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.

**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 ml min\(^{-1}\) per 1.73 m\(^2\) (interquartile range, IQR: 81, 94). GWAS were based on genotypes imputed from HaploTyping Reference Consortium\(^{25}\) or 1000 Genomes Project\(^{26}\) 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 \(\lambda_{GC} = 1.04\); genomic control factor \(\lambda_{GC} = 1.05\)). 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 estimates\(^8\), 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 A full list of members and affiliations appears at the end of the paper.
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 participants\(^{35}\). 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 (\(P = 3.5 \times 10^{-10}\)) 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 (\(P > 0.05\)); and ‘unlikely’ for 15 eGFR index SNPs showing concordant, significant association with BUN (one-sided \(P < 0.05\)). 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, DDCDC1, WDR72, UMOD–PDILT, MYO19, AQP4, Nfatc1, PSD4, HOXD8, NRP1, 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; \(P = 8.1 \times 10^{-16}\)). 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)\(^{36}\). 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 (\(P < 5 \times 10^{-4}\)) to explore enriched pathways, tissues and cell types on the basis of gene expression data with DEPICT\(^{37}\). 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 (\(P < 1 \times 10^{-8}\), 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 \(P < 5 \times 10^{-4}\) 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 regression\(^{38}\) 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 mice\(^{41}\). 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 five-significant eGFR-associated SNPs or map near known loci, while the green line indicates genome-wide significance \((P = 5 \times 10^{-8})\). 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 \(P_{\text{anceh}} < 1.6 \times 10^{-4}\) for association with CKD 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-PDILT (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 ($-\log_2(P_{\text{anc-het}})$; 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 ($\leq 500$kb up- or downstream of the trans-ancestry index SNP), and
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 \(-\log_\text{P}\) 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 gene. The dashed line corresponds to genome-wide significance (\(P = 10^{-8}\)). The remaining 7 showed \(P \leq 3.3 \times 10^{-8}\) 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 set (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   | Credible set size | SNP PP | Functional consequence | CADD score | DHSs, tissue | Brief summary of the gene’s function and relevant literature |
|-----------|-------|-------------------|--------|------------------------|------------|--------------|----------------------------------------------------------|
| CACNA1S   | rs3850625 | 1                  | 1.00   | p.(Arg1539Cys)         | 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 and after indoxyl sulfate treatment. 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. |
| CPS1      | rs1047891 | 1                  | 1.00   | p.(Thr1406Asn)         | 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, serum metabolites and urinary glycine, as well as for many other quantitative biomarkers. This variant has been reported to associate with hyperammonemia after valproate therapy. |
| EDEM3     | rs78444298 | 1                  | 1.00   | p.(Pro746Ser)          | 24.6       | -            | The gene product accelerates proteasome-mediated ER-associated degradation of glycoproteins by catalyzing mannose trimming from Man8GlcNAc2 to yield Man7GlcNAc2 on N-glycans. This variant has been identified by a previous exome chip association study with eGFR. |
| KLHDC7A   | rs11261022 | 7                  | 0.71   | p.(Arg160Ser)          | 1.1        | Roadmap + ENCODE, kidney | This gene encodes the Kelch-domain-containing 7A protein and is a paralog of KBTBD11. No specific entry in relation to kidney disease in PubMed. |
| RPL3L     | rs113956264 | 1                  | 1.00   | p.(Val262Met)          | 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. |
| SLC25A45  | rs34400381 | 1                  | 1.00   | p.(Arg285Cys)          | 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 and in a whole-genome-sequencing analysis of serum creatinine. SLC25A45 may have a role in biosynthesis of arginine, which is involved in the synthesis of creatine. |
| SLC47A1   | rs111653425 | 1                  | 1.00   | p.(Ala465Val)          | 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. Rare and common variants in the locus have been identified in exome chip studies and GWAS of eGFR, respectively. Slc47a1-knockout mice show higher levels of serum creatinine and BUN, arguing against a sole effect on creatinine transport and supporting an effect on kidney function. |
| PPM1J     | rs34611728  | 5                  | 0.02   | p.(Leu213Phe)          | 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. |
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, PPM1I, 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 depletions (CADD) score greater than 15, indicating potential deleteriousness\(^a\). 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 BUN\(^a\) 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 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\(^b\) were associated with higher urinary uromodulin concentrations\(^b\). 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.

---

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

| Gene | SNP | Credible set size | SNP PP | Functional consequence | CADD score | DHSs, tissue | Brief summary of the gene’s function and relevant literature |
|------|-----|------------------|--------|------------------------|------------|-------------|----------------------------------------------------------|
| CERS2 | rs267738 | 5 | 0.46 | p.(Glu115 Ala) (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\(^b\). Cers2-knockout mice show strongly reduced ceramide levels in the kidney and develop renal parenchyma abnormalities\(^b\). This variant has been reported as associated with the rate of albuminuria increase in individuals with diabetes\(^b\). |
| C9 | 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\(^b\). |
| SLC22A2 | 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 SLC22A2 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\(^b\), indicating that this transporter is essential in excreting toxins. The locus has been reported in previous GWAS of eGFR\(^b\). |

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. Boldface indicates the SNPs most likely to be relevant for kidney function on the basis of combined effects on eGFR and BUN.

---

\(^{a}\) 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. Boldface indicates the SNPs most likely to be relevant for kidney function on the basis of combined effects on eGFR and BUN.

\(^{b}\) 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. Boldface indicates the SNPs most likely to be relevant for kidney function on the basis of combined effects on eGFR and BUN.
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 kinogen, which is cleaved to bradykinin. Bradykinin influences blood pressure, natriuresis and diuresis and can be linked to kidney function via the renin–angiotensin–aldosterone system\(^3\). 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, hemotocrit and multiple kidney-function-related traits (Supplementary Table 13). The eGFR index SNP rs1458038 \((PP > 50\%, \text{CADD score} = 14.8; \text{Supplementary Table 13})\) 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 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 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 CKD (Fig. 6b).

**Colocalization with uromodulin protein levels in urine.** The UMOD locus is of particular clinical interest for CKD research\(^4\); rare UMOD mutations cause autosomal dominant tubulo-interstitial kidney disease\(^5\), and common variants at UMOD give rise to the strongest eGFR and CKD GWAS signals\(^6\). 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...
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 group 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 variant, 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 GWAS 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 datasets, 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 biopsies, 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 cells.

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 loci for functional follow-up.

**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 -log10(P 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; r2 = 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.
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 cisimidine across the basolateral membrane of renal tubular cells. 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 nephrotoxicity. 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 urine and hence alter CKD risk.

Strengths of this project include the large sample size with dense genotype imputation, standardized and automated phenotyping, generation and control quality, 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 GFR. 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 efforts. 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 synthesis. Conversely, the SLC14A2 locus was associated with BUN but not with eGFR, in line with the function of the encoded protein as a urea transporter. 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 predate 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](https://doi.org/10.1038/s41588-019-0407-x).

Received: 16 September 2018; Accepted: 29 March 2019;
Published online: 31 May 2019

**References**

1. Eckardt, K. U. et al. Evolving importance of kidney disease: from subspecialty to global health burden. *Lancet* **382**, 158–169 (2013).
2. Jha, V. et al. Chronic kidney disease: global dimension and perspectives. *Lancet* **382**, 260–272 (2013).
3. 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).
4. 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).
5. 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).
6. 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).
7. 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).
8. Wuttke, M. & Kottgen, A. Insights into kidney diseases from genome-wide association studies. *Nat. Rev. Nephrol.* **12**, 549–562 (2016).
9. Gorski, M. et al. 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. *Sci. Rep.* **7**, 45040 (2017).
10. Pattaro, C. et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.* **7**, 10023 (2016).
11. 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).
12. Pattaro, C. et al. Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet.* **8**, e1002584 (2012).
13. Kottgen, A. et al. New loci associated with kidney function and chronic kidney disease. *Nat. Genet.* **42**, 376–384 (2010).
14. Chambers, J. C. et al. Genetic loci influencing kidney function and chronic kidney disease. *Nat. Genet.* **42**, 373–375 (2010).
15. Kottgen, A. et al. Multiple loci associated with indices of renal function and chronic kidney disease. *Nat. Genet.* **41**, 712–717 (2009).
16. 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).
17. 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).
18. 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).
19. 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).
20. Mahajan, A. et al. Trans-ethnic fine mapping highlights kidney-function genes linked to salt sensitivity. *Am. J. Hum. Genet.* **99**, 636–646 (2016).
21. 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 human 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 genotypic imputation. Nat. Genet. 48, 1279–1283 (2016).
26. Abeasciu, 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. SO2S and ACPI loci identified through large-scale exome chip analysis regulate kidney development and function. J. Am. Soc. Nephrol. 28, 98–108 (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, 784–773 (2016).
32. Wacholder, J. A B family of measures 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 uric acid 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 (2010).
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 tubularinterstitial 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. Olden, M. et al. Common variants in UMOD associate with urinary uric acid levels: a meta-analysis. J. Am. Soc. Nephrol. 25, 1869–1882 (2014).
43. Moreau, M. E. et al. The kallikrein–kinin system: current and future pharmacological targets. J. Pharm. Sci. 99, 6–38 (2010).
44. Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017).
45. 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).
46. Eckardt, K. U. et al. Autosomal dominant tubularinterstitial kidney disease: diagnosis, classification, and management—a KDIGO consensus report. Kidney Int. 88, 676–683 (2015).
47. Gillies, C. E. et al. An eQTl landscape of kidney tissue in human nephrotic syndrome. Am. J. Hum. Genet. 103, 232–244 (2018).
48. Olden, M. et al. Common variants in UMOD associate with urinary uric acid levels: a meta-analysis. J. Am. Soc. Nephrol. 25, 1869–1882 (2014).
49. Moreau, M. E. et al. The kallikrein–kinin system: current and future pharmacological targets. J. Pharm. Sci. 99, 6–38 (2010).
50. Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017).
51. 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).
52. Eckardt, K. U. et al. Autosomal dominant tubularinterstitial kidney disease: diagnosis, classification, and management—a KDIGO consensus report. Kidney Int. 88, 676–683 (2015).
53. Gillies, C. E. et al. An eQTl landscape of kidney tissue in human nephrotic syndrome. Am. J. Hum. Genet. 103, 232–244 (2018).
54. Olden, M. et al. Common variants in UMOD associate with urinary uric acid levels: a meta-analysis. J. Am. Soc. Nephrol. 25, 1869–1882 (2014).
55. Moreau, M. E. et al. The kallikrein–kinin system: current and future pharmacological targets. J. Pharm. Sci. 99, 6–38 (2010).
56. Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017).
57. 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).
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
Christian Gieger105,106,107, Giorgia Giotto21,52, Franco Giulianini108, Martin Gögele19, Scott D. Gordon109, Daniel F. Gudbjartsson16, Vilmundur Gudnason110,111, Toomas Haller112, Pavel Hamet113,114, Tamara B. Harris115, Catharina A. Hartman116, Caroline Hayward20, Jacklyn N. Hellwege117,118,119, Chew-Kiat Heng120,121, Andrew A. Hicks19, Edith Hofer122,123, Wei Huang124,125, Nina Nutri-Kähönen126,127, Shih-Jen Hwang128,129, M. Arfan Ikram25, Olafur S. Indridason130, Erik Ingelsson131,132,133,134, Marcus Ising135, Vincent W. V. Jaddoe25,96,97, Johanna Jakobsdottir136, Jost B. Jonas137,138, Peter K. Joshi139, Navya Shilpa Josyula139, Bettina Jung6, Mika Kähönen140,141, Yoichiro Kamatani27,142, Candace M. Kammerer143, Masahiro Kanai27,144, Mika Kastarinen145, Shona M. Kerr20, Chiea-chuen Khor64,87, Wieland Kiess12,84,85, Marcus E. Kleber156, Aldi T. Kraja5, Alena Krajcoviechova69,70, Holly Kramer17,151, Bernhard K. Kräme81, Florian Kronenberg152, Michiaki Kubo153, Brigitte Kühlne105, Mikko Kuokkanen154,155, Johanna Kuusisto145,156, Martina La Bianca21, Markku Laakso145,156, Leslie A. Lange157, Carl D. Langefeld86, Jeanette Jen-Mai Lee15, Benjamin Lehne29, Terho Lehtimäki158,159, Wolfgang Lieb160, Lifelines Cohort Study161, Su-Chi Lim15,162, Lars Lind163, Cecilia M. Lindgren164,165, Jun Liu125, Jianjun Liu87,166, Markus Loeffler112, Ruth J. F. Loos148,167, Susanne Lucae135, Mary Ann Lukas168, Leo-Pekka Lyytikäinen158,159, Reedit Mägi112, Patrik E. Magnusson165, Anubha Mahajan169,170, Nicholas G. Martin109, Jade Martins171, Winfried März172,173,174, Deborah Mascalzoni19, Koichi Matsuda175, Christa Meisinger176,177, Thomas Meitinger147,178,179, Olle Melander180, Andres Metspalu112, Evgenia K. Mikaelsdottir16, Yuri Milanescchi181, Kozeta Miliku25,96,97, Pashupati P. Mishra158,159, V. A. Million Veteran Program161, Karen L. Mohlke182, Nina Mononen158,159, Grant W. Montgomery183, Dennis O. Mook-Kanamori78,184, Josyf C. Mychaleckij185, Girish N. Nadkarni148,166, Mike A. Nalls187,188, Matthias Nauck23,189, Kjell Nikus190,191, Boting Ning192, Ilja M. Nolte124, Raymond Noordam193, Jeffrey O'Connell194, Michelle L. O'Donoghue195,196, Jadealine O’Hara21,212,213, Deborah Olafsson197, Albertine J. Oldehinkel116, Marju Orho-Melander32, Willem H. Ouwehand77, Sandosh Padmanabhan198, Nicholette D. Palmer199, Runolfur Palsson111,130, Brenda W. J. H. Penninx19, Thomas Perls200, Markus Perola201, Mario Pirastu202, Nicola Pirastu58, Giorgio Pistois203, Anna I. Podgoraia10, Ozren Polasek204,205, Belen Ponte206, David J. Porteous57,207, Tanja Poulain12, Peter P. Pramstaller19, Michael H. Preuss48, Bram P. Prins55, Michael A. Province5, Ton J. Rabelink279,208, Laura M. Raffield182, Olli T. Raitakari209,210, Dermot F. Reilly10, Rainer Rettig211, Myriam Rheinberger6, Kenneth M. Rice43, Paul M. Ridker108,212, Fernando Rivadeneira25,213, Federica Rizz124,214,215, David J. Roberts216, Antonietta Robino21, Peter Rossing21, Igor Rudan58, Rico Ruedei39,40, Daniela Ruggiero18,71, Kathleen A. Ryan217, Yasaman Saba218, Charumathi Sabanayagam164, Veikko Salomaa201, Erika Salvi214,219, Kai-Uwe Saum50, Helena Schmidt220, Reinhold Schmidt122, Ben Schöttker50,51, Christina-Alexandra Schulz32, Nicole Schup221,222,223, Christian M. Shaffer59, Yuan Shi64, Albert V. Smith111, Blair H. Smith224, Nicole Soranzo225, Cassandra N. Spracklen182, Konstantin Strauch226,227, Heather M. Stringham45, Michael Stummvoll228, Per O. Svensson229,230, Silke Szymbak100, E-Shyong Tai15,166,231, Salman M. Tajuddin95, Nicholas Y. Q. Tan64, Kent D. Taylor232, Andrej Teren12,233, Yih-Chung Tham64, Joachim Thiery12,53, Chris H. L. Tho24, Hauke Thomsen234, Gudmar Thorleifsson16, Daniela Toniolo24, Anke Tönjes228, Johanne Tremblay113,235, Ioanna Tzoulaki90,236, André G. Uitterlinden219, Simona Vaccargiu202, Rob M. van Dam15,166, Pim van der Harst237,238,239, Cornelia M. van Duijn25, Digna R. Velez Edward119,240, Niek Verweij237,
Suzanne Vogelezang25,96,97, Uwe Völker23,241, Peter Vollenweider242, Gerard Waebler242, Melanie Waldenberger105,106,147, Lars Wallentin243,244, Ya Xing Wang138, Chaolong Wang67,245, Dawn M. Waterworth4, Wen Bin Wei246, Harvey White247, John B. Whitfield109, Sarah H. Wild248, James F. Wilson26,58, Mary K. Wojczynski5, Charlene Wong67, Tien-Yin Wong54,67, Liang Xu138, Qiong Yang192, Masayuki Yasuda64,249, Laura M. Yerges-Armstrong4, Weihua Zhang61,90, Alan B. Zonderman95, Jerome I. Rotter232,250,251, Murielle Bochud73, Bruce M. Psaty252,253, Veronique Vitart20, James G. Wilson254, Abbas Dehghan29,90, Afshin Parsa255,256, Daniel I. Chasman108,212, Kevin Ho257,258, Andrew P. Morris72,169, Olivier Devuyst259, Shreram Akilesh37,260, Sarah A. Pendergrass261, Xueling Sim15, Carsten A. Böger6,262, Yukinori Okada263,264, Todd L. Edwards119,265, Harold Snieder24, Kari Stefansson16, Adriana M. Hung119,266, Iris M. Heid7,268, Markus Scholz112,268, Alexander Teumer22,23,268, Anna Köttgen19,268* and Cristian Pattaro19,268*

1 Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Center–University of Freiburg, Freiburg, Germany. 2 Renal Division, Department of Medicine IV, Faculty of Medicine and Medical Center–University of Freiburg, Freiburg, Germany. 3 Division of Nephrology and Hypertension, Department of Medicine, University of Utah, Salt Lake City, USA. 4 Target Sciences–Genetics, GlaxoSmithKline, Collegeville, PA, USA. 5 Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA. 6 Department of Nephrology, University Hospital Regensburg, Regensburg, Germany. 7 Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany. 8 Department of Nephrology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. 9 Welch Center for Prevention, Epidemiology and Clinical Research, Baltimore, MD, USA. 10 Genetics, Merck & Co., Inc., Kenilworth, NJ, USA. 11 Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. 12 LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany. 13 Division 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. 14 Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA. 15 Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore. 16 deCODE Genetics/Amgen, Inc., Reykjavík, Iceland. 17 Department of Public Health Sciences, Loyola University Chicago, Maywood, IL, USA. 18 Institute of Genetics and Biophysics ‘Adriano Buzzati-Traverso’–CNR, Naples, Italy. 19 EuraC Research, Institute for Biomedicine (affiliated with the University of Lübeck), Bolzano, Italy. 20 Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. 21 Institute for Maternal and Child Health, IRCCS ‘Burlo Garofolo’, Trieste, Italy. 22 Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. 23 EZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany. 24 Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. 25 Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. 26 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. 27 Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences (IMS), Yokohama, Japan. 28 Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. 29 Department of Epidemiology and Biostatistics, Faculty of Medicine, School of Public Health, Imperial College London, London, UK. 30 Institute of Public Health & Social Sciences, Khyber Medical University, Peshawar, Pakistan. 31 Steno Diabetes Center Copenhagen, Gentofte, Denmark. 32 Diabetes and Cardiovascular Disease–Genetic Epidemiology, Department of Clinical Sciences in Malmö, Lund University, Malmö, Sweden. 33 Division of Family Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden. 34 School of Health and Social Studies, Dalarna University, Dalarna, Sweden. 35 Department of Nephrology, Department of Internal Medicine, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. 36 Division of Nephrology, University of Washington, Seattle, WA, USA. 37 Kidney Research Institute, University of Washington, Seattle, WA, USA. 38 Cardiology, Geneva University Hospitals, Geneva, Switzerland. 39 Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. 40 Swiss Institute of Bioinformatics, Lausanne, Switzerland. 41 Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa. 42 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. 43 Department of Biostatistics, University of Washington, Seattle, WA, USA. 44 Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy. 45 Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. 46 Human Genetics Center, University of Texas Health Science Center, Houston, TX, USA. 47 CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, University of Lille, Lille, France. 48 Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 49 Digital Health Center, Hasslo Plattner Institute and University of Potsdam, Potsdam, Germany. 50 Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. 51 Network Aging Research, University of Heidelberg, Heidelberg, Germany. 52 Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy. 53 Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig, Leipzig, Germany. 54 Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany. 55 MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 56 National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK. 57 Center for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. 58 Center for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. 59 Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. 60 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. 61 Department of Cardiology, Ealing Hospital, Middlesex, UK. 62 Imperial College Healthcare NHS Trust, Imperial College London, London, UK. 63 MRC–PHE Center for Environment and Health, School of Public Health, Imperial College London, London, UK. 64 Singapore Eye Research Institute, Singapore National Eye Center, Singapore, Singapore. 65 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. 66 Ophthalmology and Visual Sciences Academic Clinical Program (Eye ACP), Duke–NUS Medical School, Singapore, Singapore.
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/genpi-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. 1). Studies on adults (>18 years of age) estimated GFR with the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation1, by using the R package nephr031. Studies on individuals who were 18 years old or younger used the Schwartz formula32. eGFR was winsorized at 15 and 200 ml min\(^{-1}\) per 1.73 m\(^2\). CKD was defined as an eGFR below 60 ml min\(^{-1}\) per 1.73 m\(^2\). In studies reporting blood urea measurements, BUN was defined as blood urea x 2.8, with units expressed as mg d\(^{-1}\).

Genotyping and genotype imputation. Genotypes were imputed on the basis of the Haploype 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 \(\lambda_G\geq1\). Fixed-effects inverse-variance-weighted meta-analysis was performed with METAL61, 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 \(\geq 1\) in all analyses of eGFR, BUN and CKD)39. The genome-wide significance level was \(5 \times 10^{-8}\). Between-study heterogeneity was assessed with the \(F\) statistic41. For CKD, variants with \(F \geq 95\%\) 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, vol.1.2) 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 × 10\(^{-4}\) (referring to the corresponding P value as \(P_{\text{meta}}\)).

Variance explained and genetic heritability. The proportion of phenotypic variance explained by the index SNPs was estimated as \(\beta^2\left(1-\frac{\lambda_G}{\lambda_G+1}\right)\), with \(\beta\) being the SNP effect, \(\rho\) the allele frequency and \(\tau\) 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). Genetic heritability for age- and sex-adjusted log(eGFR) was estimated with the R package MCMCglmm37 on the Cooperative Health Research in South Tyrol (CHRIS) study42, a participating pedigree-based study with 186 pedigrees of up to five generations (\(n = 4,373\)). We fitted two models with and without inclusion of the identified index SNPs (304/308), running 1 million MCMC iterations (burn-in = 500,000).

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 MVP43, 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 equation42 after excluding subjects on dialysis, transplant patients, amputees, individuals on HIV medications and those with creatinine values of <0.4 mg d\(^{-1}\). 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 SNPTST2 V2.5.4-beta42. All GWAS were stratified by self-reported ancestry (7.9% white; 20.4% black; 0.4% Hispanic). diabates and hypertension status. Results were combined across strata by fixed-effects inverse-variance-weighted meta-analysis in METAL43. 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.

For replication, we used 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 \(r^2\geq0.8\) 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-anthracly-specific GWAS. CKDGen and MVP meta-analysis results were homologous for sample-size-weighted meta-analysis of z scores with METAL43. 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, ‘uncertain’ for eGFR index SNPs whose effect on BUN was not different from 0 (\(P > 0.05\)) and ‘unlikely’ for all eGFR index SNPs with a concurrent, 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 genetic risk score association analysis was conducted on the UK Biobank GWAS datasets, as described previously44,45. Genetic association results with diseases were obtained for 452,264 UK Biobank participants available in the GeneAtlas46 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; 8,410 cases); and ischemic heart diseases (I20–I25; 33,387 cases). Asthma (I45; 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 (\(\beta\)) corresponds to the OR for the disease depending on the relative change in eGFR, for example, OR, \(\exp(\beta)\) 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 known to correlate with eGFR and BUN. We estimated pairwise genetic correlation coefficients (\(r\)) 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 regression48. 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 × 10⁻⁶.

Pathway and tissue enrichment analysis. We used DEPICT v1 release 194 to perform DEPICT analysis, including pathway/gene set enrichment and tissue/cell-type analyses as described previously³⁴. All 14,461 gene sets were reidentified by assigning genes that were transcriptionally co-regulated with other genes in a pathway or cell-type microarray. DEPICT v1.2 was used to perform pathway enrichment analysis, 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 eGF and BUN, we included all variants associated with the trait at P < 5 × 10⁻⁶ in the trans-ancestry meta-analysis. Independent variant discovery was performed by using PLINK 1.9 (ref. ³⁷) with 500-kb flanking regions and r² > 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 algorithm³⁵ 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 eGF 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 previously³⁶. The ten cell-type groups were defined as 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 phenotypic ontology standardized release 2017 (abnormal nonciliated 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 a post-discovery association 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 eGF were queried for significantly associated SNPs mapping to the selected candidate genes. Enrichment of significant genetic associations within each candidate gene list was computed from the complementary cumulative binomial distribution. 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 eGF-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 individuals, 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 ≥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 PLINK³⁷. 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 (r² > 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 Sclt algorithm³⁶ 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-wise significant if the SNP's P value conditional on all previously identified SNPs in the same region was <5 × 10⁻⁶.

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 gtx⁸⁰, 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 publication. 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)⁸¹, on the basis of the 1000Gp3v5 and Ensembl v87 datasets. SNIPA was also used to derive the CADD Phred-like score⁸², on the basis of CADD v1.3. The Ensembl VEP⁸³ tool was used for prediction of the primary effects of SNPs.

Colocalization of eGF 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 eGF 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 study⁸⁴, as well as the 44 tissues included in the GTEx Project v6 release⁸⁵. 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 NephiQTL 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 (r² cutoff of >0.8). We limited annotation to replicated index SNPs with fine-mapping PP ≥1%. Owing to expected small effect sizes, only genome-wide trans-eQTL studies of either pulmonary blood mononuclear cell (PBMC) or ≥1,000 individuals were considered, resulting in five non-overlapping studies⁸⁷–⁸⁹ (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 × 10⁻⁶ 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 eGF 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.imibi.uni-freiburg.de.

References
60. Fuchsberger, C., Tülin, D., Pramstaller, P. P. & Pattaro, C. GWAToolbox: an R package for fast quality control and handling of genome-wide association studies meta-data analysis. 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 fine-mapping resolution. Hum. Mol. Genet. 26, 3639–3650 (2017).
67. Hadfield, J. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J. Stat. Softw. 33, 1–22 (2010).
68. Pattaro, C. et al. The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. J. Transl. Med. 13, 348 (2015).
69. Noce, D. et al. Sequential recruitment of study participants may inflate genetic heritability estimates. Hum. Genet. 136, 743–757 (2017).
70. 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).
71. Kottgen, A. et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nat. Genet. 45, 145–154 (2013).
72. 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).
73. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
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 the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
76. 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).
77. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
78. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from fine-mapping resolution. Hum. Mol. Genet. 26, 3639–3650 (2017).
79. Hadfield, J. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J. Stat. Softw. 33, 1–22 (2010).
80. Pattaro, C. et al. The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. J. Transl. Med. 13, 348 (2015).
81. Noce, D. et al. Sequential recruitment of study participants may inflate genetic heritability estimates. Hum. Genet. 136, 743–757 (2017).
82. 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).
83. Kottgen, A. et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nat. Genet. 45, 145–154 (2013).
84. 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).
85. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
86. Kottgen, A. et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nat. Genet. 45, 145–154 (2013).
87. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
88. 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).
89. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
90. 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).
91. Kottgen, A. et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nat. Genet. 45, 145–154 (2013).
92. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
93. 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).
94. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
95. 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).
96. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
97. 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).
98. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
99. 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).
100. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
101. 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).
102. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
103. 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).
104. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
105. 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).
106. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in the UK Biobank. Nat. Genet. 50, 1593–1599 (2018).
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

☑ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐ The statistical test(s) used AND whether they are one- or two-sided

☐ Only common tests should be described solely by name; describe more complex techniques in the Methods section.

☐ A description of all covariates tested

☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐ 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)

☐ For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted

☐ Give $P$ values as exact values whenever suitable.

☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☐ 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/ckdgen-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 -Slct 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

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

☑ Antibodies
☑ Eukaryotic cell lines
☑ Palaeontology
☑ Animals and other organisms
☑ Human research participants
☑ Clinical data

Methods

n/a  Involved in the study

☑ ChIP-seq
☑ Flow cytometry
☑ 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.