Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function

Citation for published version:

Digital Object Identifier (DOI):
10.1038/ncomms10023

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Nature Communications

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function

Cristian Pattaro et al.

Reduced glomerular filtration rate defines chronic kidney disease and is associated with cardiovascular and all-cause mortality. We conducted a meta-analysis of genome-wide association studies for estimated glomerular filtration rate (eGFR), combining data across 133,413 individuals with replication in up to 42,166 individuals. We identify 24 new and confirm 29 previously identified loci. Of these 53 loci, 19 associate with eGFR among individuals with diabetes. Using bioinformatics, we show that identified genes at eGFR loci are enriched for expression in kidney tissues and in pathways relevant for kidney development and transmembrane transporter activity, kidney structure, and regulation of glucose metabolism. Chromatin state mapping and DNase I hypersensitivity analyses across adult tissues demonstrate preferential mapping of associated variants to regulatory regions in kidney but not extra-renal tissues. These findings suggest that genetic determinants of eGFR are mediated largely through direct effects within the kidney and highlight important cell types and biological pathways.
Chronic kidney disease (CKD) is a global public health problem, and is associated with an increased risk for cardiovascular disease, all-cause mortality and end-stage renal disease. Few new therapies have been developed to prevent or treat CKD over the past two decades, underscoring the need to identify and understand the underlying mechanisms of CKD.

Prior genome-wide association studies (GWAS) have identified multiple genetic loci associated with CKD and estimated glomerular filtration rate (eGFR), a measure of the kidney’s filtration ability that is used to diagnose and stage CKD. Subsequent functional investigations point towards clinically relevant novel mechanisms in CKD that were derived from initial GWAS findings, providing proof of principle that locus discovery through large-scale GWAS efforts can translate to new insights into CKD pathogenesis.

To identify additional genetic variants associated with eGFR and guide future experimental studies of CKD-related mechanisms, we have now performed GWAS meta-analyses in up to 133,413 individuals, more than double the sample size of previous studies. Here, we describe multiple novel genomic loci associated with kidney function traits and provide extensive locus characterization and bioinformatics analyses, further delineating the physiologic basis of kidney function.

Results

Stage 1 discovery analysis. We analysed associations of eGFR based on serum creatinine (eGFRcrea), cystatin C (eGFRCys, an additional, complementary biomarker of renal function) and CKD (defined as eGFRcrea < 60 ml min⁻¹ per 1.73 m²) with ∼2.5 million autosomal single-nucleotide polymorphisms (SNPs) in up to 133,413 individuals of European ancestry from 49 predominantly population-based studies (Supplementary Table 1). Results from discovery GWAS meta-analysis are publicly available at http://fox.nihbi.nih.gov/CKDGen. We performed analyses in each study sample in the overall population and stratified by diabetes status, since genetic susceptibility to CKD may differ by diabetes status, since genetic susceptibility to CKD may differ in the presence of this strong clinical CKD risk factor. Population stratification did not impact our results as evidenced by low genomic inflation factors in our meta-analyses, which ranged from 1.00 to 1.04 across all our analyses (Supplementary Fig. 1).

In addition to confirming 29 previously identified loci (Fig. 1a; Supplementary Table 2), we identified 48 independent novel loci (Supplementary Table 3) where the index SNP, defined as the variant with the lowest p-value in the region, had an association p-value < 1.0 × 10⁻⁸. Of these 48 novel SNPs, 21 were genome-wide significant with p-values < 5.0 × 10⁻⁸. Overall, 43 SNPs were identified in association with eGFRcrea (nine in the non-diabetes sample), one with GFRCys and four with CKD, as reported in Supplementary Table 3. Manhattan plots for CKD, eGFRCys and eGFRcrea in diabetes are shown in Fig. 1b,c and Supplementary Fig. 2, respectively.

Stage-2 replication. Novel loci were tested for replication in up to 42,166 additional European ancestry individuals from 15 studies (Supplementary Table 1). Of the 48 novel candidate SNPs submitted to replication, 24 SNPs demonstrated a genome-wide significant combined stage 1 and 2 p-value < 5.0 × 10⁻⁸ (Table 1). Of these, 23 fulfilled additional replication criteria (q-value < 0.05 in stage 2). Only rs6795744 at the WNT7A locus demonstrated suggestive replication (p-value < 5.0 × 10⁻⁴, q-value > 0.05). Because serum creatinine is used to estimate eGFRcrea, associated genetic loci may be relevant to creatinine production or metabolism rather than kidney function per se. For this reason, we contrasted associations of eGFRcrea versus eGFRCys, the latter estimated from an alternative and creatinine-independent biomarker of GFR (Supplementary Fig. 3; Supplementary Table 4). The majority of loci (22/24) demonstrated consistent effect directions of their association with both eGFRcrea and eGFRCys.

Association plots of the 24 newly identified genomic regions that contain a replicated or suggestive index SNP appear in Supplementary Fig. 4. The odds ratio for CKD for each of the novel loci ranged from 0.93 to 1.06 (Supplementary Table 4). As evidenced by the relatively small effect sizes, the proportion of phenotypic variance of eGFRcrea explained by all new and known loci was 3.22%: 0.81% for the newly uncovered loci and 2.41% for the already known loci.

Associations stratified by diabetes and hypertension status. The effects of the 53 known and novel loci in individuals with (stage 1 + stage 2 n = 16,477) and without (stage 1 + stage 2 n = 154,881) diabetes were highly correlated (correlation coefficient: 0.80; 95% confidence interval: 0.67, 0.88; Supplementary Fig. 5) and of similar magnitude (Fig. 2; Supplementary Table 5), suggesting that identification of genetic loci in the overall population may also provide insights into loci with potential importance among individuals with diabetes. The previously identified UMOR locus showed genome-wide significant association with eGFRcrea among those with diabetes (Supplementary Fig. 2; rs12917707, P-value = 2.5 × 10⁻⁸), and six loci (NKFB1, UNCX, TSAPN9, AP5B1, SIPAL13 and PTPRO) had nominally significant associations with eGFRcrea among those with diabetes. Of the previously identified loci, 13 demonstrated nominal associations among those with diabetes, for a total of 19 loci associated with eGFRcrea in diabetes.

Exploratory comparison of the association effect sizes in subjects with and without hypertension based on our previous work showed that novel and known loci are also similarly associated with eGFRcrea among individuals with and without hypertension (Supplementary Fig. 6).

Tests for SNP associations with related phenotypes. We tested for overlap with traits that are known to be associated with kidney function in the epidemiologic literature by investigating SNP associations with systolic and diastolic blood pressure, myocardial infarction, left ventricular mass, heart failure, fasting glucose and urinary albumin excretion (CKDGen Consortium, personal communication). We observed little association of the 24 novel SNPs with other kidney function-related traits, with only 2 out of 165 tests reaching the Bonferroni significance level of 0.0003 (see Methods and Supplementary Table 6).

To investigate whether additional traits are associated with the 24 new eGFR loci, we queried the NHGRI GWAS catalog (www.genome.gov). Overall, nine loci were previously identified in association with other traits at a P-value of 5.0 × 10⁻⁸ or lower (Supplementary Table 7), including body mass index (ETV5) and serum urate (INHBC, A1CF and AP5B1).

Trans-ethnic analyses. To assess the generalizability of our findings across ethnicities, we evaluated the association of the 24 newly identified loci with eGFRcrea in 16,840 participants of 12 African ancestry population studies (Supplementary Table 8) and in up to 42,296 Asians from the AGEN consortium (Supplementary Table 9). Seven SNPs achieved nominal direction-consistent significance (P < 0.05) in AGEN, and one SNP was significant in the African ancestry meta-analysis (Supplementary Table 9). Random-effect meta-analysis showed that 12 loci (SDCCAG8, LRP2, IGFBP5, SKIL, UNCX, KBTBD2, A1CF, KCNQ1, AP5B1, PTPRO, TP53INP2 and BCAS1) had fully consistent effect direction across the three ethnic groups.
**Figure 1 | Discovery stage genome-wide association analysis.** Manhattan plots for eGFRcrea, CKD and eGFRcys. Previously reported loci are highlighted in light blue (grey labels). (a) Novel loci uncovered for eGFRcrea in the overall and in the non-diabetes groups are highlighted in blue and green, respectively. (b) Results from CKD analysis with highlighted known and novel loci for eGFRcrea. (c) Results from eGFRcys with highlighted known and novel loci for eGFRcrea and known eGFRcys loci.
The 24 novel SNPs associated with eGFRcrea in European ancestry individuals.

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Chr.</th>
<th>Position (bp)</th>
<th>Locus name</th>
<th>Effect/Non effect allele (EAF)</th>
<th>SNP function</th>
<th>Stage 1 (discovery)</th>
<th>Stage 2 (replication)</th>
<th>Combined analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Beta P value</td>
<td>Beta q-value</td>
<td>P value</td>
</tr>
<tr>
<td>rs3850625</td>
<td>1</td>
<td>201,016,296</td>
<td>CACNA1S</td>
<td>A/G (0.12)</td>
<td>Exonic, non-syn. SNV</td>
<td>0.0080 2.55E - 09</td>
<td>0.0071 5.46E - 03</td>
<td>0.0083 6.82E - 11</td>
</tr>
<tr>
<td>rs2731884</td>
<td>2</td>
<td>217,682,779</td>
<td>GABPS</td>
<td>A/C (0.58)</td>
<td>Intrinsic</td>
<td>0.0049 1.65E - 08</td>
<td>0.0055 2.06E - 03</td>
<td>0.0053 1.33E - 10</td>
</tr>
<tr>
<td>rs6802041</td>
<td>3</td>
<td>170,091,902</td>
<td>SKIL</td>
<td>T/C (0.87)</td>
<td>Intrinsic</td>
<td>0.0067 1.36E - 07</td>
<td>0.0046 2.33E - 02</td>
<td>0.0068 2.58E - 08</td>
</tr>
<tr>
<td>rs10531801</td>
<td>3</td>
<td>185,222,353</td>
<td>ETVS</td>
<td>T/G (0.87)</td>
<td>Intrinsic</td>
<td>0.0070 3.80E - 09</td>
<td>0.0046 1.79E - 02</td>
<td>0.0072 1.03E - 09</td>
</tr>
<tr>
<td>rs10994860</td>
<td>10</td>
<td>6,247,624</td>
<td>ACCT</td>
<td>T/C (0.19)</td>
<td>Intrinsic</td>
<td>0.0075 1.00E - 11</td>
<td>0.0061 5.46E - 03</td>
<td>0.0077 1.07E - 12</td>
</tr>
<tr>
<td>rs163160</td>
<td>11</td>
<td>2,789,955</td>
<td>KCNQ1</td>
<td>A/G (0.82)</td>
<td>Intrinsic</td>
<td>0.0067 9.02E - 09</td>
<td>0.0050 9.89E - 03</td>
<td>0.0065 2.26E - 14</td>
</tr>
<tr>
<td>rs1647484</td>
<td>16</td>
<td>89,708,292</td>
<td>DPF4</td>
<td>C/G (0.53)</td>
<td>Intergenic</td>
<td>0.0047 9.19E - 09</td>
<td>0.0049 1.91E - 02</td>
<td>0.0046 9.15E - 18</td>
</tr>
<tr>
<td>rs8091180</td>
<td>18</td>
<td>77,164,243</td>
<td>NAPJC1</td>
<td>A/G (0.56)</td>
<td>Intergenic</td>
<td>0.0054 1.43E - 08</td>
<td>0.0052 5.46E - 03</td>
<td>0.0060 1.28E - 09</td>
</tr>
</tbody>
</table>

The 16 loci whose smallest P value was observed in the ‘overall’ group

1The eight loci whose smallest P value was observed in the ‘no diabetes’ group

To identify additional potentially associated variants and more formally evaluate trans-ethnic heterogeneity of the loci identified through meta-analysis in European ancestry populations, we performed a trans-ethnic meta-analysis22, combining the 12 African ancestry studies with the 48 European Ancestry studies used in the discovery analysis of eGFRcrea. Of the 24 new loci uncovered for eGFRcrea, 15 were also genome-wide significant in the trans-ethnic meta-analysis (defined as log_{10} Bayes Factor > 6, Supplementary Table 10), indicating that for most of these loci, there is little to no allelic effect heterogeneity across the two ethnic groups. No additional loci were significantly associated with log_{10} Bayes Factor > 6.

Bioinformatic and functional characterization of new loci

We used several techniques to prioritize and characterize genes underlying the identified associations, uncover connections between associated regions, detect relevant tissues and assign functional annotations to associated variants. These included expression quantitative trait loci (eQTL) analyses, pathway analyses, DNAse I hypersensitivity site (DHS) mapping, chromatin mapping, manual curation of genes in each region and zebrafish knockdown.

eQTL analysis. We performed eQTL analysis using publicly available eQTL databases (see Methods). These analyses connected novel SNPs to transcript abundance of SYPL2, SDCCA8, MANBA, KBTBD2, PTPRO and SPATA33 (C16orf55), thereby supporting these as potential candidate genes in the respective associated regions (Supplementary Table 11).

Pathway analyses. We used a novel method, Data-driven Expression Prioritized Integration for Complex Traits (DEPICT)23, to prioritize genes at associated loci, to test whether genes at associated loci are highly expressed in specific tissues or cell types and to test whether specific biological pathways and gene sets are enriched for genes in associated loci. On the basis of all SNPs with eGFRcrea association P values <10^{-5} in the discovery meta-analysis, representing 124 independent regions, we identified at least one significantly prioritized gene in 49 regions, including 9 of the 24 novel genome-wide significant regions (Supplementary Table 12). Five tissue and cell type annotations were enriched for expression of genes from the associated regions, including the kidney and urinary tract, as well as hepatocytes and adrenal glands and cortex (Fig. 3a; Supplementary Table 13). Nineteen reconstituted gene sets showed enrichment of genes mapping into the associated regions at a permutation P value <10^{-5} (Supplementary Table 14; Fig. 4), highlighting processes related to renal development, kidney membrane transporter activity, kidney and urogenital system morphology, regulation of glucose metabolism, as well as specific protein complexes important in renal development.
DNase I hypersensitivity and H3K4m3 chromatin mark analyses. To evaluate whether eGFRcrea-associated SNPs map into gene regulatory regions and to thereby gain insight into their potential function, we evaluated the overlap of independent eGFRcrea-associated SNPs with DHS values $<10^{-4}$ (or their proxies) with DHSs using publicly available data from the Epigenomics Roadmap Project and ENCODE for 123 cell types (see Methods). DHSs mark accessible chromatin regions where transcription may occur. Compared with a set of control SNPs (see Methods), eGFRcrea-associated SNPs were significantly more likely to map to DHS in six specific tissues or cell types (Fig. 3b), including adult human renal cortical epithelial cells, adult renal proximal tubule epithelial cells, H7 embryonic stem cells (differentiated 2 days), adult human renal epithelial cells, adult small airway epithelial cells and amniotic epithelial cells. No significant enrichment was observed for adult renal glomerular endothelial cells, the only other kidney tissue evaluated.

Next, we analysed the overlap of the same set of SNPs with H3K4m3 sites in six specific tissues or cell types (Fig. 3b), including adult human renal cortical epithelial cells, adult renal proximal tubule epithelial cells, H7 embryonic stem cells (differentiated 2 days), adult human renal epithelial cells, adult small airway epithelial cells and amniotic epithelial cells. No significant enrichment was observed for adult renal glomerular endothelial cells, the only other kidney tissue evaluated.

Chromatin annotation maps. In addition to assessing individual regulatory marks separately, we annotated the known and replicated novel SNPs, as well as their perfect proxies in a complementary approach. Chromatin annotation maps were generated integrating $>10$ epigenetic marks from cells derived from adult human kidney tissue and a variety of non-renal tissues from the ENCODE project (see Methods). The proportion of variants to which a function could be assigned was significantly higher when using chromatin annotation maps from renal tissue compared with using maps that investigated the same epigenetic marks in other non-renal tissues (Fig. 3c), again indicating that eGFRcrea associated SNPs are, or tag, kidney-specific regulatory variants. The difference between kidney and non-renal tissues was particularly evident for marks that define enhancers: the proportion of SNPs mapping to weak and strong enhancer regions in the kidney tissue was higher than in all non-kidney tissues (Fishers’ exact test P value 0.0117), and rectal mucosa (P value 0.0445). Taken together, these findings are suggestive of cell-type-specific regulatory roles for eGFR loci, with greatest specificity for the kidney.

Functional characterization of new loci. To prioritize genes for functional studies, we applied gene prioritization algorithms including GRAIL, DEPICT and manual curation of selected genes in each region (Supplementary Table 12). For each region, gene selection criteria were as follows: (1) either GRAIL P value $<0.05$ or DEPICT false discovery rate (FDR) $<0.05$; (2) the effect of a given allele on eGFRcrea and on eGFRcys was direction-consistent and their ratio was between 0.2 and 5.
enrichment of variants mapping to enhancer regions specifically in kidney but not other non-renal tissues.

P

logistic mixed effect model. Cell types indicated with coloured lines had nominally significant enrichment (* indicate

NFATC1

particular tissues, including kidney and urinary tract. Shown are permutation test

6

NATURE COMMUNICATIONS | 7:10023 | DOI: 10.1038/ncomms10023 | www.nature.com/naturecommunications

single gene. Using this approach, nearest gene if the signal was located in a region containing a

SNPs and their perfect proxies were annotated based on genomic location using chromatin annotation maps from different tissues including adult kidney

c
tubule epithelial cells; Saec, small airway epithelial cells). (Hae, amniotic epithelial cells; Hrce, renal cortical epithelial cells; Hre, renal epithelial cells; Hrgec, renal glomerular endothelial cells; Rptec, renal proximal tubule epithelial cells; Saec, small airway epithelial cells). (c) ENCODE/Chromatin ChIP-seq mapping: known and replicated novel eGFRcrea-associated SNPs and their perfect proxies were annotated based on genomic location using chromatin annotation maps from different tissues including adult kidney epithelial cells. P values from Fishers’ exact tests for 2 × 2 tables are reported (significance level = 5.6 × 10^{-3}, see Methods). There is significant enrichment of variants mapping to enhancer regions specifically in kidney but not other non-renal tissues.

(to ensure relative homogeneity of the beta coefficients); (3) nearest gene if the signal was located in a region containing a single gene. Using this approach, NFKB1, DPEP1, TSPAN9, NFATC1, WNT7A, PTPRO, SYPL2, UNCX, KBTBD2, SKIL and A1CF were prioritized as likely genes underlying effects at the new loci (Supplementary Table 12).

We investigated the role of these genes during vertebrate kidney development by examining the functional consequences of gene knockdown in zebrafish embryos utilizing antisense morpholino oligonucleotide (MO) technology. After knockdown, we assessed the expression of established renal markers pax2a (global kidney), nephrin (podocytes) and slc20a1a (proximal tubule) at 48 hours post fertilization by in situ hybridization11. In all cases, morphant embryos did not display significant gene expression defects compared with controls (Supplementary Table 15).

Discussion

We identified 24 new loci in association with eGFR and confirmed 29 previously identified loci. A variety of complementary analytic, bioinformatic and functional approaches indicate enrichment of implicated gene products in kidney and urinary tract tissues. A greater proportion of the lead SNPs or their perfect proxies map into gene regulatory regions, specifically enhancers, in adult renal tissues compared with non-renal tissues. In addition to the importance in the adult kidney, our results indicate a role for kidney function variants during development.

We extend our previous findings, as well as those from other groups7–13 by identifying >50 genomic loci for kidney function, many of which were not previously known to be connected to kidney function and disease. Using a discovery data set that is nearly double in size to our prior effort7, we are now able to robustly link associated SNPs to kidney-specific gene regulatory function. Our work further exemplifies the continued value of increasing the sample size of GWAS meta-analyses to uncover additional loci and gain novel insights into the mechanisms underlying common phenotypes26.

There are several messages from our work. First, many of the genetic variants associated with eGFR appear to affect processes specifically within the kidney. The kidney is a highly vascular and

Figure 3 | Bioinformatic analysis of eGFR-associated SNPs. Connection of eGFR-associated SNPs to gene expression and variant function across a variety of tissues, pathways and regulatory marks was considered. (a) The DEPICT method shows that implicated eGFR-associated genes are highly expressed in particular tissues, including kidney and urinary tract. Shown are permutation test P values (see Methods). (b) Enrichment of eGFRcrea-associated SNPs in DHS according to discovery P value thresholds. SNPs from the eGFR discovery genome-wide scan meeting a series of P value thresholds in the range 10^{-4}–10^{-16} preferentially map to DHSs, when compared with a set of control SNPs, in 6 of 123 cell types. Represented are main effects odds ratios from a logistic mixed effect model. Cell types indicated with coloured lines had nominally significant enrichment (* indicate P values <0.05) at the P value

<10^{-16} threshold and/or were derived from renal tissues (H7esDiff2d: H7 embryonic stem cells, differentiated 2 days with BMP4, activin A and bFGF; Hae, amniotic epithelial cells; Hrce, renal cortical epithelial cells; Hre, renal epithelial cells; Hrgec, renal glomerular endothelial cells; Rptec, renal proximal tubule epithelial cells; Saec, small airway epithelial cells). (c) ENCODE/Chromatin ChIP-seq mapping: known and replicated novel eGFRcrea-associated SNPs and their perfect proxies were annotated based on genomic location using chromatin annotation maps from different tissues including adult kidney epithelial cells. P values from Fishers’ exact tests for 2 × 2 tables are reported (significance level = 5.6 × 10^{-3}, see Methods). There is significant enrichment of variants mapping to enhancer regions specifically in kidney but not other non-renal tissues.
metabolically active organ that receives 20% of all cardiac output, contains an extensive endothelium-lined capillary network, and is sensitive to ischemic and toxic injury. As a result, hypertension, cardiovascular diseases and diabetes each affect renal hemodynamics and contribute to kidney injury. However, many of the eGFR-associated SNPs in our GWAS could be assigned gene regulatory function specifically in the kidney and its epithelial cells, but not in human glomerular endothelial cells or the general vasculature. In addition, variants associated with eGFR were not associated with vascular traits, such as blood pressure or myocardial infarction. Taken together, these findings suggest that genetic determinants of eGFR may be mediated largely through direct effects within the kidney.

Second, despite the specificity related to renal processes, we also identified several SNPs that are associated with eGFR in diabetes, and our pathway analyses uncovered gene sets associated with glucose transporter activity and abnormal glucose homeostasis. Uncovering bona fide genetic loci for diabetic CKD has been difficult. We have now identified a total of 19 SNPs that demonstrate at least nominal association with eGFR in diabetes. The diabetes population is at particularly high risk of CKD, and identifying kidney injury pathways may help develop new treatments for diabetic CKD.

Finally, even though CKD is primarily a disease of the elderly, our pathway enrichment analyses highlight developmental processes relevant to the kidney and the urogenital tract. Kidney disease has been long thought to have developmental origins, in part related to early programming (Barker hypothesis)\(^2^7\), low birth weight, nephron endowment and early growth and early-life nutrition\(^2^8\). Our pathway enrichment analyses suggest that developmental pathways such as placental morphology, kidney weight and embryo size, as well as protein complexes of importance in renal development may in part contribute to the developmental origins of CKD.

Figure 4 | Gene set overlap analysis. The 19 reconstituted gene sets with \(P\) value < \(10^{-5}\) were considered. Their overlap was estimated by computing the pairwise Pearson correlation coefficient \(\rho\) between each pair of gene sets followed by discretization into one of three bins: \(0.3 \leq \rho < 0.5\), low overlap; \(0.5 \leq \rho < 0.7\), medium overlap; \(\rho \geq 0.7\), high overlap. Overlap is shown by edges between gene set nodes and edges representing overlap corresponding to \(\rho < 0.3\) are not shown. The network was drawn with Cytoscape\(^4^8\).

phenotype upon gene knockdown does not mean that the gene cannot be the one underlying the observed association signal in humans. Finally, our conclusions that eGFRcrea-associated SNPs regulate the expression of nearby genes specifically in kidney epithelial cells are based on DHSs, H3K4me3 chromatin marks and chromatin annotation maps. Since these analyses rely mostly on variant positions, additional functional investigation such as luciferase assay that assess transcriptional activity more directly are likely to gain additional insights into the variants’ mechanism of action.

The kidney specificity for loci we identified may have important translational implications, particularly since our DHS and chromatin annotation analyses suggest that at least a set of gene regulatory mechanisms is important in the adult kidney. Kidney-specific pathways are important for the development of novel therapies to prevent and treat CKD and its progression with minimal risk of toxicity to other organs. Finally, the biologic insights provided by these new loci may help elucidate novel mechanisms and pathways implicated not only in CKD but also of kidney function in the physiological range.

In conclusion, we have confirmed 29 genomic loci and identified 24 new loci in association with kidney function that together highlight target organ-specific regulatory mechanisms related to kidney function.

Methods

Overview. This was a collaborative meta-analysis with a distributive data model. Briefly, an analysis plan was created and circulated to all participating studies. Studies then uploaded study-specific data centrally; files were cleaned, and a specific meta-analysis for each trait was performed. Details regarding each step are provided below. All participants in all discovery and replication studies provided informed consent. Each study had its research protocol approved by the local ethics committee.

Phenotype definitions. Serum creatinine was measured in each discovery and replication study as described in Supplementary Tables 16 and 17, and statistically calibrated to the US nationally representative National Health and Nutrition Examination Study data in all studies to account for between-laboratory variation\(^2^9\).\(^3^0\). eGFRcrea was estimated using the four-variable Modification of Diet in Renal Disease Study Equation. Cystatin C, an alternative biomarker for kidney function, was measured in a sub-set of participating studies. eGFRcrea was estimated as 76.7 \times (serum cystatin C)\(^{-1.19}\) (ref. 31). eGFRcrea and eGFRcys values
Genotypes. Genotyping was conducted in each study as specified in Supplementary Tables 18 and 19. After applying appropriate quality filters, 45 studies used markers of highest quality to impute Supplementary Tables 18 and 19. After applying appropriate quality filters, 45 results was applied. Between-study heterogeneity was assessed through the and s.e. representing the effect of each SNP on the phenotype and its standard associations of the beta coefficients in stage 1 and stage 2 (one-sided Stage 1 discovery meta-analysis on SNP dosage levels adjusting for sex and age. For all traits, adjustment for diabetes, 4 from CKD and 1 from eGFRcys. Three criteria were used to ensure validity of novel loci declared as having fixed effects and using inverse-variance weighting. The genomic inflation factor was estimated for each study as the ratio between the median of all observed test values and the expected median of a value from the combined meta-analysis. The proportion of phenotypic variance explained by each SNP was calculated as the ratio of the variance explained by the SNP to the total variance of the phenotype. This value was estimated to test for consistent effect direction with the discovery stage; (3) $p$-value $< 0.05$ in the replication stage. $Q$-values were estimated using the package VALSE in R. The tuning parameter lambda for the estimation of the overall proportion of true null hypotheses, $\alpha_0$, was estimated using the bootstrap method17. When the third criterion was not satisfied, the locus was declared 'suggestive'.

Genome-wide association analysis. By following a centralized analysis plan, each study regressed sex- and age-adjusted residuals of the logarithm of eGFRcrea or eGFRcys on SNP dosage levels. Logistic regression of CKD status was performed on SNP dosage levels adjusting for sex and age. For all traits, adjustment for diabetes, 4 from CKD and 1 from eGFRcys. Three criteria were used to ensure validity of novel loci declared as having fixed effects and using inverse-variance weighting. The genomic inflation factor was estimated for each study as the ratio between the median of all observed test values and the expected median of a value from the combined meta-analysis. The proportion of phenotypic variance explained by each SNP was calculated as the ratio of the variance explained by the SNP to the total variance of the phenotype. This value was estimated to test for consistent effect direction with the discovery stage; (3) $p$-value $< 0.05$ in the replication stage. $Q$-values were estimated using the package VALSE in R. The tuning parameter lambda for the estimation of the overall proportion of true null hypotheses, $\alpha_0$, was estimated using the bootstrap method17. When the third criterion was not satisfied, the locus was declared 'suggestive'.

Stage 1 discovery meta-analysis. GWAS of eGFRcrea were contributed by 48 studies (total sample size, $N = 133,413$); 45 studies contributed GWAS data for the non-diabetes stratum ($N = 118,448$) and 39 for the diabetes stratum ($N = 13,512$). GWAS of CKD were comprised by 43 studies, for a total sample size of 117,165, including 12,385 CKD cases. GWAS of eGFRcys were comprised by 16 studies for a total sample size of 32,834. All GWAS files underwent quality control using the GWAtoolbox package21 in R, before including them into the meta-analysis. Genomic-control (GC) correction was performed with the software METAL35, assuming fixed effects and using inverse-variance weighting. The genomic inflation factor $\lambda$ was estimated for each study as the ratio of the median of all observed test statistics ($b_i/s_i^2$) and the expected median of a $\chi^2$ with 1 degree of freedom, with $b_i$ and $s_i$ representing the effect of each SNP on the phenotype and its standard error, respectively. GC correction was applied to $P$ values and $s_i$'s in case of $\lambda > 1$ (first GC correction). SNPs with an average minor allele frequency (MAF) of $\geq 0.01$ were used for the meta-analysis. To limit the possibility of false positives, after the meta-analysis, a second GC correction on the aggregated results was applied. Between-study heterogeneity was assessed through the $P$ statistic.

After removing SNPs with MAF of $< 0.05$ and which were available in $< 50$% of the studies, SNPs with a $P$ value of $\leq 10^{-6}$ were selected and clustered into independent loci through LD pruning based on an $r^2$ of $\leq 0.2$ within a window of $\pm 1$ MB to each side of the index SNP. After removing loci containing variants that have been previously replicated at a $P$ value of $5.0 \times 10^{-8}$ (refs 7,8), the SNP with the lowest $P$ value within each locus was selected for replication (‘index SNP’). If a SNP had an association $P$ value of $\leq 10^{-6}$ with more than one trait, the trait where the SNP had the lowest $P$ value was selected as discovery trait/stratum. Altogether, this resulted in 48 SNPs: 34 from eGFRcrea, 9 from eGFRcys among those without diabetes, 4 from CKD and 1 from eGFRcys.

Stage 2 replication analysis. In silico replication analysis for any of the studied traits was carried out using eight independent studies whose genotyping platforms are provided in Supplementary Table 19. De novo genotyping was performed in seven additional studies ($N = 22,850$ individuals) of European ancestry (Supplementary Table 20), including the Bus Santé, ESTHER, KORA-F3 (subset of F3 without GWAS), KORA-F4 (subset of F4 without GWAS), Ogliastra Genetic (Supplementary Table 20), including the Bus Santé, ESTHER, KORA-F3 (subset of F3 without GWAS), KORA-F4 (subset of F4 without GWAS), Ogliastra Genetic (Supplementary Table 20). Significance was evaluated at the Bonferroni corrected level of 0.05 in the replication stage. Q-values were estimated using the package VALSE in R. The proportion of phenotypic variance explained by each novel and known loci was estimated as $\sum R_i^2$, where $R_i^2 = b_i^2/\text{var(SNP)}$. $b_i$ is the coefficient of determination for each of the 53 individual SNPs associated with eGFRcrea uncovered to date (24 novel and 29 known ones), $\text{var(SNP)} = 2 \times MAF_{\text{SNP}} \times (1 - MAF_{\text{SNP}})$20. Var($y$) was estimated in the ARIC study and all loci were assumed to have independent effects on the phenotype.

Test for SNP associations with related traits. We performed evaluations of SNP association with results generated from consortia investigating other traits. Specifically, we evaluated systolic and diastolic blood pressure in ICBP17, myocardial infarction in CARDIOGRAM18, left ventricular mass19, heart failure20, the urinary albumin to creatinine ratio (CKDGen consortium, personal communication) and fasting plasma glucose in MAGIC21. In total, we performed 165 tests, corresponding to 7 traits tested for association against each of the 24 novel SNPs, with the exception of myocardial infarction for which results from 3 SNPs were not available (Supplementary Table 6). Significance was evaluated at the Bonferroni corrected level of 0.05/165 = 0.0003.

Lookup of replicated loci in the NHGRI GWAS catalog. All replicated SNPs, as well as SNPs in LD ($r^2 > 0.2$) within $\pm 1$ MB distance were checked for their association with other traits according to the NHGRI GWAS catalog40 (accessed April 14, 2014).

SNP assessments in other ethnic groups. We performed cross-ethnicity SNP evaluations in participants of African ancestry from a meta-analysis of African ancestry individuals and from participants of Asian descent from the AGEN consortium41.

African ancestry meta-analysis. We performed fixed-effect meta-analysis of the genome-wide association data from 12 African ancestry studies (Supplementary Table 8) with imputation to HapMap reference panel, based on inverse-variance weighting using METAL. Only SNPs with MAF $\geq 0.01$ and imputation quality $r^2 \geq 0.3$ were considered for the meta-analysis. After meta-analysis, we removed SNPs with MAF $< 0.05$ and which were available in $< 50$% of the studies. Statistical significance was assessed at the standard threshold of $5.0 \times 10^{-8}$. Genomic control correction was applied at both the individual study level before meta-analysis and after the meta-analysis. Transethnic meta-analysis. We performed a trans-ethnic meta-analysis of GWAS data from cohorts of different ethnic backgrounds using MANTRA (Meta-Analysis
of Trans-ethnic Association studies) software. We combined the 48 European ancestry studies that contributed eGFRcrea, which were included in stage 1 discovery meta-analysis, and the 12 African ancestry studies mentioned above for a total sample size of 150,253 samples. We limited our analysis to biallelic SNPs with MAF ≥ 0.01 and imputation quality r2 ≥ 0.3. Relatedness between the 60 studies was estimated using default settings from up to 5.9 million SNPs. Only SNPs that were present in more than 25 European ancestry studies and 6 African ancestry studies (total sample size ≥ 120,000) were considered after meta-analysis. Genomic-wide significance was defined as a log10 Bayes’ Factor (log[BF]) ≥ 6 (ref. 41).

Gene Relationships Across Implicated Loci (GRLAI). To prioritize the gene(s) most likely to give rise to association signals in a given region, the software GRLAI was used.22 The index SNP of all previously known kidney function associated regions, as well as the novel SNPs identified here was used as input, using the CEU HapMap (hg18 assembly) and the functional datasource text_2009_03, established before the publication of kidney function-related GWAS. Results from GRLAI were used to prioritize genes for follow-up functional work.

Expression quantitative trait loci analysis. We identified alias rsID and proxies (r2 > 0.8) for our index SNPs using SNAP software across 4 HapMap builds. SNAP rsIDs and aliases were searched for primary SNPs and LD proxies against a reconstituted gene sets that are enriched in genes from associated regions and DEPICT software program on the 124 regions to prioritize genes that may be experimental replicates (typically duplicates) for the majority of cells and tissues.

DEPICT analysis. In this work, we first used PLINK42 to identify independently associated SNPs using all SNPs with eGFRcrea association P values < 10−5 (HapMap release 27 CEU data41; LD r2 threshold = 0.01; physical kb threshold = 1,000). We then used the DEPICT method23 to construct associated regions by mapping to independently associated SNPs if they overlapped or resided within 5 kb of a given associated region. After mapping overlapping regions and discarding regions that mapped within the major histocompatibility complex locus (chromosome 6, base pairs 20,000,000–40,000,000), 124 non-overlapping regions remained that covered a total of 320 genes. Finally, we ran the DEPICT software program on the 124 regions to prioritize genes that may represent promising candidates for experimental follow up studies, identify reconstituted gene sets that are enriched in genes from associated regions and therefore may provide insight into general kidney function biology, and identify tissue and cell-type annotations in which genes from associated regions are highly expressed. Specifically, for each tissue, the DEPICT method performs a t-test comparing the tissue-specific expression of eGFRcrea-associated genes and all other genes. Next, for each tissue, empirical enrichment P values are computed by repeatedly sampling random sets of loci (matched to the actual eGFRcrea loci by gene density) from the entire genome to estimate the empirical mean and s.d. of the enrichment statistic’s null distribution. To visualize the nineteen reconstituted gene sets with enrichment statistic’s null distribution. To visualize the nineteen reconstituted gene sets, we repeatedly sampling random sets of loci (matched to the actual eGFRcrea loci by gene density) from the entire genome to estimate the empirical mean and s.d. of the enrichment statistic’s null distribution.

DNase I hypersensitivity analysis. The overlap of SNPs associated with eGFRcrea at P < 10−4 with DHSs was examined using publicly available data from the ENCODE Project and Epigenomaps. In all, DHS mappings were available for 123 mostly adult cells and tissues44. The overlap of DNase I hypersensitive sites from the 1st Genomes EUR panel (phase 1 v3, 1/14/2014 haplotypes) using a series of P value thresholds (10−4, 10−6, 10−8, ... and 10−16); LD proxies of the index SNPs from the clumping procedure were then identified by LD tagging in PLINK with r2 = 0.8 in windows of 100 kb, again using LD relationships in the 1000G EUR panel, restricted to SNPs with MAF ≥ 0.01 and also present in HapMap2 CEU population. A reference set of control SNPs was constructed using the same clumping and tagging procedures applied to NHGRI GWAS catalog SNPs (available at http://www.genome.gov/gwastudies/, accessed 13 March 2013) with discovery P values ≤ 5.0 × 10−8 in European populations. In total, there were 1,204 such reference SNPs after LD pruning. A small number of reference SNPs or their proxies overlapping with the eGFRcrea SNPs or their proxies were excluded. For each cell-type and P value threshold, the enrichment of eGFRcrea SNPs (or their LD proxies) mapping to DHSs relative to the GWAS catalog reference SNPs (or their LD proxies) was expressed as an odds ratio from logistic mixed effect models that treated the repeated peak determinations as random effects (limma package in R). Significance for enrichment odds ratio was derived from the significance of beta coefficients for the main effects in the mixed models.

Interrogation of human kidney chromatin annotation maps. Different articles on chromatin modification patterns can be used to generate tissue-specific chromatin-state annotation maps. These annotation maps provide valuable resource to discover regulatory regions and their cell-type-specific distributions and activities, which may help with the interpretation especially of intergenic variants identified in association studies24. We therefore investigated the genomic mapping of the known and replicated novel index SNPs, as well as their perfect LD proxies (n = 173, r2 = 1 for proxies) using a variety of resources, including chromatin maps generated from human kidney tissue cells (HKC-E cells). Chromatin immune-precipitation sequencing (ChIP-seq) data from human kidney samples were generated by NIH Roadmap Epigenomics Mapping Consortium46. Briefly, proximal tubule cells derived from adult human kidney were collected and subjected to ChIP-seq with 1% formaldehyde. Subsequently, ChIP-seq was conducted using whole-cell extract from adult kidney tissue as the input (GS621638) and assessing the following chromatin marks: H3K36me3 (GS621634), H3K4me1 (GS670025), H3K4me3 (GS621648), H3K9ac (GS772811) and H3K9me3 (GS621651). The MACS version 1.4.1 (model-based analysis of ChIP-Seq) peak finding algorithm was used to identify regions of ChIP-Seq enrichment47. A FDR threshold of enrichment of 0.01 was used for all data sets. The resulting genomic coordinates in bed format were further used in ChromHMM v1.06 for chromatin annotation48. For comparison, the same genomic coordinates were investigated in chromatin annotation maps of regulatory regions as well as nine different cell lines from the ENCODE Project: umbilical vein endothelial cells (HUVEC), mammary epithelial cells (HMEC), normal epidermal keratinocytes (NHKEK), B-lymphoblastoid cells (GM12878), erythroytic leukemia cells (K562), normal lung fibroblasts (NHFL), skeletal muscle myoblasts (HSMIM), embryonic stem cells (H1 ES) and hepatocellular carcinoma cells (HepG2). We tested whether the proportion of SNPs pointing to either strong or weak enhancers in the human kidney cell line was different from that of the other nine tissues by means of a Fisher’s exact test for 2 × 2 tables, contrasting each of the nine cell lines listed above against the reference kidney cell line, at a Bonferroni-corrected significance level of 0.05/9 = 5.6 × 10−3.

Functional characterization of new loci. Replicated gene regions were prioritized for functional studies using the following criteria: (1) GRLAI identification of a gene in each region of P value <0.05 or DEPICT, FDR <0.05; (2) an eGFRcrea to eGFRcys ratio between 0.2 and 5 with direction consistency between the beta coefficients; (3) nearest gene if the signal was located in a gene-poor region. The list of genes selected for functional work can be found in Supplementary Table 12. This same prioritization scheme was also used to assign locus names. Morpholino knockdowns were performed in zebrafish.

Zebrafish (strain Tubingen, ZF) were maintained according to established Harvard Medical School Institutional Animal Care and Use Committee protocols (protocol # 04626). Male and female fish were mated (age 6–12 months) for embryo production. Embryos were injected at the one-cell stage with MOs (GeneTools) designed to block either the ATG start site or an exon–intron splice site of the target gene (Supplementary Table 21). In cases where human loci are duplicated in zebrafish, both orthologues were knocked down simultaneously by combination MO injection. MOs were injected in escalating doses at concentrations up to 250 μM. Embryos were fixed in 4% paraformaldehyde at 48 hpostfertilization for in situ hybridization using probes (Sigma-Aldrich) and in Zen2/Zeiss/ThiaseProtocol.html. Gene expression was visualized using established renal markers pa2a (global kidney), nephrin (podocytes) and scl20a1a (proximal tubule). The number of morphant embryos displaying abnormal gene expression was compared with control embryos by means of a Fisher’s exact test.

References


Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: J.T. and P.H. are consultants for Servier. J.C. received research grants and honoraria from Servier. K.S. obtained research support from Boehringer Ingelheim. The remaining authors declared no competing financial interests.

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

How to cite this article: Pattaro, C. et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat. Commun. 7:10023 doi: 10.1038/ncomms10023 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

168 Center for Statistical Genetics, Department of Biostatistics, University of Washington, Seattle, WA 98195, USA. 169 Department of Nutrition, University of North Carolina, Chapel Hill, North Carolina 27599, USA. 170 Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK. 171 Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. 172 Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. 173 Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. 174 Center for Human Genetic Research, Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. 175 Geriatric Rehabilitation Unit, Azienda Sanitaria Firenze (ASF), 50125 Florence, Italy. 176 Département de Génétique Médicale, Université de Lausanne, Lausanne 1015, Switzerland. 177 Pathology and Laboratory Medicine, University of Western Australia, 6009 Crawley, Western Australia, Australia. 178 Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. 179 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle 98195, Washington, USA. 180 National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. 181 Department of Medicine III, Medical Faculty Carl Gustav Carus at the Technical University of Dresden, Dresden 01307, Germany. 182 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Heidelberglaan 100, Utrecht 3508 GA, The Netherlands. 183 Leibniz-Institute for Artherosclerosis Research, Department of Molecular Genetics of Cardiovascular Disease, University of Münster, Münster, Germany. 184 University Hospital Münster, Internal Medicine D, Münster, Germany. 185 Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester LE3 9QP, UK. 186 Clinical Pharmacology Unit, University of Cambridge, Addenbrookes Hospital, Hills Road, Cambridge CB2 0QQ, UK. 187 Department of Health Sciences, University of Leicester, University Rd, Leicester LE1 7RH, UK. 188 Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. 189 Clinical Pharmacology and The Genome Centre, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. 190 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK. 191 Centre for Cellular and Molecular Biology (CCMB), Council of Scientific and Industrial Research (CSIR), Uppal Road, Hyderabad 500 007, India. 192 University of Maryland, School of Medicine, Baltimore, Maryland 21201, USA. 193 School of Science and Engineering, University of Ballarat, Ballarat 3353, Australia. 194 International Centre for Circulatory Health, National Heart & Lung Institute, Imperial College, London, UK. 195 Center for Genome Science, Imperial College, London, UK. 196 Genome Institute of Singapore, National Institute of Health, Singapore. 197 School of Medical Sciences, University of New South Wales, Sydney, NSW 2052, Australia. 198 Department of Cardiovascular and Pulmonary Medicine, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. 199 Department of Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. 200 Department of Epidemiology, University of Washington, Seattle, Washington 98195, USA. 201 Department of Clinical Epidemiology, University of Southern Denmark, Odense, Denmark. 202 Department of Clinical Epidemiology, University of Southern Denmark, Odense, Denmark. 203 Medical Faculty Carl Gustav Carus, Technische Universität Dresden, 01307 Dresden, Germany. 204 Department of Public Health and Community Medicine, Institute of Medicine, University of Western Australia, 6009 Crawley, Western Australia, Australia. 205 Cardiovascular Genetics, University of Utah, School of Medicine, Salt Lake City, UT 84132, USA. 206 Division of Community Health Sciences, St George's University, London SW17 0RE, UK. 207 Cardiovascular Research Unit, Department of Medicine, University of Oxford, Oxford OX3 7BN, UK. 208 Department of Cardiovascular Medicine, Institute Municipal d’Investigacio Medica, Barcelona Biomedical Research Park, 88 Doctor Aiguader, Barcelona 08003, Spain. 209 Medizinische Klinik II, Universität zu Lübeck, 23562 Lübeck, Germany. 210 Department of Statistics, Pontificia Universidad Catolica de Chile, Vicuña Mackena, Santiago 8460, Chile. 211 Department of Cardiovascular Medicine, The Welcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. 212 Department of Clinical Sciences, Lund University, SE-205 02 Malmö, Sweden. 213 Department of Internal Medicine, University of the West Indies, Mona, Jamaica. 214 Department of Medicine, University of Mississippi Medical Center, 2500 North State St, Jackson, Mississippi 39216, USA. 215 Department of Preventive Medicine and Epidemiology, Loyola University Medical School, Maywood, Illinois, USA. 216 Tuscany Regional Health Agency, Florence, Italy. 217 Department of Internal Medicine, The University of Chicago, Chicago, Illinois 60611, USA. 218 Occupational and Environmental Medicine, Department of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg 40530, Sweden. 219 MRC Centre for Causal Analyses in Translational Epidemiology, School of Social & Community Medicine, University of Bristol, Bristol BS8 2BN, UK. 220 BHF Glasgow Cardiovascular Research Centre, University of Glasgow, 126 University Place, Glasgow G12 8TA, UK. 221 Georgia Prevention Institute, Department of Pediatrics, Medical College of Georgia, 30912 Augusta, Georgia, USA. 222 Cardiovascular Epidemiology and Genetics, Institut Municipal d’Investigacio Medica, Barcelona Biomedical Research Park, 88 Doctor Aiguader, Barcelona 08003, Spain. 223 Cardiovascular Medicine, University of Washington, Seattle, Washington 98195, USA. 224 Cardiovascular Genetics, University of Utah, School of Medicine, Salt Lake City, UT 84132, USA. 225 Cardiovascular Medicine, University of Michigan, School of Medicine, South Quad, 9100 E. Ann Arbor, Michigan 48109, USA. 226 Department of Internal Medicine, Division of Cardiovascular Medicine, University of Utah, School of Medicine, Salt Lake City, UT 84132, USA. 227 Department of Public Health and Community Medicine, Institute of Medicine, University of Western Australia, 6009 Crawley, Western Australia, Australia. 228 Cardiovascular Genetics, University of Utah, School of Medicine, Salt Lake City, UT 84132, USA. 229 Cardiovascular Medicine, University of Utah, School of Medicine, Salt Lake City, UT 84132, USA. 230 Cardiovascular Research Institute, Department of Medicine, University of Utah, School of Medicine, Salt Lake City, UT 84132, USA.
Biostatistics, Washington University School of Medicine, Saint Louis, Missouri 63110, USA. 275Center for Non-Communicable Diseases, 74800 Karachi, University of Technology, Institute of Biomedical Engineering, Ehitajate tee 5, Tallinn 19086, Estonia. 291Department of Epidemiology and Public Health, UCL, London WC1E 6BT, UK. 241Department of Medicine, University of Kuopio, Kuopio University Hospital, Kuopio 70220, Finland. 242Studies Coordinating Centre, Division of Hypertension and Cardiac Rehabilitation, Division of Cardiovascular Diseases, Campus Immuno RIST, P.O. Box 100, 02029 Kuopio, Finland. 243Studies Coordinating Centre, Division of Hypertension and Cardiac Rehabilitation, Division of Cardiovascular Diseases, Campus Immuno RIST, P.O. Box 100, 02029 Kuopio, Finland. 244Finnish Institute of Occupational Health, Department of Epidemiology and Biostatistics, Case Western Reserve University, 2103 Cornell Road, Cleveland, Ohio 44106, USA. 245Gerontological Research Center, National Institute on Aging, King’s College London, SE1 7EH London, UK. 253Department of Genetics, Biochemistry, University of Toronto, Via Sanenta 19, Toronto M5J 2N4, Canada. 254ALSPAC Laboratory, University of Bristol, Bristol BS8 2BN, UK. 255Institute of Human Genetics, Helmholtz Zentrum Munich, German Research Center for Environmental Health, Neuherberg 82154, Germany. 256U872 Institut National de la Santé et de la Recherche Médicale, Centre de Recherches du Cordeliers, 75006 Paris, France. 257Mayo Clinic, Jacksonville, 32224 Florida, USA. 258Department of Basic Medical Research and Education, Ethene University Graduate School of Medicine, Toon 791-0295, Japan. 259Department of Geriatric Medicine, Ehime University Graduate School of Medicine, Toon 791-0295, Japan. 260Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA. 261Division of Genetics, GlaxoSmithKline, Philadelphia, Pennsylvania 19101, USA. 262Department of Primary Care and Population Health, UCL, London NW3 2PF, UK. 263Department of Medicine (Geriatrics), University of Mississippi Medical Center, Jackson, 38216 Mississippi, USA. 264Laboratory of Cardiovascular Science, Intramural Research Program, National Institute on Aging, NIH, Baltimore, 21224-6825 Maryland, USA. 265Department of Geriatric Medicine, Osaka University Graduate School of Medicine, Suita 565-0871, Japan. 266Tokoh University, Graduate School of Pharmaceutical Sciences and Medicine, Sendai 980-8578, Japan. 267Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore. 268Istituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche, Cittadella Universitaria di Monserrato, 09042 Monserrato, Cagliari, Italy. 269Columbia University, New York, 10027 New York, USA. 270Armen, 1 Kendall Square, Building 100, Cambridge, Massachusetts 02139, USA. 271National Institute of Health and Welfare, Helsinki 00271, Finland. 272Merck Research Laboratory, 126 East Lincoln Avenue, Rahway, New Jersey 07065, USA. 273South Asia Network for Chronic Disease, Public Health Foundation of India, C-1/5, SDA, New Delhi 1100016, India. 274Division of Biostatistics, Washington University School of Medicine, Saint Louis, Missouri 63110, USA. 275Center for Non-Communicable Diseases, 74800 Karachi, University of Technology, Institute of Biomedical Engineering, Ehitajate tee 5, Tallinn 19086, Estonia. 276Department of Biostatistics, University of Washington, Seattle, 98105 Washington, USA. 277Department of Emergency and Cardiovascular Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg 40685, Sweden. 278Epidemiology & Prevention, Division of Public Health Sciences, Wake Forest University, School of Medicine, Winston-Salem, North Carolina 27157, USA. 279University of Ibadan, PMB 5077 Ibadan, Nigeria. 280Prevention and Care of Diabetes, Department of Medicine III, Medical Faculty Carl Gustav Carus at the Technical University of Dresden, Dresden 01307, Germany. 281Department of Laboratory Medicine, Institute of Human Genetics, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, California 94143, USA. 282Imperial College Cardiovacular Unit (ICCRU), Imperial College, London W6 8RF, UK. 283Centre for Molecular Epidemiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore. 284Department of Medicine, Yong Loo Lin School of Medicine, National University Hospital, Singapore 119074, Singapore. 285Faculty of Medicine, University of Split, 21000 Split, Croatia. 286Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Oslo 0317, Norway. 287Diabetes Prevention Unit, National Institute for Health and Welfare, Helsinki 00271, Finland. 288Lifestyle-related Disease Prevention Center, Shiga University of Medical Science, Otsu 520-2192, Japan. 289Department of Medical Science and Cardiorenal Medicine, Yokohama City University School of Medicine, Yokohama 236-0044, Japan. 290Tallinn University of Technology, Institute of Biomedical Engineering, Ehitajate tee 5, Tallinn 109086, Estonia. 291Department of Epidemiology and Public Health, Imperial College, Norfolk Place, London W1 2PG, UK. 292Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, Lausanne 1011, Switzerland. 293Department of Medicine, Johns Hopkins University, Baltimore, 21205 Maryland, USA. 294Lapland Central Hospital, Department of Physics, Box 8041, Roivaniemi 96101, Finland. 295Department of Internal Medicine, Diabetology, and Nephrology, Medical University of Silesia, Zabrze 41-800, Poland.

**AGEN Consortium**


296Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan. 297Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo 113-8510, Japan. 298Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan. 299School of Chinese Medicine, China Medical University, Taichung 404, Taiwan. 300Cardiovascular Institute and Fu Wai Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100037, China. 301Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo 162-8655, Japan. 302Laboratory for Endocrinology and Metabolism, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan. 303Laboratory for Medical Informatics, RIKEN
Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital, Boston, 02114 Massachusetts, USA.

Division for Human Genetics, Massachusetts General Hospital, Boston, 02114 Massachusetts, USA.

Division of Research, Kaiser Permanente, Oakland, 94611 California, USA.

Institute for Human Genetics, University of California, San Francisco, San Francisco, 94143 California, USA.

Department of Cardiovascular Medicine, Cleveland Clinic, 7255 Old Oak Blvd, Cleveland, Ohio 44130, USA.

Medizinische Klinik und Poliklinik, Johannes-Gutenberg Universität Mainz, Universitätsmedizin, 55122 Mainz, Germany.

Institut für Klinische Chemie und Laboratoriumsmedizin, Johannes-Gutenberg Universität Mainz, Universitätsmedizin, 55122 Mainz, Germany.

INSERM UMR 937, Pierre et Marie Curie University (UPMC, Paris 6) and Medical School, 75005 Paris, France.

Boston University, School of Public Health, Boston, 02118 Massachusetts, USA.

University of Minnesota School of Public Health, Division of Epidemiology and Community Health, School of Public Health (A.R.F.), Minneapolis, 55455 Minnesota, USA.

University of Washington, Department of Internal Medicine, Seattle, 98195-6420 Washington, USA.

University of Texas, School of Public Health, Houston, 77030 Texas, USA.

Department of Medicine, Landspitali University Hospital, Reykjavik 101, Iceland.

Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany.

Division of Endocrinology and Diabetes, Graduate School of Molecular Endocrinology and Diabetes, University of Ulm, 89069 Ulm, Germany.

Division of Endocrinology, Department of Medicine, Medical University of Graz, 8010 Graz, Austria.

Synlab Center of Laboratory Diagnostics Heidelberg, 69037 Heidelberg, Germany.

Division of Clinical Chemistry, Department of Medicine, Albert Ludwigs University, 79085 Freiburg, Germany.

Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University Graz, 8010 Graz, Austria.

Cardio Group Frankfurt-Sachsenhausen, 60594 Frankfurt, Germany.

The Center for Applied Genomics, Children’s Hospital of Philadelphia, 19104 Philadelphia, Pennsylvania, USA.

Cardiovascular Research Institute, Medstar Health Research Institute, Washington Hospital Center, Washington, DC 20010, USA.

Genetics Division and Drug Discovery, GlaxoSmithKline, King of Prussia, Pennsylvania 19406, USA.

The Institute for Translational Medicine and Therapeutics, School of Medicine, University of Pennsylvania, Philadelphia, 19104-5158 Pennsylvania, USA.

Department of Cardiovascular Surgery, University of Leicester, Leicester LE1 7RH, UK.

Division of Cardiovascular and Diabetes Research, Multidisciplinary Cardiovascular Research Centre, Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds LS2 9JT, UK.

Institute for Community Medicine, Ernst-Moritz-Arndt-Universität Greifswald, 17475 Germany.

Department of Cardiology, University of Maryland Hospital, Baltimore, Maryland 21201, USA.

Institute of Medical Biometry, Epidemiology, and Informatics, Johannes Gutenberg University, Mainz 55101, Germany.

Institute for Community Medicine, Ernst-Moritz-Arndt-Universität Greifswald, 17475, Germany.