally and provided to all participating studies (mostly population-based studies; Supplementary Table 1). Data processing, analysis and troubleshooting instructions were distributed to all studies via a wiki system (https://ckdgen.eurac.edu/mediawiki/index.php/CKDGen_Round_4_EPACTS_analysis_plan). Automatically generated summary files were checked centrally. Upon phenotype approval, studies ran their GWAS and uploaded results and imputation quality (IQ) information to a common calculation server. GWAS quality control was performed with GWAToolbox⁶⁰ and custom scripts to assess ancestry-matched allele frequencies and variant positions. All studies had their own research protocols approved by the respective local ethics committees. All participants in all studies provided written informed consent.

Phenotype definition. Each study measured serum creatinine and BUN concentrations as described in Supplementary Table 1. Creatinine values obtained with a Jaffé assay before 2009 were calibrated by multiplying by 0.95 (ref. ⁶¹). Studies on adults (>18 years of age) estimated GFR with the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation⁶², by using the R package *nephro*⁶³. Studies on individuals who were 18 years old or younger used the Schwartz formula⁶⁴. eGFR was winsorized at 15 and 200 ml min⁻¹ per 1.73 m². CKD was defined as an eGFR below 60 ml min⁻¹ per 1.73 m². In studies reporting blood urea measurements, BUN was derived as blood urea × 2.8, with units expressed as mg dl⁻¹.

Genotyping and genotype imputation. Genotypes were imputed on the basis of the Haplotype Reference Consortium v1.1 or 1000 Genomes Project phase 3 v5 (1000Gp3v5) ALL or phase 1 v3 (1000Gp1v3) ALL panel. Imputed variants were coded as allelic dosages accompanied by the corresponding IQ scores (IMPUTE2 info score, MACH/minimac RSQ or as applicable) and annotated on the NCBI b37 (hg19) reference build (see Supplementary Table 2 for study-specific genotyping arrays, haplotype phasing and genotype imputation methods).

Genome-wide association studies. Each study fitted sex- and age-adjusted linear regression models to log(eGFR) and BUN. Regression residuals were regressed on SNP dosage, assuming an additive genetic model. Study site, genetic principal components, relatedness and other study-specific features were accounted for in the study-specific models as appropriate (Supplementary Table 2). Logistic regression models were fitted for CKD.

Trans-ancestry GWAS meta-analysis. Studies contributed 121 GWAS summary statistics files for eGFR (total post-quality-control $n = 765,348$), 60 GWAS files for CKD (total post-quality-control $n = 625,219$, including 64,164 CKD cases) and 65 GWAS files for BUN (total post-quality-control $n = 416,178$). Ancestry-specific details for eGFR, CKD and BUN are given in Supplementary Table 1.

Before meta-analysis, study-specific GWAS files were filtered to retain only variants with IQ score > 0.6 and minor allele count (MAC) > 10, and genomic control (GC) correction was applied in the case where GC factor $\lambda_{GC} > 1$. Fixed-effects inverse-variance-weighted meta-analysis was performed with METAL⁶⁵, which was adapted to increase the precision of effect estimates and their standard errors (seven decimal places instead of four).

After meta-analysis of 43,994,957 SNPs, only SNPs present in ≥50% of the GWAS files and with total MAC ≥ 400 were retained. Across ancestry groups, this yielded 8,221,591 variants for eGFR (8,834,748 in European ancestry), 8,176,554 variants for BUN (8,358,347 in European ancestry) and 9,585,923 variants for CKD. Post-meta-analysis GC correction was not applied (LD score regression intercept ≈ 1 in all analyses of eGFR, BUN and CKD)⁶⁶. The genome-wide significance level was set at 5×10^{-8} . Between-study heterogeneity was assessed with the I^2 statistic⁶⁷. For CKD, variants with $I^2 \geq 95\%$ were removed to moderate the influence of single large studies. Variants were assigned to loci by selecting the SNP with the lowest P value across the genome as the index SNP, defining the corresponding locus as the 1-Mb segment centered on the index SNP, and repeating the procedure until no further genome-wide-significant SNPs remained. The extended major histocompatibility complex (MHC) region was considered as a single locus. A locus was considered to be new if not containing any variant identified by previous GWAS of eGFR.

Meta-regression analysis of trans-ancestry GWAS. For eGFR, we evaluated ancestry-related heterogeneity by using the software Meta-Regression of Multi-Ethnic Genetic Association (MR-MEGA, v0.1.2)⁶⁸ with study-specific GWAS results. Meta-regression models included three axes of genetic variation. Genomic control correction was applied to the meta-regression results. The 308 genome-wide-significant index SNPs from the trans-ancestry GWAS meta-analysis were tested for ancestry-related heterogeneity of the allelic effects at a significance level of $0.05/308 = 1.6 \times 10^{-4}$ (referring to the corresponding P value as $P_{\text{anc-het}}$).

Variance explained and genetic heritability. The proportion of phenotypic variance explained by the index SNPs was estimated as $\beta^2 \left(\frac{2p(1-p)}{\text{var}} \right)$, with β being

the SNP effect, p the effect allele frequency and var the variance of the sex- and age-adjusted log(eGFR) residuals (assumed to be 0.016 on the basis of data from 11,827 European-ancestry participants of the population-based ARIC study)⁹. Genetic heritability for age- and sex-adjusted log(eGFR) was estimated with the R package MCMCglmm⁶⁹ on the Cooperative Health Research in South Tyrol (CHRIS) study⁷⁰, a participating pedigree-based study with 186 pedigrees of up to five generations ($n = 4,373$)⁷¹. We fitted two models with and without inclusion of the identified index SNPs (304/308), running 1 million MCMC iterations (burn-in = 500,000)⁷¹.

Comparison with and replication of results in the MVP. The eGFR-associated SNPs identified in the discovery GWAS meta-analyses were tested for replication in a GWAS from the MVP²³, an independent trans-ancestry study with participants recruited across 63 US Veterans Administration (VA) medical facilities. Written informed consent was obtained and all documents and protocols were approved by the VA Central Institutional Review Board. After genotyping and quality control, genotypes were phased and imputed on the 1000Gp3v5 reference panel. Serum creatinine was assessed up to 1 year before MVP enrollment by isotope dilution mass spectrometry. GFR was estimated by using the CKD-EPI equation⁶² after excluding subjects on dialysis, transplant patients, amputees, individuals on HIV medications and those with creatinine values of <0.4 mg dl⁻¹. GWAS of eGFR on SNP dosage were performed by fitting linear regression models adjusted for age at creatinine measurement, age², sex, body-mass index and the first ten genetic principal components, by using SNPTEST v2.5.4-beta⁷². All GWAS were stratified by self-reported ancestry (79.6% white non-Hispanic, 20.4% black non-Hispanic), diabetes and hypertension status. Results were combined across strata by fixed-effects inverse-variance-weighted meta-analysis in METAL⁶⁵. This analysis encompassed a total of 280,722 individuals across all strata, of whom 216,518 were non-Hispanic whites (European ancestry). The MVP is described more extensively in the Supplementary Note 3.

Of the 308 eGFR index SNPs identified in the CKDGen trans-ancestry analysis, 305 variants or their good proxies were available in the MVP GWAS (proxies were required to have $r^2 \geq 0.8$ with the index SNP and were selected by maximum r^2 followed by minimum distance in the case of ties). Replication testing of the 256 European-ancestry-specific index SNPs was restricted to the MVP European-ancestry GWAS. CKDGen and MVP meta-analysis results were pooled via sample-size-weighted meta-analysis of z scores with METAL⁶⁵. In both the trans-ancestry and European-ancestry-specific analyses, replication was defined by one-sided $P < 0.05$ in the MVP and genome-wide significance of the CKDGen and MVP meta-analysis result.

Assessment of relevance to kidney function with BUN. We used genetic associations with BUN to assess replicated eGFR-associated SNPs with respect to their potential relevance to kidney function. Support for kidney function relevance was categorized as 'likely' for all eGFR index SNPs with an inverse, significant (one-sided $P < 0.05$) association with BUN for a given reference allele, 'inconclusive' for eGFR index SNPs whose effect on BUN was not different from 0 ($P \geq 0.05$) and 'unlikely' for all eGFR index SNPs with a concordant, significant (one-sided $P < 0.05$) association with BUN for a given reference allele.

Genetic risk score analysis in the UK Biobank dataset. To test the combined effect of eGFR-associated SNPs on outcomes related to clinically diagnosed CKD, a GRS-based association analysis was conducted on the basis of summary GWAS results, as described previously^{73,74}. Genetic association results with diseases were obtained for 452,264 UK Biobank participants available in the GeneAtlas⁷⁵ database for glomerular diseases (ICD-10 codes N00–N08; 2,289 cases); acute renal failure (N17; 4,913 cases); chronic renal failure (N18; 4,905 cases); urolithiasis (N20–N23; 7,053 cases); hypertensive diseases (I10–I15; 84,910 cases); and ischemic heart diseases (I20–I25; 33,387 cases). Asthma (J45; 28,628 cases) was included as a negative control. The log(estimated OR) value provided by the GeneAtlas PheWAS interface was used as the effect size, and its standard error was calculated from the corresponding effect size and P value. When OR = 1, the standard error was imputed by the median value of the remaining associations of the trait. Of the 147 eGFR index SNPs from the trans-ancestry GWAS meta-analysis that were replicated and showed likely relevance to kidney function, 144 were available in the UK Biobank dataset, and 259 of all 264 replicated trans-ancestry GWAS meta-analysis SNPs were available. The effect of the GRS association (β) corresponds to the OR for the disease depending on the relative change in eGFR, for example, OR = 1.10^β for a 10% change in eGFR. Alternatively, $\exp(\beta)$ can be interpreted as the OR for the disease per unit change in log(eGFR).

Genetic correlations with other complex traits and diseases. Genome-wide genetic correlation analysis was performed to investigate evidence of co-regulation or shared genetic basis between eGFR and BUN concentrations and other complex traits and diseases, both known and not known to correlate with eGFR and BUN. We estimated pairwise genetic correlation coefficients (r_g) between the results of our trans-ancestry meta-analyses of eGFR and BUN and each of 748 precomputed and publicly available GWAS summary statistics for complex traits and diseases available through LD Hub v1.9.0 by using LD score regression²⁸. An overview of

the sources of these summary statistics and their corresponding sample sizes is available at <http://ldsc.broadinstitute.org/>. Statistical significance was assessed at the Bonferroni-corrected level of $0.05/748 = 6.7 \times 10^{-5}$.

Pathway and tissue enrichment analysis. We used DEPICT v1 release 194 to perform DEPICT analysis²⁹, including pathway/gene set enrichment and tissue/cell-type analyses as described previously^{3,10}. All 14,461 gene sets were reconstituted by identifying genes that were transcriptionally co-regulated with other genes in a panel of 77,840 gene expression microarrays³⁶, from mouse knockout studies, and molecular pathways from protein–protein interaction screening. In the tissue and cell-type enrichment analysis, we tested whether genes in associated regions were highly expressed in 209 MeSH annotation categories for 37,427 microarrays (Affymetrix U133 Plus 2.0 array platform). For both eGFR and BUN, we included all variants associated with the trait at $P < 5 \times 10^{-8}$ in the trans-ancestry meta-analysis. Independent variant clumping was performed by using PLINK 1.9 (ref. 77) with 500-kb flanking regions and $r^2 > 0.01$ in the 1000Gp1v3 dataset. After excluding the MHC region, DEPICT was run with 500 repetitions to estimate the FDR and 5,000 permutations to compute P values adjusted for gene length by using 500 null GWAS. All significant gene sets were merged into meta gene sets by running an affinity propagation algorithm⁷⁸ implemented in the Python scikit-learn package (<http://scikit-learn.org/>). The resulting network was visualized with Cytoscape (<http://cytoscape.org/>).

Enrichment of heritability by cell-type group. We used stratified LD score regression to investigate important tissues and cell types on the basis of the trans-ancestry eGFR and BUN meta-analysis results. Heritability enrichment in ten cell-type groups was assessed by using the default options of stratified LD score regression described previously³⁰. The ten cell-type groups were collapsed from 220 cell-type-specific regulatory annotations for the four histone marks H3K4me1, H3K4me3, H3K9ac and H3K27ac. Enrichment in a cell-type category was defined as the proportion of SNP heritability in that group divided by the proportion of SNPs in the same cell-type group.

Analysis of genes causing kidney phenotypes in mice. A nested candidate gene analysis was performed with GenToS⁷⁹ to identify additional genetic associations that were not genome-wide significant. Candidate genes that when manipulated cause kidney phenotypes in mice were selected with the comprehensive MGI phenotype ontology in September 2017 (abnormal renal glomerular filtration rate (MP:0002847); abnormal kidney morphology (MP:0002135); abnormal kidney physiology (MP:0002136)). The human orthologs of these genes were obtained, when available, with the Human–Mouse: Disease Connection webtool (<http://www.informatics.jax.org/humanDisease.html>). Statistical significance was defined as Bonferroni correction of a type I error level of 0.05 for the number of independent common SNPs across all genes in each of the three candidate gene lists plus their flanking regions, derived from an ancestry-matched reference population. The GWAS meta-analysis summary statistics for eGFR were queried for significantly associated SNPs mapping to the selected candidate genes. Enrichment of significant genetic associations in genes within each candidate list was computed from the complementary cumulative binomial distribution⁷⁹. GenToS was used with default parameters on each of the three candidate gene lists, with the 1000 Genomes phase 3 release 2 ALL dataset as reference.

Independent variant identification in the European-ancestry meta-analysis. To identify additional independent eGFR-associated variants within the European-ancestry-specific and replicated loci, approximate conditional analyses were performed on the basis of genome-wide discovery summary statistics that incorporated LD information from an ancestry-matched reference population. These analyses were restricted to participants of European ancestry because an LD reference sample scaled to the size of our meta-analysis could only be constructed from publicly available data for European-ancestry individuals²⁴, for which we randomly selected 15,000 UK Biobank participants (dataset ID 8974). Individuals who withdrew consent and those not meeting data cleaning requirements were excluded, keeping only those who passed a sex-consistency check, had a $\geq 95\%$ call rate and did not represent outliers with respect to SNP heterozygosity. For each pair of individuals, the proportion of variants shared identical by descent (IBD) was computed with PLINK⁸⁰. Only one member of each pair with an IBD coefficient ≥ 0.1875 was retained. Individuals were restricted to those of European ancestry by excluding outliers along the first two principal components from a principal-component analysis seeded with the HapMap phase 3 release 2 populations as reference. The final dataset to estimate LD included 13,558 European-ancestry individuals and 16,969,363 SNPs.

The basis for statistical fine-mapping was the 228 1-Mb genome-wide-significant loci identified in the European-ancestry meta-analysis, clipping at chromosome borders. Overlapping loci as well as pairs of loci whose respective index SNPs were correlated ($r^2 > 0.1$ in the UK Biobank LD dataset described above) were merged. A single SNP was chosen to represent the MHC region, resulting in a final list of 189 regions before fine-mapping. Within each region, the GCTA COJO Slct algorithm⁸¹ was applied to identify independent variants by using a stepwise forward selection approach. We used the default collinearity cutoff

of 0.9 (sensitivity analyses showed no major influence of alternative cutoff values; data not shown). We deemed an additional SNP as independently genome-wide significant if the SNP's P value conditional on all previously identified SNPs in the same region was $< 5 \times 10^{-8}$.

Fine-mapping and credible sets in the European-ancestry meta-analysis. For each region containing multiple independent SNPs and for each independent SNP in such regions, approximate conditional analyses were conducted with the GCTA COJO-Cond algorithm to generate approximate conditional association statistics conditioned on the other independent SNPs in the region. By using Wakefield's formula implemented in the R package gtx⁸², we derived approximate Bayes factors (ABFs) from conditional estimates in regions with multiple independent SNPs and from the original estimates for regions with a single independent SNP. Given that 95% of the SNP effects on log(eGFR) fell within the range -0.01 to 0.01 , the standard deviation prior was chosen as 0.0051 on the basis of formula (8) in the original publication³². Sensitivity analyses showed that results were robust when higher values were used for the standard deviation prior (data not shown). For each variant within an evaluated region, the ABF obtained from the association β values and their standard errors for the marginal (single-signal regions) or conditional (multi-signal regions) estimates was used to calculate the PP for a SNP of driving the association signal ('causal variant'). We derived 99% credible sets, representing the SNP sets containing the variant(s) driving the association signal with 99% probability, by ranking variants by their PPs and adding them to the set until cumulative PP $> 99\%$ was reached in each region.

Variant annotation. Functional annotation of SNPs mapping to credible sets was performed with SNIpA v3.2 (March 2017)⁸³, on the basis of the 1000Gp3v5 and Ensembl v87 datasets. SNIpA was also used to derive the CADD Phred-like score⁸⁴, on the basis of CADD v1.3. The Ensembl VEP tool⁸⁵ was used for prediction of the primary effects of SNPs.

Colocalization of eGFR signal and gene expression in cis. As the great majority of gene expression datasets are generated on the basis of European-ancestry samples, colocalization analysis was based on genetic associations with eGFR in the European-ancestry sample and with gene expression (eQTLs) quantified from microdissected human glomerular and tubulo-interstitial kidney portions from 187 individuals from the NEPTUNE study⁴¹, as well as the 44 tissues included in the GTEx Project v6p release³⁸. The eQTL and GWAS effect alleles were harmonized. For each locus, we identified tissue gene pairs with reported eQTL data within ± 100 kb of each GWAS index SNP. The region for each colocalization test was the eQTL *cis* window defined in the underlying GTEx and NephQTL studies. We used the coloc.fast function, with the default setting, from the R package gtx (<https://github.com/tobyjohnson/gtx>), which is an adaptation of Giambartolomei's colocalization method⁸⁶. The gtx package was also used to estimate the direction of effect over the credible sets as the ratio of the average PP-weighted GWAS effects over the PP-weighted eQTL effects.

Trans-eQTL analysis. We performed *trans*-eQTL annotation through LD mapping on the basis of the 1000Gp3v5 European reference panel (r^2 cutoff of > 0.8). We limited annotation to replicated index SNPs with fine-mapping PP $\geq 1\%$. Owing to expected small effect sizes, only genome-wide *trans*-eQTL studies of either peripheral blood mononuclear cells or whole blood with $n \geq 1,000$ individuals were considered, resulting in five non-overlapping studies^{87–91} (Supplementary Table 14). For one study⁹¹, we had access to an update with larger sample size ($n = 6,645$) obtained by combining two non-overlapping studies (LIFE-Heart⁹² and LIFE-Adult⁹³). To improve the stringency of results, we focused the analysis on interchromosomal *trans*-eQTLs with $P < 5 \times 10^{-8}$ in ≥ 2 studies.

Colocalization with urinary uromodulin concentrations. Association of genetic variants with UUCR at the *UMOD*–*PDILT* locus was evaluated in the German Chronic Kidney Disease (GCKD) study⁹⁴. Uromodulin concentrations were measured from frozen stored urine by an established ELISA with excellent performance⁹⁵. Concentrations were indexed to creatinine to account for urine dilution. Genetic associations were assessed with the same software and settings as for eGFR association (Supplementary Table 2). Colocalization analyses were performed with identical software and settings as described above for the association with gene expression.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Genome-wide summary statistics for this study have been made publicly available at <http://ckdgen.imbi.uni-freiburg.de>.

References

- Fuchsberger, C., Taliun, D., Pramstaller, P. P. & Pattaro, C. GWAtoolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics* **28**, 444–445 (2012).

61. Coresh, J. et al. Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. *J. Am. Med. Assoc.* **311**, 2518–2531 (2014).
62. Levey, A. S. et al. A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150**, 604–612 (2009).
63. Pattaro, C. et al. Estimating the glomerular filtration rate in the general population using different equations: effects on classification and association. *Nephron Clin. Pract.* **123**, 102–111 (2013).
64. Schwartz, G. J. et al. Improved equations estimating GFR in children with chronic kidney disease using an immunonephelometric determination of cystatin C. *Kidney Int.* **82**, 445–453 (2012).
65. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
66. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
67. Higgins, J. P. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
68. Magi, R. et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639–3650 (2017).
69. Hadfield, J. MCMC methods for multi-response generalized linear mixed models: the MCMC glmm R package. *J. Stat. Softw.* **33**, 1–22 (2010).
70. Pattaro, C. et al. The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. *J. Transl. Med.* **13**, 348 (2015).
71. Noce, D. et al. Sequential recruitment of study participants may inflate genetic heritability estimates. *Hum. Genet.* **136**, 743–757 (2017).
72. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
73. Kottgen, A. et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat. Genet.* **45**, 145–154 (2013).
74. Dastani, Z. et al. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.* **8**, e1002607 (2012).
75. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).
76. Fehrmann, R. S. et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
77. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
78. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976 (2007).
79. Hoppmann, A. S., Schlosser, P., Backofen, R., Lausch, E. & Kottgen, A. GenToS: use of orthologous gene information to prioritize signals from human GWAS. *PLoS One* **11**, e0162466 (2016).
80. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
81. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
82. Wakefield, J. Bayes factors for genome-wide association studies: comparison with *P*-values. *Genet. Epidemiol.* **33**, 79–86 (2009).
83. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmuller, G. SNIIPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334–1336 (2015).
84. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
85. McLaren, W. et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
86. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
87. Zeller, T. et al. Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).
88. Fehrmann, R. S. et al. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.* **7**, e1002197 (2011).
89. Westra, H. J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
90. Joehanes, R. et al. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol.* **18**, 16 (2017).
91. Kirsten, H. et al. Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci. *Hum. Mol. Genet.* **24**, 4746–4763 (2015).
92. Beutner, F. et al. Rationale and design of the Leipzig (LIFE) Heart Study: phenotyping and cardiovascular characteristics of patients with coronary artery disease. *PLoS One* **6**, e29070 (2011).
93. Loeffler, M. et al. The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. *BMC Public Health* **15**, 691 (2015).
94. Eckardt, K. U. et al. The German Chronic Kidney Disease (GCKD) study: design and methods. *Nephrol. Dial. Transpl.* **27**, 1454–1460 (2012).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|--------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

To maximize phenotype standardization across studies, an analysis plan and a command line script (<https://github.com/genepi-freiburg/kdngen-pheno>) were created centrally and provided to all participating studies. Data processing, analysis and troubleshooting instructions were distributed to all studies via a Wiki system (https://ckdgen.eurac.edu/mediawiki/index.php/CKDGen_Round_4_EPACTS_analysis_plan). Automatically generated summary files were checked centrally. Upon phenotype approval, studies run their GWAS and uploaded results and imputation quality (IQ) information to a common calculation server. GWAS QC was performed using GWAToolbox and custom (R, Bash) scripts to assess ancestry-matched allele frequencies and variant positions.

Data analysis

See Supplementary Table 2 for study-specific genotyping arrays, haplotype phasing and genotype imputation methods. Fixed effects inverse-variance weighted meta-analysis was performed using METAL (version 2011-03-25), which was adapted to increase the precision of effect estimates and their standard errors (SE; seven decimal places instead of four). We evaluated ancestry-related heterogeneity using the software Meta-Regression of Multi-Ethnic Genetic Association (MR-MEGA v0.1.2). Genetic heritability was estimated using the R package 'MCMCglmm'. We estimated pairwise genetic correlation coefficients through LD Hub v1.9.0 using LD Score regression. We used DEPICT v1 release 194 to perform Data-Driven Expression Prioritized Integration for Complex Traits analysis. A nested candidate gene analysis was performed using GenToS (<https://github.com/genepi-freiburg/GenToS>). Approximate conditional analyses were conducted using the GCTA v1.91.6beta COJO-Cond and -Slt algorithms. Using the Wakefield's formula implemented in the R package 'gtx', we derived approximate Bayes factors (ABF). Functional annotation of SNPs mapping into credible sets was performed with SNiPA v3.2 (March 2017), based on the 1000Gp3v5 and Ensembl v87 datasets. SNiPA was also used to derive the Combined Annotation Dependent Depletion (CADD) PHRED-like score, based on CADD v1.3. The Ensembl VEP tool was used for SNP's primary effect prediction. For co-localization analyses, we used the 'coloc.fast' function from the R package 'gtx' (<https://github.com/tobyjohnson/gtx>), which is an adaption of Giambartolomei's co-localization method.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Genome-wide summary statistics for this study are made publicly available through dbGaP accession number phs000930.v7.p1 and at <http://ckdgen.imbi.uni-freiburg.de>.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Studies contributed 121 GWAS summary statistics files for eGFR (total post-QC n=765,348), 60 GWAS files for CKD (total post-QC n=625,219, including 64,164 CKD cases), and 65 GWAS files for BUN (total post-QC n=416,178).
Data exclusions	Before meta-analysis, study-specific GWAS files were filtered to retain only variants with IQ score>0.6 and minor allele count (MAC)>10. After meta-analysis of 43,994,957 SNPs, only SNPs present in ≥50% of the GWAS files and with total MAC≥400 were retained. Across ancestries, this yielded 8,221,591 variants for eGFR (8,834,748 in EA), 8,176,554 for BUN (8,358,347 in EA), and 9,585,923 for CKD.
Replication	The eGFR-associated SNPs identified in the discovery GWAS meta-analyses were tested for replication in a GWAS from the MVP, an independent trans-ethnic study with participants recruited across 63 U.S. Veteran's Administration (VA) medical facilities.
Randomization	This does not apply to our study. All samples were analyzed together.
Blinding	This does not apply to our study. No group allocation was necessary.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	See Supplementary Table 1 for population characteristics and phenotype distributions of all participating studies, including covariates such as age and sex.
Recruitment	See Supplementary Table 1 for a description of all participating studies and their study design.
Ethics oversight	All studies had their own research protocols approved by the respective local ethics committees. Participants provided written informed consent.

Note that full information on the approval of the study protocol must also be provided in the manuscript.