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

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

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

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

We therefore conducted a trans-ancestry GWAS meta-analysis in the CKD Genetics (CKDGen) Consortium (n = 765,348) and replicated findings in the Million Veteran Program (MVP; n = 280,722), for a combined sample size of greater than 1 million participants. The first aim of this study was to identify new globally important loci for kidney function through maximizing statistical power (Supplementary Fig. 1). Results from GWAS of the complementary kidney function marker blood urea nitrogen (BUN; n = 416,178) were used to prioritize the eGFR-associated loci on the basis of those most likely to be relevant for kidney function. A genetic risk score (GRS) for low eGFR was used to test relevance for clinically diagnosed CKD among 452,264 independent individuals. The second aim was to characterize replicated eGFR-associated loci through complementary computational approaches, including various enrichment and network analyses, fine-mapping, and colocalization with gene expression in 46 tissues and protein levels (Supplementary Fig. 1). We focused this aim on European-ancestry individuals, as fine-mapping based on summary statistics requires linkage disequilibrium (LD) reference panels whose sample size scales with that of the GWAS. The resulting list of prioritized variants and genes provides a rich resource of potential therapeutic targets to improve CKD treatment and prevention.

Results

Discovery trans-ancestry meta-analysis. We performed 121 GWAS encompassing 765,348 individuals of European (n = 567,460), East Asian (n = 165,726), African-American (n = 13,842), South Asian (n = 13,359) and Hispanic (n = 4,961) ancestry (median age, 54 years; 50% female; Supplementary Table 1). The median of the study-specific mean eGFR values was 89 ml min⁻¹ per 1.73 m² (interquartile range, IQR: 81, 94). GWAS were based on genotypes imputed from Haploetype Reference Consortium or 1000 Genomes Project reference panels (Methods and Supplementary Table 2). Following study-specific variant filtering and quality-control procedures, we performed a fixed-effects inverse-variance-weighted meta-analysis, finding no evidence of unmodeled population structure (LD score regression intercept = 1.04; genomic control factor λGC = 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, and 19.6% of eGFR genetic heritability (h² = 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 Appendix Fig. 3).
ARTICLES

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\(^\text{15}\). 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 × 10\(^{-12}\)) and SH3YL1 (P = 1.2 × 10\(^{-11}\)) were also considered to be replicated on the basis of previous evidence\(^{13,14}\), 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-ancestposer analysis of 65 GWAS for BUN (n = 416,178; Supplementary Table 1) showed no evidence of unmodeled population structure (\(\lambda_{\text{IC}} = 1.03\); 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 \(P < 0.05\); Supplementary Table 6). This comparative analysis of complementary biomarkers supports the idea that signals at the majority of eGFR-associated loci probably reflect kidney function.

Next, we investigated the effects of the eGFR index SNPs on CKD in CKDGen studies (n = 625,219, including 64,164 CKD cases; Methods). GWAS meta-analysis of CKD identified 23 genome-wide-significant loci, including 17 likely relevant for kidney function (SDCCAG8, LARP4B, DDCDC1, WDR72, UMOD–PDILT, MYO19, AQP4, NFATC1, PSID4, 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)\(^{19}\). We observed 37 significant correlations (\(P < 10^{-5} = 0.05/748\); Supplementary Table 6 and Supplementary Table 7). After serum creatinine, the largest negative correlations were observed between eGFR and serum citrate (\(r_g = -0.27\) and urate (\(r_g = -0.23\), followed by anthropometric traits including lean mass and physical fitness (for example, \(r_g = -0.20\) for left hand grip strength). While the inverse correlation with muscle-mass-related traits probably reflects higher creatinine generation leading to lower creatinine-based eGFR, the correlations with citrate and urate levels probably reflect reduced filtration function, as does the positive correlation with GFR estimated from cystatin C (\(r_g = 0.53\)).

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

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

Functional enrichment and pathway analyses. To identify molecular mechanisms and tissues of importance for kidney function, we assessed the enrichment of the eGFR and BUN genetic associations by using tissue-specific gene expression, regulatory annotations, and gene sets and pathways (Methods). First, we used eGFR-associated SNPs (\(P < 5 \times 10^{-8}\)) to explore enriched pathways, tissues and cell types on the basis of gene expression data with DEPICT\(^{10}\). 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 polynuria, dilated renal tubules and expanded mesangial matrix, as well as signaling and transcription, and energy metabolism (Supplementary Fig. 7c). Tissue and cell-type enrichment analysis of BUN-associated SNPs associated at \(P < 5 \times 10^{-8}\) highlighted a very similar pattern (Supplementary Fig. 8) but without enrichment for muscle tissues, further supporting the use of BUN to prioritize the loci most likely to be related to kidney function.

Second, we used stratified LD score regression\(^{18}\) 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\(^{11}\). 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.
(enrichment $P=1.1 \times 10^{-5}$) 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
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 (≤500kb up- or downstream of the trans-ancestry index SNP), and
The remaining 7 showed $P \leq 3.3 \times 10^{-5}$ 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, with a median set size of 26 SNPs (IQR: 6, 60). We observed 58 small credible sets ($\leq 5$ SNPs), including 20 single-SNP sets: EDEM3, CACNA1S, HOXD11, CPS1, DAB2, SLC34A1, LINCOI512, LARP4B, DCDC1, SLC25A45, SLC6A13, GATM, CGNL1, CYP1A1, NRG4, RLPL3, UMOD-PDILT, SLC47A1 and two independent sets at BCL2L14 (Fig. 4 and Supplementary Table 11). Of the 58 small credible sets, 33 were likely relevant for kidney function and contain genes and SNPs that can now be prioritized for further study (Supplementary Table 11).
### Table 1 | Genes implicated as causal via identification of missense SNPs with high probability of driving the eGFR association signal

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Credible set size</th>
<th>SNP PP</th>
<th>Functional consequence</th>
<th>CADDScore</th>
<th>DHSs, tissue</th>
<th>Brief summary of the gene's function and relevant literature</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CACNA1S</strong></td>
<td>rs3850625</td>
<td>1</td>
<td>1.00</td>
<td>p.(Arg1539Cys) <strong>(NP_000060.2)</strong></td>
<td>34.0</td>
<td>-</td>
<td>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.</td>
</tr>
<tr>
<td><strong>CPS1</strong></td>
<td>rs1047891</td>
<td>1</td>
<td>1.00</td>
<td>p.(Thr1406Asn) <strong>(NP_001866.2)</strong></td>
<td>22.1</td>
<td>-</td>
<td>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.</td>
</tr>
<tr>
<td><strong>EDEM3</strong></td>
<td>rs78444298</td>
<td>1</td>
<td>1.00</td>
<td>p.(Pro746Ser) <strong>(NP_079467.3)</strong></td>
<td>24.6</td>
<td>-</td>
<td>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.</td>
</tr>
<tr>
<td><strong>KLHDC7A</strong></td>
<td>rs11261022</td>
<td>7</td>
<td>0.71</td>
<td>p.(Arg160Ser) <strong>(NP_689588.2)</strong></td>
<td>1.1</td>
<td>Roadmap + ENCODE, kidney</td>
<td>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.</td>
</tr>
<tr>
<td><strong>RPL3L</strong></td>
<td>rs113956264</td>
<td>1</td>
<td>1.00</td>
<td>p.(Val262Met) <strong>(NP_005052.1)</strong></td>
<td>27.2</td>
<td>-</td>
<td>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.</td>
</tr>
<tr>
<td><strong>SLC25A45</strong></td>
<td>rs34400381</td>
<td>1</td>
<td>1.00</td>
<td>p.(Arg285Cys) <strong>(NP_001070709.2)</strong></td>
<td>26.0</td>
<td>ENCODE, kidney</td>
<td>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.</td>
</tr>
<tr>
<td><strong>SLC47A1</strong></td>
<td>rs111653425</td>
<td>1</td>
<td>1.00</td>
<td>p.(Ala465Val) <strong>(NP_060712.2)</strong></td>
<td>24.6</td>
<td>-</td>
<td>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.</td>
</tr>
<tr>
<td><strong>PPM1J</strong></td>
<td>rs34611728</td>
<td>5</td>
<td>0.02</td>
<td>p.(Leu213Phe) <strong>(NP_005158.5)</strong></td>
<td>13.1</td>
<td>ENCODE, kidney</td>
<td>This gene encodes a serine/threonine protein phosphatase. The variant has been reported in association with eGFR in an exome chip association study.</td>
</tr>
</tbody>
</table>

Continued
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 were associated with higher urinary uromodulin concentrations. 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)

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNPa</th>
<th>Credible set size</th>
<th>SNP PP</th>
<th>Functional consequence</th>
<th>CADD score</th>
<th>DHSs, tissue</th>
<th>Brief summary of the gene’s function and relevant literature</th>
</tr>
</thead>
</table>
| CERS2 | rs267738 | 5 | 0.46 | p.(Glu15Ala) (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. Cers2-knockout mice show strongly reduced ceramide levels in the kidney and develop renal parenchyma abnormalities. This variant has been reported as associated with the rate of albuminuria increase in individuals with diabetes.
| 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.
| SLC22A2 | rs316019 | 4 | 0.04 | p.(Ser270Ala) (NP_003049.2) | 12.7 | – | Encodes the polycrystalline 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, indicating that this transporter is essential in excreting toxins. The locus has been reported in previous GWAS of eGFR.

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, KHLDC7A, PPM11, CERS2, C9 and SLC22A2; Supplementary Table 12), of which CACNA1S, RPL3L, CERS2 and C9 were likely relevant for kidney function (Fig. 4a). The majority of the 11 variants had a combined annotation-dependent depletion (CADD) score greater than 15, indicating potential deleteriousness. 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 argues against a sole effect on creatinine transport.

To evaluate the regulatory potential of SNPs from small credible sets in the kidney, we annotated them to open chromatin regions identified from primary human tubular and glomerular cell cultures, as well as from publicly available kidney cell types (Methods). We identified 72 SNPs mapping to one of these annotations, which may thus represent causal regulatory variants (Supplementary Table 12). A particularly interesting finding was the intronic rs77924615 SNP in PDILT, which showed PP > 99% of driving the association at the UMOD locus and mapped to open chromatin in all evaluated resources (native kidney cells, ENCODE and Roadmap kidney cell types; Fig. 4b).
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. 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, hemocrit and multiple kidney-function-related traits (Supplementary Table 13). The eGFR index SNP rs1458038 (PP $> 50\%$, CADD score = 14.8; Supplementary Table 13) colocalized with the eGFR signal only in the tubulo-interstitial kidney portion (Fig. 5), supporting its regulatory potential in controlling the expression levels of FGF5 in this compartment. Both KNG1 and FGF5 index SNPs were associated with BUN and CKD and are thus probably related to kidney function.

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

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

**Colocalization with uromodulin protein levels in urine.** The UMOD locus is of particular clinical interest for CKD research; rare UMOD mutations cause autosomal dominant tubulo-interstitial kidney disease, and common variants at UMOD give rise to the strongest eGFR and CKD GWAS signals. 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

---

**Fig. 6 | Colocalization of independent eGFR association signals at the UMOD–PDILT locus with urinary uromodulin concentrations (UUCR) supports UMOD as the effector gene.** Association plots show association -log_{10}(P 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 \( \leq 5 \) 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 cinmetidine across the basolateral membrane of renal tubular cells\(^{41}\). 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\(^{42}\). 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\(^{43}\) and hence alter CKD risk.

Strengths of this project include the large sample size with dense genotype imputation, standardized and automated phenotype generation and quality control, and independent replication, as well as the advanced and comprehensive downstream bioinformatics analyses. Further strengths are the use of BUN to prioritize eGFR-associated loci likely relevant for kidney function and to provide genome-wide BUN summary statistics as an annotation resource for other studies of eGFR. Moreover, we evaluated a GRS for eGFR for association with clinically diagnosed CKD in a large independent study. Among the limitations, non-European populations are still under-represented in our study, as in many other genomic efforts\(^{44}\). 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\(^{45}\). 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\(^{46}\). Even so, lack of association for a SNP with one kidney function marker based on a combination of \( P \) value and effect direction may not necessarily mean that the locus is not relevant to kidney function. Our categorization of the eGFR loci into three classes on the basis of direction of effect and significance of BUN association should be interpreted with caution, with ‘likely’ and ‘unlikely’ reflecting uncertainty of the assignment. Factors complicating the comparison of eGFR and BUN associations at the locus level are differential statistical power, differential ancestry distribution and potential allelic heterogeneity. Further large-scale studies with multiple kidney function markers measured in the same individuals are therefore warranted.

To identify broadly representative and generalizable association signals, we focused on SNPs that were present in the majority of the participating studies. This choice might have limited our ability to uncover new variants or to fine-map low-frequency or population-specific variants, which represents a complementary avenue of research. Moreover, even with well-powered fine-mapping approaches, potentially causal SNPs need to be confirmed as functional variants in experimental studies. Although colocalization with gene expression can help prioritize effector genes, these associations are based on measures from a single time point and hence cannot answer whether changes in gene expression precede or follow changes in kidney function.

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

**Online content**

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

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

**References**

A T U


Ron E Barry, Maria Pina Janine F. Felix, Kai-E Daniele S 968 Deutschland. D.O.M.-K. is a part-time research consultant at Metabolon. M.A.N. work. W.M. is employed with Synlab Services and holds shares of Synlab Holding Amgen, AstraZeneca, Novartis, Pfizer and Sanofi, all outside the scope of the submitted DalCor, Kowa, Novartis, Pfizer and Sanofi and modest personal fees for lectures from C P.V., A.P.d.V., G.W., L. Wallentin, H.W., J.B.W., S.H.W., J.G.W., A.B.Z., J.Ä. P.O.S., N.Y.Q.T., A. Teren, Y.-C.T., J. Tremblay, I.T., A. Tönjes, S. Vaccargiu, S. Vogelezang, D.J.P., T. Poulain, M.A.P., T.J.R., O.T.R., M.R., P.M.R., P.R., I.R., D.R., V.S., R.S., B.H.S., J.B.J., B.J., C.M.K., M. Kastarinen, J.S.K., A. Krajcoviechova, F.K., M. Kubo, M. Kähönen, G.S., P.O.S., S.M.T., N.Y.Q.T., B.O.T., A. Teumer, H.T., A. Tin, J. Tremblay, I.A. Tönjes, M.R., K.M.R., P.M.R., F. Rivadeneira, D.F.R., P.R., I.R., D.R., V.S., E.S., B.H.S., C.N.S., S.M.T., K.T., A. Teumer, D.T., J. Tremblay, A.G.U., S.P., N.D.P., B.W.P., M. Perola, D.J.P., M.H.P., O.T.R., D.F.R., F. Rivadeneira, F. Rizzi, Matthias Wuttke 12,267, Yoshiko Li 1,267, Andrew Giri 3,267, J. Martin 3,267, Tai-Fung Li 1,267, Janie F. Felix 25,96,97, John He 25,96,97, John C. Chambers 29,60,61,62,63, Miaolin Chee 64,65,66,67, Yurong Cheng 64,66,67, Yurong Cheng 64,66,67, Kaare Christensen 68, Renata Cifkova 69,70, Marina Ciuollo 18,71, Maria Pina Concà 21, James P. Cook 22, Josep Coresh 23, Tanguy Corre 39,40,73, Cinzia Felicita Sala 74, Daniele Cusi 75,76, John Danesh 77, E. Warren Daw 78, Martin H. de Borst 35, Alessandro De Grandi 19, Renée de Mutsert 79, Aiko P. J. de Vries 79, Frauke Degenhard 80, Graciala Delgado 81, Ayse Demirkan 20, Damaris Vergnaud 82,38, Katalin Dittrich 84,85, Jasmin Diers 86, Rajkumar Dorajoo 87, Ricardo Díez-Costa 88,89, Georg Ehret 38, Paul Elliott 90,91,92,93, Karhans Endl 23,94, Michele K. Evans 95, Janine F. Felix 25,96,97, Valencia Hui Xian Foo 64, Oscar H. Franco 25,98, Andre Franke 80, Barry I. Friedman 99, Sandra Freitag-Wolf 100, Yechiel Friedlander 100, Philippe Froug 47,102, Ron T. Gansevoort 35, He Gao 90, Paolo Gasparini 21,252, J. Michael Gaziano 103, Vilhantzas Giedraitis 104, is supported by a consulting contract between Data Tecnica International and the National Institute on Aging (NIA), National Institutes of Health (NIH) and consults for illumina, the Michael J. Fox Foundation and University of California HealthCare. O.H.E. works in ErasmusMC for a center of research across the life course funded by Nestlé Nutrition (Nestec); Metagenics; and AXA. K.B.S., L.Y.A., D.M.W. and M.A.L. are full-time employees of GlaxoSmithKline. M.Ö.D. received grant support from GlaxoSmithKline, MSD, Eisai, AstraZeneca, MedCo and Janssen. H.W. received grants and non-financial support from GlaxoSmithKline, during the conduct of the study; grants from Sanofi-Aventis, Eli Lilly, the National Institute of Health, Omthera Pharmaceuticals, Pfizer New Zealand, Elsevier Inc. and Dalcour Pharma UK; honoraria and non-financial support from AstraZeneca; and is on advisory boards for Sirtex and Acadion and received personal fees from CSL Behring and American Regent outside the scope of the submitted work. L. Wallentin received institutional grants from GlaxoSmithKline, AstraZeneca, BMS, Boehringer-Ingelheim, Pfizer, MSD and Roche Diagnostics. D.F.R. and A.I.P. are employees of MSD. M. Scholz received consultancy of and grant support from Merck Serono not related to this project. B.M.P serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the steering committee of the Yale Open Data Access Project funded by Johnson & Johnson. J. Danesh is a member of the Novartis Cardiovascular and Metabolic Advisory Board and received grant support from Novartis. A.S.B. received grants from MSD, Pfizer, Novartis, Biogen and Bioverativ and personal fees from Novartis. V.S. has participated in a conference trip sponsored by Novo Nordisk and received a honorarium from the same source for participating in an advisory board meeting. A. Kötting received grant support from Gruenenthal. All other authors declare no conflicts of interest.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0407-x.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to A.K. or C.P.
Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

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

Matthias Wuttke 12,267, Yong Li 1,267, Man Li 3,267, Karsten B. Sieber 4,267, Mary F. Feitosa 5,267, Mathias Gorsk 6,267, Adrienne Tin 8,9, Lihua Wang 5, Audrey Y. Chu 10, Anselm Hoppmann 11, Holger Kirsten 11,12, Ayush Giri 13,14, Jin-Fang Chai 15, Gardar Sveinbjornsson 16, Bamidele O. Tayo 17, Teresa Nutile 18, Christian Fuchsberger 19, Jonathan Marten 20, Massimiliano Cocca 21, Sahar Ghasemi 22,23, Yizhe Xu 3, Katrin Horn 11,12, Damia Noce 19, Peter J. van der Most 24, Sanaz Sedaghat 25, Zhi Yu 8,26, Masato Akiyama 27,28, Salima Afaq 29,30, Tarunveer S. Ahuwalia 31, Peter Almgren 32, Najaf Amin 25, Johan Arnlöv 33,34, Stephen J. L. Bakker 35, Nisha Bansal 36,37, Daniela Baptista 38, Sven Bergmann 39,40,41, Mary L. Biggs 42,43, Ginevra Biino 44, Michael Boehnke 45, Eric Boerwinkle 46, Mathilde Boissel 47, Erwin P. Bottinger 48,49, Thibaud S. Boutin 20, Hermann Brenner 50,51, Marco Brumata 52, Ralph Burkhardt 12,53,54, Adam S. Butterworth 55,56, Eric Campana 52, Archie Campbell 57, Edward Campbell 58, Mickaël Canouil 47, Robert J. Carroll 59, Eulalia Catamo 21, John C. Chambers 29,60,61,62,63, Miaolin Chee 64, Miaolin Chee 64, Xu Chen 65, Chen-Yu Cheng 64,66,67, Yungkin Cheng 1, Kaare Christensen 68, Renata Cifkova 69,70, Marina Ciuollo 18,71, Maria Pina Concà 21, James P. Cook 22, Josep Coresh 23, Tanguy Corre 39,40,73, Cinzia Felicita Sala 74, Daniele Cusi 75,76, John Danesh 77, E. Warren Daw 78, Martin H. de Borst 35, Alessandro De Grandi 19, Renée de Mutsert 79, Aiko P. J. de Vries 79, Frauke Degenhard 80, Graciala Delgado 81, Ayse Demirkan 25, Emanuela Di Angelantonio 82,83, Katalin Dittrich 84,85, Jasmin Diers 86, Rajkumar Dorajoo 87, Kai-Uwe Eckardt 88,89, Georg Ehret 38, Paul Elliott 90,91,92,93, Karhans Endl 23,94, Michele K. Evans 95, Janine F. Felix 25,96,97, Valencia Hui Xian Foo 64, Oscar H. Franco 25,98, Andre Franke 80, Barry I. Friedman 99, Sandra Freitag-Wolf 100, Yechiel Friedlander 100, Philippe Froug 47,102, Ron T. Gansevoort 35, He Gao 90, Paolo Gasparini 21,252, J. Michael Gaziano 103, Vilhantzas Giedraitis 104,
Methods

Overview. We set up a collaborative meta-analysis based on a distributive data model and quality-control procedures. To maximize phenotype standardization across studies, an analysis plan and a command line script (https://github.com/genepi-freiburg/cdkg-round4-phenos) 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://cdkgem연구.육리보/ckdgen_round4_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 GWAToolbox (https://github.com/ETICS/GWAToolbox) 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. 11). Studies on adults (≥18 years of age) estimated GFR with the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation (12), by using the R package nepho (13). Studies on individuals who were 18 years old or younger used the Schwartz formula (14). eGFR was winsorized at 15 and 200 ml min⁻¹ per 1.73 m². CKD was defined as an eGFR below 60 ml min⁻¹ per 1.73 m². In studies reporting blood urea measurements, BUN was defined as blood urea x 2.8, with units expressed as mgdl⁻¹.

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

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

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

Before meta-analysis, study-specific GWAS files were filtered to retain only variants with IQ score > 0.6 and minor allele count (MAC) > 10, and genomic control (GC) correction was applied in the case where GC factor λGC > 1. Fixed-effects inverse-variance-weighted meta-analysis was performed with METAL (15), which was adapted to increase the precision of effect estimates and their standard errors (seven decimal places instead of four).

After meta-analysis of 43,990,857 SNPs, only SNPs present ≥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 n = 1 in all analyses of eGFR, BUN and CKD) (16). The genome-wide significance level was 5 x 10⁻⁸. Between-study heterogeneity was assessed with the F statistic (17). For CKD, variants with F ≥ 0.95 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) (18) 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 x 10⁻⁴ (referring to the corresponding P value as P₃₀₈,₀⁻⁴).

Variance explained and genetic heritability. The proportion of phenotypic variance explained by the index SNPs was estimated as \( \hat{\beta}^2 \) \( \frac{\text{var}(\text{SNP})}{\text{var}(\text{trait})} \), with \( \hat{\beta} \) being the SNP effect, \( p \) the effect allele frequency and var the variance of the sex- and age-adjusted log(eGFR) residuals (assumed to be 0.016 on the basis of data from 11,827 European-ancestry participants of the population-based ARIC study). Genetic heritability for age- and sex-adjusted log(eGFR) was estimated with the R package MCMCglmm (19) on the Cooperative Health Research in South Tyrol (CHRIS) study (20), 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 MVP (21), an independent trans-ancestry study with participants recruited across 63 US Veterans Administration (VA) medical facilities. Written informed consent was obtained, 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 equation (22) after excluding subjects on dialysis, transplant patients, amputees, individuals on HIV medications and those with creatinine values of <0.4 mgdl⁻¹. GWAS of eGFR on SNP dosage were performed by fitting linear regression models adjusted for age at creatinine measurement, age², sex, body-mass index and the first ten genetic principal components, by using SNPTEST v2.5.4-beta (23). All GWAS were stratified by self-reported ancestry (7.6% white non-Hispanic, 20.4% black non-Hispanic), diabetes and hypertension status. Results were combined across strata by fixed-effects inverse-variance-weighted meta-analysis (24,25). 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 the 308 eGFR index SNPs identified in the CKDGen trans-ancestry analysis, 305 variants or their good proxies were available in the MVP GWAS (proxies were required to have r² > 0.8 with the index SNP and were selected by maximum r² followed by minimum distance in the case of ties). Replication testing of the 256 European-ancestry-specific index SNPs was restricted to the MVP European-ancestry-specific GWAS. CKDGen and MVP meta-analysis results were pooled via sample-size-weighted meta-analysis of z scores with METAL (26). 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, ‘unlikely’ for eGFR index SNPs whose effect on BUN was not different from 0 (P > 0.05) and ‘unlike’ for all eGFR index SNPs with a Concordant, significant (one-sided P < 0.05) association with BUN for a given reference allele.

Genetic risk score analysis in the UK Biobank dataset. To test the combined effect of eGFR-associated SNPs on outcomes related to clinically diagnosed CKD, a genetic risk score analysis was conducted on the 147 eGFR index SNPs identified in the discovery GWAS results, as described previously (27). Genetic association results with diseases were obtained for 452,264 UK Biobank participants available in the GeneAtlas (28) 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; 8,910 cases); and ischemic heart diseases (I20–I25; 33,387 cases). Asthma (I45; 28,629 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 = 1.10 for a 10% change in eGFR. Alternatively, \( \exp(\beta) \) can be interpreted as the OR for the disease per unit change in log(eGFR).

Genetic correlations with other complex traits and diseases. Genome-wide genetic correlation analysis was performed to investigate evidence of co-regulation or shared genetic basis between eGFR and BUN concentrations and other complex traits and diseases, both known to be related to kidney function with eGFR and BUN. We estimated pairwise genetic correlation coefficients (\( \rho \)) 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 regression (18). An overview of
ARTICLES

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 previously45. All 14,461 gene sets were reidentified by identifying genes that were transcriptionally co-regulated with other genes in a panel of 37,427 microarrays46. Enrichment was performed by using PLINK 1.9 (ref. 37) with 500-kb flanking regions and r^2 > 0.01 in the 1000Gp1v3 dataset. After excluding the MHC region, DEPICT was run with 500 repetitions to estimate the FDR and 5,000 permutations to compute P values adjusted for gene length by using 500 null GWAS. All significant gene sets were merged into meta-gene sets by running an affinity propagation algorithm33 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 eQTL and LD 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 comprised of 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 GenToS83 to identify additional genetic associations that were not genome-wide significant. Candidate genes that when manipulated cause kidney phenotypes in mice were selected with the comprehensive MGI phenotype ontology. For the 144 mouse knockout studies, and molecular pathways from protein–protein interaction networks as the proportion of SNP heritability in that group divided by the proportion of

**Independent variant identification in the European-ancestry meta-analysis.** To identify additional independent eQTL-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. The GWAS meta-analysis summary statistics for eQTLs 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 distribution76. GenToS was used with default parameters on each of the three candidate gene lists, with the 100 Genomes phase 3 release 2 ALL dataset as reference.

**Variant annotation.** Functional annotation of SNPs mapping to credible sets was performed with SNIPA v3.2 (March 2017)88, on the basis of the 1000Gp3v5 and Ensembl v87 datasets. SNIPA was also used to derive the CADD Phred-like score89, on the basis of CADD v1.3. The Ensembl VEP tool90 was used for prediction of the primary effects of SNPs.

**Colocalization of eQTL gene and signal 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 eQTL 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 study91, as well as the 44 tissues included in the GTEx Project v6 release56. 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/obrynjuso/gtx), which is an adaptation of Giambaritoluemi’s colocalization method77. The gtx package was also used to estimate the direction of effect over the credible sets as the ratio of the average PP-weighted GWAS effects over the PP-weighted eQTL effects.

**Trans-eQTL analysis.** We performed trans-eQTL annotation through LD mapping on the basis of the 1000Gp3v5 European reference panel (r^2 cutoff of >0.8). We limited annotation to replicated index SNPs with fine-mapping PP ≥ 1%. Owing to expected small effect sizes, only genome-wide trans-eQTL studies of either polymorphically blood mononuclear cells n ≥ 1,000 individuals were considered, resulting in five non-overlapping studies72-74 (Supplementary Table 14). For one study81, we had access to an update with larger sample size (n = 6,645) obtained by combining two non-overlapping studies (LIFE-Heart and LIFE-Adult). To improve the stringency of results, we focused the analysis on interchromosomal trans-eQTLs with P ≤ 5 × 10^−8 in ≥2 studies.

**Colocalization with urinary uromodulin concentrations.** Association of genetic variants with UUCR at the UMOD–PDILT locus was evaluated in the German Chronic Kidney Disease (GCKD) study82. Uromodulin concentrations were measured from frozen stored urine by an established ELISA with excellent performance83. Concentrations were indexed to creatinine to account for urine dilution. Genetic associations were assessed with the same software and settings as for eQTL association (Supplementary Table 2). Colocalization analyses were performed with identical software and settings as described above for the association with gene expression.

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

**Data availability**

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

**References**


NATURE GENETICS | www.nature.com/naturegenetics
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

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 -Slt algorithms. Using the Wakefield’s formula implemented in the R package ‘gtx’, we derived approximate Bayes factors (ABF). Functional annotation of SNPs mapping into credible sets was performed with SNIPA v3.2 (March 2017), based on the 1000Gp3v5 and Ensembl v87 datasets. SNIPA was also used to derive the Combined Annotation Dependent Depletion (CADD) PHRED-like score, based on CADD v1.3. The Ensembl VEP tool was used for SNP’s primary effect prediction. For co-localization analyses, we used the ‘coloc.fast’ function from the R package ‘gtx’ (https://github.com/tobyjohnson/gtx), which is an adaptation 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

<table>
<thead>
<tr>
<th>n/a</th>
<th>Involved in the study</th>
</tr>
</thead>
<tbody>
<tr>
<td>☒</td>
<td>Antibodies</td>
</tr>
<tr>
<td>☒</td>
<td>Eukaryotic cell lines</td>
</tr>
<tr>
<td></td>
<td>Palaeontology</td>
</tr>
<tr>
<td>☒</td>
<td>Animals and other organisms</td>
</tr>
<tr>
<td>☒</td>
<td>Human research participants</td>
</tr>
<tr>
<td></td>
<td>Clinical data</td>
</tr>
</tbody>
</table>

Methods

<table>
<thead>
<tr>
<th>n/a</th>
<th>Involved in the study</th>
</tr>
</thead>
<tbody>
<tr>
<td>☒</td>
<td>ChIP-seq</td>
</tr>
<tr>
<td>☒</td>
<td>Flow cytometry</td>
</tr>
<tr>
<td>☒</td>
<td>MRI-based neuroimaging</td>
</tr>
</tbody>
</table>

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

Public Health and Community Medicine, Hebrew University of Jerusalem, Jerusalem, Israel. 101Department of Genomics of Common Disease, Imperial College London, London, UK. 102Massachusetts Veterans Epidemiology Research and Information Center, VA Cooperative Studies Program, VA Boston Healthcare System, Boston, MA, USA. 103Molecular Geriatrics, Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden. 104Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. 105Institute of Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. 106German Center for Diabetes Research (DZD), Neuherberg, Germany. 107Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA. 108QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. 109Icelandic Heart Association, Kopavogur, Iceland. 110Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavík, Iceland. 111Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia. 112Montreal University Hospital Research Center, CHUM, Montreal, Quebec, Canada. 113Medipharm, Montreal, Quebec, Canada. 114Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, US National Institutes of Health, Bethesda, MD, USA. 115Interdisciplinary Center of Psychopathology and Emotion Regulation (ICPE), University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. 116Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. 117Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. 118Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. 119Khto Teck Puan–National University Children's Medical Institute, National University Health System, Singapore, Singapore. 

120Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria. 121Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria. 122Department of Genetics, Shanghai–MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center, Shanghai, China. 123Shanghai Industrial Technology Institute, Shanghai, China. 124Department of Pediatrics, Tampere University Hospital, Tampere, Finland. 125Department of Pediatrics, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. 126NHHLBI's Framingham Heart Study, Framingham, MA, USA. 127The Center for Population Studies, NHLBI, Framingham, MA, USA. 128Division of Nephropathy, Internal Medicine Services, Landskaps–The National University Hospital of Iceland, Reykjavík, Iceland. 129Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. 130Stanford Cardiovascular Institute, Stanford University, Stanford, CA, USA. 131Molecular Epidemiology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden. 132Stanford Diabetes Research Center, Stanford University, Stanford, CA, USA. 133Max Planck Institute of Psychiatry, Munich, Germany. 134The Center of Public Health Sciences, University of Iceland, Reykjavík, Iceland. 135Department of Ophthalmology, Medical Faculty Mannheim, University Heidelberg, Mannheim, Germany. 136Beijing Institute of Ophthalmology, Beijing Key Laboratory of Ophthalmology and Visual Sciences, Beijing Tongren Hospital, Capital Medical University, Beijing, China. 137Geisinger Research, Biomedical and Translational Informatics Institute, Rockville, MD, USA. 138Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland. 139Department of Clinical Physiology, Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. 140Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. 141Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. 142Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. 143Department of Medicine, Kuoip University Hospital, Kuopio, Finland. 144Deutsches HerzZentrum München, Technische Universität München, Munich, Germany. 145DZHK (German Center for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany. 146Institute of Epidemiology and Biostatistics, University of Ulm, Ulm, Germany. 147National Heart and Lung Institute, Imperial College London, London, UK. 148Integrated Research and Treatment Center Adiposity Diseases, University of Leipzig, Leipzig, Germany. 149Division of Nephropathy and Hypertension, Loyola University Chicago, Chicago, IL, USA. 150Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria. 151RIKEN Center for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan. 152The Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland. 153Clinical Division of Diabetes, Department of Public Health, University of Helsinki, Helsinki, Finland. 154DIABETES Research Program, University of Helsinki, Helsinki, Finland. 155Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland. 156Division of Biomedical Informatics and Personalized Medicine, School of Medicine, University of Colorado Denver-Anschutz Medical Campus, Aurora, CO, USA. 157Department of Clinical Chemistry, FimmLab, Tampere, Finland. 158Department of Clinical Chemistry, Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. 159Institute of Epidemiology and Biobank Pogen, Kiel University, Kiel, Germany. 160A list of members and affiliations appears in the Supplementary Note. 161Department of Medicine, Tho Teck Puan Hospital, Singapore, Singapore. 162Cardiovascular Epidemiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden. 163Nuffield Department of Medicine, University of Oxford, Oxford, UK. 164Broad Institute of Harvard and MIT, Cambridge, MA, USA. 165Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. 166The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 167Target Sciences—Genetics, GlaxoSmithKline, Albuquerque, NM, USA. 168Wellcome Trust Center for Human Genetics, University of Oxford, Oxford, UK. 169Oxford Center for Diabetes, Endocrinology and Metabolism, Oxford, Oxford, UK. 170Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. 171Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany. 172Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria. 173Medical Clinic V, Medical Faculty Mannheim, University Heidelberg, Mannheim, Germany. 174Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. 175Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. 176Chair of Epidemiology, Ludwig-Maximilians-Universität München at UNIKA-T Augsburg, Augsburg, Germany. 177Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany. 178Institute of Human Genetics, Technische Universität München, Munich, Germany. 179Hypertension and Cardiovascular Disease, Department of Clinical Sciences Malmö, Lund University, Malmö, Sweden. 180Department of Psychiatry, VU University Medical Center, Amsterdam, the Netherlands. 181Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. 182Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland, Australia. 183Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands. 184Center for Public Health Genomics, University of Virginia, Charlottesville, Charlottesville, VA, USA. 185Division of Nephropathy, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 186Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. 187Data Tectica International, Glen Echo, MD, USA. 188Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. 189Department of Cardiology, Heart Center, Tampere University Hospital, Tampere, Finland. 190Department of Cardiology, Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, Tampere University, Tampere, Finland. 191Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. 192Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Center, Leiden, the Netherlands. 193University of Maryland School of Medicine, Baltimore, MD, USA. 194Cardiovascular Division, Brigham and Women's Hospital, Boston, MA, USA. 195TIMI Study Group, Boston, MA, USA. 196Department of Clinical Biochemistry, Landskaps–The National University Hospital of Iceland, Reykjavik, Iceland. 197Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. 198Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. 199Department of Medicine, Geriatrics Section, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA. 200National Institute for Health and Welfare, Helsinki, Finland. 201Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Li Punti, Sassari, Italy. 202Department of Psychiatry, University Hospital of
A catalogue of genetic loci associated with kidney function from analyses of a million individuals

Supplementary Materials

Matthias Wuttke, Yong Li, Man Li, Karsten B. Sieber, Mary F. Feitosa, Mathias Gorski et al.

Supplementary Note 1: Between-study heterogeneity and ancestry-related heterogeneity in the discovery meta-analysis

Supplementary Note 2: Trans-eQTL analysis

Supplementary Note 3: Details on the replication study: the Million Veteran’s Program (MVP)

Supplementary Figure 1: Analysis Flowchart

Supplementary Figure 2: Regional Association Plots

Supplementary Figure 3: Genetic Heritability

Supplementary Figure 4: BUN Manhattan plot

Supplementary Figure 5: Genetic Risk Score analysis

Supplementary Figure 6: Genetic correlation plot for eGFR

Supplementary Figure 7: Pathway and tissue enrichment analysis with DEPICT (eGFR)

Supplementary Figure 8: Pathway and tissue enrichment analysis with DEPICT (BUN)

Supplementary Figure 9: Co-localization of eGFR-association signals with gene expression across 44 GTEx tissues and two kidney tissues

Extended acknowledgements and study funding information

Supplementary Tables are provided separately as a spreadsheet.
Supplementary Note 1: Between-study heterogeneity and ancestry-related heterogeneity in the discovery meta-analysis

Before seeking replication, we evaluated results from the discovery meta-analysis for heterogeneity by design and heterogeneity related to ancestry. Most of the 308 SNPs showed homogeneous effects across studies (median $I^2=5\%$, interquartile range: 0-13%; Supplementary Table 3; Figure 2A). Only one index SNP had $I^2>50\%$ (UMOD-PDILT locus, $I^2=60\%$), where previously described heterogeneity\(^1,2\) is suspected to be age-related.\(^3\) We then investigated the heterogeneity of genetic effects that was correlated with ancestry using meta-regression\(^4\) (Methods) and identified three index SNPs with significant ancestry-related heterogeneity at the LINC01362, GATM, and PSD4 loci (ancestry heterogeneity p-value (p-anc-het) <0.05/308; Figure 2A, Supplementary Table 3). The index SNP at UMOD-PDILT did not show evidence for ancestry-related heterogeneity (p-anc-het=0.59). These results do not support large differences in estimated effects across ancestries for the majority of the identified SNPs. Ancestry-specific results for all 308 index SNPs are reported in Supplementary Table 4.

Supplementary Note 2: Trans-eQTL analysis

Trans-eQTL annotation of the index SNPs was only performed using whole blood and peripheral blood mononuclear cells, for which eQTL datasets with large sample size were available (Methods). Based on the analysis of 5 non-overlapping EA genome-wide eQTL studies (sample size range 1469 - 6645, Supplementary Table 14), we identified, among others, a reproducible link of rs10774625 (12q24.11) with several transcripts, including one for the calcium-binding protein gene S100A10 (1q21.3) and STAT1 (2q32.2). S100A10 encodes a subunit of annexin A2, which co-localizes with PLA2R at the cell surface and in extracellular vesicles from podocytes.\(^5\) Inhibition of STAT1 has been reported to protect from glomerular mesangial cell senescence\(^6\) and to ameliorate renal oxidative stress\(^7\) (Supplementary Table 15).
Supplementary Note 3: Details on the replication study: the Million Veteran’s Program (MVP)

Study definition. The MVP\(^8\) is an independent trans-ethnic study whose participants were recruited across 63 U.S. Veteran’s Administration (VA) medical facilities. Written informed consent was obtained and all documents and protocols were approved by the VA Central Institutional Review Board.

Genotypes. DNA was genotyped using a customized Affymetrix Axiom Biobank Array chip with additional content added to provide coverage of African and Hispanic haplotypes, as well as markers for common diseases in the VA population. After QC, genotypes were pre-phased using EAGLE v2\(^9\) and imputed based on the 1000Gp3v5 reference panel using minimac3.\(^{10}\) Genotype PCs were estimated using FlashPCA.\(^{11}\)

Phenotype. Serum creatinine was assessed up to one year prior to MVP enrollment using isotope dilution mass spectrometry. GFR was estimated using the CKD-EPI equation\(^{12}\) after excluding subjects on dialysis, transplant patients, amputees, individuals on HIV medications, and those with creatinine values of <0.4 mg/dl.

Additional epidemiological information. Diabetes was defined as use of anti-diabetic medications or by assignment of an International Classification of Diseases 9 (ICD-9) code for diabetes during the baseline period. Hypertension was defined as having an ICD-9 code for hypertension, being on antihypertensive drug or having ≥2 measures of systolic or diastolic blood pressure >140 mmHg or >90 mmHg, respectively.

Genome-wide association study. GWAS of eGFR on SNP dosage levels were performed by fitting linear regression models adjusted for age at creatinine measurement, age\(^2\), sex, body mass index, and the first 10 genetic PCs, using SNPTEST v2.5.4-beta.\(^{13}\) All GWAS were stratified by self-reported ethnicity (79.6% White non-Hispanic; 20.4% Black non-Hispanic), diabetes, and hypertension status. Results were combined across strata using fixed effects inverse-variance weighted meta-analysis in METAL.\(^{14}\) This analysis encompassed 280,722 individuals across all strata, of whom 216,518 were non-Hispanic Whites (EA).
References (Supplementary Material)


Supplementary Figure 1: Analysis Flowchart

Locus discovery analysis in the trans-ethnic sample and fine-mapping analysis in EA participants.

**Locus discovery analysis**
(maximal sample size)

- **Trans-ethnic GWAMA of eGFR**
  - 5 ancestries: EAS, SAS, AA, HIS, EA
  - 121 studies, N = 765,348
  - No. discovered loci: **308**

- **Generalizability and Replication**
  - Ancestry-related heterogeneity
  - Independent replication (N = 280,722)
  - No. replicated loci: **264**

- **Clinical & Epidemiological Landscape**
  - Kidney function relevance: BUN analysis (N = 416,178)
  - Effects on CKD (N = 625,219)
  - eGFR genetic risk score and clinical CKD (independent N = 452,264)
  - Genetic correlation with 748 traits

- **Functional enrichment**
  - Pathways
  - Cell-type & tissue expression
  - Cell-type specific heritability
  - Mouse kidney disease genes

**Fine-mapping analysis**
(ancestry with large LD reference panel)

- **European-ancestry GWAMA**
  - 85 studies, N = 567,460
  - No. discovered loci: **256**

- **Generalizability and Replication**
  - Independent replication (N = 216,518)
  - No. replicated loci: **228**

- **Kidney function relevance**
  - BUN analysis (N = 243,029)

- **Fine-mapping**
  - No. independent SNPs: **253**
  - Generation of 253 credible sets
  - Functional annotation of SNPs in credible sets

- **Gene prioritization**
  - Co-localization with gene expression in cis across 46 tissues
  - Co-localization with gene expression in trans in whole blood

**Catalogue** of genes and variants for functional follow-up
Supplementary Figure 2: Regional Association Plots
Regional Association Plot Booklet for all 308 loci identified in association with eGFR through trans-ethnic meta-analyses.

*The PDF booklet is available online as a separate download with the Supplementary Information.*
Supplementary Figure 3: Genetic Heritability

Distribution of the genetic heritability \( (h^2) \) estimates of age- and sex-adjusted log(eGFR) residuals in the Cooperative Health Research In South Tyrol (CHRIS) study, for index SNPs from the trans-ethnic GWAS. \( h^2 \) distribution is shown before (gray) and after (green) inclusion of the index SNPs into the model, with the shift representing the amount of \( h^2 \) explained by the index SNPs.
Supplementary Figure 4: BUN Manhattan plot

Manhattan plot of results from the GWAS meta-analysis of blood urea nitrogen (BUN).
Supplementary Figure 5: Genetic Risk Score analysis

Association between lower genetically-predicted eGFR based on a genetic risk score (GRS) and ICD-10 based clinical diagnoses from 452,264 individuals from the UK Biobank. Asthma is included as a negative control. The GRS was derived as described in the Methods. Displayed are odds ratios and their 95% CIs per 10% lower GRS-predicted eGFR. Dark gray is used for results from the 147 SNPs likely to be most relevant for kidney function (same as in Figure 2D), light gray is used for results from all 264 replicated eGFR-associated index SNPs.

- **N17 Acute renal failure**: 1.30 (1.16,1.47), p = 1.46e-05, 1.20 (1.10,1.31), p = 5.92e-05
- **N00-N08 Glomerular diseases**: 1.45 (1.22,1.71), p = 1.84e-05, 1.42 (1.25,1.62), p = 6.97e-08
- **N18 Chronic renal failure**: 2.13 (1.90,2.39), p = 8.05e-38, 2.01 (1.84,2.19), p = 1.40e-55
- **N20-N23 Urolithiasis**: 0.75 (0.69,0.83), p = 7.72e-09, 0.73 (0.67,0.78), p = 5.59e-18
- **I10-I15 Hypertensive diseases**: 1.07 (1.04,1.10), p = 6.27e-06, 1.04 (1.02,1.06), p = 1.47e-04
- **I20-I25 Ischaemic heart diseases**: 0.96 (0.92,1.00), p = 7.96e-02, 0.99 (0.96,1.02), p = 4.82e-01
- **J45 Asthma**: 0.99 (0.95,1.04), p = 7.98e-01, 0.99 (0.95,1.02), p = 5.29e-01

Odds ratio per 10% lower GRS-predicted eGFR
Supplementary Figure 6: Genetic correlation plot for eGFR

Genetic correlation plot based on the summary statistics from the trans-ethnic GWAS meta-analysis of eGFR and 748 other complex traits and diseases available through LD Hub.

The genetic correlations with citrate and cystatin C were not significant ($P=6.0 \times 10^{-4}$ and $4.0 \times 10^{-4}$, respectively, Supplementary Table 7), because these traits were measured in a limited number of studies, resulting in smaller GWAS sample sizes.
Supplementary Figure 7: Pathway and tissue enrichment analysis with DEPICT (eGFR)

Shown is the barplot of the results of the tissue and cell type enrichment analysis in Panel A. Cells, tissues and physiological systems are highlighted in blue if the association false discovery rate (FDR) was <0.05 and are summarized in the table in Panel B. The strongest enrichment was observed for urogenital and renal physiological systems and tissues: kidney, kidney cortex, and urinary tract. We additionally found significant enrichment for mucous membrane, respiratory mucosa, nasal mucosa, and nose (enrichment p-values from 3.1×10^-4 to 1.2×10^-3), possibly reflecting epithelial cell processes including transport mechanisms shared with the kidney. Panel C illustrates the highly correlated and strongly associated meta gene sets (P <1×10^-6, FDR<0.05) from the pathway and gene-set enrichment analysis clustered according to their biological relevance for kidney function, energy metabolism and signaling and transcription.
Supplementary Figure 8: Pathway and tissue enrichment analysis with DEPICT (BUN)

Shown is the barplot of the results of the tissue and cell type enrichment analysis. Cells, tissues and physiological systems are highlighted in blue if the association FDR was <0.05. Tissue and cell-type enrichment analysis of BUN-associated SNPs with $P<5 \times 10^{-8}$ highlighted a very similar pattern to the one observed for eGFR: the strongest enrichment was observed for urogenital and renal physiological systems and tissues, and significant enrichment was also observed for mucous membrane, respiratory mucosa, nasal mucosa, nose, epithelial cells and the epithelium.
**Supplementary Figure 9: Co-localization of eGFR-association signals with gene expression across 44 GTEx tissues and two kidney tissues**

All eGFR loci were tested for co-localization with all eQTLs where the eQTL cis-window overlapped (±100 kb) the sentinel genetic variants. Genes with at least one positive co-localization (posterior probability of one common causal variant, H4, ≥0.80) in any of the 44 tissues for which eQTL data was released by the GTEx Project or in two renal tissue are illustrated with the respective sentinel variants (Y-axis). Co-localizations across all tissues (X-axis) are illustrated as dots, where the size of the dots indicates the posterior probability of the co-localization. Negative co-localizations (posterior probability of H4 <0.80) are grey, while the positive co-localizations are color-coded based on the predicted change in expression relative to the allele associated with lower eGFR.
Supplementary Figure 9 – continued (1/4)

Posterior probability of colocalization
- [0.0, 0.2]
- [0.2, 0.4]
- [0.4, 0.6]
- [0.6, 0.8]
- [0.8, 1.0]

Change in gene expression with decreased eGFR
- 0.05
- 0.03
- 0.01
- 0.00
- -0.01
- -0.02
- -0.03
Supplementary Figure 9 – continued (3/4)
Supplementary Figure 9 – continued (4/4)
Extended acknowledgements and study funding information

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute, the National Institutes of Health, or the US Department of Health and Human Services.

**AA-DHS**  NIH R01 DK071891 (Barry I. Freedman, PI); NIH R01 NS075107 (Barry I. Freedman, PI).

**ADVANCE**  ADVANCE genomic sub-studies were supported by grants from the Ministry of Science and Innovation from the Quebec Government, from Genome Quebec, from the Consortium Québécois du Médicament, from the Canadian Institutes of Health Research and from Medpharmgene, OPTITHERA Inc and Les Laboratoires Servier.

**AFTER EU**  The AFTER EU study is the Danish part of the EURAGEDIC study which was supported by the European Commission (contract QLG2-CT-2001-01669). The genotyping for this study was part of the Genetics of Diabetic Nephropathy (GenDN) study, primarily funded by Juvenile Diabetes Research Foundation (JDRF) International Prime Award Number 17-2013-8. Tarunveer S Ahluwalia was also funded by the GenDN study grant and Lundbeck foundation Travel Grant (Ref. Number 2013-14471).

**AGES**  This study has been funded by NIA contract N01-AG-12100 and HHSN271201200022C with contributions from NEI, NIDCD and NHLBI, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, VSN: 00-063. The researchers thank the participants for their willingness to participate in the study.

**Airwave**  We thank all participants in the Airwave Health Monitoring Study. The Airwave Health Monitoring Study is funded by the UK Home Office (780-TETRA) with additional support from the National Institute for Health Research (NIHR), and the Imperial College Biomedical Research Centre in collaboration with Imperial College NHS Healthcare Trust. The views expressed are those of the authors and not necessarily those of the sponsors. PE acknowledges support from the Medical Research Council and Public Health England (MR/L01341X/1) for the MRC-PHE Centre for Environment and Health; and the NIHR Health Protection Research Unit in Health Impact of Environmental Hazards (HPRU-2012-10141). This work used computing resources of the UK MEDical BIOinformatics partnership (UK MED-BIO) supported by the Medical Research Council (MR/L01632X/1). PE is supported by the UK Dementia Research Institute which receives its funding from UK DRI Ltd funded by the UK Medical Research Council, Alzheimer’s Society and Alzheimer’s Research UK. PE is associate director of the Health Data Research UK London funded by a consortium led by the UK Medical Research Council.

**Amish**  We thank the Amish research volunteers for their long-standing partnership in research, and the research staff at the Amish Research Clinic for their work and dedication. The Amish contribution was supported by NIH grants R01 AG18728, R01 HL088119, U01 GM074518, U01 HL072515, U01 HL084756, and NIH K12RR023250, and P30 DK072488. Additional
support was provided by the University of Maryland General Clinical Research Center, grant M01 RR 16500, the Baltimore Veterans Administration Medical Center Geriatrics Research, and the Paul Beeson Physician Faculty Scholars in Aging Program.

**ARIC** The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I), R01HL087641, R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The work of Anna Köttgen was supported by a Heisenberg Professorship (KO 3598/3-1) as well as CRCs 1140 and 992 of the German Research Foundation.

**ASPS, ASPS-Fam** The authors thank the staff and the participants for their valuable contributions. We thank Birgit Reinhart for her long-term administrative commitment, Elfi Hofer for the technical assistance at creating the DNA bank, Ing. Johann Semmler and Anita Harb for DNA sequencing and DNA analyses by TaqMan assays and Irmgard Poelzl for supervising the quality management processes after ISO9001 at the biobanking and DNA analyses. The research reported in this article was funded by the Austrian Science Fond (FWF) grant number P20545-P05, P13180 and P1904 as well as by the Austrian National Bank (OeNB) Anniversary Fund grant number P15435 and the Austrain Federal Ministry of Science, Research and Economy under the aegis of the EU Joint Programme-Neurodegenerative Disease Research (JPND)-www.jpnd.eu. The Medical University of Graz supports the databank of the ASPS.

**BBJ** The BioBank Japan project is supported by the Ministry of Education, Culture, Sports, Sciences and Technology of Japanese government and the Japan Agency for Medical Research and Development.

**BES** BES was supported by the National Key Laboratory Fund, Beijing, China.

**BioMe** The Mount Sinai IPM Biobank Program is supported by The Andrea and Charles Bronfman Philanthropies. Ruth Loos is funded by R01DK110113, U01HG007417, R01DK101855, and R01DK107786.

**CHNS** The China Health and Nutrition Survey (CHNS) was supported by the China National Institute of Nutrition and Food Safety; the China Center for Disease Control; the National Institutes of Health (R01HD30880, R01HL108427, and R01DK104371); the Fogarty International Center of the National Institutes of Health; the China-Japan Friendship Hospital; the Chinese Ministry of Health; and the Carolina Population Center (R24 HD050924). Cassandra N. Spracklen was supported by the American Heart Association Postdoctoral Fellowship (17POST3650016).

**CHRIS** Full acknowledgements for the CHRIS study are reported at [http://translational-medicine.biomedcentral.com/articles/10.1186/s12967-015-0704-9#Declarations](http://translational-medicine.biomedcentral.com/articles/10.1186/s12967-015-0704-9#Declarations). The CHRIS study was funded by the Department of Innovation, Research, and University of the Autonomous
Province of Bolzano-South Tyrol.

**CHS**
Cardiovascular Health Study: This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN26820080007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, and U01HL130114 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Cilento**
We thank the populations of Cilento for their participation in the study. This work was supported by grants from the Italian Ministry of Universities and Research and CNR (Interomics Flagship Project, PON03PE_00060_7), the Assessorato Ricerca Regione Campania, the Fondazione con il SUD (2011-PDR-13), and the Istituto Banco di Napoli - Fondazione to Marina Ciullo.

**CoLaus**
The CoLaus|PsyCoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 3200B0–105993, 3200B0-118308, 33CSCO-122661, 33CS30-139468 and 33CS30-148401).

**CROATIA**
10001 Dalmatians: The Croatian Biobank (CROATIA) The CROATIA-Vis, CROATIA-Korcula and CROATIA-Split were funded by grants from the Medical Research Council (UK), from the Republic of Croatia Ministry of Science, Education and Sports (108-1080315-0302; 216-1080315-0302) and the Croatian Science Foundation (8875); and the CROATIA-Korčula genotyping was funded by the European Union framework program 6 project EUROSPAN (LSHGCT2006018947). We thank the staff of several institutions in Croatia that supported the field work, including Zagreb Medical Schools, the Institute for Anthropological Research in Zagreb, the recruitment team from the Croatian Centre for Global Health, University of Split and all the study participants. We are grateful to the Helmholtz Zentrum Munchen (Munich, Germany), AROS Applied Biotechnology, (Aarhus, Denmark) and the Edinburgh Clinical Research facility, University of Edinburgh (Edinburgh, United Kingdom) for SNP array genotyping. Genetic analyses were supported by the MRC HGU “QTL in Health and Disease” core programme.

**CZECH POST**
The study was supported by research grant 15-27109A provided by the Health Research Agency of the Ministry of Health, Czech Republic; Krka, tovarna zdravil, d.d., Novo mesto, Slovenia; Servier s.r.o., Czech Republic.

**DECODE**
The study was funded by deCODE Genetics/Amgen inc. We thank the study subjects for their valuable participation and our colleagues, who contributed to data collection, sample handling, and genotyping.
The D.E.S.I.R. study has been funded by INSERM contracts with Caisse nationale de l'assurance Maladie des Travailleurs Salariés (CNAMTS), Lilly, Novartis Pharma, and Sanofi-Aventis; INSERM (Réseaux en Santé Publique, Interactions entre les déterminants de la santé, Cohortes Santé TGIR 2008); the Association Diabète Risque Vasculaire; the Fédération Française de Cardiologie; La Fondation de France; Association de Langue Française pour l’Etude du Diabète et des Maladies Métaboliques (ALFEDIAM)/Société Francophone de Diabétologie (SFD); l’Office National Interprofessionnel des Vins (ONIVINS); Ardex Medical; Bayer Diagnostics; Becton Dickinson; Cardionics; Merck Santé; Novo Nordisk; Pierre Fabre; Roche; Topcon.


The Diabetic Cohort (DC) was supported by the individual research grant from the National Medical Research Council (NMRC) and the Biomedical Research Council (BMRC) of Singapore. The Genome Institute of Singapore provided services for genotyping.

DIACORE Cohort recruiting and management was funded by the KfH Stiftung Präventivmedizin e.V. (Carsten A. Böger). Genome-wide genotyping was funded the Else Kröner-Fresenius-Stiftung (2012_A147), the KfH Stiftung Präventivmedizin and the University Hospital Regensburg. Data analysis was funded by the Else Kröner-Fresenius Stiftung (2012_A147) and by the Deutsche Forschungsgemeinschaft with grant DFG BO 3815/1-4. The work of Iris M. Heid and Carsten A. Böger was supported by DFG CRC 1350 / C6 and and DFG BO 3815/4-1.

EGCUT The EGCUT studies were financed by Estonian Government (grants IUT20-60 and IUT24-6) and by European Commission through the European Regional Development Fund in the frame of grant Estonian Center of Genomics/Roadmap II (project No. 2014-2020.4.01.16-0125) and grant GENTRANSMED (Project No. 2014-2020.4.01.15-0012) and through H2020 grant no 692145 (ePerMed).

ERF The Erasmus Rucphen Family (ERF) has received funding from the Centre for Medical Systems Biology (CMSB) and Netherlands Consortium for Systems Biology (NCSB), both within the framework of the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO). ERF study is also a part of EUROSPAN (European Special Populations Research Network) (FP6 STRP grant number 018947 (LSHG-CT-2006-01947)); European Network of Genomic and Genetic Epidemiology (ENGAGE) from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413; "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254); FP7 project
EUROHEADPAIN (nr 602633), the Internationale Stichting Alzheimer Onderzoek (ISAO); the Hersenstichting Nederland (HSN); and the JNPD under the project PERADES (grant number 733051021, Defining Genetic, Polygenic and Environmental Risk for Alzheimer’s Disease using multiple powerful cohorts, focused Epigenetics and Stem cell metabolomics). We thank all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work, and P. Snijders M.D. for his help in data collection of the data. Jun Liu, Cornelia M. van Duijn, and Ayse Demirkan have used exchange grants from the Personalized pREvention of Chronic Diseases consortium (PRECeDI). Ayse Demirkan is supported by a Veni grant (2015) from ZonMw.

**ESTHER**
The ESTHER study was funded by the Saarland state Ministry for Social Affairs, Health, Women and Family Affairs (Saarbrücken, Germany), the Baden-Württemberg state Ministry of Science, Research and Arts (Stuttgart, Germany), the Federal Ministry of Education and Research (Berlin, Germany) and the Federal Ministry of Family Affairs, Senior Citizens, Women and Youth (Berlin, Germany).

**FamHS**
The study was supported by grant R01-DK-089256 from NIDDK and grant R01HL117078 from NHLBI.

**FHS**
The Framingham Heart Study is supported by HHSN268201500001.

**FINCAVAS**
The Finnish Cardiovascular Study (FINCAVAS) has been financially supported by the Competitive Research Funding of the Tampere University Hospital (Grant 9M048 and 9N035), the Finnish Cultural Foundation, the Finnish Foundation for Cardiovascular Research, the Emil Aaltonen Foundation, Finland, the Tampere Tuberculosis Foundation, and EU Horizon 2020 (grant 755320 for TAXINOMISIS). The authors thank the staff of the Department of Clinical Physiology for collecting the exercise test data.

**Finrisk**
We thank all individuals who participated in Finrisk surveys. Veikko Salomaa and Mikko Kuokkanen were supported by the Finnish Foundation for Cardiovascular Research. Mikko Kuokkanen was also supported by Päiviikki and Sakari Sohlberg Foundation. We thank the CSC – IT center for science, Finland, for providing computational resources.

**GCKD**
The GCKD study was funded by the German Ministry of Research and Education (Bundesministerium für Bildung und Forschung, BMBF), by the Foundation KfH Stiftung Präventivmedizin and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 246781735 – SFB 1140. Unregistered grants to support the study were provided by Bayer, Fresenius Medical Care and Amgen. Genotyping was supported by Bayer Pharma AG. Uromodulin measurements in GCKD were supported by the Swiss National Centre of Competence in Research Kidney Control of Homeostasis program and the Swiss National Science Foundation grant 31003A_169850. The work of Matthias Wuttke was supported by DFG CRC 1140 and the Else Kroener Fresenius Forschungskolleg NAKSYS. The work of Yong Li was supported by DFG KO 3598/4-1. The work of Anselm Hoppmann was partly supported by NIDDK R01 DK087635-09.
**Generation R**

The Generation R Study is conducted by the Erasmus Medical Center in close collaboration with the School of Law and Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service Rotterdam area, Rotterdam, the Rotterdam Homecare Foundation, Rotterdam and the Stichting Trombosedienst & Arsenlaboratorium Rijnmond (STAR-MDC), Rotterdam. We acknowledge the contribution of children and parents, general practitioners, hospitals, midwives and pharmacies in Rotterdam. The generation and management of GWAS genotype data for the Generation R Study was done at the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. We would like to thank Karol Estrada, Dr. Tobias A. Knoch, Anis Abuseiris, Luc V. de Zeeuw, and Rob de Graaf, for their help in creating GRIMP, BigGRID, MediGRID, and Services@MediGRID/D-Grid, [funded by the German Bundesministerium fuer Forschung und Technology; grants 01 AK 803 A-H, 01 IG 07015 G] for access to their grid computing resources. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Manoushka Ganesh, Lizbeth Herrera and Marjolein Peters for their help in creating, managing and QC of the GWAS database. The general design of Generation R Study is made possible by financial support from the Erasmus MC, University Medical Center, Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw) and the Ministry of Health, Welfare and Sport. Janine Felix and Vincent Jaddoe received additional funding from the European Union’s Horizon 2020 research and innovation programme (733206, LIFECYCLE). Fernando Rivadeneira received additional funding from the Netherlands Organization for Health Research and Development (VIDI 016. 136. 367). This project received additional funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 633595 (DynaHEALTH). Vincent Jaddoe received additional funding from the Netherlands Organization for Health Research and Development (VIDI 016.136.361), European Research Council (ERC Consolidator Grant, ERC-2014-CoG-648916) and European Union’s FP-7 programme (Early Nutrition).

**GS:SFHS**

Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. We thank all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award “STRatifying Resilience and Depression Longitudinally” (STRADL) Reference 104036/Z/14/Z).

**GSK**

Funding Source: Max-Planck Society, German Federal Ministry of Education and Research (BMBF) in the framework of the National Genome Research Network (NGFN), Foerderkennzeichen 01GS0481.

**HANDLS**

The authors thank all study participants and the Healthy Aging in
Neighborhoods of Diversity across the Life Span (HANDLS) study medical staff for their contributions. The HANDLS study was supported by the Intramural Research Program of the NIH, National Institute on Aging and the National Center on Minority Health and Health Disparities (project # Z01-AG000513 and human subjects protocol number 09-AG-N248). Data analyses for the HANDLS study utilized the high-performance computational resources of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. (http://biowulf.nih.gov; http://hpc.nih.gov).

**HYPERGENES**
Funding Source: HYPERGENES project (FP7-HEALTH-F4-2007-201550) and InterOmics (PB05 MIUR-CNR Italian Flagship Project).

**INGI-CARL**
The project was approved by the local administration of Carlantino, the Health Service of Foggia Province, Italy, and ethical committee of the IRCCS Burlo-Garofolo of Trieste. We thank the people of Carlantino for their support.

**INGI-FVG**
Project co-financed by the European Regional Development Fund under the Regional Operational Programme of Friuli Venezia Giulia - Objective "Regional Competitiveness and Employment" 2007/2013, Telethon Foundation (GGP09037), Fondo Trieste (2008), Regione FVG (L.26.2008), and Italian Ministry of Health (RC16/06, ART. 13 D.LGS 297/99) (to Paolo Gasparini). We thank the people of Friuli Venezia Giulia Region for their support.

**INGI-VBI**
The research was supported by funds from Compagnia di San Paolo, Torino, Italy; Fondazione Cariplo, Italy and Ministry of Health, Ricerca Finalizzata 2011 and CCM 2010 to Daniela Toniolo. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank the inhabitants of the VB that made this study possible, the local administrations, the Tortona and Genova archdiocese and the ASL-22, Novi Ligure (Al) for support. We thank Clara Camaschella for data collection supervision and organization of the clinical data collection, Fiammetta Viganò for technical help, and Corrado Masciullo for the analysis platform.

**INTERVAL**
We thank all INTERVAL voluntary participants. We thank the INTERVAL study co-ordination teams (at the Universities of Cambridge and Oxford and at NHS Blood and Transplant [NHSBT]), including the blood donation staff at the 25 static centers, for their help with INTERVAL participant recruitment and study fieldwork, as well as the Cambridge BioResource and NHSBT staff for their help with volunteer recruitment. We thank members of the Cambridge BioResource Scientific Advisory Board and Management Committee for their support of our study and the NIH Research Cambridge Biomedical Research Centre for funding (RG64219). The INTERVAL academic coordinating centre receives core support from the UK Medical Research Council (G0800270), the BHF (SP/09/002), the NIHR, and Cambridge Biomedical Research Centre, as well as grants from the European Research Council (268834), the European Commission Framework Programme 7 (HEALTH-F2-2012-279233), MSD, and Pfizer. The INTERVAL study is funded by NHSBT (11-01-GEN) and has been supported by the NIHR-BTRU in Donor Health and Genomics (NIHR BTRU-2014-10024) at the University of Cambridge in partnership with
The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health of England, or NHSBT. The Cardiovascular Epidemiology Unit at the University of Cambridge is funded by UK MRC (G0800270), BHF (SP/09/002), UK NIHR Cambridge Biomedical Research Centre, ERC (268834), and European Commission Framework Programme 7 (HEALTH-F2-2012-279233). Willem H. Ouwehand is a NIHR Senior Investigator. David J. Roberts was supported by the NIHR Programme “Erythropoiesis in Health and Disease” (NIHR-RP-PG-0310-1004). John Danesh is a BHF Professor, European Research Council Senior Investigator, and NIHR Senior Investigator. Nicole Soranzo is supported by the Wellcome Trust (WT098051 and WT091310), the EU FP7 (EPiGENESYS 257082 and BLUEPRINT HEALTH-F5-2011-282510).

**JHS** The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I/HHSN26800001) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD). The authors also wish to thank the staffs and participants of the JHS. James G. Wilson is supported by U54GM115428 from the National Institute of General Medical Sciences. Laura M. Raffield is supported by T32 HL129982.

**JUPITER** The JUPITER trial and its genetic substudy were funded by AstraZeneca.

**KORA** The KORA research platform (KORA, Cooperative Health Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München - German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMIinnovativ. Statistical KORA analyses were supported by DFG BO-3815/4-1 (to Carsten A. Böger) and BMBF 01ER1206 and 01ER1507 (to Iris M. Heid) and by University of Regensburg.

**LIFE-Adult, LIFE-Heart, LIFE-Child** LIFE-Adult, LIFE-Heart, and LIFE-Child, are funded by the Leipzig Research Center for Civilization Diseases (LIFE). LIFE is an organizational unit affiliated to the Medical Faculty of the University of Leipzig. LIFE is funded by means of the European Union, by the European Regional Development Fund (ERDF) and by funds of the Free State of Saxony within the framework of the excellence initiative.

**Lifelines** The Lifelines Cohort Study, and generation and management of GWAS genotype data for the Lifelines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the
University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. The authors wish to acknowledge the services of the Lifelines Cohort Study, the contributing research centers delivering data to Lifelines, and all the study participants.

**Living-biobank**
The genotyping was funded by the Agency for Science, Technology and Research, Singapore (https://www.a-star.edu.sg/), and by Merck Sharp & Dohme Corp., Whitehouse Station, NJ USA (http://www.merck.com). The participants were from the MEC study (funded by the Biomedical Research Council (BMRC 03/1/27/18/216), National Medical Research Council (0838/2004), National Research Foundation (through BMRC 05/1/21/19/425 and 11/1/21/19/678), Ministry of Health, Singapore, National University of Singapore and National University Health System, Singapore), and The SH2012 study (funded by the Ministry of Health, Singapore, National University of Singapore and National University Health System, Singapore).

**LLFS**
The study was supported by grants: U01AG023746, U01AG023712, U01AG023749, U01AG023755, and U01AG023744 from NIA.

**LOLIPOP**
The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966, G0700931), the Wellcome Trust (084723/Z/08/Z, 090532 & 098381) the NIHR (RP-PG-0407-10371), the NIHR Official Development Assistance (ODA, award 16/136/68), the European Union FP7 (EpiMigrant, 279143), and H2020 programs (iHealth-T2D, 643774). We acknowledge support of the MRC-PHE Centre for Environment and Health, and the NIHR Health Protection Research Unit on Health Impact of Environmental Hazards. The work was carried out in part at the NIHR/Wellcome Trust Imperial Clinical Research Facility. The views expressed are those of the author(s) and not necessarily those of the Imperial College Healthcare NHS Trust, the NHS, the NIHR or the Department of Health. We thank the participants and research staff who made the study possible. JC is supported by the Singapore Ministry of Health’s National Medical Research Council under its Singapore Translational Research Investigator (STaR) Award (NMRC/STaR/0028/2017).

**LURIC**
LURIC was supported by the 7th Framework Program of the EU (AtheroRemo, grant agreement number 201668 and RiskyCAD, grant agreement number 305739). The work of W.M. and M.E.K. is supported as part of the Competence Cluster of Nutrition and Cardiovascular Health (nutriCARD) which is funded by the German Ministry of Education and Research (grant agreement no. 01EA1411A). The authors thank the LURIC study team involved in patient recruitment as well as sample and data handling, and the laboratory staff at the Ludwigshafen General Hospital and the universities of Freiburg, Ulm and Heidelberg, Germany.

**MDC-CC**
This study was supported by the European Research Council (Consolidator grant nr 649021, Orho-Melander), the Swedish Research Council, the Swedish Heart and Lung Foundation, the Novo Nordic Foundation, the Swedish Diabetes Foundation, and the Påhlsson Foundation, and by equipment grants from the Knut and Alice Wallenberg
MESA and the MESA SHARE project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. MESA Family is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support is provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071258, R01HL071259, by the National Center for Research Resources, Grant UL1RR033176, and the National Center for Advancing Translational Sciences, Grant UL1TR001881. This publication was developed under a STAR research assistance agreement, No. RD831697 (MESA Air), awarded by the U.S Environmental protection Agency. It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors and the EPA does not endorse any products or commercial services mentioned in this publication.

METSIM Michael Boehnke was supported by NIH grant DK062370.

MICROS We thank all study participants, all primary care practitioners, and the personnel of the Hospital of Silandro (Department of Laboratory Medicine) for their participation and collaboration in the research project. The study was supported by the Ministry of Health and Department of Educational Assistance, University and Research of the Autonomous Province of Bolzano, the South Tyrolean Sparkasse Foundation, and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947).

MVP This work was supported by the MVP-VA Award #I01BX003360 (PI Adriana M. Hung).

MyCode (Geisinger) We would like to acknowledge the participants, staff, and our colleagues associated with the Geisinger MyCode Community Health Initiative. We also thank the staff of the PACDC of Geisinger for assistance with the phenotypic data, and the staff of the Biomedical & Translational Informatics and Kidney Health Research Institute.

NEO The authors of the NEO study thank all individuals who participated in the Netherlands Epidemiology in Obesity study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group, Pat van Beelen, Petra Noordijk and Ingeborg de Jonge for the coordination, lab and data management of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research
Profile Area Vascular and Regenerative Medicine. Dennis Mook-Kanamori is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023).

**NESDA**
The infrastructure for the NESDA study ([www.nesda.nl](http://www.nesda.nl)) is funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development (ZonMw, grant no. 10-000-1002) and financial contributions by participating universities and mental health care organizations (VU University Medical Center, GGZ inGeest, Leiden University Medical Center, Leiden University, GGZ Rivierduinen, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, and Rob Giel Onderzoekscentrum). Statistical analyses were carried out on the Genetic Cluster Computer ([http://www.geneticcluster.org](http://www.geneticcluster.org)), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation.

**OGP**
The Ogliastra Genetic Park study was supported by grant from the Italian Ministry of Education, University and Research (MIUR) no. 5571/DSPAR/2002. We thank all study participants for their contributions and the municipal administrations for their economic and logistic support.

**ORCADES**
The Orkney Complex Disease Study (ORCADES) was supported by the Chief Scientist Office of the Scottish Government (CZB/4/276, CZB/4/710), a Royal Society URF to James F. Wilson, the MRC Human Genetics Unit quinquennial programme “QTL in Health and Disease”, Arthritis Research UK and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). DNA extractions were performed at the Wellcome Trust Clinical Research Facility in Edinburgh. We would like to acknowledge the invaluable contributions of the research nurses in Orkney, the administrative team in Edinburgh, the people of Orkney, and the data analysts in particular Dr Thibaud Boutin for the genotype imputation to the HRC reference panel.

**PIVUS**
The ULSAM study was supported by Wellcome Trust grants WT098017, WT064890, WT090532, Uppsala University, Uppsala University Hospital, the Swedish Research Council and the Swedish Heart-Lung Foundation. Cecilia M. Lindgren is supported by the Li Ka Shing Foundation, WT-SSI/John Fell funds and by the NIHR Biomedical Research Centre, Oxford, by Widenlife and NIH (5P50HD028138-27). Johan Ärnlöv was supported by The Swedish Research Council, Swedish Heart-Lung Foundation, Dalarna University and Uppsala University.

**PREVEND**
The Prevention of Renal and Vascular Endstage Disease Study (PREVEND) genetics is supported by the Dutch Kidney Foundation (Grant E033), the EU project grant GENECURE (FP-6 LSHM CT 2006 037697), the National Institutes of Health (grant LM010098), the Netherlands organization for health research and development (NWO VENI grant 916.761.70), and the Dutch Inter University Cardiology Institute Netherlands (ICIN). Niek Verweij was supported by NWO VENI grant 016.186.125.

**POPGEN**
The PopGen 2.0 network was supported by a grant from the German Ministry for Education and Research (01EY1103). Sandra Freitag-Wolf was supported by German Research Foundation, Clusters of Excellence 306,
Inflammation at Interfaces.

**QIMR adolescent**

We acknowledge funding by Australian National Health and Medical Research Council (NHMRC) grants 241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485, 552498 and Australian Research Council grants A7960034, A79906588, A79801419, DP0770096, DP0212016, DP0343921 for building and maintaining the adolescent twin family resource through which samples were collected.

**QIMR adult**

We acknowledge the contributions of many staff in the Genetic Epidemiology Unit, Queensland Institute of Medical Research, in interviewing study participants, sample processing and DNA extraction, and data management. Funding for aspects of this work was provided by the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485, 552498), the Australian Research Council (A7960034, A79906588, A79801419, DP0770096, DP0212016, DP0343921), the EU 5th Framework Programme GenomEUtwin Project (QLG2-CT-2002-01254), and the U.S. National Institutes of Health (AA07535, AA10248, AA11998, AA13320, AA13321, AA13326, AA14041, AA17688, DA12854, MH66206). G.W.M. was supported by National Health and Medical Research Council (NHMRC) Fellowship Schemes.

**RS**

The Rotterdam Study (RS) has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The RS has been entered into the Netherlands National Trial Register (NTR; www.trialregister.nl) and into the WHO International Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp/network/primary/en/) under shared catalogue number NTR6831. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians. The generation and management of GWAS genotype data for the RS (RS I, RS II, RS III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Karol Estrada, PhD, Yuri Aulchenko, PhD, and Carolina Medina-Gomez, MSc, for the creation and analysis of imputed data. The RS is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study
participants, the staff from the RS and the participating general practitioners and pharmacists.

**SCES**
The Singapore Chinese Eye Study (SCES) was supported by grants from the National Medical Research Council (STaR/0003/2008), the Singapore Bio Imaging Consortium (C-011/2006) and the Biomedical Research Council (08/1/35/19/550). The Genome Institute of Singapore provided services for genotyping.

**SHIP/SHIP-Trend**
SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network ‘Greifswald Approach to Individualized Medicine (GANI_MED)’ funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthineers, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania. The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH.

**SiMES**
The Singapore Malay Eye Study (SiMES) was funded by the National Medical Research Council (NMRC 0796/2003 and NMRC/STaR/0003/2008) and Biomedical Research Council (BMRC, 09/1/35/19/616). The Genome Institute of Singapore provided services for genotyping.

**SINDI**
The Singapore Indian Eye Study (SINDI) was funded by grants from the Biomedical Research Council of Singapore (BMRC 09/1/35/19/616 and 08/1/35/19/550), and the National Medical Research Council of Singapore (NMRC/STaR/0003/2008). The Genome Institute of Singapore provided services for genotyping.

**SCHS - CHD cases and controls**
The Singapore Chinese Health Study was supported by the National Institutes of Health, USA (RO1 CA144034 and UM1 CA182876), the nested case-control study of myocardial infarction by the Singapore National Medical Research Council (NMRC 1270/2010), and genotyping by the HUJ-CREATE Programme of the National Research Foundation, Singapore (Project Number 370062002).

**SKIPOGH**
The SKIPOGH study was supported by a Swiss national science foundation grant (FN33CM30-124087 and FN33CM30-140331), by the Swiss National Centres of Competence in Research Kidney and TransCure. Olivier Devuyst was supported by Swiss National Science Foundation (project grant 31003A-169850) and the Swiss National Centre of Competence in Research (NCCR) Kidney Control of Homeostasis (Kidney.CH).

**SOLID-TIMI 52**
The SOLID-TIMI 52 trial was supported and funded by grants from GlaxoSmithKline.

**Sorbs**
This work was supported by grants from the German Research Foundation (SFB-1052 “Obesity mechanisms” A01, B03, SPP 1629 TO 718/2-1), from the German Diabetes Association, from the DHFD (Diabetes Hilfs- und Forschungsfonds Deutschland) and from IFB Adiposity Diseases (AD2-060E, AD2-06E95, AD2-06E99). IFB Adiposity Diseases is supported by
the Federal Ministry of Education and Research (BMBF), Germany, FKZ: 01EO1501.

**SP2** The Singapore Prospective Study Program (SP2) were supported by the individual research grant and clinician scientist award schemes from the National Medical Research Council (NMRC) and the Biomedical Research Council (BMRC) of Singapore. The Genome Institute of Singapore provided services for genotyping.

**STABILITY** The STABILITY trial was supported and funded by grants from GlaxoSmithKline.

**TRAILS** TRAILS (TRacking Adolescents’ Individual Lives Survey) is a collaborative project involving various departments of the University Medical Center and University of Groningen, the University of Utrecht, the Radboud Medical Center Nijmegen, and the Parnassia Bavo group, all in the Netherlands. TRAILS has been financially supported by grants from the Netherlands Organization for Scientific Research NWO (Medical Research Council program grant GB-MW 940-38-011; ZonMW Brainpower grant 100-001-004; ZonMw Risk Behavior and Dependence grant 60-60600-97-118; ZonMw Culture and Health grant 261-98-710; Social Sciences Council medium-sized investment grants GB-MaGW 480-01-006 and GB-MaGW 480-07-001; Social Sciences Council project grants GB-MaGW 452-04-314 and GB-MaGW 452-06-004; NWO large-sized investment grant 175.010.2003.005; NWO Longitudinal Survey and Panel Funding 481-08-013 and 481-11-001); the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-NL (CP 32), the participating universities, and Accare Center for Child and Adolescent Psychiatry. Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation. We are grateful to all adolescents who participated in this research and to everyone who worked on this project and made it possible.

**TwinGene** TwinGene is a sub-study of the Swedish Twin Registry which is managed by Karolinska Institutet and receives funding through the Swedish Research Council under the grant no 2017-00641. TwinGene was supported by grants from the Swedish Research Council (M-2005-1112), GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254), NIH DK U01-066134, The Swedish Foundation for Strategic Research (SSF). Heart and Lung foundation no. 20070481. Analysis on renal biomarkers was supported by grants from Stockholm County Council, The Swedish Society of Medicine (SLS-412071), the Serafimer Hospital Foundation and Stiftelsen för Njursjuka.

**ULSAM** The ULSAM study was supported by Wellcome Trust grants WT098017, WT064890, WT090532, Uppsala University, Uppsala University Hospital, the Swedish Research Council and the Swedish Heart-Lung Foundation. Johan Ärnlöv was supported by The Swedish Research Council, Swedish Heart-Lung Foundation, Dalarna University and Uppsala University.

**Vanderbilt** The data used for the analyses were obtained from Vanderbilt University Medical Center’s BioVU, which is supported by numerous sources:
institutional funding, private agencies, and federal grants. These include
the NIH funded Shared Instrumentation Grant S10RR025141; and CTSA
grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data
are also supported by investigator-led projects that include U01HG004798,
R01NS032830, RC2GM092618, P50GM115305, U01HG006378,
U19HL065962, R01HD074711; and additional funding sources listed at
https://victr.vanderbilt.edu/pub/biovu/. Jacklyn N. Hellwege is supported by
the Vanderbilt Molecular and Genetic Epidemiology of Cancer training
program, funded by T32CA160056.

VIKING  The Viking Health Study – Shetland (VIKING) was supported by the MRC
Human Genetics Unit quinquennial programme grant “QTL in Health and
Disease”. DNA extractions and genotyping were performed at the
Edinburgh Clinical Research Facility, University of Edinburgh. We
acknowledge the contributions of the research nurses in Shetland, the
administrative team in Edinburgh, the people of Shetland, and the data
analysts, in particular Dr Thibaud Boutin for the genotype imputation to the
HRC reference panel.

WGHS  The Women's Genome Health Study (WGHS) is supported by the National
Heart, Lung, and Blood Institute (HL043851 and HL080467) and the
National Cancer Institute (CA047988 and UM1CA182913) with funding for
genotyping provided by Amgen.

YFS  The Young Finns Study has been financially supported by the Academy of
Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378
(Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution
of Finland; Competitive State Research Financing of the Expert
Responsibility area of Kuopio, Tampere and Turku University Hospitals
(grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish
Foundation for Cardiovascular Research; Finnish Cultural Foundation; The
Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil
Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane
Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes
Association; and EU Horizon 2020 (grant 755320 for TAXINOMISIS); and
European Research Council (grant 742927 for MULTIEPIGEN project);
Tampere University Hospital Supporting Foundation. We thank the teams
that collected data at all measurement time points; the persons who
participated as both children and adults in these longitudinal studies; and
biostatisticians Irina Lisinen, Johanna Ikonen, Noora Kartiosuo, Ville Aalto,
and Jarno Kankaanranta for data management and statistical advice.

Variant annotation was supported by software resources provided via the Caché Campus
program of the InterSystems GmbH to Alexander Teumer.
LifeLines Cohort Study – Author Group

Behrooz Z Alizadeh (1), H Marike Boezen (1), Lude Franke (2), Pim van der Harst (3), Gerjan Navis (4), Marianne Rots (5), Morris Swertz (2), Bruce HR Wolffenbuttel (6), Cisca Wijmenga (2)

(1) Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands

(2) Department of Genetics, University of Groningen, University Medical Center Groningen, The Netherlands

(3) Department of Cardiology, University of Groningen, University Medical Center Groningen, The Netherlands

(4) Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, The Netherlands

(5) Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, The Netherlands

(6) Department of Endocrinology, University of Groningen, University Medical Center Groningen, The Netherlands
Million Veteran Program: Consortium Acknowledgement

MVP Executive Committee
- Co-Chair: J. Michael Gaziano, M.D., M.P.H.
- Co-Chair: Rachel Ramoni, D.M.D., Sc.D.
- Jim Breeling, M.D. (ex-officio)
- Kyong-Mi Chang, M.D.
- Grant Huang, Ph.D.
- Sumitra Muralidhar, Ph.D.
- Christopher J. O’Donnell, M.D., M.P.H.
- Philip S. Tsao, Ph.D.

MVP Program Office
- Sumitra Muralidhar, Ph.D.
- Jennifer Moser, Ph.D.

MVP Recruitment/Enrollment
- Recruitment/Enrollment Director/Deputy Director, Boston – Stacey B. Whitbourne, Ph.D.; Jessica V. Brewer, M.P.H.
- MVP Coordinating Centers
  - Clinical Epidemiology Research Center (CERC), West Haven – John Concato, M.D., M.P.H.
  - Cooperative Studies Program Clinical Research Pharmacy Coordinating Center, Albuquerque - Stuart Warren, J.D., Pharm D.; Dean P. Argyres, M.S.
  - Genomics Coordinating Center, Palo Alto – Philip S. Tsao, Ph.D.
  - Massachusetts Veterans Epidemiology Research Information Center (MAVERIC), Boston - J. Michael Gaziano, M.D., M.P.H.
  - MVP Information Center, Canandaigua – Brady Stephens, M.S.
- Core Biorepository, Boston – Mary T. Brophy M.D., M.P.H.; Donald E. Humphries, Ph.D.
- MVP Informatics, Boston – Nhan Do, M.D.; Shahpoor Shayan
- Data Operations/Analytics, Boston – Xuan-Mai T. Nguyen, Ph.D.

MVP Science
- Genomics - Christopher J. O’Donnell, M.D., M.P.H.; Saiju Pyarajan Ph.D.; Philip S. Tsao, Ph.D.
- Phenomics - Kelly Cho, M.P.H, Ph.D.
- Data and Computational Sciences – Saiju Pyarajan, Ph.D.
- Statistical Genetics – Elizabeth Hauser, Ph.D.; Yan Sun, Ph.D.; Hongyu Zhao, Ph.D.

MVP Local Site Investigators
- Atlanta VA Medical Center (Peter Wilson)
- Bay Pines VA Healthcare System (Rachel McArdle)
- Birmingham VA Medical Center (Louis Dellitalia)
- Cincinnati VA Medical Center (John Harley)
- Clement J. Zablocki VA Medical Center (Jeffrey Whittle)
- Durham VA Medical Center (Jean Beckham)
- Edith Nourse Rogers Memorial Veterans Hospital (John Wells)
- Edward Hines, Jr. VA Medical Center (Salvador Gutierrez)
- Fayetteville VA Medical Center (Gretchen Gibson)
- VA Health Care Upstate New York (Laurence Kaminsky)
- New Mexico VA Health Care System (Gerardo Villareal)
- VA Boston Healthcare System (Scott Kinlay)
- VA Western New York Healthcare System (Junzhe Xu)
- Ralph H. Johnson VA Medical Center (Mark Hamner)
- Wm. Jennings Bryan Dom VA Medical Center (Kathlyn Sue Haddock)
- VA North Texas Health Care System (Sujata Bhushan)
- Hampton VA Medical Center (Pran Iruvanti)
- Hunter Holmes McGuire VA Medical Center (Michael Godschalk)
- Iowa City VA Health Care System (Zuhair Ballas)
- Jack C. Montgomery VA Medical Center (Malcolm Buford)
- James A. Haley Veterans’ Hospital (Stephen Mastorides)
- Louisville VA Medical Center (Jon Klein)
- Manchester VA Medical Center (Nora Ratcliffe)
- Miami VA Health Care System (Hermes Florez)
- Michael E. DeBakey VA Medical Center (Alan Swann)
- Minneapolis VA Health Care System (Maureen Murdoch)
- N. FL/S. GA Veterans Health System (Peruvemba Sriram)
- Northport VA Medical Center (Shing Shing Yeh)
- Overton Brooks VA Medical Center (Ronald Washburn)
- Philadelphia VA Medical Center (Darshana Jhala)
- Phoenix VA Health Care System (Samuel Aguayo)
- Portland VA Medical Center (David Cohen)
- Providence VA Medical Center (Satish Sharma)
- Richard Roudebush VA Medical Center (John Callaghan)
- Salem VA Medical Center (Kris Ann Oursler)
- San Francisco VA Health Care System (Mary Whooley)
- South Texas Veterans Health Care System (Sunil Ahuja)
- Southeast Louisiana Veterans Health Care System (Amparo Gutierrez)
- Southern Arizona VA Health Care System (Ronald Schifman)
- Sioux Falls VA Health Care System (Jennifer Greco)
- St. Louis VA Health Care System (Michael Rauchman)
- Syracuse VA Medical Center (Richard Servatius)
- VA Eastern Kansas Health Care System (Mary Oehlert)
- VA Greater Los Angeles Health Care System (Agnes Wallborn)
- VA Loma Linda Healthcare System (Ronald Fernando)
- VA Long Beach Healthcare System (Timothy Morgan)
- VA Maine Healthcare System (Todd Stapley)
- VA New York Harbor Healthcare System (Scott Sherman)
- VA Pacific Islands Health Care System (Gwenevere Anderson)
- VA Palo Alto Health Care System (Philip Tsao)
- VA Pittsburgh Health Care System (Elif Sonel)
- VA Puget Sound Health Care System (Edward Boyko)
- VA Salt Lake City Health Care System (Laurence Meyer)
- VA San Diego Healthcare System (Samir Gupta)
- VA Southern Nevada Healthcare System (Joseph Fayad)
- VA Tennessee Valley Healthcare System (Adriana Hung)
- Washington DC VA Medical Center (Jack Lichy)
- W.G. (Bill) Hefner VA Medical Center (Robin Hurley)
- White River Junction VA Medical Center (Brooks Robey)
- William S. Middleton Memorial Veterans Hospital (Robert Striker)
rs12183971

rs11743174

rs3812036

rs1175524

16 GWAS hits omitted

16 GWAS hits omitted

15 GWAS hits omitted

15 GWAS hits omitted

rs12183971

rs11743174

rs3812036

rs1175524

rs12183971

rs11743174

rs3812036

rs1175524

rs12183971

rs11743174

rs3812036

rs1175524
22 GWAS hits omitted

rs3705502

rs1165406

rs2855159

rs35984974

Position on chr6 (Mb)

Position on chr6 (Mb)

Position on chr6 (Mb)

Position on chr6 (Mb)
Response to fenofibrate (adiponectin levels)

1 GWAS hit

Position on chr6 (Mb)

Renal function-related traits (eGFR area)

- Position on chr6 (Mb)

- Position on chr6 (Mb)

- Position on chr6 (Mb)

- Position on chr6 (Mb)

- Position on chr6 (Mb)

- Position on chr6 (Mb)

- Position on chr6 (Mb)

- Position on chr6 (Mb)
2 GWAS hits

Position on chr8 (Mb)

rs1913641

Breast cancer

r^2

0.8

0.6

0.4

0.2

rs2976178

Adverse response to lemotigine and phenyl

REKO 1.2P

PSM4H2

SLCTA13

WBP4

CPNE4

rs79346194

Interstitial lung disease

Blood pressure

Height

rs2954017

Cholesterol total

HDL cholesterol

LDL cholesterol

Position on chr8 (Mb)
Serum dimethylarginine levels

- NFATC3
- CRYAB
- CRYAB
- LOC10391397
- LOC64902
- AKTIP
- CHD9
- RBL2
- RPRMPL

Position on chr16 (Mb)

- 52.8
- 53
- 53.2
- 53.4
- 53.6

Tetraploidy of Patiss

- IgG glycosylation
- Fibrogen
- Wilgp (non-seg)

Position on chr16 (Mb)

- 68
- 68.2
- 68.4
- 68.6
- 68.8

m silenced 338

- chr16 (Mb)
- Position on chr16 (Mb)
- Position on chr16 (Mb)