THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

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

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Nature Genetics

OPEN ACCESS

1 **A catalogue of genetic targets for kidney function from analyses of a**

2 **million individuals**

3

4 **Authors for Correspondence**:

5

6 Anna Köttgen, MD MPH

7 Institute of Genetic Epidemiology

8 Medical Center - University of Freiburg

9 Hugstetter Str. 49, 79106 Freiburg, Germany

10 +49 (0)761 270-78050

11 anna.koettgen@uniklinik-freiburg.de

12

13 Cristian Pattaro PhD

14 Eurac Research

15 Institute for Biomedicine

16 Via Galvani 31, 39100 Bolzano, Italy

17 +39 0471 055 527

18 cristian.pattaro@eurac.edu

1 **Abstract**

2 Chronic kidney disease (CKD) is a public health burden with multi-systemic

3 complications. Through trans-ethnic meta-analysis of genome-wide association studies

4 of estimated glomerular filtration rate (eGFR) and independent replication (n=1,046,070),

5 we identified 264 associated loci (166 novel). Of these, 147 were likely to be relevant for

6 kidney function based on associations with the alternative kidney function marker blood

7 urea nitrogen (n=416,178). Pathway and enrichment analyses, including mouse models

8 with renal phenotypes, support the kidney as the main target organ. A genetic risk score

9 for lower eGFR was associated with clinically diagnosed CKD in 452,264 independent

10 individuals. Co-localization analyses of associations with eGFR among 783,978

11 European-ancestry individuals and gene expression across 46 human tissues, including

12 tubulo-interstitial and glomerular kidney compartments, identified 17 genes differentially

13 expressed in kidney. Fine-mapping highlighted missense driver variants in 11 genes and

14 kidney-specific regulatory variants. These results provide a comprehensive priority list of

15 molecular targets for translational research.

Chronic kidney disease (CKD) is a major public health issue, with increasing incidence and prevalence worldwide.[1] Its associated burden of disease encompasses metabolic disturbances, end-stage kidney disease, and multi-systemic complications such as cardiovascular disease.[1-4] CKD is a leading cause of death[5] and shows one of the highest increases in disease-attributable mortality over the last decade.[2] Nevertheless, public and clinical awareness remains low.[3] Moreover, clinical trials in nephrology are still underrepresented,[6] resulting in scarce therapeutic options to alter disease progression and high costs for health systems.[7] A major barrier to developing new therapeutics is the limited understanding of the mechanisms underlying kidney function in health and disease, with 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 one hundred eGFR-associated genetic loci[8] in samples of European[9-15], Asian[16-19], and multiple[20] ancestries. However, similar to other complex traits and diseases, identifying causal genes and molecular mechanisms implicated by genetic associations is challenging and has only been successful for few kidney function-associated loci.[21,22] Advanced statistical fine-mapping approaches and newly emerging multi-tissue gene expression data provide new opportunities for prioritizing putative causal variants, effector genes, and target tissues from results of large-scale GWAS meta-analyses.

We therefore conducted a trans-ethnic GWAS meta-analysis in the CKD Genetics Consortium (CKDGen, n=765,348), and replicated findings in the Million Veteran Program (MVP, n=280,722),[23] for a combined sample size of >1 million participants. The first aim of this study was to identify novel, globally important loci for kidney function through maximizing statistical power (**Supplementary Figure 1**). Results from GWAS of the complementary kidney function marker blood urea nitrogen (BUN, n=416,178) were used to prioritize eGFR-associated loci 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 co-localization with gene

expression in 46 tissues and protein levels (**Supplementary Figure 1**). We focused this aim on European ancestry (EA) individuals, as fine-mapping based on summary statistics requires linkage disequilibrium (LD) reference panels whose sample size should scale with that of the GWAS.[24] 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-ethnic 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) ancestries (median age: 54 years; 50% females; **Supplementary Table 1**). The median of the study-specific mean eGFR was 89 ml/min/1.73m² (interquartile range, IQR: 81, 94). GWAS were based on genotypes imputed from the Haplotype Reference Consortium[25] or 1000 Genomes Project[26] reference panels (Methods, **Supplementary Table 2**). Following study-specific variant filtering and quality control (QC) procedures, we performed a fixed-effects inverse-variance weighted meta-analysis, showing no evidence of unmodeled population structure (LD Score regression intercept =1.04; λ_{GC}=1.05). After variant filtering, 8,221,591 single nucleotide polymorphisms (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 novel and 108 contained an index SNP reported by previous eGFR GWAS (**Figure 1**; **Supplementary Table 3**). Regional association plots are shown in **Supplementary Figure 2**. The minor alleles across index SNPs showed both decreasing and increasing effects on eGFR, with larger effects observed for lower frequency SNPs (**Figure 1**, inset). The 308 index SNPs explained 7.1% of the eGFR variance, nearly doubling recent GWAS-based estimates,[9] and 19.6% of eGFR genetic heritability (h^2=39%, 95% credible interval: 32%, 47%), estimated in a participating general-population-based pedigree-study (**Supplementary Figure 3**; Methods). Index SNPs' effects were largely homogeneous across studies (**Figure 2A**;

Supplementary Table 3) and ancestries (**Supplementary Table 4**; **Supplementary Material**).

Replication and meta-analysis of >1 million individuals

We assessed replication in an independent trans-ethnic GWAS meta-analysis of eGFR performed among 280,722 MVP participants.[23] Effect estimates, available for 305/308 SNPs, showed almost perfect direction consistency (302/305 SNPs, 99%) and very strong correlation with the discovery results (**Figure 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 (**Figure 1**; **Supplementary Table 3**; Methods). Of the three SNPs not available in MVP, the index SNPs at *SHROOM3* (P=3.5×10^{-120}) and *SH3YL1* (P=1.2×10^{-11}) were also considered replicated based on prior evidence[15,27], resulting in a total of 264 replicated SNPs (166 novel). Of these, 74 SNPs were genome-wide significant in MVP alone (**Supplementary Table 3**).

Association of eGFR-associated loci with BUN and CKD

To evaluate whether associations with creatinine-based eGFR were likely 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-ethnic meta-analysis of 65 BUN GWAS (n=416,178; **Supplementary Table 1**) showed no evidence of unmodeled population structure (λ_{GC}=1.03; LD Score regression intercept=0.98) and yielded 111 genome-wide significant loci (15 known, 96 novel, **Supplementary Figure 4**, **Supplementary Table 5**).

Of the 264 replicated eGFR index SNPs, 34 and 146 showed genome-wide and nominally (P<0.05) significant association with BUN, respectively (**Supplementary Table 6**). SNP effects were inversely correlated (r=-0.65; **Figure 2C**). Their 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 SNPs not associated

with BUN ($P \geq 0.05$); and *unlikely* for 15 eGFR index SNPs showing concordant, significant association with BUN (one-sided $P<0.05$; **Supplementary Table 6**). This comparative analysis of complementary biomarkers supports that signals at the majority of eGFR-associated loci likely reflect kidney function.

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

Lastly, we tested if 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 a genetically lower eGFR, was associated with higher odds ratios (ORs) of chronic renal failure, glomerular diseases, acute renal failure, and hypertensive diseases (**Figure 2D; Supplementary Figure 5**). The OR of chronic renal failure per 10% lower GRS-predicted eGFR was 2.13 (95%CI: 1.90, 2.39; $P=8.1 \times 10^{-38}$). A significant protective association with urolithiasis may reflect a reduced ability to concentrate urine at lower eGFR.

Genetic correlations of eGFR and BUN with other phenotypes

We assessed genome-wide genetic correlations (r_g) of eGFR associations with each of 748 complex traits and diseases (Methods).[28] We observed 37 significant correlations ($P<6.7 \times 10^{-5}=0.05/748$, **Supplementary Figure 6**; **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 (e.g., $r_g=-0.20$ with left hand grip strength). While the inverse correlation with muscle mass-related traits likely reflects higher creatinine generation leading to lower creatinine-based eGFR, the correlations with citrate and urate levels

1   likely reflect reduced filtration function, as does the positive correlation with GFR

2   estimated from cystatin C (r_g=0.53).

3   A very similar pattern of genetic correlations was observed for BUN

4   (**Supplementary Table 7**), but the genetic correlations with muscle mass-related traits

5   were generally lower than those for eGFR. The largest genetic correlations of BUN was

6   observed with CKD (r_g=0.47), as compared to creatinine-based (r_g=-0.29) and cystatin-

7   based eGFR (r_g=-0.26).

8   In summary, significant genetic correlations with eGFR reflect the two biological

9   components that govern serum creatinine concentrations: its excretion via the kidney and

10   its generation in muscle. The fact that genetic correlations between BUN and muscle-

11   mass related traits are generally lower than those observed for eGFR underscores the

12   value of genetic associations with BUN to help prioritize eGFR-associated loci most likely

13   to be relevant for kidney function.

14

15   *Functional enrichment and pathway analyses*

16   To identify molecular mechanisms and tissues of importance for kidney function, we

17   assessed the enrichment of the eGFR and BUN genetic associations using tissue-specific

18   gene expression, regulatory annotations, and gene sets and pathways (Methods). First,

19   we used eGFR-associated SNPs (P<5×10^{-8}) to explore enriched pathways, tissues and

20   cell types based on gene expression data using DEPICT.[29] We identified 16 significantly

21   enriched physiological systems, cell types and tissues highlighting several aspects of

22   kidney function, physiology and disease. The strongest enrichment was observed for

23   urogenital and renal physiological systems and tissues (kidney, kidney cortex, and urinary

24   tract; false discovery rate (FDR) <0.05; **Supplementary Figure 7A** and **B**). Pathway and

25   gene set enrichment analysis identified three highly correlated and strongly associated

26   meta gene-sets (P<1×10^{-6}, FDR<0.05), including some relevant to kidney such as

27   polyuria, dilated renal tubules, and expanded mesangial matrix, as well as signaling and

28   transcription, and energy metabolism (**Supplementary Figure 7C**). Tissue and cell-type

29   enrichment analysis of BUN-associated SNPs with P<5×10^{-8} highlighted a very similar

pattern (**Supplementary Figure 8**) but without enrichment for muscle tissues, further supporting the use of BUN to prioritize loci most likely to be related to kidney function.

Second, we used stratified LD Score regression[30] on the genome-wide eGFR and BUN summary statistics to identify cell-type groups with enriched heritability based on data from diverse, cell-type specific functional genomic elements. The strongest enrichment for eGFR was observed for kidney (13.2-fold), followed by liver (7.3-fold) and adrenal/pancreas (5.7-fold enrichment; **Supplementary Table 8**). Kidney was also the most enriched cell-type group for BUN (11.5-fold enrichment; **Supplementary Table 8**).

Lastly, using a complementary approach, we assessed enrichment of eGFR-associated variants in genes resulting in kidney phenotypes in genetically manipulated mice.[31] From the Mouse Genome Informatics database, we selected all genes causing 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 10 genes causing abnormal GFR in mice (enrichment p-value=8.9×10^{-4}), 55 causing abnormal kidney physiology (enrichment p-value=1.1×10^{-4}) and 96 causing abnormal kidney morphology (enrichment p-value=1.8×10^{-5}; **Figure 3**; Methods). Of these, 25 genes represent novel eGFR candidate genes in humans, i.e. they were not previously 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 second signal analysis in EA individuals

Conditional and fine-mapping analyses were restricted to EA participants, for whom data to construct a large enough LD reference panel was publicly available (Methods). The meta-analysis of 85 EA CKDGen GWAS identified 256 genome-wide significant loci (**Supplementary Table 10**). Replication among 216,518 EA MVP participants confirmed 228 SNPs, including 227 index SNPs that met replication criteria and the *SHROOM3* index SNP (Methods, **Supplementary Table 10**). Of these 228 SNPs, 221 mapped into one of the 264 replicated loci from the trans-ethnic analysis (≤±500 kb from the trans-

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

To perform 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 into 189 regions (Methods). For each independent variant, we computed a 99% credible set,[32] with median set size of 26 SNPs (IQR: 6, 60). We observed 58 *small credible sets* (≤5 SNPs), including 20 single-SNP sets: *EDEM3*, *CACNA1S*, *HOXD11*, *CPS1*, *DAB2*, *SLC34A1*, *LINC01512*, *LARP4B*, *DCDC1*, *SLC25A45*, *SLC6A13*, *GATM*, *CGNL1*, *CYP1A1*, *NRG4*, *RPL3L*, *UMOD-PDILT, SLC47A1,* and two independent sets at *BCL2L14* (**Supplementary Table 11; Figure 4**). 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**).

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 into a small credible set are of particular interest because they directly implicate the affected gene. Such missense SNPs were identified in 11 genes (*SLC47A1*, *RPL3L*, *SLC25A45*, *CACNA1S*, *EDEM3*, *CPS1*, *KLHDC7A*, *PPM1J*, *CERS2*, *C9,* and *SLC22A2;* **Supplementary Table 12**), of which *CACNA1S*, *RPL3L*, *CERS2*, and *C9* were likely relevant for kidney function (**Figure 4A**). The majority of the 11 variants had CADD score>15, indicating potential deleteriousness.[33] 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 MATE1 knockout mice have higher blood levels of both creatinine and BUN[34] argues against a sole effect on creatinine transport.

To evaluate the regulatory potential of small credible set SNPs in kidney, we annotated them to open chromatin regions identified from primary human tubular and glomerular cell cultures,[35] as well as from publicly available kidney cells types (Methods). We identified 72 SNPs mapping into one of these annotations, which may thus represent causal regulatory variants (**Supplementary Table 12**). A particularly interesting finding was the intronic rs77924615 in *PDILT*, which showed PP>99% of driving the association at the *UMOD* locus, and mapped into open chromatin in all evaluated resources (native kidney cells, ENCODE and Roadmap kidney cell types; **Figure 4B**).

Gene prioritization: co-localization with gene expression

We performed co-localization analyses for each eGFR-associated locus with gene expression in *cis* across 46 tissues including kidney glomerular and tubulo-interstitial compartments (Methods). A PP>80% of co-localization in at least one kidney tissue was observed for 17 transcripts mapping into 16 of the 228 replicated loci (**Figure 5**), pointing towards a shared underlying SNP associated with both eGFR and gene expression, and implicating the gene encoding the co-localized transcript as the locus' effector gene(s).

Novel 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 (**Figure 5**), consistent with previous GWAS of urinary uromodulin concentration, in which alleles associated with lower eGFR at *UMOD*[15] were associated with higher urinary uromodulin concentrations.[36] The lead SNP at this locus was rs77924615, highlighted above as the candidate causal regulatory variant mapping into the 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.

Second, novel, biologically plausible candidates emerged. For example, our results suggest *KNG1* and *FGF5* as effector genes in the respective eGFR-associated loci (**Figure 5**, **Supplementary Table 13**). *KNG1* encodes for high-molecular weight kininogen, which is cleaved to bradykinin. Bradykinin influences blood pressure,

natriuresis and diuresis, and can be linked to kidney function via the renin-angiotensin-aldosterone system.[37] *FGF5* encodes for Fibroblast Growth Factor 5, and the index SNPs for eGFR or highly correlated SNPs (r^2>0.9) have been identified in multiple GWAS of blood pressure, atrial fibrillation, coronary artery disease, hematocrit and multiple kidney-function related traits (**Supplementary Table 13**). The eGFR index SNP rs1458038 (PP>50%, CADD score=14.8; **Supplementary Table 13**), co-localized with the eGFR signal only in tubulo-interstitial kidney portions (**Figure 5**), supporting its regulatory potential on the expression levels of *FGF5* in this compartment. Both *KNG1* and *FGF5* index SNPs were associated with BUN and CKD, and are thus likely related to kidney function.

Third, co-localization of eGFR with gene expression across multiple tissues revealed that for kidney-co-localized transcripts, some showed the same direction on transcript levels across all tissues with lower eGFR (e.g. *METTL10*), while others showed higher transcript levels in some tissues but lower levels in others (e.g. *SH3YL1*; **Figure 5**). These observations were also reflected broadly across all transcripts with evidence of co-localization in any tissue (**Supplementary Figure 9**), and highlight tissue-shared and tissue-specific signals.[38,39]

Finally, *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 Material**; **Supplementary Tables 14** and **15**).

Co-localization with uromodulin protein levels in urine

The *UMOD* locus is of particular clinical interest for CKD research:[21] rare *UMOD* mutations cause autosomal-dominant tubulo-interstitial kidney disease[40], and common variants at *UMOD* give rise to the strongest eGFR and CKD GWAS signals.[15] We therefore performed conditional analyses based on the EA-specific summary statistics and found two independent variants: rs77924615, mapping into upstream *PDILT*, and rs34882080 mapping into an intron of *UMOD* (**Figure 6A**). SNP association with the urinary uromodulin-to-creatinine ratio (UUCR) in one participating cohort (**Figure 6B**) matched the eGFR association pattern. Co-localization of the conditional eGFR and

UUCR associations was evaluated separately for the rs34882080 (**Figure 6C**) and rs77924615 (**Figure 6D**). Both regions showed high probability of a shared underlying variant driving the respective associations with eGFR and UUCR levels (PP=0.97 and 0.96, 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 EA analysis is shown in **Supplementary Table 16**, to facilitate selection of the most promising candidates for further experimental studies.

Discussion

This trans-ethnic study is 5-fold larger than previous eGFR GWAS meta-analyses and identified 264 replicated loci, 166 of which are reported here for the first time. By also analyzing BUN, an established complementary kidney function marker, 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. Co-localization 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 EA participants implicate single potentially causal variants at 20 loci.

Most previous eGFR GWAS meta-analyses were limited to a single ancestry group[8] and did not prioritize causal variants or effector genes in associated loci. While underpowered to uncover novel loci, one previous trans-ethnic study employed fine-mapping, resolving one signal to a single variant,[20] rs77924615 at *UMOD-PDILT*, 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 locus to single potentially causal variant with its effector gene,

protein and target tissue, represents a critical advance over 10 years of eGFR GWAS,[15] 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 GTEx Project and not included in its tissue-specific eQTL datasets,[38] which emphasizes 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 micro-dissected human kidney biopsies,[41] kidney-specific regulatory information from the ENCODE and Roadmap resources, and by obtaining regulatory information from primary cultures of human glomerular and tubulo-interstitial cells.[35]

Functional follow-up studies of potentially causal variants will 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- or ≤5-SNP resolution, some of which may have broader clinical relevance. For example, the OCT2 protein encoded by *SLC22A2* transports several cationic drugs such as metoprolol, cisplatin, metformin and cimetidine across the basolateral membrane of renal tubular cells.[42] The prioritized missense SNP encodes p.(Ser270Ala), a known pharmaco-genomic variant that alters the transport of these drugs and their side-effects, such as cisplatin-induced nephrotoxicity.[43] 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[44] and hence alter CKD risk.

Strengths of this project include the large sample size with dense genotype imputation, standardized and automated phenotype generation and QC, independent replication, as well as 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 a

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 underrepresented in our study, like many other genomic efforts.[45] Statistical fine-mapping using trans-ethnic data with different LD structures can potentially narrow down association signals. However, a sufficiently large reference dataset to compute ancestry-matched LD structure for summary-statistics based fine-mapping was only available for EA, highlighting the potential of future large-scale efforts with trans-ethnic fine-mapping and the need to generate data from non-EA populations enabling such endeavors. Lastly, 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-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 their synthesis, active secretion, or reabsorption, as illustrated by loci detected in our study: for example, the *GATM* locus was associated with eGFR but not with BUN, consistent with the function of the encoded protein as a rate-limiting enzyme in creatine synthesis.[46] Conversely, the *SLC14A2* locus was associated with BUN but not with eGFR, consistent with the function of the encoded protein as a urea transporter.[47] Even so, lack of a SNP's association 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 based on effect direction and significance of the BUN associations 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 novel 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 co-localization 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 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.

References

1. 1. Eckardt, K.U. *et al.* Evolving importance of kidney disease: from subspecialty to global health burden. *Lancet* **382**, 158-69 (2013).
2. 2. Jha, V. *et al.* Chronic kidney disease: global dimension and perspectives. *Lancet* **382**, 260-72 (2013).
3. 3. Ene-Iordache, B. *et al.* Chronic kidney disease and cardiovascular risk in six regions of the world (ISN-KDDC): a cross-sectional study. *Lancet Glob Health* **4**, e307-19 (2016).
4. 4. Go, A.S., Chertow, G.M., Fan, D., McCulloch, C.E. & Hsu, C.Y. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* **351**, 1296-305 (2004).
5. 5. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1151-1210 (2017).
6. 6. Inrig, J.K. *et al.* The landscape of clinical trials in nephrology: a systematic review of Clinicaltrials.gov. *Am J Kidney Dis* **63**, 771-80 (2014).
7. 7. Levin, A. *et al.* Global kidney health 2017 and beyond: a roadmap for closing gaps in care, research, and policy. *Lancet* **390**, 1888-1917 (2017).
8. 8. Wuttke, M. & Kottgen, A. Insights into kidney diseases from genome-wide association studies. *Nat Rev Nephrol* **12**, 549-62 (2016).
9. 9. Gorski, M. *et al.* 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. *Sci Rep* **7**, 45040 (2017).
10. 10. Pattaro, C. *et al.* Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun* **7**, 10023 (2016).
11. 11. Chasman, D.I. *et al.* Integration of genome-wide association studies with biological knowledge identifies six novel genes related to kidney function. *Hum Mol Genet* **21**, 5329-43 (2012).
12. 12. Pattaro, C. *et al.* Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet* **8**, e1002584 (2012).
13. 13. Kottgen, A. *et al.* New loci associated with kidney function and chronic kidney disease. *Nat Genet* **42**, 376-84 (2010).
14. 14. Chambers, J.C. *et al.* Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet* **42**, 373-5 (2010).
15. 15. Kottgen, A. *et al.* Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet* **41**, 712-7 (2009).
16. 16. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* **50**, 390-400 (2018).
17. 17. Okada, Y. *et al.* Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet* **44**, 904-9 (2012).
18. 18. Hishida, A. *et al.* Genome-Wide Association Study of Renal Function Traits: Results from the Japan Multi-Institutional Collaborative Cohort Study. *Am J Nephrol* **47**, 304-316 (2018).
19. 19. Lee, J. *et al.* Genome-wide association analysis identifies multiple loci associated with kidney disease-related traits in Korean populations. *PLoS One* **13**, e0194044 (2018).
20. 20. Mahajan, A. *et al.* Trans-ethnic Fine Mapping Highlights Kidney-Function Genes Linked to Salt Sensitivity. *Am J Hum Genet* **99**, 636-646 (2016).
21. 21. Devuyst, O. & Pattaro, C. The UMOD Locus: Insights into the Pathogenesis and Prognosis of Kidney Disease. *J Am Soc Nephrol* **29**, 713-726 (2018).
22. 22. Yeo, N.C. *et al.* Shroom3 contributes to the maintenance of the glomerular filtration barrier integrity. *Genome Res* **25**, 57-65 (2015).
23. 23. Gaziano, J.M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* **70**, 214-23 (2016).

24. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am J Hum Genet* **101**, 539-551 (2017).

25. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).

26. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).

27. Li, M. *et al.* SOS2 and ACP1 Loci Identified through Large-Scale Exome Chip Analysis Regulate Kidney Development and Function. *J Am Soc Nephrol* **28**, 981-994 (2017).

28. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).

29. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).

30. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).

31. Jing, J. *et al.* Combination of mouse models and genomewide association studies highlights novel genes associated with human kidney function. *Kidney Int* **90**, 764-73 (2016).

32. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* **81**, 208-27 (2007).

33. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* **24**, 2125-37 (2015).

34. Tsuda, M. *et al.* Targeted disruption of the multidrug and toxin extrusion 1 (mate1) gene in mice reduces renal secretion of metformin. *Mol Pharmacol* **75**, 1280-6 (2009).

35. Sieber, K.B. *et al.* Integrated Functional Genomic Analysis Enables Annotation of Kidney Genome-Wide Association Study Loci. *J Am Soc Nephrol* (2019).

36. Olden, M. *et al.* Common variants in UMOD associate with urinary uromodulin levels: a meta-analysis. *J Am Soc Nephrol* **25**, 1869-82 (2014).

37. Moreau, M.E. *et al.* The kallikrein-kinin system: current and future pharmacological targets. *J Pharmacol Sci* **99**, 6-38 (2005).

38. Battle, A., Brown, C.D., Engelhardt, B.E. & Montgomery, S.B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).

39. Gamazon, E.R. *et al.* Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet* **50**, 956-967 (2018).

40. Eckardt, K.U. *et al.* Autosomal dominant tubulointerstitial kidney disease: diagnosis, classification, and management--A KDIGO consensus report. *Kidney Int* **88**, 676-83 (2015).

41. Gillies, C.E. *et al.* An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *Am J Hum Genet* **103**, 232-244 (2018).

42. Dudley, A.J., Bleasby, K. & Brown, C.D. The organic cation transporter OCT2 mediates the uptake of beta-adrenoceptor antagonists across the apical membrane of renal LLC-PK(1) cell monolayers. *Br J Pharmacol* **131**, 71-9 (2000).

43. Filipski, K.K., Mathijssen, R.H., Mikkelsen, T.S., Schinkel, A.H. & Sparreboom, A. Contribution of organic cation transporter 2 (OCT2) to cisplatin-induced nephrotoxicity. *Clin Pharmacol Ther* **86**, 396-402 (2009).

44. Motohashi, H. & Inui, K. Organic cation transporter OCTs (SLC22) and MATEs (SLC47) in the human kidney. *AAPS J* **15**, 581-8 (2013).

45. Popejoy, A.B. & Fullerton, S.M. Genomics is failing on diversity. *Nature* **538**, 161-164 (2016).

46. Humm, A., Huber, R. & Mann, K. The amino acid sequences of human and pig L-arginine:glycine amidinotransferase. *FEBS Lett* **339**, 101-7 (1994).

1   47.   Olives, B. *et al.* Molecular characterization of a new urea transporter in the human kidney. *FEBS*
2   *Lett* **386**, 156-60 (1996).
3
4

Online Methods

Overview

We set up a collaborative meta-analysis based on a distributive data model and QC procedures. To maximize phenotype standardization across studies, an analysis plan and a command line script (https://github.com/genepi-freiburg/ckdgen-pheno) were created centrally and provided to all participating studies (mostly population-based, **Supplementary Table 1**). Data processing, analysis and troubleshooting instructions were distributed to all studies via a Wiki system (https://ckdgen.eurac.edu/mediawiki/index.php/ CKDGen_Round_4_EPACTS_analysis_plan). Automatically generated summary files were checked centrally. Upon phenotype approval, studies run their GWAS and uploaded results and imputation quality (IQ) information to a common calculation server. GWAS QC was performed using GWAtoolbox[48] 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 as described in **Supplementary Table 1**. Creatinine values obtained with a Jaffé assay before 2009 were calibrated by multiplying by 0.95.[49] Studies on >18 year-old adults estimated GFR with the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation,[50] using the R package 'nephro'.[51] Studies on ≤18 year-old individuals used the Schwartz formula.[52] eGFR was winsorized at 15 and 200 ml/min/1.73m^2. CKD was defined as eGFR<60 ml/min/1.73m^2. In studies reporting blood urea measurements, BUN was derived as blood urea×2.8, with units expressed as mg/dl.

Genotyping and genotype imputation

Genotypes imputed based on the Haplotype Reference Consortium (HRC) v1.1 or the 1000 Genomes Project phase 3 v5 (1000Gp3v5) ALL or phase 1 v3 (1000Gp1v3) ALL panels. 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 (GWAS)

Each study fitted sex- and age-adjusted linear regression models to log(eGFR) and BUN. Regression residuals were regressed on SNP dosage levels, assuming an additive genetic model. Study site, genetic principal components (PCs), relatedness, or 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-ethnic GWAS meta-analysis

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). 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 applied in case of GC factor λ_{GC}>1. Fixed effects inverse-variance weighted meta-analysis was performed using METAL,[53] which was adapted to increase the precision of effect estimates and their standard errors (SE; seven decimal places instead of four).

After meta-analysis of 43,994,957 SNPs, only SNPs present in ≥50% of the GWAS files and with total MAC≥400 were retained. Across 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. Post-meta-analysis GC correction was not applied (LD Score regression intercept≈1 in all analyses of eGFR, BUN, and CKD).[54] The genome-wide significance level was set at $5×10^{-8}$. Between-study heterogeneity was assessed using the I^2 statistic.[55] For CKD, variants with I^2≥95% were removed to moderate influence of single large studies. Variants were assigned to loci by selecting the SNP with the lowest p-value genome-wide 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 MHC region was considered as a single locus. A locus was considered novel if not containing any variant identified by previous GWAS of eGFR.

Meta-regression analysis of trans-ethnic GWAS

For eGFR, we evaluated ancestry-related heterogeneity using the software Meta-Regression of Multi-Ethnic Genetic Association (MR-MEGA v0.1.2)[56] using study-specific GWAS results. Meta-regression models included three axes of genetic variation. GC correction was applied to the meta-regression results. The 308 genome-wide significant index SNPs from the trans-ethnic GWAS meta-analysis were tested for ancestry-related heterogeneity of the allelic effects at a significance level of $0.05/308=1.6×10^{-4}$ (referring to the corresponding p-value as p-anc-het).

Variance explained and genetic heritability

The proportion of phenotypic variance explained by the index SNPs was estimated as $\beta^2 \left(\frac{2p(1-p)}{var} \right)$, with β being the SNP effect, p the effect allele frequency, and *var* the variance of the sex- and age-adjusted log(eGFR) residuals (assumed as 0.016 based on data from 11,827 EA participants of the population-based ARIC study).[9] Genetic heritability of age- and sex-adjusted log(eGFR) was estimated using the R package 'MCMCglmm'[57] on the Cooperative Health Research In South Tyrol (CHRIS) study,[58] a participating pedigree-based study with 186 up-to-5 generation pedigrees (n=4373).[59] We

1 fitted two models, with and without the inclusion of the identified index SNPs (304/308),

2 running 1,000,000 MCMC iterations (*burn in*=500,000).[59]

3

4 *Comparison with and replication of results in the MVP*

5 The eGFR-associated SNPs identified in the discovery GWAS meta-analyses, were

6 tested for replication in a GWAS from the MVP,[23] an independent trans-ethnic study with

7 participants recruited across 63 U.S. Veteran's Administration (VA) medical facilities.

8 Written informed consent was obtained and all documents and protocols were approved

9 by the VA Central Institutional Review Board. After genotyping and QC, genotypes were

10 phased and imputed on the 1000Gp3v5 reference panel. Serum creatinine was assessed

11 up to one year prior to MVP enrollment using isotope dilution mass spectrometry. GFR

12 was estimated using the CKD-EPI equation[50] after excluding subjects on dialysis,

13 transplant patients, amputees, individuals on HIV medications, and those with creatinine

14 values of <0.4 mg/dl. GWAS of eGFR on SNP dosage levels were performed by fitting

15 linear regression models adjusted for age at creatinine measurement, age^2, sex, body

16 mass index, and the first 10 genetic PCs, using SNPTEST v2.5.4-beta.[60] All GWAS were

17 stratified by self-reported ethnicity (79.6% White non-Hispanic; 20.4% Black non-

18 Hispanic), diabetes, and hypertension status. Results were combined across strata using

19 fixed effects inverse-variance weighted meta-analysis in METAL.[53] This analysis

20 encompassed a total of 280,722 individuals across all strata, of whom 216,518 were non-

21 Hispanic Whites (EA). The MVP is described more extensively in the **Supplementary**

22 **Material**.

23   Of the 308 eGFR index SNPs identified in the CKDGen trans-ethnic analysis, 305

24 variants or their good proxies were available in the MVP GWAS (proxies had to have

25 r^2≥0.8 with the index SNP and were selected by maximum r^2 followed by minimum

26 distance in case of ties). Replication testing of the 256 EA-specific index SNPs was

27 restricted to the MVP EA GWAS. CKDGen and MVP meta-analysis results were pooled

28 via sample size weighted meta-analysis of z-scores using METAL.[53] In both the trans-

29 ethnic and EA-specific analyses, replication was defined as a one-sided p-value<0.05 in

30 the MVP and genome-wide significance of the CKDGen and MVP meta-analysis result.

Assessment of kidney function relevance using BUN

We used genetic associations with BUN to assess replicated eGFR-associated SNPs with respect to their potential kidney function relevance. Support for kidney function relevance was categorized as "likely" (1) for all eGFR index SNPs with an inverse, significant (one-sided $P<0.05$) association with BUN for a given reference allele, "inconclusive" (2) for eGFR index SNPs whose effect on BUN was not different from 0 ($P≥0.05$), and "unlikely" (3) 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 (GRS) analysis in the UK Biobank dataset

To test the combined effect of eGFR-associated SNPs on clinically diagnosed CKD related outcomes, a GRS-based association analysis was conducted based on summary GWAS results, as described before.[61,62] The genetic association results with the diseases were obtained for 452,264 UK Biobank participants available in the GeneAtlas[63] database for glomerular diseases (ICD-10 codes N00-N08; 2289 cases); acute renal failure (N17; 4913 cases); chronic renal failure (N18; 4905 cases); urolithiasis (N20-N23; 7053 cases); hypertensive diseases (I10-I15; 84,910 cases); and ischemic heart diseases (I20-I25; 33,387 cases). Asthma (J45; 28,628 cases) was included as a negative control. The log(estimated OR) provided by the GeneAtlas phewas interface was used as effect size, and its SE was calculated from the corresponding effect size and p-value. When OR=1, the SE was imputed by the median value of the remaining associations of the trait. Of the 147 eGFR index SNPs from the trans-ethnic GWAS meta-analysis that were replicated and showed likely kidney function relevance, 144 were available in the UK Biobank dataset, and 259 out of all 264 replicated trans-ethnic GWAS meta-analysis SNPs. The effect *beta* of the GRS association corresponds to the OR of the disease depending on the relative change in eGFR, e.g. OR=1.10^*beta* for a 10% change in eGFR. Alternatively, exp(*beta*) can be interpreted as the OR of the disease per unit change of log(eGFR).

1   *Genetic correlations with other complex traits and diseases*

2   Genome-wide genetic correlation analysis was performed to investigate evidence of co-

3   regulation or shared genetic bases between eGFR and BUN and other complex traits and

4   diseases, both known and not known to correlate with eGFR and BUN. We estimated

5   pairwise genetic correlation coefficients (r_g) between the results of our trans-ethnic meta-

6   analyses of eGFR and BUN and each of 748 pre-computed and publicly available GWAS

7   summary statistics of complex traits and diseases available through LD Hub v1.9.0 using

8   LD Score regression.[28] An overview of the sources of these summary statistics and their

9   corresponding sample sizes is available at http://ldsc.broadinstitute.org. Statistical

10   significance was assessed at the Bonferroni corrected level of $0.05/748=6.7×10^{-5}$.

11

12   *Pathway and tissue enrichment analysis*

13   We used DEPICT v1 release 194 to perform Data-Driven Expression Prioritized

14   Integration for Complex Traits analysis,[29] including pathway/gene-set enrichment and

15   tissue/cell type analyses as described previously.[9,10] All 14,461 gene sets were

16   reconstituted by identifying genes that were transcriptionally co-regulated with other

17   genes in a panel of 77,840 gene expression microarrays,[64] from mouse knock-out studies,

18   and molecular pathways from protein-protein interaction screening. In the tissues and cell

19   type enrichment analysis, we tested whether genes in associated regions were highly

20   expressed in 209 MeSH annotation categories for 37,427 microarrays (Affymetrix U133

21   Plus 2.0 Array platform). For both eGFR and BUN, we included all variants associated

22   with the trait at $P<5×10^{-8}$ in the trans-ethnic meta-analysis. Independent variant clumping

23   was performed using Plink 1.9[65] with 500 kb flanking regions and $r^2>0.01$ in the

24   1000Gp1v3 dataset. After excluding the MHC region, DEPICT was run with 500

25   repetitions to estimate the FDR and 5000 permutations to compute p-values adjusted for

26   gene length by using 500 null GWAS. All significant gene sets were merged into meta

27   gene sets by running an affinity propagation algorithm[66] implemented in the Python 'scikit-

28   learn' package (http://scikit-learn.org/). The resulting network was visualized using

29   Cytoscape (http://cytoscape.org/).

30   *Enrichment of heritability by cell type group*

1   We used stratified LD Score regression to investigate important tissues and cell types

2   based on the trans-ethnic eGFR and BUN meta-analysis results. Heritability enrichment

3   in 10 cell type groups was assessed using the default options of stratified LD Score

4   regression described previously.[30] The 10 cell type groups were collapsed from 220 cell-

5   type specific regulatory annotations for the four histone marks H3K4me1, H3K4me3,

6   H3K9ac, and H3K27ac. The enrichment of a cell type category was defined as the

7   proportion of SNP heritability in that group divided by the proportion of SNPs in the same

8   cell type group.

9

Analysis of genes causing kidney phenotypes in mice

11   A nested candidate gene analysis was performed using GenToS[67] to identify additional

12   genetic associations that were not genome-wide significant. Candidate genes that when

13   manipulated cause kidney phenotypes in mice were selected using the comprehensive

14   Mouse Genome Informatics (MGI) phenotype ontology in September 2017 (abnormal

15   renal glomerular filtration rate [MP:0002847]; abnormal kidney morphology

16   [MP:0002135]; abnormal kidney physiology [MP:0002136]). The human orthologs of

17   these genes were obtained, when available, using the Human-Mouse: Disease

18   Connection webtool (http://www.informatics.jax.org/humanDisease.html). Statistical

19   significance was defined as a Bonferroni correction of a type I error level of 0.05 for the

20   number of independent common SNPs across all genes in each of the three candidate

21   gene lists plus their flanking regions, derived from an ancestry-matched reference

22   population. The GWAS meta-analysis summary statistics for eGFR were queried for

23   significantly associated SNPs mapping into the selected candidate genes. Enrichment of

24   significant genetic associations in genes within each candidate list was computed from

25   the complementary cumulative binomial distribution.[67] GenToS was used with default

26   parameters on each of the three candidate gene lists, using the 1000 Genomes phase 3

27   release 2 ALL dataset as reference.

28

29

1 *Independent variant identification in the EA meta-analysis*

2 To identify additional, independent eGFR-associated variants within the EA-specific and

3 replicated loci, approximate conditional analyses were performed based on genome-wide

4 discovery summary statistics that incorporated LD information from an ancestry-matched

5 reference population. These analyses were restricted to participants of EA, because an

6 LD reference sample scaled to the size of our meta-analysis could only be constructed

7 from publicly available data for EA individuals,[24] for which we randomly selected 15,000

8 UK Biobank participants (dataset ID 8974). Individuals who withdrew consent and those

9 not meeting data cleaning requirements were excluded, keeping only those who passed

10 sex-consistency check, had ≥95% call rate, and did not represent outliers with respect to

11 SNP heterozygosity. For each pair of individuals, the proportion of variants shared

12 identical-by-descent (IBD) was computed using PLINK.[68] Only one member of each pair

13 with IBD coefficient ≥0.1875 was retained. Individuals were restricted to those of EA by

14 excluding outliers along the first two PCs from a PC analysis seeded with the HapMap

15 phase 3 release 2 populations as reference. The final dataset to estimate LD included

16 13,558 EA individuals and 16,969,363 SNPs.

17 The basis for statistical fine-mapping were the 228 1-Mb genome-wide significant

18 loci identified in the EA meta-analysis, clipping at chromosome borders. Overlapping loci

19 as well as pairs of loci whose respective index SNPs were correlated (r^2 >0.1 in the UKBB

20 LD dataset described above) were merged. A single SNP was chosen to represent the

21 MHC region, resulting in a final list of 189 regions prior to fine-mapping. Within each

22 region, the GCTA COJO Slct algorithm[69] was used to identify independent variants

23 employing a step-wise forward selection approach. We used the default collinearity cut-

24 off of 0.9 (sensitivity analyses showing no major influence of alternative cutoff values;

25 data not shown). We deemed an additional SNP as independently genome-wide

26 significant if the SNPs' p-value conditional on all previously identified SNPs in the same

27 region was $<5\times10^{-8}$.

Fine-mapping and credible sets in the EA meta-analysis

For each region containing multiple independent SNPs and for each independent SNP in such regions, approximate conditional analyses were conducted using the GCTA COJO-Cond algorithm to generate approximate conditional association statistics conditioned on the other independent SNPs in the region. Using the Wakefield's formula implemented in the R package 'gtx',[70] we derived approximate Bayes factors (ABF) from conditional estimates in regions with multiple independent SNPs and from the original estimates for regions with a single independent SNP. Given that 95% of the SNP effects on log(eGFR) fell within -0.01 to 0.01, the standard deviation prior was chosen as 0.0051 based on formula 8 in the original publication.[32] Sensitivity analyses showed that results were robust when higher values were used for the standard deviation prior (data not shown). For each variant within an evaluated region, the ABF obtained from the association betas and their SEs of the marginal (single signal region) or conditional estimates (multi-signal regions) was used to calculate the PP for a SNP of driving the association signal ("causal variant"). We derived 99% credible sets, representing the SNP sets containing the variant(s) driving the association signal with 99% probability, by ranking variants by their PP and adding them to the set until reaching a cumulative PP>99% in each region.

Variant annotation

Functional annotation of SNPs mapping into credible sets was performed with SNiPA v3.2 (March 2017),[71] based on the 1000Gp3v5 and Ensembl v87 datasets. SNiPA was also used to derive the Combined Annotation Dependent Depletion (CADD) PHRED-like score,[72] based on CADD v1.3. The Ensembl VEP tool[73] was used for SNP's primary effect prediction.

Co-localization of eGFR and gene expression in cis

As the great majority of gene expression datasets is generated based on EA ancestry samples, co-localization analysis was based on the genetic associations with eGFR in the EA sample and with gene expression (eQTL) quantified from micro-dissected human

glomerular and tubulo-interstitial kidney portions from 187 individuals from the NEPTUNE study,[41] as well as from the 44 tissues included in the GTEx Project v6p release.[38] 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 co-localization test was defined as the eQTL *cis* window defined in the underlying GTEx and NephQTL studies. We used the 'coloc.fast' function, using default setting, from the R package 'gtx' (https://github.com/tobyjohnson/gtx), which is an adaption of Giambartolomei's co-localization method.[74] 'gtx' 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 based on the 1000Gp3v5 European reference panel (r^2 cut-off >0.8). We limited annotation to replicated index SNPs with a fine-mapping PP≥1%. Due to expected small effect sizes, only genome-wide *trans*-eQTL studies of either peripheral blood mononuclear cells or whole blood with n≥1000 individuals were considered, resulting in five non-overlapping studies[75-79] (**Supplementary Table 14**). For one study,[79] we had access to an update with larger sample size (n=6645) obtained by combining two non-overlapping studies (LIFE-Heart[80] and LIFE-Adult[81]). To improve stringency of results, we focused the analysis on inter-chromosomal *trans*-eQTLs with $P<5\times10^{-8}$ in ≥2 studies.

Co-localization with urinary uromodulin concentrations

Association between concentrations of the urinary uromodulin-to-creatinine ratio with genetic variants at the *UMOD-PDILT* locus was evaluated in the German Chronic Kidney Disease (GCKD) study.[82] Uromodulin concentrations were measured from frozen stored urine using an established ELISA assay with excellent performance.[36] Concentrations were indexed to creatinine to account for urine dilution. Genetic associations were assessed using the same software and settings as for the eGFR association

1   (**Supplementary Table 2**). Co-localization analyses were performed using identical

2   software and settings as described above for the association with gene expression.

3

4   *Life Sciences Reporting Summary*

5   Further information on research methods are available in the Nature Research Life

6   Sciences Reporting Summary linked to this article.

References (Online Methods)

48. Fuchsberger, C., Taliun, D., Pramstaller, P.P. & Pattaro, C. GWAtoolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics* **28**, 444-5 (2012).

49. Coresh, J. *et al.* Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. *JAMA* **311**, 2518-2531 (2014).

50. Levey, A.S. *et al.* A new equation to estimate glomerular filtration rate. *Ann Intern Med* **150**, 604-12 (2009).

51. Pattaro, C. *et al.* Estimating the glomerular filtration rate in the general population using different equations: effects on classification and association. *Nephron Clin Pract* **123**, 102-11 (2013).

52. Schwartz, G.J. *et al.* Improved equations estimating GFR in children with chronic kidney disease using an immunonephelometric determination of cystatin C. *Kidney Int* **82**, 445-53 (2012).

53. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).

54. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).

55. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**, 1539-58 (2002).

56. Magi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* **26**, 3639-3650 (2017).

57. Hadfield, J. MCMC methods for multi-response generalized linear mixed models: the MCMC glmm R Package. *J Stat Softw* **33**, 1-22 (2010).

58. Pattaro, C. *et al.* The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. *J Transl Med* **13**, 348 (2015).

59. Noce, D. *et al.* Sequential recruitment of study participants may inflate genetic heritability estimates. *Hum Genet* **136**, 743-757 (2017).

60. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).

61. Kottgen, A. *et al.* Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet* **45**, 145-54 (2013).

62. Dastani, Z. *et al.* Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet* **8**, e1002607 (2012).

63. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat Genet* **50**, 1593-1599 (2018).

64. Fehrmann, R.S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* **47**, 115-25 (2015).

65. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

66. Frey, B.J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972-6 (2007).

67. Hoppmann, A.S., Schlosser, P., Backofen, R., Lausch, E. & Kottgen, A. GenToS: Use of Orthologous Gene Information to Prioritize Signals from Human GWAS. *PLoS One* **11**, e0162466 (2016).

68. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).

69. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).

70. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol* **33**, 79-86 (2009).

71. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmuller, G. SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334-6 (2015).

72. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).

73. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069-70 (2010).

74. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).

75. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).

76. Fehrmann, R.S. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet* **7**, e1002197 (2011).

77. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-1243 (2013).

78. Joehanes, R. *et al.* Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol* **18**, 16 (2017).

79. Kirsten, H. *et al.* Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding locidagger. *Hum Mol Genet* **24**, 4746-63 (2015).

80. Beutner, F. *et al.* Rationale and design of the Leipzig (LIFE) Heart Study: phenotyping and cardiovascular characteristics of patients with coronary artery disease. *PLoS One* **6**, e29070 (2011).

81. Loeffler, M. *et al.* The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. *BMC Public Health* **15**, 691 (2015).

82. Eckardt, K.U. *et al.* The German Chronic Kidney Disease (GCKD) study: design and methods. *Nephrol Dial Transplant* **27**, 1454-60 (2012).

8

9   **Data Availability**

10 Genome-wide summary statistics for this study are made publicly available through

11 dbGaP accession number phs000930.v7.p1 and at http://ckdgen.imbi.uni-freiburg.de.

Author contributions

Manuscript writing group: MWu, YL, MLi, KSi, MF, MGo, ATi, LWan, HKi, TA, KHo, IH, MSc, ATeu, AKöt, CP

Design of this study: CAB, CF, MGo, AKöt, APM, CP, ATeu, ATi, MWu

Management of an individual contributing study: TSA, EdA, SAk, SJB, GB, MBoc, MBoe, EB, MHdB, HB, ASB, CAB, AC, RJC, JCC, DIC, CYC, KC, RC, MCi, JC, DC, RMvD, JDa, OD, CMvD, KUE, GE, PE, MKE, JFF, OHF, BIF, YF, RTG, HG, PG, JMG, VGi, CG, FG, ADG, VGu, TBH, PvdH, CAH, CH, CKH, AAH, KHo, AHu, MAI, OSI, EI, VWJ, JBJ, BJ, CMK, CCK, WKi, MEK, WKo, JSK, HKr, FK, BKK, MKub, JK, MKä, AKör, AKöt, TL, YL, SCL, MLo, RJL, SL, MAL, PKM, NGM, DM, KMa, OM, AMe, EKM, YM, KLM, GWM, APM, RdM, WM, GNN, JO, MLO, AJO, MOM, WHO, AP, CP, SAP, BWP, TPe, MPe, MPi, OP, BP, PPP, MAP, BMP, TJR, OTR, DFR, RRe, MR, PMR, DJR, PR, IR, CS, VS, KUS, HSc, RS, MSc, BS, XS, HSn, NSo, CNS, KSte, KStr, MSt, GS, POS, EST, BOT, YCT, JTh, ATi, DT, JTr, IT, ATö, PV, APdV, UV, GW, LWal, YXW, DMW, WBW, HW, JBW, SHW, JGW, CWo, TYW, MWu, LX, QY, MY, WZ, ABZ"

Statistical Methods and Analysis: TSA, MA, PA, MLB, GB, MBoi, TSB, MBr, CAB, MCa, RJC, JFC, DIC, Miao-Li Chee, XC, YC, AYC, MCo, MPC, JPC, TC, ADeh, GD, ADem, JDi, RD, DRVE, TLE, MFF, JFF, BIF, SFW, CF, SG, AG, MGo, DFG, MGö, TH, PH, PvdH, IMH, JNH, EH, AHo, KHor, SJH, JJ, PKJ, NSJ, BJ, YK, MKan, CCK, HKi, MEK, AKr, HKr, MKuo, AKöt, BK, LAL, CDL, MLi, YL, JiLiu, JuLiu, LPL, AMa, JoMar, JaMar, KMi, PPM, NM, APM, PJvdM, WM, MAN, MN, BN, DN, IMN, RN, TN, YO, CP, SAP, NP, MHP, BPP, LMR, MR, KMR, FRiv, FRiz, RRu, KAR, YSa, ES, MSc, CAS, SSe, YSh, KBS, XS, AVS, CNS, HMS, GS, SSz, SMT, BOT, ATeu, CHT, HT, GT, JTr, NV, VV, SVo, CWa, LWan, JFW, MKW, MWu, YX, QY, LMYA, WZ

Bioinformatics: TSA, SAk, PA, DB, SB, ASB, CAB, EC, RJC, XC, AYC, MCo, MPC, TC, EWD, FD, ADeh, JDi, RD, GE, AF, HG, SG, AG, SDG, MGo, PH, IMH, EH, AHo, KHor, JJ, NSJ, CCK, HKi, MEK, AKr, AKöt, CDL, BL, MLi, YL, JiLiu, LPL, JoMar, JaMar, YM, PPM, KLM, DOMK, PJvdM, RM, WM, RN, TN, SAP, NP, GP, AIP, MHP, BPP, FRiz, RRu, YSa, ES, MSc, CAS, SSe, CMS, KBS, AVS, CNS, SSz, HT, JTr, CWa, JFW, MWu, YX, LMYA, ZY, WZ

Interpretation of Results: TSA, EdA, CAB, CYC, KD, JDi, RD, KE, MFF, JFF, BIF, SG, CG, AG, MGo, PH, PvdH, HT, IMH, KHo, KHor, WH, SJH, BJ, HKi, WKo, AKr, AKöt, MLa, CDL, MLi, YL, PKM, JoMar, KMi, KLM, APM, NDP, CP, SAP, BPP, DFR, MR, PMR, MSc, SSe, KBS, CNS, POS, BOT, ATeu, ATi, JTr, IT, AGU, NV, VV, SVo, LWal, HW, MWu, YX, MY, LMYA

Genotyping: NA, DB, RB, ASB, CAB, AC, HC, DIC, CYC, EWD, ADem, RD, CMvD, GE, MKE, MFF, AF, YF, CF, RTG, HG, SDG, PH, PvdH, HT, CH, CKH, WH, EI, CCK, MEK, WKo, JSK, PK, ATK, AKr, FK, MKub, MKä, AKör, LAL, TL, LPL, PKM, TM, OM, YM, KLM, NM, GWM, DOMK, APM, JCM, WM, MAN, MOM, SP, NDP, BWP, MPe, DJP, MHP, OTR, DFR, FRiv, FRiz, JIR, DR, VS, ES, BHS, CNS, SMT, KT, ATeu, DT, JTr, AGU, SVa, UV, MWa, CWa, LWan, YXW, JGW, MKW, ABZ, JÄ

Critical review of manuscript: TSA, SAk, PA, EdA, SJB, NB, MLB, GB, MHdB, EPB, TSB, HB, ASB, CAB, HC, DIC, XC, YC, AYC, MCi, JC, RMvD, GD, OD, JDi, RD, KUE, DRVE, TLE, PE, KE, MKE, MFF, JFF, OHF, AF, BIF, YF, CF, HG, SG, CG, AG, SDG, MGo, DFG, PH, TBH, PvdH, CAH, CH, IMH, JNH, CKH, KHo, AHo, WH, NHK, SJH, OSI, EI, VWJ, JJ, JBJ, PKJ, BJ, MKas, SMK, MEK, WKo, ATK, HKr, FK, BKK, MKuo, MKä, AKör, AKöt, BK, MLa, LAL, CDL, JJML, TL, MLi, YL, WL, LL, CML, MLo, RJL, LPL, PKM, AMa, JoMar, NGM, DM, CM, TM, OM, EKM, KMi, KLM, GWM, DOMK, RdM, WM, GNN, MAN, MN, KN, BN, IMN, RN, TN, MLO, AJO, MOM, NDP, RP, AP, CP, SAP, BWP, MPe, OP, MHP, BPP, BMP, TJR, LMR, OTR, RRe, MR, KMR, PMR, FRiv, DJR, PR, IR, DR, CS, VS, KUS, MSc, CAS, NSc, BS, SSe, KBS, XS, AVS, HSn, CNS, KStr, GS, POS, SMT, NYQT, BOT, ATeu, HT, ATi, JTr, IT, ATö, AGU, NV, VV, SVo, APdV, UV, MWa, LWal, DMW, HW, JBW, SHW, JGW, MWu, QY, ZY, ABZ

Subject Recruitment: SAf, EPB, HB, CAB, AC, HC, JCC, Miao-Ling Chee, KC, RC, MCi, DC, KD, MKE, VHXF, BIF, RTG, VGu, CAH, WH, NHK, OSI, MI, VWJ, JBJ, BJ, CMK, MKas, JSK, AKr, FK, MKub, MKä, AKöt, MLa, JJML, TL, WL, LL, NGM, KMa, CM, AMe, RdM, WM, KN, MLO, IO, AJO, SP, CP, SAP, BWP, MPe, OP, BP, DJP, TPo, MAP, TJR, OTR, MR, PMR, PR, IR, DR, VS, RS, BHS, POS, NYQT, ATer, YCT, JTr, IT, ATö, SVa, SVo, PV, APdV, GW, LWal, HW, JBW, SHW, JGW, ABZ, JÄ

Competing interests

Wolfgang Koenig reports modest consultation fees for advisory board meetings from Amgen, DalCor, Kowa, Novartis, Pfizer and Sanofi, and modest personal fees for lectures from Amgen, AstraZeneca, Novartis, Pfizer and Sanofi, all outside the submitted work. Winfried März is employed with Synlab Services GmbH and holds shares of Synlab Holding Deutschland GmbH. Dennis O. Mook-Kanamori is a part time research consultant at Metabolon Inc. Mike A. Nalls is supported by a consulting contract between Data Tecnica International LLC and the National Institute on Aging (NIA), National Institutes of Health (NIH), Bethesda, MD, USA and consults for Illumina Inc., the Michael J. Fox Foundation, and the University of California Healthcare. Oscar H. Franco works in ErasmusAGE, a center for aging research across the life course funded by Nestlé Nutrition (Nestec Ltd.); Metagenics Inc.; and AXA. Karsten B. Sieber, Laura Yerges-Armstrong, Dawn M. Waterworth, and Mary Ann Lukas, are full-time employees of GlaxoSmithKline. Michelle L. O'Donoghue received grant support from GlaxoSmithKline, MSD, Eisai, AstraZeneca, MedCo and Janssen. Harvey White: Grants and non-financial support from GlaxoSmithKline, during the conduct of the study; grants from Sanofi-Aventis, Eli Lilly, National Institute of Health, George Institute, Omthera Pharmaceuticals, Pfizer New Zealand, Intarcia Therapeutics Inc., Elsai Inc., Dalcor Pharma UK Inc,; Honoraria and non-financial support from AstraZeneca and; Advisory Boards for Sirtex and Acetilion and personal fees from CSL Behring LLC and Luitpold Pharmaceuticals Inc. outside the submitted work. Lars Wallentin: Institutonal grants from GlaxoSmithKline, AstraZeneca, BMS, Boehringer-Ingelheim, Pfizer. MSD, Roche Diagnostics. Dermot F. Reilly and Anna I. Podgornaia are employees of Merck Sharp Dohme Corp., Whitehouse Station, NJ, USA. Markus Scholz: Consultancy of and grant support from Merck Serono not related to this project. Bruce M. Psaty 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. John Danesh is member of the Novartis Cardiovascular and Metabolic Advisory Board, received grant support from Novartis. Adam S. Butterworth received grants from MSD, Pfizer, Novartis, Biogen and Bioverativ and personal fees from Novartis. Veikko Salomaa has participated in a conference trip sponsored by Novo Nordisk and received a honorarium from the same source for

1 participating in an advisory board meeting. Anna Köttgen received grant support from

2 Gruenenthal. All other authors declare no conflicts of interest.

Figure Legends

Figure 1 – Trans-ethnic GWAS meta-analysis identifies 308 loci associated with
eGFR

Circos plot: Red band: $-\log_{10}(P)$ for association with eGFR, by chromosomal position.
Blue line indicates genome-wide significance ($P=5\times10^{-8}$). Black gene labels indicate novel
loci, blue labels known loci. Non-replicating loci are colored in gray (novel) or light blue
(known). Green band: Measures of heterogeneity related to the index SNPs associated
with eGFR. Dot sizes are proportional to I² or ancestry-related heterogeneity (p-anc-het).
Blue band: $-\log_{10}(P)$ for association with CKD, by chromosomal position. Red line
indicates genome-wide significance ($P=5\times10^{-8}$). Radial lines mark regions with p-anc-het
$<1.6\times10^{-4} = 0.05/308$ or I² $>25\%$. Inset: Effects of all 308 index SNPs on log(eGFR) by
minor allele frequency, color-coded by the associated odds ratio (OR) of CKD (red scale
for OR≤1, blue scale for OR>1). The largest effects on CKD were observed for
rs77924615 at *UMOD-PDILT* (odds ratio (OR) = 0.81, 95%CI: 0.80, 0.83), rs187355703
at *HOXD8* (OR=0.82, 95%CI: 0.77, 0.87), and rs10254101 at *PRKAG2* (OR=1.11,
95%CI: 1.09, 1.11). Triangles highlight SNPs that were associated with CKD (one-sided
P<0.05).

Figure 2 – Generalizability with respect to other populations and other kidney
function markers

Panel A: Measures of heterogeneity for 308 eGFR-associated index SNPs.
Comparison of each variant's heterogeneity quantified as I² from the trans-ethnic meta-
analysis (Y-axis) vs. ancestry-related heterogeneity from meta-regression (-$\log_{10}$(p-anc-
het), X-axis). Histograms summarize the distribution of the heterogeneity measures on
both axes. SNPs with ancestry-related heterogeneity (p-anc-het<$1.6\times10^{-4} = 0.05/308$) are
marked in blue and labeled; SNPs with I²>50% are labeled. **Panel B: Comparison of**
genetic effect estimates between CKDGen Consortium discovery (X-axis) and MVP
replication (Y-axis). Blue font indicates one-sided *P*<0.05 in the MVP. Error bars indicate

95% confidence intervals (CIs). Dashed line: line of best fit. Pearson's correlation coefficient r=0.92, 95%CI: 0.90, 0.94. **Panel C: Comparison of the magnitude of genetic effects on eGFR (X-axis) vs. BUN (Y-axis). for the 264 replicated eGFR-associated index SNPs**. Color coding reflects the evidence of kidney-function relevance (see Methods), which is coded as *likely* (blue), *inconclusive* (gray), or *unlikely* (green). Error bars indicate 95% CIs. Dashed line: line of best fit. Correlation: r=-0.65, 95%CI: -0.72, -0.58. **Panel D**: **Association between lower genetically-predicted eGFR based on a genetic risk score (GRS) of 147 SNPs likely to be most relevant for kidney function and ICD-10 based clinical diagnoses from 452,264 individuals from the UK Biobank**. Asthma was included as a negative control. Displayed are odds ratios and their 95% CIs per 10% lower GRS-predicted eGFR (Methods).

Figure 3 – Human orthologs of genes with renal phenotypes in genetically manipulated mice are enriched for association signals with eGFR

Signals in candidate genes identified based on the murine phenotypes abnormal GFR (**Panel A**), abnormal kidney physiology (**Panel B**), and abnormal kidney morphology (**Panel C**). Y-axis: $-\log_{10}(P)$ for association with eGFR in the trans-ethnic meta-analysis for the variant with the lowest p-value in each candidate gene. Dashed line indicates genome-wide significance ($P=5\times10^{-8}$), solid gray line indicates the experiment-wide significance threshold for each nested candidate gene analysis (included in lower right corner in each panel). Orange color indicates genome-wide significance, red color experiment-wide but not genome-wide significance, and blue color indicates genes with no significantly associated SNPs. Genes are labeled when reaching experiment- but not genome-wide significance; black font for genes not mapping into loci reported in the main analysis, gray font otherwise. Enrichment p-value reported for observed number of genes with association signals below the experiment-wide threshold against the expected number based on the complementary cumulative binomial distribution (Methods).

Figure 4 – Credible set size (X-axis) against variant posterior probability (Y-axis) of 3655 SNPs in 253 99% credible sets by annotation

Panel A: Exonic variants. SNPs are marked by triangles, with size proportional to their CADD score. Red triangles and variant labeling indicate missense SNPs mapping into small (≤5 SNPs) credible sets or with high individual posterior probability of driving the association signal (>0.5). **Panel B: Regulatory potential.** Symbol colors identify variants with regulatory potential as derived from DNAse hypersensitivity analysis in target tissues (Methods). Variant annotation was restricted to variants with variant PP>1%; SNPs with PP≥90% contained in credible sets with ≤10 SNPs were labeled.

Figure 5 – Co-localization of eGFR-association signals with gene expression in 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 ≥1 positive co-localization (posterior probability of one common causal variant, H4, ≥0.80) in a kidney tissue are illustrated with the respective sentinel SNP (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 marked in gray, while the positive co-localizations are color-coded based on the predicted change in expression relative to the allele associated with lower eGFR.

Figure 6 – Co-localization of independent eGFR-association signals at the *UMOD-PDILT* locus with urinary uromodulin concentrations supports *UMOD* as the effector gene.

Association plots: association $-\log_{10}$(p-value) (Y axis) vs. chromosomal position (X axis). Approximate conditional analyses among EA individuals support the presence of two independent eGFR-associated signals (**Panel A**). The association signal with urinary uromodulin/creatinine levels looks similar (**Panel B**); r^2=0.93 between rs34882080 and rs34262842. Co-localization of association with eGFR (upper sub-panel) and urinary uromodulin/creatinine levels (lower sub-panel) for the independent regions centered on

1   *UMOD* (**Panel C**) and *PDILT* (**Panel D**) support a shared underlying variant in both

2   regions with high posterior probability.

Table 1 – **Genes implicated as causal via identification of missense SNPs with high probability of driving the eGFR association signal.** Genes are included if they contain a missense SNP with posterior probability of association of >50% or mapping into a small credible set (≤5 SNPs).

Gene	SNP[1]	Credible set size	SNP PP[2]	Functional consequence	CADD score[3]	DHS[4], tissue	Brief summary of the gene's function and relevant literature (OMIM entries are indicated as #number)
CACNA1S	**rs3850625**	1	1.00	p.(Arg1539Cys) (NP_000060.2)	34.0	-	Encodes a subunit of the slowly inactivating L-type voltage-dependent calcium channel in skeletal muscle. Reports of altered expression in kidney cancer (PMID 28781648) and after indoxyl sulfate treatment (PMID: 27550174). Rare variants can cause autosomal dominant hypokalemic periodic paralysis, type 1 (#170400) or malignant hyperthermia susceptibility (#601887). Common variation at this locus has been reported as associated with eGFR in previous GWAS (PMID: 24029420, PMID: 26831199).
CPS1	rs1047891	1	1.00	p.(Thr1406Asn) (NP_001866.2)	22.1	-	Encodes a key mitochondrial enzyme of the urea cycle that catalyzes the synthesis of carbamoyl phosphate from ammonia and bicarbonate to remove excess urea. Rare mutations cause autosomal recessive carbamoylphosphate synthetase I deficiency (#237300). GWAS locus for eGFR (PMID: 20383146), serum metabolites (PMID: 23378610), and urinary glycine (PMID: 26352407), as well as for many other quantitative biomarkers. This variant has been reported to associate with hyperammonemia after valproate therapy (PMID: 23997965).
EDEM3	rs78444298	1	1.00	p.(Pro746Ser) (NP_079467.3)	24.6	-	The gene product accelerates the glycoprotein ER-associated degradation by proteasomes by catalyzing mannose trimming from Man8GlcNAc2 to Man7GlcNAc2 in the N-glycans. This variant has been identified by a previous exome chip association study with eGFR (PMID: 27920155).
KLHDC7A	rs11261022	7	0.71	p.(Arg160Ser) (NP_689588.2)	1.1	Roadmap, ENCODE kidney	Kelch Domain Containing 7A is a protein coding gene and a paralog of *KBTBD11*. No specific entry in relation to kidney disease in PubMed.
RPL3L	**rs113956264**	1	1.00	p.(Val262Met) (NP_005052.1)	27.2	-	The gene product shares sequence similarity with ribosomal protein L3. It has a tissue-specific expression pattern, with highest levels in skeletal muscle and heart.
SLC25A45	rs34400381	1	1.00	p.(Arg285Cys) (NP_001070709.2)	26.0	ENCODE kidney	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 (PMID: 24159190) and in a whole-genome sequence (WGS) analysis of serum creatinine (PMID: 25082825). *SLC25A45* may play a role in biosynthesis of arginine, which is involved in the synthesis of creatine.
SLC47A1	rs111653425	1	1.00	p.(Ala465Val) (NP_060712.2)	24.6	-	Encodes the 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 WGS analysis of serum creatinine from Iceland (PMID: 25082825). Rare and common variants in the locus have been identified in exome chip (PMID: 27920155) and in GWAS (PMID: 20383146) studies of eGFR, respectively. MATE1 knockout (KO) mice show higher levels of serum creatinine and BUN (PMID: 19332510), arguing against a sole effect on creatinine transport and supporting an effect on kidney function.
PPM1J	rs34611728	5	0.02	p.(Leu213Phe) (NP_005158.5)	13.1	ENCODE kidney	This gene encodes the serine/threonine protein phosphatase. The variant has been reported in association with eGFR in an exome chip association study (PMID: 27920155).
CERS2	**rs267738**	5	0.46	p.(Glu115Ala) (NP_071358.1)	32.0 / 28.2	-	Encodes Ceramide Synthase 2, which may be involved in sphingolipid synthesis. Changes in ceramides were reported as essential in renal Madin-Darby Canine Kidney (MDCK) cell differentiation (PMID: 28515139). *CERS2* KO mice show strongly reduced ceramide levels in the kidney and develop renal parenchyma abnormalities (PMID: 19801672). This variant has been reported as associated with the rate of albuminuria increase in patients with diabetes (PMID: 25238615).
C9	**rs700233**	5	0.32	p.(Arg5Trp) (NP_001728.1)	6.6	-	Encodes a constituent of the membrane attack complex that plays a key role in the innate and adaptive immune response. Rare mutations can cause C9 deficiency (#613825). *C9* is mentioned in several kidney disease case reports, including patients with congenital factor 9 deficiency showing IgA nephropathy (PMID: 1453611).
SLC22A2	rs316019	4	0.04	p.(Ser270Ala) (NP_003049.2)	12.7	-	Encodes the polyspecific organic cation transporter (OCT2) that is primarily expressed in the kidney, where it mediates tubular uptake of organic compounds including creatinine from the circulation. Many publications relate *SLC22A2* to kidney function. rs316019 is a known pharmacogenomics variant associated with response to metformin and other drugs such as cisplatin. Carriers of the risk allele have a higher risk of cisplatin-induced nephrotoxicity (PMID: 19625999), indicating that this transporter is essential in excreting toxins. The locus has been reported in previous GWAS of eGFR (PMID: 20383146).

[1]Boldface indicates SNPs most likely to be relevant for kidney function based on the combined effects on eGFR and BUN; [2]PP: posterior probability. [3]CADD score: Combined Annotation Dependent Depletion (CADD) PHRED-like score (Methods); [4]DHS: DNAse Hypersensitivity Site

Author List

Matthias Wuttke*[1,2], Yong Li*[1], Man Li*[3], Karsten B. Sieber*[4], Mary F. Feitosa*[5], Mathias Gorski*[6,7], Adrienne Tin[8,9], Lihua Wang[5], Audrey Y. Chu[10], Anselm Hoppmann[1], Holger Kirsten[11,12], Ayush Giri[13,14], Jin-Fang Chai[15], Gardar Sveinbjornsson[16], Bamidele O. Tayo[17], Teresa Nutile[18], Christian Fuchsberger[19], Jonathan Marten[20], Massimiliano Cocca[21], Sahar Ghasemi[22,23], Yizhe Xu[3], Katrin Horn[11,12], Damia Noce[19], Peter J. van der Most[24], Sanaz Sedaghat[25], Zhi Yu[8,26], Masato Akiyama[27,28], Saima Afaq[29,30], Tarunveer S. Ahluwalia[31], Peter Almgren[32], Najaf Amin[25], Johan Ärnlöv[33,34], Stephan J.L. Bakker[35], Nisha Bansal[36,37], Daniela Baptista[38], Sven Bergmann[39,40,41], Mary L. Biggs[42,43], Ginevra Biino[44], Michael Boehnke[45], Eric Boerwinkle[46], Mathilde Boissel[47], Erwin P. Bottinger[48,49], Thibaud S. Boutin[20], Hermann Brenner[50,51], Marco Brumat[52], Ralph Burkhardt[12,53,54], Adam S. Butterworth[55,56], Eric Campana[52], Archie Campbell[57], Harry Campbell[58], Mickaël Canouil[47], Robert J. Carroll[59], Eulalia Catamo[21], John C. Chambers[29,60,61,62,275], Miao-Ling Chee[63], Miao-Li Chee[63], Xu Chen[64], Ching-Yu Cheng[63,65,66], Yurong Cheng[1], Kaare Christensen[67], Renata Cifkova[68,69], Marina Ciullo[18,70], Maria Pina Concas[21], James P. Cook[71], Josef Coresh[8], Tanguy Corre[39,40,72], Sala Cinzia Felicita[73], Daniele Cusi[74,75], John Danesh[76], E. Warwick Daw[5], Martin H. de Borst[35], Alessandro De Grandi[19], Renée de Mutsert[77], Aiko P.J. de Vries[78], Frauke Degenhardt[79], Graciela Delgado[80], Ayse Demirkan[25], Emanuele di Angelantonio[81,82,83], Katalin Dittrich[84,85], Jasmin Divers[86], Rajkumar Dorajoo[87], Kai-Uwe Eckardt[88,89], Georg Ehret[38], Paul Elliott[90,91,92,93], Karlhans Endlich[23,94], Michele K. Evans[95], Janine F. Felix[25,96,97], Valencia Hui Xian Foo[63], Oscar H. Franco[25,98], Andre Franke[79], Barry I. Freedman[99], Sandra Freitag-Wolf[100], Yechiel Friedlander[101], Philippe Froguel[47,102], Ron T. Gansevoort[35], He Gao[90], Paolo Gasparini[21,52], J. Michael Gaziano[103], Vilmantas Giedraitis[104], Christian Gieger[105,106,107], Giorgia Girotto[21,52], Franco Giulianini[108], Martin Gögele[19], Scott D. Gordon[109], Daniel F. Gudbjartsson[16], Vilmundur Gudnason[110,111], Toomas Haller[112], Pavel Hamet[113,114], Tamara B. Harris[115], Catharina A. Hartman[116], Caroline Hayward[20], Jacklyn N. Hellwege[117,118,119], Chew-Kiat Heng[120,121], Andrew A. Hicks[19], Edith Hofer[122,123], Wei Huang[124,125], Nina Hutri-Kähönen[126,127], Shih-Jen Hwang[128,129], M. Arfan Ikram[25], Olafur S. Indridason[130], Erik Ingelsson[131,132,133,134], Marcus Ising[135], Vincent W.V. Jaddoe[25,96,97], Johanna Jakobsdottir[136,137], Jost B. Jonas[138,139], Peter K. Joshi[58], Navya Shilpa Josyula[140], Bettina Jung[6], Mika Kähönen[141,142], Yoichiro Kamatani[27,143], Candace M. Kammerer[144], Masahiro Kanai[27,145], Mika Kastarinen[146], Shona M. Kerr[20], Chiea-Chuen Khor[63,87], Wieland Kiess[12,84,85], Marcus E. Kleber[80], Wolfgang Koenig[147,148,149], Jaspal S. Kooner[61,62,150,151], Antje Körner[12,84,85], Peter Kovacs[152], Aldi T. Kraja[5], Alena Krajcoviechova[68,69], Holly Kramer[17,153], Bernhard K. Krämer[80], Florian Kronenberg[154], Michiaki Kubo[155], Brigitte Kühnel[105], Mikko Kuokkanen[156,157], Johanna Kuusisto[158,159], Martina La Bianca[21], Markku Laakso[158,159], Leslie A. Lange[160], Carl D. Langefeld[86], Jeannette Jen-Mai Lee[15], Benjamin Lehne[29], Terho Lehtimäki[161,162], Wolfgang Lieb[163], Lifelines Cohort Study[164], Su-Chi Lim[15,165], Lars Lind[166], Cecilia M. Lindgren[167,168], Jun Liu[25], Jianjun Liu[87,169], Markus Loeffler[11,12], Ruth J.F. Loos[48,170], Susanne Lucae[135], Mary Ann Lukas[171], Leo-Pekka Lyytikäinen[161,162], Reedik Mägi[112], Patrik K.E. Magnusson[64], Anubha Mahajan[172,173], Nicholas G. Martin[109], Jade Martins[174], Winfried März[175,176,177], Deborah Mascalzoni[19], Koichi Matsuda[178], Christa Meisinger[179,180], Thomas Meitinger[148,181,182], Olle Melander[183], Andres Metspalu[112], Evgenia K. Mikaelsdottir[16], Yuri Milaneschi[184], Kozeta Miliku[25,96,97], Pashupati P. Mishra[161,162], VA Million Veteran

Program[185], Karen L. Mohlke[186], Nina Mononen[161,162], Grant W. Montgomery[187], Dennis O. Mook-Kanamori[77,188], Josyf C. Mychaleckyj[189], Girish N. Nadkarni[48,190], Mike A. Nalls[191,192], Matthias Nauck[23,193], Kjell Nikus[194,195], Boting Ning[196], Ilja M. Nolte[24], Raymond Noordam[197], Jeffrey O'Connell[198], Michelle L. O'Donoghue[199,200], Isleifur Olafsson[201], Albertine J. Oldehinkel[116], Marju Orho-Melander[32], Willem H. Ouwehand[76], Sandosh Padmanabhan[202], Nicholette D. Palmer[203], Runolfur Palsson[111,130], Brenda W.J.H. Penninx[184], Thomas Perls[204], Markus Perola[205,206], Mario Pirastu[207], Nicola Pirastu[58], Giorgio Pistis[208], Anna I. Podgornaia[10], Ozren Polasek[209,210], Belen Ponte[211], David J. Porteous[57,212], Tanja Poulain[12], Peter P. Pramstaller[19], Michael H. Preuss[48], Bram P. Prins[55], Michael A. Province[5], Ton J. Rabelink[78,213], Laura M. Raffield[186], Olli T. Raitakari[214,215], Dermot F. Reilly[10], Rainer Rettig[216], Myriam Rheinberger[6], Kenneth M. Rice[43], Paul M. Ridker[108,217], Fernando Rivadeneira[25,218], Federica Rizzi[219,220], David J. Roberts[81,221], Antonietta Robino[21], Peter Rossing[31], Igor Rudan[58], Rico Rueedi[39,40], Daniela Ruggiero[18,70], Kathleen A. Ryan[222], Yasaman Saba[223], Charumathi Sabanayagam[63,224], Veikko Salomaa[205], Erika Salvi[219,225], Kai-Uwe Saum[50], Helena Schmidt[226], Reinhold Schmidt[122], Ben Schöttker[50,51], Christina-Alexandra Schulz[32], Nicole Schupf[227,228,229], Christian M. Shaffer[59], Yuan Shi[63,224], Albert V. Smith[111], Blair H. Smith[230], Nicole Soranzo[231], Cassandra N. Spracklen[186], Konstantin Strauch[232,233], Heather M. Stringham[45], Michael Stumvoll[234], Per O. Svensson[235,236], Silke Szymczak[100], E-Shyong Tai[15,169,237], Salman M. Tajuddin[95], Nicholas Y. Q. Tan[63], Kent D. Taylor[238], Andrej Teren[12,239], Yih-Chung Tham[63], Joachim Thiery[12,53], Chris H.L. Thio[24], Hauke Thomsen[240], Gudmar Thorleifsson[16], Daniela Toniolo[73], Anke Tönjes[234], Johanne Tremblay[113,241], Ioanna Tzoulaki[90,242], André G. Uitterlinden[218], Simona Vaccargiu[207], Rob M. van Dam[15,169,243], Pim van der Harst[244,245,246], Cornelia M. van Duijn[25], Digna R. Velez Edward[119,247,248], Niek Verweij[244], Suzanne Vogelezang[25,96,97], Uwe Völker[23,249], Peter Vollenweider[250], Gerard Waeber[250], Melanie Waldenberger[105,106,148], Lars Wallentin[251,252], Ya Xing Wang[139], Chaolong Wang[87,253], Dawn M. Waterworth[4], Wen Bin Wei[254], Harvey White[255], John B. Whitfield[109], Sarah H. Wild[256], James F. Wilson[20,58], Mary K. Wojczynski[5], Charlene Wong[66], Tien-Yin Wong[63,66,224], Liang Xu[139], Qiong Yang[196], Masayuki Yasuda[63,257], Laura M. Yerges-Armstrong[4], Weihua Zhang[61,90], Alan B. Zonderman[95], Jerome I. Rotter[238,258,259], Murielle Bochud[72], Bruce M. Psaty[260,261], Veronique Vitart[20], James G. Wilson[262], Abbas Dehghan[29,90], Afshin Parsa[263,264], Daniel I. Chasman[108,217], Kevin Ho[265,266], Andrew P. Morris[71,172], Olivier Devuyst[267], Shreeram Akilesh[37,268], Sarah A. Pendergrass[269], Xueling Sim[15], Carsten A. Böger[6,270], Yukinori Okada[271,272], Todd L. Edwards[119,273], Harold Snieder[24], Kari Stefansson[16], Adriana M. Hung[119,274], Iris M. Heid**[7], Markus Scholz**[11,12], Alexander Teumer**[22,23], Anna Köttgen**†[1,8], Cristian Pattaro**†[19]

* Indicates joint contribution

** Indicates joint oversight

† Indicates corresponding author

Author affiliations

1 Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Center - University of Freiburg, Freiburg, Germany

2 Department of Medicine IV, Renal Division, Faculty of Medicine and Medical Center - University of Freiburg, Freiburg, Germany

3 Department of Medicine, Division of Nephrology and Hypertension, University of Utah, Salt Lake City, USA

4 Target Sciences - Genetics, GlaxoSmithKline, Collegeville (Pennsylvania), USA

5 Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis (Missouri), USA

6 Department of Nephrology, University Hospital Regensburg, Regensburg, Germany

7 Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany

8 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore (Maryland), USA

9 Epidemiology and Clinical Research, Welch Centre for Prevention, Baltimore (Maryland), USA

10 Genetics, Merck & Co., Inc., Kenilworth, New Jersey, USA

11 Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany

12 LIFE Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany

13 Division of Quantitative Sciences, Department of Obstetrics & Gynecology, Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, USA

14 Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA

15 Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore

16 deCODE Genetics, Amgen Inc., Reykjavik, Iceland

17 Department of Public Health Sciences, Loyola University Chicago, Maywood (Illinois), USA

18 Institute of Genetics and Biophysics "Adriano Buzzati-Traverso" - CNR, Naples, Italy

19 Eurac Research, Institute for Biomedicine (affiliated to the University of Lübeck), Bolzano, Italy

20 Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

21 Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy

22 Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

23 DZHK (German Center for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany

24 Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

25 Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

26 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore (Maryland), USA

27 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan

28 Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

29 Department of Epidemiology and Biostatistics, Faculty of Medicine, School of Public Health, Imperial College London, London, UK

30 Institute of Public health & social sciences, Khyber Medical University, Pakistan

31 Steno Diabetes Centre Copenhagen, Gentofte, Denmark

32 Diabetes and Cardiovascular Disease - Genetic Epidemiology, Department of Clincial Sciences in Malmö, Lund University, Malmö, Sweden

33 Department of Neurobiology, Care Sciences and Society, Division of Family Medicine and Primary Care, Karolinska Institutet, Stockholm, Sweden

34 School of Health and Social Studies, Dalarna University, Sweden

35 Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

36 Division of Nephrology, University of Washington, Seattle (Washington), USA

37 Kidney Research Institute, University of Washington, Seattle (Washington), USA

38 Cardiology, Geneva University Hospitals, Geneva, Switzerland

39 Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

40 Swiss Institute of Bioinformatics, Lausanne, Switzerland

41 Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa

42 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle (Washington), USA

43 Department of Biostatistics, University of Washington, Seattle (Washington), USA

44 Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy

45 Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA

46 Human Genetics Centre, University of Texas Health Science Centre, Houston (Texas), USA

47 CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, University of Lille, Lille, France

48 The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York (New York), USA

49 Digital Health Centre, Hasso Plattner Institute and University of Potsdam, Potsdam, Germany

50 Division of Clinical Epidemiology and Aging Research, German Cancer Research Centre (DKFZ), Heidelberg, Germany

51 Network Aging Research, University of Heidelberg, Heidelberg, Germany

52 University of Trieste, Department of Medicine, Surgery and Health Sciences, Trieste, Italy

53 Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig, Leipzig, Germany

54 Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany

55 MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

56 National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK

57 Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh,

Edinburgh, UK
58 Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
59 Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville (Tennessee), USA
60 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
61 Department of Cardiology, Ealing Hospital, Middlesex UB1 3HW, UK
62 Imperial College Healthcare NHS Trust, Imperial College London, London, UK
63 Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore
64 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
65 Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore
66 Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore
67 Unit of Epidemiology, Biostatistics and Biodemography, Department of Public Health, Southern Denmark University, Odense, Denmark
68 Centre for Cardiovascular Prevention, First Faculty of Medicine, Department of Medicine, Charles University in Prague, Prague, Czech Republic
69 Thomayer Hospital, Prague, Czech Republic
70 IRCCS Neuromed, Pozzilli, Italy
71 Department of Biostatistics, University of Liverpool, Liverpool, UK
72 Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland
73 San Raffaele Research Institute, Milano, Italy
74 Institute of Biomedical Technologies, Italy National Research Council, Bresso (Milano), Italy
75 Bio4Dreams - business nursery for life sciences, Bresso (Milano), Italy
76 University of Cambridge, Cambridge, UK
77 Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands
78 Section of Nephrology, Department of Internal Medicine, Leiden University Medical Centre, Leiden, The Netherlands
79 Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
80 5th Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
81 Blood and Transplant Research Unit in Donor Health and Genomics, National Institute of Health Research, Cambridge, UK
82 Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
83 NHS Blood and Transplant, Cambridge, UK
84 Department of Women and Child Health, Hospital for Children and Adolescents, University of Leipzig, Leipzig, Germany
85 Centre for Pediatric Research, University of Leipzig, Leipzig, Germany
86 Public Health Sciences - Biostatistics, Wake Forest School of Medicine, Winston-Salem (North Carolina), USA
87 Genome Institute of Singapore, Agency for Science Technology and Research, Singapore, Singapore
88 Intensive Care Medicine, Charité, Berlin, Germany
89 Department of Nephrology and Hypertension, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany
90 Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK
91 Imperial College NIHR Biomedical Research Centre, Imperial College London, London, UK
92 Dementia Research Institute, Imperial College London, London, UK
93 Health Data Research UK-London, London, UK
94 Department of Anatomy and Cell Biology, University Medicine Greifswald, Greifswald, Germany
95 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Baltimore (Maryland), USA
96 The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands
97 Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands
98 Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland
99 Internal Medicine - Section on Nephrology, Wake Forest School of Medicine, Winston-Salem (North Carolina), USA
100 Institute of Medical Informatics and Statistics, Kiel University, University Hospital Schleswig-Holstein, Kiel
101 School of Public Health and Community Medicine, Hebrew University of Jerusalem, Jerusalem, Israel
102 Department of Genomics of Common Disease, Imperial College London, London, UK
103 Massachusetts Veterans Epidemiology Research and Information Center, VA Cooperative Studies Program, VA Boston Healthcare System, Boston (Massachusetts), USA
104 Department of Public Health and Caring Sciences, Molecular Geriatrics, Uppsala University, Uppsala, Sweden
105 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
106 Institute of Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
107 German Center for Diabetes Research (DZD), Neuherberg, Germany
108 Division of Preventive Medicine, Brigham and Women's Hospital, Boston, USA
109 QIMR Berghofer Medical Research Institute, Brisbane, Australia
110 Icelandic Heart Association, Kopavogur, Iceland
111 Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland
112 Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
113 Montreal University Hospital Research Centre, CHUM, Montreal, Canada
114 Medpharmgene, Montreal, Canada
115 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda (Maryland), USA

116 Interdisciplinary Centre Psychopathology and Emotion regulation (ICPE), University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
117 Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, Nashville, USA
118 Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, USA
119 Department of Veteran's Affairs, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA
120 Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
121 Khoo Teck Puat-National University Children's Medical Institute, National University Health System, Singapore, Singapore
122 Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria
123 Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria
124 Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Centre, Shanghai, China
125 Shanghai Industrial Technology Institute, Shanghai, China
126 Department of Pediatrics, Tampere University Hospital, Tampere, Finland
127 Department of Pediatrics, Faculty of Medicine and Life Sciences, University of Tampere, Finland
128 NHLBI's Framingham Heart Study, Framingham (Massachusetts), USA
129 The Centre for Population Studies, NHLBI, Framingham (Massachusetts), USA
130 Division of Nephrology, Internal Medicine Services, Landspitali - he National University Hospital of Iceland, Reykjavik, Iceland
131 Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, USA
132 Stanford Cardiovascular Institute, Stanford University, USA
133 Molecular Epidemiology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
134 Stanford Diabetes Research Center, Stanford University, Stanford, USAs
135 Max Planck Institute of Psychiatry, Munich, Germany
136 Icelandic Heart Association, Holtasmari 1, Kopavogur IS-201, Iceland
137 The Centre of Public Health Sciences, University of Iceland, Sturlugata 8, Reykjavík IS-101, Iceland
138 Department of Ophthalmology, Medical Faculty Mannheim, University Heidelberg, Mannheim, Germany
139 Beijing Institute of Ophthalmology, Beijing Key Laboratory of Ophthalmology and Visual Sciences, Beijing Tongren Hospital, Capital Medical University, Beijing, China
140 Geisinger Research, Biomedical and Translational Informatics Institute, Rockville, USA
141 Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland
142 Department of Clinical Physiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
143 Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
144 Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh (Pennsylvania), USA
145 Department of Biomedical Informatics, Harvard Medical School, Boston, USA
146 Kuopio University Hospital, Kuopio, Finland
147 Deutsches Herzzentrum München, Technische Universität München, Munich, Germany
148 DZHK (German Centre for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany
149 Institute of Epidemiology and Biostatistics, University of Ulm, Ulm, Germany
150 MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK
151 National Heart and Lung Institute, Imperial College London, London, UK
152 Integrated Research and Treatment Centre Adiposity Diseases, University of Leipzig, Leipzig, Germany
153 Division of Nephrology and Hypertension, Loyola University Chicago, USA
154 Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria
155 RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan
156 The Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland
157 Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland
158 University of Eastern Finland, Kuopio, Finland
159 Kuopio University Hospital, Finland
160 Division of Biomedical Informatics and Personalized Medicine, School of Medicine, University of Colorado Denver - Anschutz Medical Campus, Aurora (Colorado), USA
161 Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland
162 Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
163 Institute of Epidemiology and Biobank Popgen, Kiel University, Kiel, Germany
164 A list of members and affiliations appears in the Supplementary Material.
165 Diabetes Centre, Khoo Teck Puat Hospital, Singapore, Singapore
166 Cardiovascular Epidemiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
167 Nuffield Department of Medicine, University of Oxford, Oxford, UK
168 Broad Institute of Harvard and MIT, USA
169 Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore
170 The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York (New York), USA
171 Target Sciences - Genetics, GlaxoSmithKline, Albuquerque (New Mexico), USA
172 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
173 Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, UK
174 Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany
175 Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany

176 Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Austria
177 Medical Clinic V, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
178 Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan
179 Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany
180 Chair of Epidemiology Ludwig- Maximilians-Universität München at UNIKA-T Augsburg, Augsburg, Germany
181 Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany
182 Institute of Human Genetics, Technische Universität München, Munich, Germany
183 Hypertension and Cardiovascular Disease, Department of Clincial Sciences Malmö, Lund University, Malmö, Sweden
184 Department of Psychiatry, VU University Medical Centre, Amsterdam, The Netherlands
185 Department of Veterans Affairs. Office of Research and Development, Washington, DC, USA. A list of members and affiliations appears in the Supplementary Material.
186 Department of Genetics, University of North Carolina, Chapel Hill (North Carolina), USA
187 University of Queensland, St Lucia, Australia
188 Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, The Netherlands
189 Centre for Public Health Genomics, University of Virginia, Charlottesville (Virginia), USA
190 Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York (New York), USA
191 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda (Maryland), USA
192 Data Tecnica International, Glen Echo (Maryland), USA
193 Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
194 Department of Cardiology, Heart Center, Tampere University Hospital, Tampere, Finland
195 Department of Cardiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, Tampere University, Tampere, Finland
196 Department of Biostatistics, Boston University School of Public Health, Boston (Massachusetts), USA
197 Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Centre, Leiden, The Netherlands
198 University of Maryland School of Medicine, Baltimore, USA
199 Cardiovascular Division, Brigham and Women's Hospital, Boston, USA
200 TIMI Study Group, USA
201 Department of Clinical Biochemistry, Landspitali University Hospital, Reykjavik, Iceland
202 Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK
203 Biochemistry, Wake Forest School of Medicine, Winston-Salem (North Carolina), USA
204 Department of Medicine, Geriatrics Section, Boston Medical Center, Boston University School of Medicine, Boston (Massachusetts), USA
205 National Institute for Health and Welfare, Helsinki, Finland
206 The Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland
207 Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Li Punti (Sassari), Italy
208 Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland
209 Faculty of Medicine, University of Split, Split, Croatia
210 Gen-info Ltd, Zagreb, Croatia
211 Service de Néphrologie, Geneva University Hospitals, Geneva, Switzerland
212 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
213 Einthoven Laboratory of Experimental Vascular Research, Leiden University Medical Centre, Leiden, The Netherlands
214 Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland
215 Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
216 Institute of Physiology, University Medicine Greifswald, Karlsburg, Germany
217 Harvard Medical School, USA
218 Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands
219 Department of Health Sciences, University of Milan, Milano, Italy
220 ePhood Scientific Unit, ePhood SRL, Milano, Italy
221 NHS Blood and Transplant; BRC Oxford Haematology Theme; Nuffield Division of Clinical Laboratory Sciences; University of Oxford, Oxford, UK
222 Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, USA
223 Molecular Biology and Biochemistry, Gottfried Schatz Research Centre for Cell Signaling, Metabolism and Aging, Medical University of Graz, Graz, Austria
224 Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore
225 Neuroalgology Unit, Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milan, Italy
226 Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Graz, Austria
227 Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, USA
228 Gertrude H. Sergievsky Centre, Columbia University Medical Centre, USA
229 Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Centre, New York, USA
230 Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK
231 Wellcome Sanger Institute, South Cambridgeshire, UK
232 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
233 Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Germany
234 Department of Endocrinology and Nephrology, University of Leipzig, Leipzig, Germany
235 Department of Clinical Science and Education, Karolinska Institutet, Södersjukhuset (KI SÖS), Stockholm, Sweden
236 Department of Cardiology, Södersjukhuset, Stockholm, Sweden
237 Duke-NUS Medical School, Singapore, Singapore

238 The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA
239 Heart Centre Leipzig, Leipzig, Germany
240 Division of Molecular Genetic Epidemiology, German Cancer Research Centre (DKFZ), Heidelberg, Germany
241 CRCHUM, Montreal, Canada
242 Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece
243 Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, USA
244 Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
245 Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
246 Durrer Centre for Cardiovascular Research, The Netherlands Heart Institute, Utrecht, The Netherlands
247 Department of Obstetrics and Gynecology, Institute for Medicine and Public Health, Vanderbilt University Medical Centre, Nashville (TN), USA
248 Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, USA
249 Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
250 Internal Medicine, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland
251 Cardiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
252 Uppsala Clinical Research Centre, Uppsala University. Uppsala, Sweden
253 School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
254 Beijing Tongren Eye Centre, Beijing Tongren Hospital, Capital Medical University, Beijing, China
255 Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, New Zealand
256 Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
257 Department of Ophthalmology, Tohoku University Graduate School of Medicine, Japan
258 Department of Pediatrics, Harbor-UCLA Medical Centre, USA
259 Department of Medicine, Harbor-UCLA Medical Centre, Torrance, USA
260 Cardiovascular Health Research Unit, Department of Medicine, Department of Epidemiology, Department of Health Service, University of Washington, Seattle (Washington), USA
261 Kaiser Permanente Washington Health Research Institute, Seattle (Washington), USA
262 Department of Physiology and Biophysics, University of Mississippi Medical Centre, Jackson (Mississippi), USA
263 Division of Kidney, Urologic and Hematologic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, USA
264 Department of Medicine, University of Maryland School of Medicine, Baltimore, USA
265 Kidney Health Research Institute (KHRI), Geisinger, Danville (Pennsylvania), USA
266 Department of Nephrology, Geisinger, Danville (Pennsylvania), USA
267 Institute of Physiology, University of Zurich, Zurich, Switzerland
268 Anatomic Pathology, University of Washington Medical Center, Seattle, USA
269 Geisinger Research, Biomedical and Translational Informatics Institute, Danville (Pennsylvania), USA
270 Department of Nephrology and Rheumatology, Kliniken Südostbayern AG, Regensburg, Germany
271 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Osaka, Japan
272 Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan
273 Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, Nashville, TN, USA
274 Vanderbilt University Medical Centre, Division of Nephrology & Hypertension, Nashville, TN, USA
275 MRC-PHE Centre for Environment and Health, Imperial College London, London W2 1PG, UK