Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis

Sarah E. Graham¹, Jonas B. Nielsen¹, Matthew Zawistowski², Wei Zhou³, Lars G. Fritsche², Maiken E. Gabrielsen⁴,⁵, Anne Heidi Skogholt⁴,⁵,⁶, Ida Surakka¹, Whitney E. Hornsby¹, Damian Fermin⁷, Daniel B. Larach⁸, Sachin Kheterpal⁸, Chad M. Brummett⁸, Seunggeun Lee², Hyun Min Kang², Goncalo R. Abecasis², Solfrid Romundstad⁶,⁹, Stein Hallan⁶,¹⁰, Matthew G. Sampson⁷, Kristian Hveem⁴,⁵,¹¹ & Cristen J. Willer¹,³,¹²

Chronic kidney disease (CKD) is a growing health burden currently affecting 10–15% of adults worldwide. Estimated glomerular filtration rate (eGFR) as a marker of kidney function is commonly used to diagnose CKD. We analyze eGFR data from the Nord-Trøndelag Health Study and Michigan Genomics Initiative and perform a GWAS meta-analysis with public summary statistics, more than doubling the sample size of previous meta-analyses. We identify 147 loci (53 novel) associated with eGFR, including genes involved in transcriptional regulation, kidney development, cellular signaling, metabolism, and solute transport. Additionally, sex-stratified analysis identifies one locus with more significant effects in women than men. Using genetic risk scores constructed from these eGFR meta-analysis results, we show that associated variants are generally predictive of CKD with only modest improvements in detection compared with other known clinical risk factors. Collectively, these results yield additional insight into the genetic factors underlying kidney function and progression to CKD.

¹Department of Internal Medicine: Cardiology, University of Michigan, Ann Arbor 48109 MI, USA. ²Department of Biostatistics: Center for Statistical Genetics, University of Michigan, Ann Arbor 48109 MI, USA. ³Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor 48109 MI, USA. ⁴K.G. Jebsen Center for Genetic Epidemiology, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim 7491, Norway. ⁵Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim 7491, Norway. ⁶Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim 7491, Norway. ⁷Department of Pediatrics: Pediatric Nephrology, University of Michigan, Ann Arbor 48109 MI, USA. ⁸Department of Anesthesiology, University of Michigan, Ann Arbor 48109 MI, USA. ⁹Department of Internal Medicine, Levanger Hospital, Health Trust Nord-Trøndelag, Levanger 7600, Norway. ¹⁰Department of Nephrology, St Olav Hospital, Trondheim 7491, Norway. ¹¹HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger 7600, Norway. ¹²Department of Human Genetics, University of Michigan, Ann Arbor 48109 MI, USA. Correspondence and requests for materials should be addressed to K.H. (email: kristian.hveem@ntnu.no) or to C.J.W. (email: cristen@umich.edu)
C

hronic kidney disease (CKD) is a common condition affecting ~1% of adults in Norway and ~15% in the United States. Due to specific comorbidities (namely diabetes) and an aging population, CKD is expected to continue to rise in global prevalence. Estimated glomerular filtration rate (eGFR) provides an assessment of kidney function and it is estimated based on serum creatinine levels with adjustment for age, race, and sex. eGFR levels below 60 ml/min/1.73 m² generally characterize chronic kidney disease, with varying severity classified by both albuminuria and eGFR levels. A subset of individuals with CKD have accelerated renal function decline and progress to end stage renal disease (ESRD).

Several other health conditions affect kidney function. Chronic diseases such as diabetes and hypertension directly influence the development of CKD, with environmental factors such as smoking accelerating disease progression. Advanced stages of CKD/ESRD necessitate dialysis or transplantation and are associated with an increased risk of cardiovascular disease and death.

It has been estimated that about one-third of the variation in eGFR levels can be attributed to genetic factors, with the remaining variability due to environmental effects. Previous genome-wide association studies (GWAS) and meta-analyses have identified a number of loci associated with serum creatinine, eGFR, or CKD. However, the introduction of denser imputation panels, including the Haplotype Reference Consortium (HRC), and the recent rise in large-scale biobanks has enabled larger sample sizes and a greater number of variants than previously studied. Analysis of these new and more densely imputed datasets are expected to identify genetic regions influencing these traits not previously found.

We analyze samples from the Michigan Genomics Initiative (MGI) and the Nord-Trøndelag Health Study (HUNT), and impute using HRC and a combined HRC and ancestry-specific panel, respectively, for association with eGFR. Finally, we perform a meta-analysis of eGFR associations with two other cohorts to uncover additional genetic variants contributing to kidney function. We identify 147 loci associated with eGFR, including 53 novel loci and one locus with a significantly larger effect in women than in men. The index variants in these loci are further associated with related traits, including diabetes, hypertension, and cardiovascular disease. Lastly, we demonstrate that genetic risk scores constructed from significantly associated eGFR variants are correlated with CKD on a population level.

Kidney-specific eQTL associations. To identify variants that may be acting through regulation of gene expression within the kidneys, we examined which eGFR index variants were significant eQTLs (p-value < 6.7 × 10⁻⁶, Bonferroni correction for 51 tissue types and 147 index variants) for a given gene in human kidney cortex, glomerulus, or tubulointerstitium. This identified 16 genes whose expression was associated with the eGFR index variants in kidney tissues, including 7 genes identified from normal kidney cortex tissue samples and 10 genes identified from kidney glomerulus or tubulointerstitium samples from individuals with nephrotic syndrome (1 gene overlapped both datasets, Supplementary Data 7). Five of these genes had expression levels associated with the eGFR index variants specifically in kidney tissues (p-value < 6.7 × 10⁻⁶) but not in other tissues in GTEx. Additional enriched tissues (FDR < 0.05, Supplementary Data 3). In addition, some genes demonstrated a colocalizing eQTL association in non-kidney tissues but are also thought to cause Mendelian kidney diseases when disrupted (ALMS1, DCDC2, MUC1, RPS10, SDCCAG8, SLC34A1).

DEPICT analysis. DEPICT analysis was performed to identify tissues and gene sets enriched for genes in the loci identified from eGFR meta-analysis. Consistent with the role of the identified genes in kidney function, the most significant tissues (FDR < 0.01) identified by DEPICT were the urinary tract and kidney (Supplementary Data 8). Additional enriched tissues (FDR < 0.05) included the exocrine glands, liver, epithelial cells, prostate, kidney cortex, macrophages, membranes, and adrenal cortex. DEPICT analysis identified 482 significant gene sets (FDR < 0.05, Supplementary Data 9). The top gene sets (p-value < 3.46 × 10⁻⁶, 0.05/14462 gene sets) primarily included those associated with kidney morphology, the activity of transport channels, and with monosaccharide metabolic processes as shown in Fig. 1 (Supplementary Fig. 3).

Overlap of eGFR loci with related traits. As individuals with CKD often have coexisting heart disease or diabetes, we examined the identified eGFR variants for evidence of pleiotropic effects. A PheWAS analysis of the eGFR index variants across 23 cardiovascular and diabetes-related phenotypes in UK Biobank, excluding individuals with CKD, identified 7 phenotypes for which a subset of the index variants was also significant (p-value < 1.48 × 10⁻⁵, blood types (ABO, FUT2). Together, these results explain 7.6% of eGFR heritability, as calculated from LD score regression. We were not able to directly test these variants for replication of the eGFR associations since a similarly-sized cohort with eGFR measurements was not available. Instead, we tested for association of the index variants in kidney-related traits in the UK Biobank (CKD, hypertensive CKD, renal failure, acute renal failure, renal failure NOS, renal dialysis, or other disorders of kidney and ureters). Seven of the 48 novel variants (5 were lost due to poor imputation) and 27 of the 85 lead variants in known loci that were available in the UK Biobank were at least nominally associated and had corresponding direction of effect with one or more UK Biobank kidney-related phenotypes, providing initial support for the biological validity of the eGFR results (Supplementary Data 5, Supplementary Fig. 2). In addition, we compared the results from the current meta-analysis with previously reported eGFR index variants. Excluding the previously published datasets, 56 of the 118 available variants were at least nominally significant in a meta-analysis of HUNT and MGI alone (Supplementary Data 6).
Bonferroni correction for 23 phenotypes and 147 index variants): diabetes, coronary atherosclerosis, hypertension, essential hypertension, pulmonary heart disease, phlebitis and thrombophlebitis, and ischemic heart disease (Supplementary Data 5, 10, Fig. 2).

Colocalization analysis with these phenotypes identified 7 loci (prioritized genes: FGF5, PRKAG2, TRIB1, DCDCS/MPPED2, L2HGDH/SOS2, UMOD, SALL1) having significantly colocalized association signals with hypertension, essential hypertension, and/or coronary atherosclerosis and 1 locus (prioritized gene: GCKR) that colocalized with association of type 2 diabetes (Supplementary

Table 1 Lead variants for novel eGFR loci from meta-analysis

| Chr | Pos (hg19) | rsID | Ref | Alt | Freq | N  | P-value | Direction | Prioritized genes |
|-----|------------|------|-----|-----|------|----|---------|------------|------------------|
| 1   | 10733081   | rs284316 | T   | C   | 0.333 | 196273 | 1.50 × 10^{-9} | +         | CASZ1           |
| 100808369 | rs1166440 | A, G  | 0.419 | 350504 | 7.07 × 10^{-9} | -         | CDC14A        |
| 180905694 | rs3795503 | T   | C   | 0.601 | 350504 | 7.54 × 10^{-13} | -         | KIAA1614        |
| 227085824 | rs1800674 | A, G  | 0.565 | 350504 | 4.23 × 10^{-8} | +         | ADCK3         |
| 18679586 | rs10856778 | C, G  | 0.831 | 350504 | 1.04 × 10^{-9} | +         | LOC105373454   |
| 43441169 | rs3513692 | T, C  | 0.456 | 177995 | 2.74 × 10^{-11} | -         | THADA          |
| 54574942 | rs1405833 | C, G  | 0.273 | 350504 | 5.10 × 10^{-10} | -         | C2orf73        |
| 178146362 | rs17581255 | C, G  | 0.189 | 350504 | 6.11 × 10^{-11} | +         | AC0742861       |
| 191728341 | rs6725814 | A, G  | 0.261 | 350504 | 3.39 × 10^{-8} | +         | MFSD6          |
| 230612642 | rs6756038 | C, T  | 0.732 | 350504 | 4.72 × 10^{-9} | -         | TRIP1         |
| 38479475 | rs7429308 | T, C  | 0.495 | 350504 | 4.90 × 10^{-11} | -         | ACVR2B, XLYB   |
| 193816778 | rs10937314 | A, T  | 0.523 | 350504 | 2.81 × 10^{-9} | +         | LINCO2028       |
| 195477711 | rs2291652 | A, G  | 0.438 | 347321 | 3.36 × 10^{-8} | -         | SDHAP2, MUC20, RPT1-1141C7, 4, SDHAPI, MUC4, MIR570 |

*Reported frequency and direction of effect is with respect to the alternate allele for the combined meta-analysis. Study specific associations and allele frequencies are given in Supplementary Data 1. Gene names are italicized.
Six of the seven index variants within loci that showed significant colocalization with the cardiovascular traits were associated with essential hypertension and/or hypertension, underscoring the connection between high blood pressure and CKD. In addition, the index variants were examined for association with 1,400 traits phenome-wide (without exclusion of CKD cases). As shown in Fig. 2, the index variants are significantly associated (p-value < 5 × 10^{-8}) with additional traits including hypothyroidism and lipid metabolism disorders.

**Construction of genetic risk scores.** We developed genetic risk scores (GRS) from the meta-analysis results to assess the relationship between the identified variants and the likelihood of having CKD. These scores were then tested as predictors of CKD in a white British subset of UK Biobank. Several p-value and r² clumping thresholds were tested to select variants for inclusion in the GRS, but all yielded relatively similar predictions of CKD status (AUC range: 0.500–0.543). The best prediction was obtained using all independent markers (r² < 0.4) with p-value < 5 × 10^{-6} and the European subset of 1000 Genomes for LD clumping (Supplementary Fig. 4). The 1189 variants included in this risk score explain an estimated 25.3% of the variance in eGFR levels, while the GRS constructed using only the significantly associated independent variants (p-value < 5 × 10^{-8}, r² < 0.2) is estimated to explain 9.4% of the variance in eGFR. The GRS alone was associated with CKD (p-value = 1.8 × 10^{-15}, logistic regression), but it did not improve prediction of CKD status compared with birth year and sex alone (AUC: 0.700), or birth year, sex, and CKD clinical risk factors (diabetes, hypertension, and hyperlipidemia) (AUC: 0.865). Of all the tested models, inclusion of the GRS in addition to birth year, sex, and clinical risk factors provided the best predictor of CKD (AUC: 0.868). We also tested prediction of CKD using the best-performing GRS from the overall meta-analysis (without birth year or additional risk factors) separately in men and women. The GRS was slightly more predictive in women (AUC: 0.552) than in men (AUC: 0.538), possibly due to differing lifestyle or hormonal factors between the sexes influencing the development of CKD. In summary, these results show that the variants identified from association studies of eGFR are correlated with the presence of CKD on a population level (Supplementary Fig. 4). However, the results are not sufficient to identify individuals with CKD from those without. This is consistent with findings from prior studies examining GRS of eGFR.

**Sex-specific analysis in HUNT.** We aimed to determine whether any of the eGFR index variants showed sex-specific association as there are known differences in the prevalence of CKD between men and women (44). Association tests of eGFR in HUNT stratified by sex identified one locus (Fig. 3, Supplementary Fig. 5, Supplementary Data 5). Of the seven index variants in loci that showed significant colocalization with the cardiovascular traits, one locus was associated with CKD (AUC range: 0.500–0.543). We aimed to determine whether any of the eGFR index variants showed sex-specific association as there are known differences in the prevalence of CKD between men and women. Association tests of eGFR in HUNT stratified by sex (Fig. 3, Supplementary Fig. 5, Supplementary Data 5). Of the seven index variants in loci that showed significant colocalization with cardiovascular traits, one locus was associated with CKD (AUC range: 0.500–0.543).
Data 11) that was significantly associated with eGFR in women but not in men. Interaction tests on the unrelated subset of individuals (N = 26,235, 37.7%) in HUNT confirmed a significant sex interaction for this variant (p-value = 1.4 × 10^{-5}, p-value < 0.05/147). We obtained the summary statistics from the CKDGen consortium sex-stratified eGFR analysis to test this variant for replication13. Within the CKDGen results, the proxy variant of rs2440164, (rs2453580, r^2 = 0.98) showed greater significance and effect in women than in men (p-value_{women} = 3.12 × 10^{-5}, effect_{women} = −0.0066, p-value_{men} = 0.014, effect_{men} = −0.0055).

**Discussion**

In summary, analysis of the HUNT and MGI biobanks and meta-analysis of eGFR across more than 350,000 individuals identified 147 loci, of which 53 were novel. Novel eGFR index variants were common with relatively small effect sizes. Despite the small effect sizes of individual variants, the identified loci give new insight into the genes underlying kidney function and the development of CKD. In support of this, many of the prioritized genes cluster into known kidney associated pathways. For example, Wnt signaling has been implicated in kidney development and disease25. Variants in DCD2 were associated with eGFR levels. Knockdown or overexpression of DCD2 is known to alter β-catenin activation of TCF transcription factors30, thereby altering Wnt signaling. Likewise, variants in a gene associated with the epidermal growth factor receptor (ErbB) family were also observed. ErbB receptors are involved in kidney development31, control of solute levels (e.g., Ca, Na)32,33, and play a role in hypertension34,35. Our meta-analysis results identified variants in MUC4 associated with eGFR levels. The beta chain of Mucin-4 (MUC4) interacts with ErbB36. Lastly, we identified variants associated with both decreased eGFR and increased tubulointerstitial kidney CDKL5 expression. CDKL5 overexpression has been shown to impair ciliogenesis37. Defects in cilia are known to cause polycystic kidney disease and nephronophthisis, among other disorders38. In addition, prioritized genes included those known to cause Mendelian kidney disease35: ALMS1, CNNM2, CYP24A1, CACNA1S, DACH1, DCD2, GNAS, LRP2, MUC1, RPS10, SALL1, SCARB2, SDCCAG8, SHH, SLC34A1, SLC7A9, SMAD3, and UMOD. These clues provide an initial link to how these identified genetic regions may lead to changes in kidney function.

Recent single-cell transcriptomic studies have classified kidney cell types in mice based on differential expression of specific genes relative to other cell types39. Interestingly, several of the prioritized genes from the present study exhibited cell-type specific expression in mouse kidney: CDC41A, DACH1, and VEGFA in podocytes, CGNL1, IRX1, PPPIR1B, and UMOD in the loop of Henle, LRP2, NAT8, SLC34A1, SLC4A1, and XYL2 in the proximal tubule, RASIP1 in the endothelial, vascular, and descending loop of Henle, and STC1 in collecting duct principal cells. These findings may help to identify kidney cell types whose function is affected by the genetic variants found in the eGFR GWAS. Recent studies have further examined the role of gene expression in eGFR and CKD. Xu et al., performed Mendelian randomization analysis using gene expression data to identify a causal role for MUC1 expression on eGFR40. Single nucleus RNA-sequencing using cells from a human kidney donor identified expression of DPEP1 that was specific to the proximal tubule41. Moreover, glomerular and tubular specific gene expression associations have been found to be significantly enriched for CKD and eGFR GWAS results42, emphasizing the need to consider eQTLs in kidney tissue when prioritizing genes from kidney-related GWAS.

Experimental evidence also supports hormonal regulation of SLC4A1 expression, the gene prioritized from the sex-stratified analysis of eGFR. SLC4A1 is also known as MATE1 (multidrug toxin and extrusion protein 1). Experimental studies of MATE1 identified higher levels of expression in the kidneys of 30–45-day old male mice compared to female43. Furthermore, He et al., found that kidney expression of MATE1 could be modified by treatment with testosterone or estradiol, compared to olive oil as a control44.

It is also interesting to consider the interplay between kidney function and other related traits based on the overlap between identified genetic regions. For example, the GCKR gene was prioritized based on eGFR meta-analysis results. GCKR encodes glucokinase regulatory protein, which regulates glucose metabolism, and has been previously associated with the development of diabetes45. The ABO gene, responsible for determination of an individual’s ABO blood type, was also prioritized based on eGFR meta-analysis results. Associations near this gene have been previously identified for other phenotypes, including LDL and total cholesterol46, coronary artery disease47, and type 2 diabetes48. As diabetes is a significant risk factor for the development of CKD, these shared associations may help to identify potential common mechanisms. Comparison of association results with cardiovascular disease-related traits also identified shared

---

**Fig. 3** LocusZoom plots of region showing differential association between sexes. eGFR meta-analysis results in HUNT stratified by sex were filtered to identify regions significant in one sex (P < 5 × 10^{-8}) but not significant (P > 0.05) in the other. Within each panel, we show regional results for eGFR association for variants near rs2440165, which is an eQTL for SLC4A1, in women (a) and men (b).
associations with hypertension, the second major risk factor for CKD. Six of the 147 loci identified from meta-analysis showed significant colocalization with hypertension, which may help to identify additional shared pathways between high blood pressure and kidney function.

While additional studies are needed to understand eGFR associations that are specific to disease subtypes, the present results build upon the previous studies to increase the number of eGFR associated loci and identify pleiotropic associations with cardiovascular disease. Limitations in the present study include of eGFR associated loci and identify pleiotropic associations with kidney transplant.

Methods

Description of cohorts. The HUNT study49 is a longitudinal, repetitive population-based health survey conducted in the county of Nord-Trøndelag, Norway in which kidney-related phenotypes have not previously been tested for genetic association. Since 1984, the adult population in the county has been examined three times, through HUNT1 (1984–86), HUNT2 (1995–97), and HUNT3 (2006–08). A fourth survey, HUNT4 (2017–2019), is ongoing. HUNT was approved by the Data Inspectorate and the Regional Ethics Committee for Medical Research in Norway. All HUNT participants gave informed consent. Approximately 120,000 individuals have participated in HUNT1–HUNT3 with extensive phenotypic measurements and biological samples. A subset of these participants have been genotyped (~70,000) using Illumina HumanCoreExome v1.0 and 1.1 and imputed with Minimac3 using a combined HRC and HUNT-specific WGS reference panel. Variants with imputation r2 > 0.3 were excluded from further analysis.

We analyzed available kidney-related phenotypes within the HUNT study, including creatinine (N = 69,591), eGFR (N = 69,591), urea (N = 20,700), and CKD (Ncases = 2044 and Ncontrols = 65,575). eGFR values were calculated using the MDRD equation65. We also calculated eGFR using the CKD-EPI equation and after adjustment for covariates including age, sex, and batch followed by inverse normal transformation, the resultant eGFR phenotype values were highly correlated with those derived in the same manner after eGFR was calculated using MDRD (Supplementary Fig. 6, r2 = 0.995). CKD status was derived from ICD-9 codes 583 and 586 and ICD-10 code N18. Association testing (24,961,484 variants) of quantitative traits was performed using BOLT-LMM22, v2.2 on the inverse normal transformed residuals of the traits adjusted for genotyping batch, sex, 4 principle components, and age (Supplementary Fig. 7). Association testing for CKD was performed using SAIGE with batch, sex, 4 principle components, and birth year as covariates. Associations stratified by sex for eGFR were also performed. For the stratified analyses, phenotypes were separately inverse-normalized and were adjusted for batch, age, and 4 principle components. Linkage disequilibrium within HUNT was calculated using PLINK v1.906. To identify independent variants, conditional analysis for eGFR was performed within the HUNT dataset using BOLT-LMM v2.3.1, conditioning on the lead variant within the identified loci until no variants with MAF > 0.5% had p-value < 5 × 10−8. To identify eGFR index variants, eGFR index variants were examined separately in women and men and were filtered to those that were significant in one sex but not significant (p-value > 0.05) in the other. Differences in effect sizes between males and females were tested using Z = (βmale−βfemale)/(SEβmale+SEβfemale−2×SEβmale×SEβfemale)/r2, where r is the Pearson correlation for male and female effect size across all variants44. Interaction tests for these variants were performed on an unrelated subset of HUNT participants (N = 26,235) in PLINK v1.9 with sex, age, batch, and 4 principle components as covariates and a sex interaction term. Significance between sexes was determined using Bonferroni correction for the number of tested loci. HUNT-specific results are provided in Supplementary Tables 3 and 4 in Supplementary Data 12.

BioBank Japan (BJI) is a registry of patients from 12 medical centers across Japan who are diagnosed with at least one of 47 common diseases. Summary statistics for 58 quantitative traits, including eGFR, are publicly available16.

Participating individuals were genotyped with either the Illumina HumanCoreExome BeadChip or HumanExome BeadChips. Imputation was performed with Minimac using the East Asian reference panel from 1000 Genomes phase 1β6. Variants with imputation r2 < 0.7 were excluded. BBJ eGFR values were calculated using the Japanese ancestry modified version of the CKD-EPI equation7 and were available on 143,658 of BBJ individuals. With BBJ eGFR values of less than 15 ml/min/1.73 m2 were excluded from the analysis. Values were standardized using rank-based inverse normalization. Association analysis (6,108,953 variants) was performed using mach2qtl with sex, age, the top 10 principle components, and disease status of all studied diseases (N = 47) included as covariates. eGFR was estimated from creatinine levels using the MDRD equation50,51. Variants with imputation quality ≤ 0.4, and those found in less than half of individuals were excluded from further analysis. Meta-analysis (10,154,908 variants) was performed using the inverse-variance method in METAL23, Pre and post-meta-analysis genetic control (GC) correction was performed.

The Michigan Genomics Initiative (MGI) is a repository of electronic medical record and genetic data at Michigan Medicine with 7,398,298 participants. In total, 21,089,590 cases and controls were identified. MGI participants are enrolled during pre-surgical encounters at Michigan Medicine and consent to linking of genetic and clinical data from electronic medical records. MGI participants were genotyped using Illumina Infinium CoreExome-24 bead arrays. Genotype data was then imputed to the Haploype Reference Consortium using the Michigan Imputation Server, providing 17 million imputed variants after standard quality control and filtering. Using European and East Asian subsets of 1000 Genomes to construct a reference panel from 1000 Genomes phase I reference panel. Variants with imputation r2 < 0.8 were included. LD information from the 1000 Genomes phase I reference panel8 was used to construct loci for association testing. Association testing was performed using the inverse-variance method in METAL23. Pre and post-meta-analysis genetic control (GC) correction was performed.

Meta-analysis. Meta-analysis was performed using the p-value based approach in METAL23. This approach was chosen to account for differing units between the effect sizes of the CKDGen (log-transformed) and MGI/BBJ/HUNT (inverse-normalized) summary statistics. This approach was validated by comparison to traditional standard error-based meta-analysis of the cohorts with available inverse-normalized summary statistics; the results showed an extremely high correlation of p-values (Pearson r = 0.966, Supplementary Fig. 8). Summary statistics from contributing studies were GC corrected prior to meta-analysis and were not filtered by minor allele frequency or sample size. Lead index variants were determined as the most significant variant in 1 Mb windows that were found in at least 2 studies. Adjacent windows were merged if the LD r2 between lead variants was ≥ 0.2. Identified variants were considered to be novel if the most significant variant was more than 1 Mb away from previously reported lead variants. Linkage disequilibrium between variants was calculated using LDLink60 or PLINK with the European and East Asian subsets of 1000 Genomes Phase III reference panel56. Score intercepts and heritability were calculated using LDSC version 1.0.21 with the European 1000 Genomes reference LD Scores using variants with minor allele frequency greater than 1%.

Variant and gene annotation. Variants were annotated using WGSAS18 and dbSNP196. Annotation of variants with associated biological processes was performed using the UniProt10 and NCBI gene database18. Genes for identified loci were prioritized based on the consensus between significantly colocalized eQTLs, missense variants within 1 Mb and in LD (r2 > 0.8) with the lead variant, and the gene prioritized by Data-driven Expression-Prioritized Integration for Complex Traits (DEPICT)55. In cases where there was no gene identified from the different annotation methods, the gene was prioritized as the nearest gene to the lead variants. Analysis was performed using the DEPICT 1.1 1000 Genomes version. Variants from meta-analysis that were found in two or more studies with p-value < 5 × 10−8 were included. LD information from the European and East Asian subsets of 1000 Genomes was used to construct loci within DEPICT. DEPICT results with FDR < 0.05 were considered significant. Gene sets with more than 25% overlap were collapsed into a single set for construction of the network diagram, as previously done66.

eQTL analysis. Publicly available eQTL association datasets from GTEX V724, NephiQTL25, and Ko et al22 were each used separately to identify overlap between gene expression and identified eQFR association results. Specific tissue types,
Analysis was performed using the R package coloc67. Priors for p1, p2, and p12 in other tissues available (from GTEx, Supplementary Fig. 9). Colocalization associated with expression of a given gene in any of the kidney-specific tissues was considered to be kidney-specific if an index variant was significantly associated with expression of a given gene in any of the kidney-specific datasets but not in other tissues available (from GTEx, Supplementary Fig. 9). Colocalization analysis was performed using the R package coloc67. Priors for p1, p2, and p12 within the coloc analysis were set to 1 × 10^-4, 1 × 10^-4, and 1 × 10^-6, respectively. Variants in the ±500 kb region surrounding each eGRF index variant were used for input into coloc. Within this region, we required at least one genome-wide significant (p-value < 5 × 10^-8) eQTL variant prior to testing for colocalization. Following the criteria published by Giambartolomei et al.67, eQTLs were considered to colocalize with the eGRF association results if the posterior probability (PP) for a shared variant was > 80%.

Determination of nearby Mendelian kidney disease genes. Genes associated with Mendelian forms of kidney disease were taken from those identified by Groopman et al.25 and from genes included in the KidneySeq v3 testing panel (Iowa Institute of Human Genetics). Nearby genes were identified as all genes having any overlap with the ±250 kb region surrounding the index variants. Gene start and end positions were taken from the HAVANA gene annotations from the GENCODE consortium68.

Comparison with related traits. Association results for phenome-wide lookups were taken from the publicly available analysis of UK Biobank69 using SAIGE17, which accounts for relatedness and population stratification by using a kinship matrix [http://phweb.sph.umich.edu/50031/]. Phenotypes were grouped for analysis based on ICD-9 and 10 codes into pedigrees following a similar strategy used by other groups for phenome-wide studies11,21. Additionally, we re-analyzed cardiovascular and diabetics traits in the white British subset of UK Biobank excluding CKD cases as has been previously suggested for identifying pleiotropic effects80. We performed analysis using individuals in the white British subset of UK Biobank that were included in the kinship calculation, excluding those identified as outliers based on the missingness rate and heterozygosity, and those missing from the UK Biobank phasing calculations. At the genotype level, we included all variants used for calculation of the kinship matrix and excluded variants after imputation with INFO score < 0.3 and variants not in the HRC imputation panel. Association testing was performed using SAIGE with sex, birth year, and 4 PCs as covariates. Sample sizes for all related traits are given in Supplementary Data 10. Colocalization analysis was performed using the R package coloc, with the priors for p1, p2, and p12 set to 1 × 10^-4, 1 × 10^-4, and 1 × 10^-6, respectively. Genetic regions for colocalization testing were defined as the most significant variant for eGRF in each locus ±500 kb. Within this region, we required at least one genome-wide significant (p-value < 5 × 10^-8) variant for each trait prior to testing for colocalization. Variants were considered to colocalize if the probability for a common variant was greater than 80%.

Genetic risk scores. Variants were selected for inclusion in the genetic risk score (GRS) using the clumping procedure in PLINK with varying r2 thresholds of 0.2, 0.4, 0.6, and 0.8 and p-value thresholds of 5 × 10^-8, 5 × 10^-6, 5 × 10^-4, and 5 × 10^-3 on the meta-analysis results from all cohorts. Because the meta-analysis included both European and East Asian samples, but the validation set included primarily European samples, we separately constructed the GRS using either the European-only or European and East Asian subsets of 1000 Genomes Phase 3 for clumping in PLINK. Effect sizes were estimated from meta-analysis of the BioBank Japan, MGI, and HUNT results only due to the differing effect size units of the CKDGen consortium results. The proportion of variance explained by the GRS was calculated as the sum of the squared effect size of all variants included in the risk score, where β and f are the effect size and frequency from meta-analysis of BioBank Japan, MGI, and HUNT. GRS were then calculated within UK Biobank as the sum of risk alleles across all variants included in the risk score, where

\[ \beta_f = \sum_{i=1}^{n} \beta_i f_i \]

and

\[ \sum_{i=1}^{n} f_i = 1 \]

where \( \beta_i \) is the effect size of variant \( i \) and \( f_i \) is its frequency. The proportion of variance explained by the GRS was calculated as

\[ \frac{\sum_{i=1}^{n} \beta_i f_i^2}{\sum_{i=1}^{n} f_i} \]

where \( \beta_i \) is the effect size of variant \( i \) and \( f_i \) is its frequency.

Data availability. Data generated during analysis is available from the corresponding author upon reasonable request. Meta-analysis eGRF summary statistics are available here: http://csg.sph.umich.edu/willer/public/eGRF2018/.

References

1. United States Renal Data System. 2017 USRDS annual data report: epidemiology of kidney disease in the United States. (National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2017).
2. Hallan, S. I. et al. Long-term trends in the prevalence of chronic kidney disease and the influence of cardiovascular risk factors in Norway. *Kidney Int. 90*, 665–673 (2016).
3. Yos, T. et al. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 386, 743–800 (2015).
4. National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Am. J. Kidney Dis. 39*, 51–266 (2002).
5. Johnson, C. A. et al. Clinical practice guidelines for chronic kidney disease in adults: Part I. Definition, disease stages, evaluation, treatment, and risk factors. *Am. Fam. Physician* 70, 869–876 (2004).
6. Go, A. S., Chertow, G. M., Fan, D., McCulloch, C. E. & Hsu, C. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N. Engl. J. Med.* 351, 1296–1304 (2004).
7. Fox, C. S. et al. Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: the Framingham Heart Study. *J. Am. Soc. Neph.* 15, 2457–2461 (2004).
8. Gorksi, M. et al. 1000 genomes-based meta-analysis identifies 10 novel loci for kidney function. *Sci. Rep.* 7, 45040 (2017).
9. Li, M. et al. SOS2 and ACPI loci identified through large-scale exome chip analysis regulate kidney development and function. *J. Am. Soc. Neph.* 28, 981–994 (2017).
10. Pattaro, C. et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.* 7, 10023 (2016).
11. Kottgen, A. et al. Multiple loci associated with indices of renal function and chronic kidney disease. *Nat. Genet.* 41, 712–717 (2009).
12. Kottgen, A. et al. New loci associated with kidney function and chronic kidney disease. *Nat. Genet.* 42, 376–384 (2010).
13. Pattaro, C. et al. Genomewide association and functional follow-up reveals new loci for kidney function. *PloS Genet.* 8, e1002584 (2012).
14. Okada, Y. et al. Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat. Genet.* 44, 904–909 (2012).
15. Sveinbjörnsson, G. et al. Rare mutations associating with serum creatinine and chronic kidney disease. *Hum. Mol. Genet.* 23, 6935–6943 (2014).
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. Chambers, J. C. et al. Genetic loci influencing kidney function and chronic kidney disease. *Nat. Genet.* 42, 373 (2010).
18. Mahajan, A. et al. Trans-ethnic fine mapping highlights kidney-function genes linked to salt sensitivity. *Am. J. Hum. Genet.* 99, 636–646 (2016).
19. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* 48, 1279 (2016).
20. Iglesias, A. I. et al. Halepoty reference consortium panel: practical implications of imputations with large reference panels. *Hum. Mutat.* 38, 1025–1032 (2017).
21. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47, 291 (2015).
22. Ko, Y. A. et al. Genetic-variation-driven gene-expression changes highlight genes with important functions for kidney disease. *Am. J. Hum. Genet.* 100, 940–953 (2017).
23. Gillies, C. E. et al. An eQTL landscape of kidney tissue in human nephritic syndrome. *Am. J. Hum. Genet.* 103, 232–244 (2018).
24. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* 550, 204 (2017).
25. Groopman, E. E. et al. Diagnostic utility of exome sequencing for kidney disease. *N. Engl. J. Med.* 380, 142–151 (2019).
26. Kummer, S., von Gersdorff, G., Kemper, M. J. & Oh, J. The influence of gender and sexual hormones on incidence and outcome of chronic kidney disease. *Pediatr. Nephrol.* 27, 1213–1219 (2012).
27. Ma, J., Yang, Q., Hwang, S. J., Fox, C. S. & Chu, A. Y. Genetic risk score and risk of stage 3 chronic kidney disease. *RMC Nephrol.* 18, 32 (2017).
M.Z., W.Z., L.G.F., I.S. and D.F. performed the statistical analysis of association data. G.R.A., S.L., H.M.K. and C.J.W. provided statistical oversight. J.B.N., M.E.G., A.H.S., W.E.H., S.K., C.M.B., K.H. and C.J.W. applied for and maintained regulatory (IRB) and other approvals. C.M.B., G.R.A. and K.H. are Principal Investigators of cohorts.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-09861-z.

Competing interests: G.R.A. is an employee of Regeneron Pharmaceuticals. The remaining authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available

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