Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies

Haojia Wu
Washington University School of Medicine in St. Louis
Benjamin D. Humphreys
Washington University School of Medicine in St. Louis
Andrew C. Heath
Washington University School of Medicine in St. Louis
Pamela A. F. Madden
Washington University School of Medicine in St. Louis

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation
https://digitalcommons.wustl.edu/open_access_pubs/7632

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies

Andrew P. Morris et al. #

Chronic kidney disease (CKD) affects ~10% of the global population, with considerable ethnic differences in prevalence and aetiology. We assemble genome-wide association studies of estimated glomerular filtration rate (eGFR), a measure of kidney function that defines CKD, in 312,468 individuals of diverse ancestry. We identify 127 distinct association signals with homogeneous effects on eGFR across ancestries and enrichment in genomic annotations including kidney-specific histone modifications. Fine-mapping reveals 40 high-confidence variants driving eGFR associations and highlights putative causal genes with cell-type specific expression in glomerulus, and in proximal and distal nephron. Mendelian randomisation supports causal effects of eGFR on overall and cause-specific CKD, kidney stone formation, diastolic blood pressure and hypertension. These results define novel molecular mechanisms and putative causal genes for eGFR, offering insight into clinical outcomes and routes to CKD treatment development.
Chronic kidney disease (CKD) affects ~10% of the global population, with considerable racial/ethnic differences in prevalence and risk factors. CKD is associated with premature cardiovascular disease and mortality, and has enormous healthcare costs for treatment, prescriptions and hospitalizations. The underlying mechanisms for CKD predisposition and development are unknown, limiting progress in the identification of prognostic biomarkers or the advancement of treatment interventions.

Large-scale genome-wide association studies (GWAS) of estimated glomerular filtration rate (eGFR), a measure of kidney function used to define CKD, have mostly been undertaken in populations of European and East Asian ancestry. Despite the success of these GWAS in identifying loci contributing to kidney function and risk of CKD, the common single nucleotide variants (SNVs) driving the association signals explain no more than ~4% of the observed-scale heritability of eGFR, and efforts to replicate these findings in other ancestry groups have been limited. Furthermore, efforts to localise the variants driving eGFR association signals at these loci, and the putative causal genes through which their effects are mediated, have been hampered by the extensive linkage disequilibrium (LD) across common variation in European and East Asian ancestry populations.

To enhance understanding of the genetic contribution to kidney function and CKD across diverse populations, and to inform global public health and personalised medicine, we recently established the Continental Origins and Genetic Epidemiology Network Kidney (COGENT-Kidney) Consortium. We undertook trans-ethnic meta-analysis of eGFR GWAS in 71,638 individuals ascertained from populations of African, East Asian, European and Hispanic/Latino ancestry. These investigations provided no evidence of heterogeneity in allelic effects on eGFR association signals between ancestry groups, emphasizing the power of trans-ethnic GWAS meta-analysis for locus discovery that will be relevant to diverse populations.

To further extend characterization of the genetic contribution to eGFR, and determine the molecular mechanisms and putative causal genes through which association signals impact on kidney function, we expand the COGENT-Kidney Consortium in this investigation by assembling GWAS in up to 312,468 individuals of diverse ancestry. With these data, we identify novel loci and distinct associations for kidney function, assess the evidence for heterogeneity in their allelic effects on eGFR, and determine genomic annotations in which these signals are enriched. We identify high-confidence variants driving eGFR association signals through annotation-informed trans-ethnic fine-mapping, and highlight putative causal genes through which their effects are mediated via integration with expression in kidney tissue. Finally, we evaluate the causal effects of eGFR on clinically-relevant renal and cardiovascular outcomes through Mendelian randomisation (MR) with our expanded catalogue of kidney function loci.

Results
Study overview. We assembled GWAS in up to 312,468 individuals from three sources (Methods): (i) 19 studies of diverse ancestry from the COGENT-Kidney Consortium, expanding the previously published trans-ethnic meta-analysis to include additional individuals of Hispanic/Latino descent; (ii) a published meta-analysis of 33 studies of European ancestry from the CKDGen Consortium; and (iii) a published study of East Asian ancestry from the Biobank Japan Project. Each GWAS was imputed up to the Phase 1 integrated 1000 Genomes Project reference panel, and SNVs passing quality control were tested for association with eGFR, calculated from serum creatinine, accounting for age, sex and ethnicity, as appropriate (Methods).

The current study represented a 2.2-fold increase in sample size over the largest published GWAS of kidney function. Assuming homogeneous allelic effects on eGFR across populations, we had more than 80% power to detect an association (p < 5 × 10^{-8}) with SNVs explaining at least 0.0127% of the trait variance under an additive genetic model. This corresponded to common/low-frequency SNVs with minor allele frequency (MAF) ≥5%/≥0.5% that decrease eGFR by ≥0.0366/≥0.113 standard deviations.

Trans-ethnic meta-analysis. To discover novel loci contributing to kidney function in diverse populations, we first aggregated eGFR association summary statistics across studies through trans-ethnic meta-analysis (Methods). We employed Stouffer’s method, implemented in METAL, because allelic effect sizes were reported on different scales in each of the three sources contributing to the meta-analysis. We identified 93 loci attaining genome-wide significant evidence of association with eGFR (p < 5 × 10^{-8}), including 20 mapping outside regions previously implicated in kidney function (Supplementary Figure 1, Supplementary Table 1). The strongest novel associations (Table 1) mapped to/near MYPN (rs7475348, p = 8.6 × 10^{-19}), SHH (rs6971211, p = 6.5 × 10^{-13}), XLYL (rs36070911, p = 2.3 × 10^{-11}) and ORCA (rs13026220, p = 3.1 × 10^{-11}).

Across the 93 loci, we then delineated 127 distinct association signals (at locus-wide significance, p < 5 × 10^{-5}) through approximate conditional analyses implemented in GCTA (Methods), each arising from different underlying causal variants and/or haplotype effects (Supplementary Tables 1 and 2). The most complex genetic architecture was observed at SLC22A2 and UMD-PIDLT, where the eGFR association was delineated to four distinct signals at each locus (Supplementary Figure 2). Genome-wide, application of LD Score regression to a meta-analysis of only European ancestry studies revealed the observed scale heritability of eGFR to be 7.6%, of which 44.7%/5.4% was attributable to variability in the known/novel loci reported here (Methods).

Trans-ethnic heterogeneity in eGFR association signals. To assess the evidence for a genetic contribution to ethnic differences in CKD prevalence, we investigated differences in eGFR associations across the diverse populations contributing to our meta-analysis. We performed trans-ethnic meta-regression of allelic effect sizes obtained from GWAS contributing to the COGENT-Kidney Consortium, implemented in MR-MEGA, including two axes of genetic variation that separate population groups as covariates to account for heterogeneity that is correlated with ancestry (Methods, Supplementary Figure 3). Despite substantial differences in allele frequencies at index SNVs for the distinct associations across ethnicities, we observed no significant evidence (p < 0.00039, Bonferroni correction for 127 signals) of heterogeneity in allelic effects on eGFR that was correlated with ancestry (Supplementary Tables 2 and 3). Furthermore, all index SNVs had minor allele frequencies >1% in multiple ethnic groups, indicating that the distinct eGFR association signals were not ancestry-specific. These observations are consistent with a model in which causal variants for eGFR as a measure of kidney function are shared across global populations and arose prior to human population migration out of Africa.

Enrichment of eGFR associations for genomic annotations. To gain insight into the molecular mechanisms that underlie the genetic contribution to kidney function, we investigated genomic
signatures of functional and regulatory annotation that were enriched for eGFR associations across the 127 distinct signals. Specifically, we compared the odds of eGFR association for SNVs mapping to each annotation with those that did not map to the annotation (Methods). We began by considering genic regions, as defined by the GENCODE Project18, and observed significant enrichment (p < 0.05) of eGFR associations in protein-coding exons (p = 0.0049), but not in 3’ or 5’ UTRs. We then interrogated chromatin immuno-precipitation sequence (ChIP-seq) binding sites for 161 transcription factors from the ENCODE Project19, which revealed significant joint enrichment of eGFR associations for HDAC2 (p = 0.0088) and EZH2 (p = 0.030). Class I histone deacetylases (including HDAC2) are required for embryonic kidney gene expression, growth and differentiation20, whilst EZH2 participates in histone methylation and transcriptional repression21. Finally, we considered ten groups of cell-type-specific regulatory annotations for histone modifications (H3K4me1, H3K4me3, H3K9ac and H3K27ac)22,23. Significant enrichment of eGFR associations was observed only for kidney-specific annotations (p = 7.4 × 10⁻¹⁴). In a joint model of these four enriched annotations, the odds of eGFR association for SNVs mapping to protein-coding exons, binding sites for HDAC2 and EZH2, and kidney-specific histone modifications were increased by 3.06-, 2.13-, 1.76- and 4.29-fold, respectively (Supplementary Figure 4).

Annotation-informed trans-ethnic fine-mapping. We performed trans-ethnic fine-mapping to localise putative causal variants for distinct eGFR association signals that were shared across global populations by taking advantage of differences in the structure of LD between ancestry groups24. To further enhance fine-mapping resolution, we incorporated an annotation-informed prior model for causality, upweighting SNVs mapping to the globally enriched genomic signatures of eGFR associations (Methods). Under this prior, we derived credible sets of variants for each distinct signal, which together account for 99% of the posterior probability (π) of driving the eGFR association (Supplementary Table 4). For 40 signals, a single SNV accounted for more than 50% of the posterior probability of driving the eGFR association, which we defined as high-confidence for causality (Supplementary Table 5). We assessed the evidence of association of these high-confidence SNVs with other measures of kidney function and damage in published GWAS9,10,25 (Supplementary Table 6). Several SNVs demonstrated nominal associations (p < 0.05) with eGFR calculated from cystatin C, blood urea nitrogen and urine albumin creatinine ratio, with the expected direction of effect of the eGFR decreasing allele.

Putative causal genes at eGFR association signals. We sought to identify the most likely target gene(s) through which the effects of each of the 40 high-confidence SNVs on eGFR were mediated via functional annotation and colocalisation with expression quantitative trait loci (eQTLs) in kidney tissue.

Only four of the SNVs were missense variants (Table 2), encoding CACNA1S p.Arg1539Cys (rs3850625, p = 2.5 × 10⁻⁹, π = 99.0%), CPS1 p.Thr1406Asn (rs1047891, p = 1.5 × 10⁻¹⁰, π = 98.1%), GCKR p.Leu446Pro (rs1260326, p = 2.0 × 10⁻³⁵, π = 86.1%) and CERS2 p.Glu115Ala (rs267738, p = 1.7 × 10⁻¹⁰, π = 55.3%). Functional annotation of these high-confidence missense variants highlighted predicted deleterious impact of CPS1 p.Thr1406Asn and CERS2 p.Glu115Ala (Methods). CACNA1S (Calcium Voltage-Gated Channel Subunit Alpha 1s) encodes a subunit of L-type calcium channel located within the glomerular afferent arteriole, is the target of anti-hypertensive dihydropyridine calcium channel blockers (such as amlodipine and nifedipine), and regulates arteriolar tone and intra-glomerular pressure26. CACNA1S missense mutations cause

<table>
<thead>
<tr>
<th>Locus</th>
<th>Lead SNV</th>
<th>Chr</th>
<th>Position (bp, b37)</th>
<th>Alleles</th>
<th>EAF</th>
<th>Fixed-effects meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Effectb</td>
<td>Other</td>
<td>p-value</td>
</tr>
<tr>
<td>PMF1-BGLAP</td>
<td>rs2842870</td>
<td>1</td>
<td>156,200,671</td>
<td>C</td>
<td>0.632</td>
<td>1.2 × 10⁻⁸</td>
</tr>
<tr>
<td>NTSC1B-RH14</td>
<td>rs13417750</td>
<td>2</td>
<td>18,681,365</td>
<td>G</td>
<td>0.189</td>
<td>1.0 × 10⁻⁸</td>
</tr>
<tr>
<td>C2orf73</td>
<td>rs1527649</td>
<td>2</td>
<td>54,581,356</td>
<td>T</td>
<td>0.234</td>
<td>1.5 × 10⁻⁹</td>
</tr>
<tr>
<td>ORC4</td>
<td>rs13026220</td>
<td>2</td>
<td>148,586,459</td>
<td>A</td>
<td>0.366</td>
<td>3.1 × 10⁻¹¹</td>
</tr>
<tr>
<td>NFE2L2</td>
<td>rs39595110</td>
<td>2</td>
<td>178,143,371</td>
<td>T</td>
<td>0.435</td>
<td>3.9 × 10⁻⁹</td>
</tr>
<tr>
<td>XYL8</td>
<td>rs36070911</td>
<td>3</td>
<td>38,498,439</td>
<td>G</td>
<td>0.528</td>
<td>2.3 × 10⁻¹¹</td>
</tr>
<tr>
<td>AK125311</td>
<td>rs856563</td>
<td>7</td>
<td>46,723,510</td>
<td>T</td>
<td>0.750</td>
<td>5.1 × 10⁻¹⁰</td>
</tr>
<tr>
<td>SHH</td>
<td>rs6971211</td>
<td>7</td>
<td>155,664,686</td>
<td>C</td>
<td>0.417</td>
<td>6.5 × 10⁻¹³</td>
</tr>
<tr>
<td>NRG1</td>
<td>rs4489283</td>
<td>8</td>
<td>32,399,662</td>
<td>C</td>
<td>0.296</td>
<td>1.5 × 10⁻⁸</td>
</tr>
<tr>
<td>TRIB1</td>
<td>rs2001945</td>
<td>8</td>
<td>126,477,978</td>
<td>T</td>
<td>0.546</td>
<td>1.6 × 10⁻⁹</td>
</tr>
<tr>
<td>DCAF12</td>
<td>rs6127993</td>
<td>9</td>
<td>34,130,435</td>
<td>G</td>
<td>0.666</td>
<td>4.0 × 10⁻⁸</td>
</tr>
<tr>
<td>MYPN</td>
<td>rs7475348</td>
<td>10</td>
<td>69,965,177</td>
<td>C</td>
<td>0.607</td>
<td>8.6 × 10⁻¹⁰</td>
</tr>
<tr>
<td>CYP26A1</td>
<td>rs4418728</td>
<td>10</td>
<td>94,839,724</td>
<td>T</td>
<td>0.539</td>
<td>1.4 × 10⁻⁸</td>
</tr>
<tr>
<td>FAM53B</td>
<td>rs4962691</td>
<td>10</td>
<td>126,424,137</td>
<td>T</td>
<td>0.571</td>
<td>5.0 × 10⁻¹⁰</td>
</tr>
<tr>
<td>RASSGRP1</td>
<td>rs9920185</td>
<td>15</td>
<td>39,273,575</td>
<td>C</td>
<td>0.649</td>
<td>1.0 × 10⁻⁸</td>
</tr>
<tr>
<td>NFA5</td>
<td>rs11641050</td>
<td>16</td>
<td>69,622,104</td>
<td>C</td>
<td>0.697</td>
<td>2.6 × 10⁻⁸</td>
</tr>
<tr>
<td>JUND-LSM4</td>
<td>rs8108623</td>
<td>19</td>
<td>18,408,519</td>
<td>C</td>
<td>0.695</td>
<td>4.4 × 10⁻⁸</td>
</tr>
<tr>
<td>ARFRP1</td>
<td>rs1758206</td>
<td>20</td>
<td>63,322,336</td>
<td>C</td>
<td>0.822</td>
<td>2.4 × 10⁻⁸</td>
</tr>
<tr>
<td>NRIP1</td>
<td>rs2823139</td>
<td>21</td>
<td>156,767,783</td>
<td>T</td>
<td>0.293</td>
<td>3.7 × 10⁻⁹</td>
</tr>
<tr>
<td>ATP50</td>
<td>rs2834317</td>
<td>21</td>
<td>35,356,706</td>
<td>A</td>
<td>0.108</td>
<td>9.5 × 10⁻¹⁰</td>
</tr>
</tbody>
</table>

Chr: chromosome, EAF: effect allele frequency, SE: standard error
*Effect allele is aligned to be eGFR decreasing allele
#Beta/SE are obtained from fixed-effects meta-analysis, with inverse variance weighting of allelic effect sizes, of up to 81,829 individuals of diverse ancestry from the COGENT-Kidney Consortium, and represent absolute decrease in eGFR (ml/min per 1.73m²) per effect allele
Table 2 | High confidence SNVs driving eGFR associations and putative causal genes through which their effects on kidney function are mediated

<table>
<thead>
<tr>
<th>Locus</th>
<th>SNV</th>
<th>p-valuea</th>
<th>π</th>
<th>Gene</th>
<th>Supporting evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANXA9</td>
<td>rs267738</td>
<td>1.7 × 10^{-10}</td>
<td>55.3%</td>
<td>CERS2</td>
<td>Encodes p.Glu115Ala (possibly damaging, deleterious)b, CACNA1S, GCKR, GCKR, GP2,</td>
</tr>
<tr>
<td>CACNA1S</td>
<td>rs3850625</td>
<td>2.5 × 10^{-9}</td>
<td>99.0%</td>
<td>CACNA1S</td>
<td>Encodes p.Arg1539Cys (possibly damaging, deleterious)b, GCKR, GCKR, GP2,</td>
</tr>
<tr>
<td>GCKR</td>
<td>rs1260326</td>
<td>2.0 × 10^{-35}</td>
<td>86.1%</td>
<td>GCKR</td>
<td>GCKR, SPTBNI, LRP2, LRP2, LRP2, LRP2, LRP2,</td>
</tr>
<tr>
<td>C2orf73</td>
<td>rs10181201</td>
<td>7.4 × 10^{-8}</td>
<td>60.9%</td>
<td>SPTBNI</td>
<td>Intrinsic; differential expression across kidney cell types.</td>
</tr>
<tr>
<td>LRP2</td>
<td>rs35472707</td>
<td>1.1 × 10^{-6}</td>
<td>64.3%</td>
<td>LRP2</td>
<td>Non-coding SNV mapped to non-kidney locus (rs77924615, p = 1.5 × 10^{-5}, π = 100.0%).</td>
</tr>
<tr>
<td>CPS1</td>
<td>rs1047891</td>
<td>1.5 × 10^{-29}</td>
<td>98.1%</td>
<td>CPS1</td>
<td>Colocalises with lead eQTL SNV.</td>
</tr>
<tr>
<td>PRDMB-FGFS</td>
<td>rs12509595</td>
<td>4.7 × 10^{-16}</td>
<td>57.1%</td>
<td>FGFS</td>
<td>Intrinsic; differential expression across kidney cell types.</td>
</tr>
<tr>
<td>RGS34-SLC34A1</td>
<td>rs3821036</td>
<td>1.0 × 10^{-32}</td>
<td>65.0%</td>
<td>SLC34A1</td>
<td>Intrinsic; differential expression across kidney cell types.</td>
</tr>
<tr>
<td>PIPK1B</td>
<td>rs2039442</td>
<td>1.3 × 10^{-26}</td>
<td>50.7%</td>
<td>PIPK1B</td>
<td>Intrinsic; differential expression across kidney cell types.</td>
</tr>
<tr>
<td>WDR37</td>
<td>rs80282103</td>
<td>2.0 × 10^{-18}</td>
<td>100.0%</td>
<td>LARP4B</td>
<td>Intrinsic; differential expression across kidney cell types.</td>
</tr>
<tr>
<td>MPPED2</td>
<td>rs7930738</td>
<td>4.7 × 10^{-7}</td>
<td>51.5%</td>
<td>MPPED2</td>
<td>Intrinsic; differential expression across kidney cell types.</td>
</tr>
<tr>
<td>UMOD-PDILT</td>
<td>rs77924615</td>
<td>1.5 × 10^{-5}</td>
<td>100.0%</td>
<td>UMOD</td>
<td>Lead eQTL SNV; differential expression across kidney cell types.</td>
</tr>
<tr>
<td>DPEP1</td>
<td>rs2460449</td>
<td>4.2 × 10^{-9}</td>
<td>97.8%</td>
<td>DPEP1</td>
<td>Lead eQTL SNV; differential expression across kidney cell types.</td>
</tr>
<tr>
<td>BCAS3</td>
<td>rs9895611</td>
<td>8.9 × 10^{-28}</td>
<td>100.0%</td>
<td>BCAS3</td>
<td>Non-coding SNV mapped to non-kidney locus (rs77924615, p = 1.5 × 10^{-5}, π = 100.0%).</td>
</tr>
<tr>
<td>TBX2</td>
<td>rs887258</td>
<td>2.7 × 10^{-13}</td>
<td>62.2%</td>
<td>TBX2</td>
<td>Colocalises with lead eQTL SNV.</td>
</tr>
</tbody>
</table>

a posterior probability of association
b p-values obtained from fixed-effects meta-analysis
c PolyPhen2/SIFT predictions

hypokalemic periodic paralysis37,28, malignant hyperthermia39 and congenital myopathy40. CACNA1S is highly expressed in skeletal muscle tissue, raising the possibility that the high-confidence missense variant may influence eGFR through creatinine production. CPS1 (Carbamoyl-Phosphate Synthase 1) is involved in the urea cycle, where the enzyme plays an important role in removing excess ammonia from cells31. GCKR (Glucokinase Regulator) produces a regulatory protein that inhibits glucokinase, and the p.Leu446Pro substitution is a highly pleiotropic variant with reported effects on a wide range of health and disease states.

CERS2 (Ceramide Synthase 2) variants have previously been associated with albuminuria in individuals with diabetes33, and interrogation of the Human Protein Atlas34 revealed that the CERS2 protein is abundantly expressed in the glomerulus and tubules of the kidney. Cers2-deficient mice exhibit changes in the structure of the kidney35. We verified that Cers2 mRNA is expressed in primary podocytes isolated from the mouse using a previously published method36 (Methods, Supplementary Figure 5). To gain insight into the potential role of CERS2 in podocyte motility and function, we isolated and grew primary murine podocytes in culture, and exposed them to the CERS2 inhibitor, ST-107437,38 (Methods). We compared the podocyte migration rate among treated and untreated cells using the scratch wound-healing assay (Supplementary Figure 6). Primary podocytes treated with 3 μM concentration of the CESRS2 inhibitor had a lower migration rate than untreated cells, with significantly higher percentages of uncovered areas remaining at 18 h after wound-scratch. Podocytes treated with ST-1074 appeared much more elongated at 18 h. Although we cannot rule out off-target effects of the inhibitor, these preliminary results suggest that CERS2 may have a functional impact on podocyte biology. However, further studies are needed to determine the specific role of the gene in the kidney, in vivo, in health and disease states.

The remaining 36 high-confidence SNVs mapped to non-coding regions, which we assessed for colocalisation with eQTL from two resources: (i) non-cancer affected healthy kidney tissue obtained from 260 individuals from the TRANSCRIPTome of renal humAn TissuE (TRANSLATE) Study39,40 and The Cancer Genome Atlas (TCGA)41; and (ii) kidney biopsies obtained from 134 healthy donors from the TransplantLines Study42 (Methods). We observed that high-confidence eGFR SNVs colocalised with lead renal eQTL variants in the TRANSLATE Study and TGCA (Table 2, Supplementary Table 7) for FGFS (rs12509595, p = 4.7 × 10^{-16}, π = 57.1%), TBX2 (rs887258, p = 2.7 × 10^{-13}, π = 62.2%), and both UMOD and GP2 for the same signal at the UMOD-PDILT locus (rs77924615, p = 1.5 × 10^{-5}, π = 100.0%). Of these three high-confidence SNVs, rs887258 was a significant eQTL (defined by 5% false discovery rate) for TBX2 across multiple tissues in the GTEX project43, whilst the associations of rs12509595 and rs77924615 with an expression of FGFS and UMOD/GP2, respectively, were specific to kidney. FGFS (Fibroblast Growth Factor 5) is expressed during kidney development, but knockout models have not shown a kidney phenotype44. FGFS has been implicated in GWAS of blood pressure and hypertension45, and other fibroblast growth factors are increasingly recognised as contributors to blood pressure regulation through renal mechanisms46. TBX2 (T-Box 2) plays a role in defining the pronephric nephron in experimental models46. UMOD encodes uromodulin (Tamm-Horsfall protein), the most abundant urinary protein. The eGFR lowering allele at the high-confidence SNV is associated with increased UMOD expression (Supplementary Figure 7), which is consistent with previous investigations that demonstrated uromodulin overexpression in transgenic mice leads to salt-sensitive hypertension and the presence of age-dependent renal lesions47.

Mapping genes to kidney cells. Kidney cells are highly specialised in function based on their location in nephron segments. Previous investigations in mouse and human have revealed that genes at kidney trait-related loci are expressed in a cell-specific manner48,49. To provide insight into cellular specificity of the signals at the UMOD-PDILT, FGFS and TBX2 loci, we mapped the four genes identified through eQTL analyses to cell types from single nucleus RNA-sequencing (snRNA-seq) data obtained from a healthy human kidney donor (4254 cells, with an average of 1803 detected genes per cell)49. UMOD and GP2 demonstrated expression specific to epithelial cells of the ascending loop of Henle (Fig. 1). Uromodulin is involved in protection against...
urinary tract infections\(^5\), and the global distribution of UMOD regulatory variants in humans correlates with pathogen diversity and prevalence in urine\(^5\). Glycoprotein 2 is a protein involved in innate immunity. These findings suggest a role for these two proteins in kidney physiology and potential host defence immunity to uropathogens at the \textit{UMOD-PDILT} locus.

By localising high-confidence SNVs to introns and UTRs (Methods), we identified eight additional genes with differential expression across nephron single cell-types (Fig. 1, Table 2): \textit{LRP2}, \textit{SLC34A1} and \textit{DPEP1} (specific to proximal tubule); \textit{SPTBN1} (specific to glomeruli endothelial cells); \textit{PIP5K1B} (specific to glomeruli mesangial cells); and \textit{LARP4B}, \textit{BCAS3}, and \textit{MPPED2} (multiple cell types in the distal nephron). Of these, \textit{DPEP1}, which encodes the protein dipeptidase 1, is implicated in the renal metabolism of glutathione and its conjugates, and regulates leukotriene activity. This localisation fits with the previously suggested connection between glutathione metabolism and its conjugates, and regulates leukotriene activity. This localisation fits with the previously suggested connection between glutathione metabolism and defence against chemical injury in proximal tubule cells\(^5\). Taken together, these findings suggest a potential role of these genes in influencing kidney structure and function through regulation of: (i) glomerular capillary pressure, determining intra-glomerular pressure and glomerular filtration; (ii) proximal tubular reabsorption, affecting tubuloglomerular feedback; or (iii) distal nephron handling of sodium or acid load, influencing kidney disease progression. Additional laboratory-based functional studies will be required to delineate the mechanistic pathways that determine kidney function in healthy and disease states, and potential routes to therapeutic targets for pharmacologic development.

\textbf{Causal effects of eGFR on clinically-relevant outcomes.} We sought to evaluate the causal effect of eGFR on clinically-relevant kidney and cardiovascular outcomes via two-sample MR\(^5\) (Methods, Supplementary Tables 8, 9 and 10). Analyses were performed separately in each of the three components of the trans-ethnic meta-analysis because allelic effect sizes were measured on different scales in each. For each trait, we accounted for heterogeneity in causal effects of eGFR via modified Q-statistics\(^5\), excluding outlying genetic instruments that may reflect pleiotropic SNVs and violate the assumptions of MR (Methods, Supplementary Tables 9 and 10).

In each component, we detected a significant \((p < 0.0042,\) Bonferroni correction for 12 traits) causal effect of lower eGFR on higher risk of all-cause CKD, glomerular diseases and CKD stage 5, based on reported association summary statistics from the CKDGen Consortium\(^8\) and the UK Biobank (Fig. 2, Supplementary Table 8). We also detected a significant causal effect of lower eGFR on lower risk of calculus of the kidney and ureter, in each component, based on reported association summary statistics from the UK Biobank (Fig. 3, Supplementary Table 8). The lead eGFR SNV at the \textit{UMOD-PDILT} locus (rs77924615) has been previously associated with kidney stone formation\(^5\) and is
consistent with the role of uromodulin in the inhibition of urine calcium crystallisation. However, this SNV was excluded from the MR analysis due to heterogeneity in effect size and was therefore not driving the causal eGFR association with risk of calculus of the kidney and ureter (Supplementary Table 9).

We also detected a novel causal effect of lower eGFR (at nominal significance, \( p < 0.05 \), in each component of the trans-ethnic meta-analysis) on higher diastolic blood pressure (DBP) and higher risk of essential (primary) hypertension, but not on systolic blood pressure, based on reported association summary statistics from automated readings and ICD10 codes from primary care data available in the UK Biobank (Fig. 4, Supplementary Table 8). These results are consistent with a role for reduced functional nephron mass on increased peripheral
Fig. 3 Two-sample MR of eGFR on calculus of kidney and ureter. Results are presented separately for each component of the trans-ethnic meta-analysis. Each point corresponds to a lead SNV (instrumental variable) across 94 kidney function loci, plotted according to the MR effect size of eGFR on calculus of kidney and ureter (Wald ratio). Bars correspond to the standard errors of the effect sizes. The red point and bar in each plot represents the MR effect size of eGFR on calculus of kidney and ureter across all SNVs under inverse variance weighted regression. The p-values are obtained under inverse variance weighted regression. Results for other methods are presented in Supplementary Table 8.

Fig. 4 Two-sample MR of eGFR on diastolic blood pressure and hypertension. Results are presented separately for each component of the trans-ethnic meta-analysis for diastolic blood pressure (top) and essential (primary) hypertension (bottom). Each point corresponds to a lead SNV (instrumental variable) across 94 kidney function loci, plotted according to the MR effect size of eGFR on outcome (Wald ratio). Bars correspond to the standard errors of the effect sizes. The red point and bar in each plot represents the MR effect size of eGFR on outcome across all SNVs under inverse variance weighted regression. The p-values are obtained under inverse variance weighted regression. Results for other methods are presented in Supplementary Table 8.
arterial resistance and confirm previous findings from observational studies. Although the causal association with DBP could not be replicated using published meta-analysis association summary statistics from the International Consortium for Blood Pressure (ICBP) (Supplementary Table 11), we note that their blood pressure measures were corrected for body-mass index (in addition to age and sex), and there was significant evidence of heterogeneity in effects of eGFR on outcome across SNVs, indicating potential pleiotropy due to collider bias, and consequently invalidating MR estimates. Despite the large sample sizes available for MR analyses from the CardiogramplusCAD Consortium and MEGASTROKE, there was no significant evidence of a causal association of eGFR on cardiovascular disease outcomes: coronary heart disease, myocardial infarction or ischemic stroke (Supplementary Table 8).

Discussion

We identified 20 novel loci for eGFR through trans-ethnic meta-analysis, and dissected 127 distinct association signals that together explain an additional 5.3% of the genome-wide observed scale heritability. The effects of index SNVs for these distinct eGFR association signals were homogeneous across major ancestry groups, which is consistent with a model in which the underlying causal variants are shared across diverse populations, and therefore amenable to trans-ethnic fine-mapping. The localisation of causal variants at eGFR association signals was further enhanced through integration with enriched signatures of genomic annotation that included kidney-specific histone modifications.

We localised high-confidence causal variants driving 40 distinct eGFR association signals, the majority of which have not been previously reported. Through a variety of approaches, including colocalisation with eQTLs in human kidney, and identification of differential expression between human kidney cell types through snRNA-seq, these high-confidence variants implicated several putative causal genes that account for eGFR variation at kidney function loci. Therefore, our strategy of utilising multiple kidney tissue-specific resources to uncover likely causal variants and the genes through which their effects are mediated, followed by mapping of these genes to specific cells in the nephron, provides important biological insight and potential targets for drug development. Knowledge of the specificity of gene expression in nephron segments should also inform future experiments to elucidate the function of some of these genes and potentially define causal molecular mechanisms underlying CKD.

MR analyses of lead SNVs at kidney function loci highlighted previously unreported causal effects of lower eGFR on higher risk of primary glomerular diseases, lower risk of kidney stone formation, and higher DBP and risk of hypertension. The causal relationships of eGFR to these outcomes have been demonstrated to be consistent across ancestries, which is essential for the development of potential interventions that would be relevant to diverse global populations. Our MR analyses also identified lead eGFR SNVs with heterogeneous causal effects on these outcomes, indicating potential pleiotropy. However, further work will be required to determine the specific pathways through which these pleiotropic SNVs act, including non-eGFR determinants of serum creatinine-based eGFR estimating equations.

In conclusion, we have undertaken the most comprehensive trans-ethnic GWAS of eGFR, which has significantly enhanced knowledge of the genetic contribution to kidney function. Our investigation emphasizes the importance of genetic studies of eGFR in diverse populations and their integration with cell-type specific kidney expression data for maximising gains in discovery and fine-mapping of kidney function loci. Taken together, these strategies offer the most promising route to treatment development for a disease with major public health impact across the globe.

Methods

Ethics statement. All human research was approved by the relevant institutional review boards and conducted according to the Declaration of Helsinki. All participants provided written informed consent. All mice were maintained on a 12-h light-dark cycle with free access to standard chow and water in the animal facility of the University of Virginia (UVA). Experiments were carried out in accordance with local and NIH guidelines, and the animal protocol was approved by the UVA Institutional Animal Care and Use Committee.

CGENT-Kidney Consortium: study-level analyses. Study sample characteristics for GWAS from the CGENT-Kidney Consortium, which incorporates 88,423 individuals of diverse ancestry (32.4% Hispanic/Latino, 28.8% European, 28.8% East Asian and 10.0% African American) are presented in Supplementary Table 12. These GWAS included those reported previously but were expanded with the addition of further studies of Hispanic/Latino ancestry to increase the diversity of represented population groups. Samples were assayed with a range of data types and cell types through snRNA-seq, these high-confidence variants were used to account for confounding due to population structure (Supplementary Table 13). Sample sets were excluded because of low genome-wide call rate, extreme heterozygosity, sex discordance, cryptic relatedness, and outlying ethnicity. SNVs were excluded because of low call rate across samples and extreme deviation from Hardy–Weinberg equilibrium. Autosomal SNVs were excluded from imputation and association analysis. Within each study, the GWAS genotype scaffold was pre-phased and imputed up to the phase 1 integrated (version 3) multi-ethnic reference panel from the 1000 Genomes Project using IMPUTEv2 or minimac.

Within each study, eGFR was calculated from serum creatinine (mg/dL), accounting for age, sex and ethnicity, using the four-variable MDRD equation. We tested the association of eGFR with each SNP in a linear regression framework, under an additive dosage model, and with adjustment for study-specific covariates to account for confounding due to population structure (Supplementary Table 13). For each SNP, the association Z-score was derived from the allele effect estimate and corresponding standard error. Z-scores and standard errors were then corrected for residual population structure via genomic control where necessary (Supplementary Table 13).

CKDGen Consortium: meta-analysis. Full details of the CKDGen Consortium meta-analysis, which incorporated GWAS in 110,517 individuals of European ancestry, have been previously published. Briefly, individuals were assayed with a range of GWAS genotyping products. After quality control, GWAS scaffolds were pre-phased and imputed up to the Phase 1 integrated (version 1 or version 3) multi-ethnic or European-specific reference panels from the 1000 Genomes Project. Imputed variants were retained for downstream association analyses if they attained IMPUTEv2 info > 0.4 or minimac r² > 0.3. Within each study, eGFR was calculated from serum creatinine (mg/dL), accounting for age and sex, using the four-variable Modification of Diet in Renal Disease (MDRD) equation. Residuals obtained after regressing ln(eGFR) on age and sex, and study-specific covariates to account for population structure where appropriate, were tested for association with each SNP in a linear regression framework under an additive dosage model. Association summary statistics within each GWAS were corrected for residual population structure via genomic control where necessary and were subsequently aggregated across studies, under a fixed-effects model, with inverse-variance weighting of allele effect sizes, as implemented in METAL.

Biobank Japan Project: study-level analysis. Full details of the Biobank Japan Project GWAS, which incorporated 145,658 individuals of East Asian ancestry, have been previously published. Briefly, individuals were assayed with the Illumina HumanOmniExpressExome BeadChip or a combination of the Illumina HumanOmniExpress BeadChip and the Illumina HumanExome BeadChip. After quality control, the GWAS scaffold was pre-phased with MaCh and imputed up to the Phase 1 integrated (version 3) East Asian-specific reference panel from the 1000 Genomes Project with minimac. Imputed variants were retained for downstream association analyses if they attained minimac r² ≥ 0.7. For each individual, eGFR was derived from serum creatinine (mg/dL) using the Japanese coefficient-modified CKD Epidemiology Collaboration (CKD-EPI) equation, and adjusted for age, sex, ten principal components of genetic ancestry, and hypertension status for 47 diseases. The resulting residuals were inverse-rank normalised and tested for association with each SNP in a linear regression framework, under an additive dosage model.
From the available GWAS summary statistics for each SNV (downloaded from http://senger.riken.jp/en/result), we derived the association Z-score from the ratio of the allelic effect estimate and corresponding standard error, and subsequently corrected for residual population structure by genomic control\(^6\), \(\lambda_{GC} = 1.252\).

Trans-ethnic meta-analysis. We aggregated eGFR association summary statistics across the three components: COGENT-Kidney Consortium GWAS, the Biobank Japan Project GWAS and the CKDGen Consortium meta-analysis. We performed fixed-effects meta-analysis, with sample size weighting of Z-scores (Stouffer’s method), as implemented in METAL\(^4\), because allelic effect estimates were on different scales in the contributing components. The COGENT-Kidney Consortium included a GWAS of a subset of 23,536 individuals from those contributing to the Biobank Japan Project, which was therefore excluded from the trans-ethnic meta-analysis. Consequently, a combined sample size of 312,468 individuals contributed to the trans-ethnic meta-analysis. SNVs represented on at least 50% of the combined sample size were retained for downstream interrogation. Meta-analysis association summary statistics were corrected for residual population structure via genomic control\(^6\), \(\lambda_{GC} = 1.113\).

Locus definition. We first selected lead SNVs retaining genome-wide significant evidence of association (\(p < 5 \times 10^{-8}\)) with eGFR in the trans-ethnic meta-analysis that were separated by at least 500kb. Loci were defined by the flanking genomic interval mapping 500kb up- and down-stream of lead SNVs. Where loci overlapped, they were combined as a single locus, and the lead SNV with minimal p-value from the meta-analysis was retained.

Dissection of association signals. To dissect distinct eGFR association signals at loci retaining genome-wide significance in the trans-ethnic meta-analysis, we used an iterative approximate conditional approach, implemented in GCTA\(^5\). Each COGENT-Kidney Consortium GWAS was first assigned to an ethnic group (Supplementary Table 12) represented in the 1000 Genomes Project reference panel (Phase 3, October 2014 release)\(^6\). The Biobank Japan Project was assigned to the East Asian ethnic group, and the CKDGen Consortium meta-analysis was assigned to the European ethnic group. Haplotypes in the 1000 Genomes Project panel that were specific to the assigned ethnic group were then used as a reference for LD between SNVs across loci for the GWAS in the approximate conditional analysis.

For each locus, we first applied GCTA to the study-level association summary statistics and matched LD reference for each GWAS (or the CKDGen Consortium meta-analysis). We adjusted for the conditional set of variants, which in the first iteration included only the lead SNV at the locus, and aggregated Z-scores across studies with sample size weighting (Stouffer’s method) under a fixed-effects model, as implemented in METAL\(^4\). The conditional meta-analysis summary statistics were corrected for residual population structure using the same genomic control adjustment\(^6\) as in the unconditional analysis (\(\lambda_{GC} = 1.113\)). We defined locus-wide significance by \(p < 5 \times 10^{-8}\), which is a Bonferroni correction for the approximate number of independent SNVs at each locus. If no SNVs at attained locus-wide significant evidence of residual association with eGFR, the iterative approximate conditional analysis for the locus was stopped. Otherwise, the SNV with the strongest residual association signal was added to the conditional set. This iterative process continued, at each stage adding the SNV with the strongest residual association from the meta-analysis to the conditional set, until no remaining SNV’s attained locus-wide significance. Note, that at each iteration, studies with missing association summary statistics for any SNV in the conditional set were excluded from the meta-analysis.

For each locus including more than one SNV in the conditional set, we then dissected each distinct association signal. We again applied GCTA to the study-level association summary statistics and matched LD reference for each GWAS (or the CKDGen Consortium meta-analysis), but this time by removing each SNV, in turn, from the conditional set of variants, and adjusting for the remainder. The conditional meta-analysis summary statistics were corrected for residual population structure using the same genomic control adjustment\(^6\) as in the unconditional analysis (\(\lambda_{GC} = 1.113\)). The SNV with the strongest residual association was defined as the index for the signal.

Estimation of observed scale heritability. We used LD Score regression\(^16\) to assess the contribution of variation to the observed scale heritability of eGFR. LD Score regression accounts for LD between SNVs on the basis of European ancestry individuals from the 1000 Genomes Project\(^7\). We therefore performed fixed-effects meta-analysis, with sample size weighting of Z-scores (Stouffer’s method), as implemented in METAL\(^4\), across European ancestry studies from the COGENT-Kidney Consortium and CKDGen Consortium (134,070 individuals), and used these association summary statistics in LD Score regression. We first calculated the contribution of genome-wide variation to the observed scale heritability of eGFR. We then disintegrated the genome into previously reported and novel loci attaining genome-wide significance in the trans-ethnic meta-analysis (Supplementary Table 1) and calculated the observed scale heritability of eGFR attributable to each.
annotations, and the Bayes’ factor approximated from the trans-ethnic meta-analysis. Specifically, for the $j$th SNV at the $i$th distinct association signal, the posterior probability $\pi_i \propto \gamma_j$. In this expression, the relative annotation informed prior is given by

$$
\gamma_j = \exp \left[ \sum_k \beta_{jk} \right]
$$

where the summation is over the selected enriched annotations, and $\beta_{jk}$ is the estimated log-fold enrichment of the $j$th annotation from the final joint model. We estimated a 99% credible set$^6$ for the $i$th distinct association signal by: (i) ranking all SNVs according to their posterior probability $\pi_i$; and (ii) including ranked SNVs until their cumulative posterior probability of driving the association attains or exceeds 0.99. Index SNV $i$’s accounting for more than 50% posterior probability of driving the eGFR association at a given signal were defined as high-confidence.

**SNV associations with measures of kidney function and damage**. We evaluated the evidence for association of high-confidence variants with measures of kidney function and damage from published GWAS: (i) eGFR calculated from cystatin C, obtained from up to 24,061 individuals of European ancestry from the CKDGen Consortium$^2$; (ii) blood urea nitrogen, obtained from up to 139,818 individuals of East Asian ancestry from the Biobank Japan Project$^5$; and (iii) urine albumin to creatinine ratio, obtained from up to 46,061 non-diabetic individuals of European ancestry from the CKDGen Consortium$^2$. Effects on these traits were aligned to the eGFR decreasing allele.

Functional annotation of high-confidence missense variants. We assessed the predicted functional impact of high-confidence missense variants across a range of databases including FATHMM (functional analyses through hidden Markov models)$^8$ and metaSVM (scores for non-synonymous variants based on SVM model)$^9$. Primary podocyte cell culture and scratch assay. The protocol for isolation and culture of primary podocytes from 129S6 mice has been previously published$^6$. Briefly, under general anaesthesia with isoflurane, mice were perfused through the heart with 6 x 10$^7$ magnetic beads/ml (Dynabeads, Invitrogen) diluted in 10 ml of phosphate-buffered saline (PBS). The kidneys were extracted, decapsulated and cut into small pieces, then digested in collagenase A (1mg/ml) at 37°C for 25 min. The digest was pressed with a syringe pestle through a 100 µm cell strainer, collected, and washed with 5 ml of PBS, repeated once, and then passed through a 40 µm cell strainer, and washed again with 5 ml of PBS. The isolated glomeruli were then washed off the cell strainer with pre-warmed cell culture media (RPMI 1640 with L-glutamine supplemented with 10% of fetal bovine serum, 100 units/ml of Pen/Strep and 1% of L-glutamine), then further isolated by magnetic particle concentrator (Invitrogen) and washed with a pre-warmed culture media 2-3 times. Glomeruli were then re-suspended in culture media and placed in the 6-well plate coated with rat tail type I collagen and incubated at 37°C. Light microscopy was used to confirm the characteristic migration of podocytes from glomeruli after 3-5 days.

Approximately after 7-9 days, the primary podocytes reached a confluent monolayer, and a rectangular wound was created by scratching the monolayer from the top to the bottom of the well using the base of a sterile 200 µl pipette tip. Images of the scratched area were immediately taken after wound creation and after an additional 18 h of incubation, using EVO XL Core Cell Imaging System ($\times$10 magnification). We utilised ST-1074, a potent inhibitor that can be used in functional studies. Normalised gene expression was modelled as a function of alternate allele dosage via linear regression, including sex, three axes of genetic variation (to account for population structure) and PEER latent factors as additional covariates. The regression coefficients of the alternate allele from the two studies were then combined in a fixed-effects meta-analysis under an inverse-variance weighting scheme. For each gene, only those SNVs in cis (within 1 Mb of the transcription start/stop sites) were included in the analysis. A total of 2000 permutations were used to derive the empirical distribution of the smallest $p$-value for each gene, which then was used to adjust the observed smallest $p$-value for the gene. The correction for testing multiple genes was based on false discovery rate (FDR) applied to permutation-adjusted $p$-values (via Storey’s method as implemented in the R package p.adjust) with a cut-off of 0.5. Further thresholds for nominal $p$-values were derived using a global permutation-adjusted $p$-value closest to FDR of 5% and the empirical distributions determined using permutations.

We identified high-confidence SNVs from the trans-ethnic fine-mapping that were colocalised with lead eQTL variants (i.e. the same SNV or in strong LD, $r^2>0.8$) at a 5% FDR, and reported the corresponding eGene.

**Kidney tissue eQTLs: TransplantLines Study.** We performed eQTL analysis using data from the TransplantLines Study$^{32}$. The study includes kidneys from donors, donated after brain death or cardiac death. Samples were genotyped on the Illumina CytosNP 12 v2 array and imputed up to the Phase 1 integrated (version 3) multi-ethnic reference panel from the 1000 Genomes Project$^{33}$ using the Michigan Imputation Server$^{34}$. Expression and genotype data were available for 236 kidney biopsies obtained from 134 donors, and analyses have been described previously$^{35}$. Briefly, residuals of gene expression for each probe were obtained after adjusting for the first 50 expression principal components to filter out environmental variation$^{37}$. A linear mixed model was used to test the association of residual expression of each probe with the allele dosage of each SNV. The $p$-values of the transcription start/stop sites using the R package lmek. Sex, age, donor type, time of biopsy and three axes of genetic variation (to account for population structure) were included in the model as fixed effects. Random effects were then included for donor to account for multiple samples obtained from the same individual. We identified genes for which high-confidence SNVs mapped to introns and untranslated regions. We mapped the genes to cell-types from smRNA-seq data generated by 10x Chromium from 12-year-old human kidney biopsies. We also incorporated the 2-year-old white male and female serum creatinine of 1.03 mg/dl$^{38}$. The kidney was dissected from the cortex and was bulk homogenized using a dounce homogenizer. The data included 4524
cells: 7.8% glomerular cells (including podocytes, endothelial cells and mesangial cells); 86.7% tubular cells (including proximal tubule, loop of Henle, distal convoluted tubule, connecting tubule, proximal intercalated cells and intercalated cells); the remaining 5.5% cells are mostly macrophages and novel cells that do not map to known cell types. An average of 1803 genes were detected per cell. We generated a differential expression gene (DEG) list by performing Wilcoxon rank sum tests on each cell-type from the single nucleus dataset. A gene was defined as mapping to a specific cell type if the expression fulfilled the following criteria: (i) present in the DEG list; (ii) expressed in >25% of the total cells in the specified cell-type; and (iii) log-fold change in expression was >0.25 in the specified cell-type when compared to all other cell-types. Gene expression values for each cell were Z-score normalized. A new gene expression matrix with mean Z-scores for each gene was obtained by averaging the Z-scores from all individual cells in the same cluster. The Z-score normalized gene expression were presented as a heatmap using the heatmap.2 function in the R package gplots.

Two-sample MR analyses. We performed a lookup of association summary statistics for lead SNVs at each of the eGFR loci across a range of clinically-relevant kidney and cardiovascular outcomes from public and proprietary data resources. These included: CKD (12,385 cases and 104,780 controls, published data from the kidney and cardiovascular outcomes from public and proprietary data resources); and (ii) the dbGaP CHARGE Summary Results site92 with accession number phs000930. The source data underlying Fig. 1 and Supplementary Figure 7 are provided as a Source Data file.

Received: 16 August 2018 Accepted: 3 December 2018
Published online: 03 January 2019

References


Acknowledgements

T.H.L. is supported by the NIH (R01-DK-113632). G.H. is supported by the Wellcome Trust (208806/Z/17/Z). G.H. and G.D.S. work in the Medical Research Council Integrative Epidemiology Unit at the University of Bristol (MC_UU_00011/1). C.M.L. is supported by the Li Ka Shing Foundation, WT-Si/Sjon Fell funds, the Oxford NIHR Biomedical Research Centre, Widenlife and the NIH (5P50-HD-028138-27). R.L.S. and T.R. are supported by the NIH (R37-NR-029993 and U54-TR-002736) and the Evelyn F McKnight Brain Institute. H.S. and A.Z. are supported by the DFG (INST 208:664-1). M.T. is supported by the British Heart Foundation (PG17/35/33001) and Kidney Research UK (RP_017_20180302). N.F. is supported by the NIH (R01-MD-012765, R56-DK-104806, and R01-DK-117445-01A1). Additional funding and acknowledgements can be found in Supplementary Note 1.

Author contributions


Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-07867-7.

Competing interests: G.N.N. has received operational funding from Goldfinch Bio. The remaining authors declare no competing interests.

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

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019
