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.

Correspondence and requests for materials should be addressed to C.P. (email: cristian.pattaro@eurac.edu) or to A.K. (email: anna.koettgen@uniklinik-freiburg.de) or to C.S.F. (email: foxca@nhlbi.nih.gov).

#A full list of authors and their affiliations appears at the end of the paper.
Chronic kidney disease (CKD) is a global public health problem\textsuperscript{1-3}, and is associated with an increased risk for cardiovascular disease, all-cause mortality and end-stage renal disease\textsuperscript{4,5}. Few new therapies have been developed to prevent or treat CKD over the past two decades\textsuperscript{6,8}, 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\textsuperscript{7-12}. Subsequent functional investigations point towards clinically relevant novel mechanisms in CKD that were derived from initial GWAS findings\textsuperscript{13}, 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\textsuperscript{-1} per 1.73 m\textsuperscript{2}) 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/CKDDGen/. We performed analyses in each study sample in the overall population and stratified 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\textsuperscript{7-9} (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 \( P \) value < 1.0 \( \times \) 10\textsuperscript{8}. Of these 48 novel SNPs, 21 were genome-wide significant with \( P \) values < 5.0 \( \times \) 10\textsuperscript{8}

Overall, 43 SNPs were identified in association with eGFRcrea (nine in the non-diabetes sample), one with eGFRcys and four with CKD, as reported in Supplementary Table 3. Manhattan plots for CKD, eGFRcrea and eGFRcys 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 \( \times \) 10\textsuperscript{-8} (Table 1). Of these, 23 fulfilled additional replication criteria (\( q \)-value < 0.05 in stage 2). Only rs67195744 at the WNT7A locus demonstrated suggestive replication (\( P \) value < 5.0 \( \times \) 10\textsuperscript{-6}, \( q \)-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 UMOD locus showed genome-wide significant association with eGFRcrea among those with diabetes (Supplementary Fig. 2; rs12917707, \( P \) value = 2.5 \( \times \) 10\textsuperscript{-8}), and six loci (NFKB1, UNCX, TSPAN9, AP5B1, SIPA1L3 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\textsuperscript{7} 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\textsuperscript{17}, myocardial infarction\textsuperscript{18}, left ventricular mass\textsuperscript{19}, heart failure\textsuperscript{20}, fasting glucose\textsuperscript{21} and urinary albumin excretion (CKDDGen 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 \( \times \) 10\textsuperscript{-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\textsuperscript{11} (Supplementary Table 9). Seven SNPs achieved nominal direction-consistent significance (\( P \leq 0.05 \)) 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.
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 to test whether specific biological pathways and gene sets are enriched for genes mapping into the associated regions at a permutation FDR of 5%.

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 eGFRc rea 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 genes in each locus given in the regional association plots (Supplementary Fig. 4). The conditions for replication were not all met (q-value >0.05 in the replication stage).

Table 1 | The 24 novel SNPs associated with eGFRcrea in European ancestry individuals.

<table>
<thead>
<tr>
<th>SNP ID&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Chr.</th>
<th>Position (bp)&lt;sup&gt;2&lt;/sup&gt;</th>
<th>Locus name&lt;sup&gt;3&lt;/sup&gt;</th>
<th>Effect/Non effect allele (EAF)</th>
<th>SNP function&lt;sup&gt;4&lt;/sup&gt;</th>
<th>Stage 1 (discovery)&lt;sup&gt;5&lt;/sup&gt;</th>
<th>Stage 2 (replication)&lt;sup&gt;5&lt;/sup&gt;</th>
<th>Combined analysis&lt;sup&gt;5&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Beta</td>
<td>P value</td>
<td>Beta</td>
</tr>
<tr>
<td>rs3850625</td>
<td>2</td>
<td>1,013,376,296</td>
<td>CACNA1S</td>
<td>A/G (0.12)</td>
<td>Exonic, nonsyn. SVN</td>
<td>0.0344</td>
<td>0.3207</td>
<td>0.0271</td>
</tr>
<tr>
<td>rs2792184</td>
<td>2</td>
<td>27,921,907</td>
<td>KMT2D2</td>
<td>A/G (0.50)</td>
<td>Intron</td>
<td>0.0097</td>
<td>0.3446</td>
<td>0.0094</td>
</tr>
<tr>
<td>rs68202041</td>
<td>3</td>
<td>170,091,902</td>
<td>SKL</td>
<td>T/C (0.87)</td>
<td>Intron</td>
<td>0.0070</td>
<td>0.4687</td>
<td>0.0066</td>
</tr>
<tr>
<td>rs10538301</td>
<td>11</td>
<td>27,921,907</td>
<td>KMT2D2</td>
<td>A/G (0.50)</td>
<td>Intron</td>
<td>0.0097</td>
<td>0.3446</td>
<td>0.0094</td>
</tr>
<tr>
<td>rs10994860</td>
<td>10</td>
<td>65,264,624</td>
<td>ALCF</td>
<td>T/C (0.19)</td>
<td>Intron</td>
<td>0.0070</td>
<td>0.4687</td>
<td>0.0066</td>
</tr>
<tr>
<td>rs163160</td>
<td>12</td>
<td>7,289,955</td>
<td>XCIQ</td>
<td>A/G (0.82)</td>
<td>Intron</td>
<td>0.0070</td>
<td>0.4687</td>
<td>0.0066</td>
</tr>
<tr>
<td>rs164748</td>
<td>13</td>
<td>89,708,292</td>
<td>DFPP1</td>
<td>G/C (0.53)</td>
<td>Intron</td>
<td>0.0047</td>
<td>0.4753</td>
<td>0.0041</td>
</tr>
<tr>
<td>rs8091180</td>
<td>18</td>
<td>77,164,243</td>
<td>AIFM3</td>
<td>A/G (0.56)</td>
<td>Intron</td>
<td>0.0054</td>
<td>0.4753</td>
<td>0.0050</td>
</tr>
</tbody>
</table>

β, baseranks; Chr, chromosome; EAF, effect allele frequency; eGFRcrea, eGFR based on serum creatinine; GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism; UTR, untranslated region.

*SNPs are grouped by the stratum where the smallest P value of the meta-analysis of the twice GC-corrected discovery meta-analysis results and replication studies. In the ‘overall’ group, the numbers for the three analyses were equal to 133,413/48, 42,116/14 and 154,881/58, in the discovery, replication and combined analyses, respectively. In the ‘overall’ group, the numbers for the three analyses were equal to 133,413/48, 42,116/14 and 154,881/58, in the discovery, replication and combined analyses, respectively. In the ‘overall’ group, the numbers for the three analyses were equal to 133,413/48, 42,116/14 and 154,881/58, in the discovery, replication and combined analyses, respectively. In the ‘overall’ group, the numbers for the three analyses were equal to 133,413/48, 42,116/14 and 154,881/58, in the discovery, replication and combined analyses, respectively. In the ‘overall’ group, the numbers for the three analyses were equal to 133,413/48, 42,116/14 and 154,881/58, in the discovery, replication and combined analyses, respectively.

**Compared to the meta-analysis of the GWAS results, the number of SNPs annotated 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 publically 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).
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 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 chromatin marks, promoter-specific histone modifications associated with active transcription, in order to gather information about cell-type specific regulatory potential of eGFRcrea-associated SNPs. Comparing 33 available adult-derived cell types, we found that eGFRcrea-associated SNPs showed the most significant overlap with H3K4m3 peaks in adult kidney (P value = 0.0029), followed by liver (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.

**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 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 variants.

**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.
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 kidney development by examining the functional consequences of gene knockdown in zebrafish embryos utilizing antisense gene knockdown in zebrafish embryos utilizing antisense.

Discussion
We identified 24 new loci in association with eGFR and confirmed 29 previously identified loci. A variety of (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 hybridization. In all cases, morphant embryos did not display significant gene expression defects compared with controls (Supplementary Table 15).

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 DHSs was considered. SNPs from the eGFR discovery genome-wide scan meeting a series of P value thresholds in the range 10^{-4} to 10^{-6} 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 with coloured lines had nominally significant enrichment (P value <0.05) at the P value threshold and/or were derived from renal tissues (H7esDiffa2d: 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; Sae, 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 x 2 tables are reported (significance level = 5.6 x 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)\(^\text{27}\), low birth weight, nephron endowment and early growth and early-life nutrition\(^\text{28}\). 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.

A limitation of our work is that causal variants and precise molecular mechanisms underlying the observed associations were not identified and will require additional experimental follow-up projects. Our attempt to gain insights into potentially causal genes through knockdown in zebrafish did not yield any clear CKD candidate gene, although the absence of a zebrafish 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\(^\text{29,30}\). 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 × (serum cystatin C)\(^{-1.19}\) (ref. 31). eGFRcrea and eGFRcys values
Genome-wide meta-analysis was performed with the software METAL33, assuming studies used markers of highest quality to impute independent loci through LD pruning based on an estimate of false positives, after the meta-analysis, a second GC correction on the aggregated genotype frequency (MAF) of 4. Genomic-control (GC) correction was applied to all 24 studies for the non-diabetes groups (N = 118,448) and 39 for the diabetes group (N = 13,522). 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 same MAF and imputation quality thresholds as described above. For all GWAS, we performed evaluations of SNPs associated with results generated from consortia investigating other traits. Specifically, we evaluated systolic and diastolic blood pressure in ICBP17, myocardial infarction in CARDIOMAP18, left ventricular mass19, heart failure20, the urinary albinom 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 in the NHGRI GWAS catalog40 (accessed April 14, 2014). We performed a trans-ethnic meta-analysis of GWAS 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 ≥ 0.01 and imputation quality r2 ≥ 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 × 10−8. For all the 24 novel and 29 known SNPs, the difference between the SNP effect on eGFRcys in the two groups, i.e. the standard error of the estimate and p(τ) indicates the correlation between effects in the two groups, which was estimated as 0.044 by sampling 100,000 independent SNPs from our DM and nonDM GWAS, after removing known and novel loci associated with eGFRcys. For a large sample size, as in our case, it approximates a standard normal distribution.

A similar analysis was performed to compare results in subjects with and without hypertension, based on results from our previous work27. The correlation between the two strata was on the order of 0.5.

Proportion of phenotypic variance explained. The percent of phenotypic variance explained by novel and known loci was estimated as: 

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

where \( b_{\text{DM}} \) and \( b_{\text{nonDM}} \) represent the proportion of phenotypic variance explained on log(eGFRcys) in the two groups, i.e. the standard error of the estimate and p(τ) indicates the correlation between effects in the two groups, which was estimated as 0.044 by sampling 100,000 independent SNPs from our DM and nonDM GWAS, after removing known and novel loci associated with eGFRcys. For a large sample size, as in our case, it approximates a standard normal distribution.

A similar analysis was performed to compare results in subjects with and without hypertension, based on results from our previous work27. The correlation between the two strata was on the order of 0.5.
of Trans-ethnic Association studies) software. We combined the 48 European ancestry studies that contributed eGFRcra, 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. It has been established before the publication of kidney function-related GWAS. Results from GRAIL were used to prioritize genes for follow-up functional work.

Expression quantitative trait loci analysis. We identified allele rSIDs and proxies (r² >0.8) for our index SNPs using SNAP software across 4 HapMap builds. SNP rSIDs and alleles were searched for primary SNPs and LD proxies against a functional datasource text_2009_03, established in the Correlation of selected eSNPs to the best eSNPs per transcript per expression of Trans-ethnic Association studies software. We combined the 48 European HapMap (hg18 assembly) and the functional datasource text_2009_03, established in the Correlation of selected eSNPs to the best eSNPs per transcript per expression (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 h post-fertilization for in situ hybridization using published methods (http://zfin.org/ZFIN/Methods/Thies_protocol.html). Gene expression was visualized using established renal markers pax2a (global kidney), nephrin (podocytes) and slc20a1a (proximal tubule). The number of morphant embryos displaying abnormal gene expression was compared with control embryos by means of a Fisher’s exact test.

Interruption of human kidney chromatin annotation maps. Different articles can be used to generate tissue-specific chromatin-state annotation maps. These new loci for kidney function.

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

References


Drafting manuscript: J Chalmers, A Tin, A Köttgen, WHL Kao, M Bochud, CA Peralta, C Pattaro, IH de Boer, CS Fox, M Garnaas, W Goessling, N Soranzo, CA Böger, IM Heid, M Gorsi, S Trompet, A Dehghan, A Teumer, KL Keene, MM Sale.


Additional information

Competing financial interests: T.J. 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: Patterson, 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).


1 Center for Biomedicine, European Academy of Bozen/Bozolo (EURAC), affiliated to the University of Lübeck, Via Galvani 13, Bolzano 39100, Italy.
2 Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-University Greifswald, Friedrich-Loeﬄer-Straße 15a, Greifswald 17487, Germany.
3 Institute for Community Medicine, University of Greifswald, Walther-Rathenau-Strasse 48, Greifswald 17487, Germany.
4 Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Franz-Josef-Strauß-Allee 11, Regensburg 93053, Germany.
5 Department of Nephrology, University Hospital Regensburg, Franz-Josef-Strauß-Allee 11, Regensburg 93053, Germany.
6 Preventive Medicine, Brigham and Women’s Hospital, 900 Commonwealth Avenue East, Boston, Massachusetts 02215, USA.
7 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, Maryland 21205, USA.
8 Department of Life and Reproduction Sciences, University of Verona, Strada Le Grazie 8, Verona 37134, Italy.
9 Division of Genetics, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, New Research Building 77 Avenue Louis Pasteur, Room 45B, Boston, Massachusetts 02215, USA.
10 Institute of Genetics and Biophysics “Adriano Buzzati-Traverso”—CNR, Via P. Castellino 11, Napoli 80131, Italy.
11 Department of Internal Medicine IV, University Hospital Freiburg, Berliner Allee 29, Freiburg 79110, Germany.
12 Division of Nephrology/Tufts Evidence Practice Center, Tufts University School of Medicine, Tufts Medical Center, Boston, Massachusetts 02111, USA.
13 Department of Biostatistics, Boston University School of Public Health, 715 Albany Street, Boston, Massachusetts 02118, USA.
14 Department of Neurology, Boston University School of Medicine, 72 East Concord ST 8603, Boston, Massachusetts 02118, USA.
15 Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children’s Hospital, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.
16 Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 2142, USA.
17 NHLBI’s Framingham Heart Study and the Center for Population Genomics, Boston Children’s Hospital, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.

This EMCC content is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

168 Center for Statistical Genetics, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA. 169 Department of Preventive Medicine, University of Southern California, Los Angeles, California 90033, USA. 170 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle 98195, Washington, USA. 171 National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. 172 Department of Medicine III, Medical Faculty Carl Gustav Carus at the Technical University of Dresden, Dresden 01307, Germany. 173 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Heidelberglaan 100, Utrecht 3508 GA, The Netherlands. 174 Leibniz-Institute for Arteriosclerosis Research, Department of Molecular Genetics of Cardiovascular Disease, University of Münster, Münster, Germany. 175 Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan Medical Center, Ann Arbor 48108, Michigan, USA. 176 Laboratory of Arteriosclerosis Research, Department of Molecular Genetics of Cardiovascular Disease, University of Münster, Münster, Germany. 177 Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester LE3 9QP, UK. 178 Clinical Pharmacology Unit, University of Cambridge, Addenbrooke’s Hospital, Hills Road, Cambridge CB2 0QQ, UK. 179 Department of Health Sciences, University of Leicester, University Rd, Leicester LE1 7RH, UK. 180 Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. 181 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. 182 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK. 183 Centre for Cellular and Molecular Biology (CCMB), Council of Scientific and Industrial Research (CSIR), Uppal Road, Hyderabad 500 007, India. 184 University of Maryland, School of Medicine, Baltimore, Maryland 21201, USA. 185 School of Science and Engineering, University of New South Wales, Building 1442, Kensington, Sydney 2033, Australia. 186 International Centre for Circulatory Health, National Heart & Lung Institute, Imperial College, London, UK. 187 Centre for Genome Science, National Institute of Health, Seoul, Korea. 188 Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford OX3 7LF, UK. 189 University of Dundee, Ninewells Hospital & Medical School, Dundee DD1 9SY, UK. 190 Centre for Cardiovascular Genetics, University College London, London WC1E 6JF, UK. 191 Centre for Epidemiology and Biostatistics, University of Western Australia, Crawley, Western Australia, Australia. 192 Department of Preventive Medicine and Epidemiology, Loyola University Medical School, Maywood, Illinois, USA. 193 Tuscany Regional Health Agency, Florence, Italy. 194 Department of Internal Medicine B, Ernst-Moritz-Arndt-University Greifswald, Greifswald 17487, Germany. 195 Occupational and Environmental Medicine, Department of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg 40530, Sweden. 196 MRC Centre for Causal Analyses in Translational Epidemiology, School of Social & Community Medicine, University of Bristol, Bristol BS8 2BN, UK. 197 BHF Glasgow Cardiovascular Research Centre, University of Glasgow, 126 University Place, Glasgow G12 8TA, UK. 198 Georgia Prevention Institute, Department of Pediatrics, Medical College of Georgia, 30912 Augusta, Georgia, USA. 199 Cardiovascular Epidemiology and Genetics, Institute Municipal d’Investigacion Medica, Barcelona Biomedical Research Park, 88 Doctor Aiguader, Barcelona 08003, Spain. 200 Medizinische Klinik II, Universität zu Lübeck, 23562 Lübeck, Germany. 201 Department of Statistics, Pontificia Universidad Catolica de Chile, Vicuña Mackenna, Santiago 8460, Chile. 202 Department of Cardiovascular Medicine, The Welcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. 203 Department of Clinical Sciences, Lund University, SE-205 02 Malmö, Sweden. 204 Tropical Medicine Research Institute, University of the West Indies, Mona, Kingston, Jamaica. 205 Department of Medicine, University of Mississippi Medical Center, 2500 North State St, Jackson, Mississippi 39216, USA. 206 Genetics of Complex Traits, Peninsula Medical School, University of Exeter, EX2 2LU Exeter, UK. 207 Institut National de la Santé et de la Recherche Médicale, U1125 Institut National de la Recherche Agronomique, Université Paris 13, F-93017 Bobigny, France. 208 Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan Medical Center, Ann Arbor 48108, Michigan, USA. 209 Laboratory of Epidemiology, Biometry, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. 210 Department of Clinical Sciences, Diabetes and Endocrinology Research Unit, University Hospital, SE-205 02 Malmö, Sweden. 211 Human Genetics Foundation (HUGEF), Via Lagrange 35, Torino 10123, Italy. 212 Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California 90048 California, USA. 213 Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger 7600, Norway. 214 Genetics of Complex Traits, Peninsula Medical School, University of Exeter, EX1 2LU Exeter, UK. 215 De´partement de Ge´ne´tique Me´dicale, Universite´ de Lausanne, Lausanne 1015, Switzerland. 216 Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK. 217 Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. 218 Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. 219 Center for Human Genetic Research, Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. 220 Geriatric Rehabilitation Unit, Azienda Sanitaria Firenze (ASF), 50125 Florence, Italy. 221 Department of Public Health and Community Medicine, Institute of Medicine, Western Australia, Crawley, Western Australia, Australia. 222 Division of Epidemiology and Biostatistics, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. 223 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle 98195, Washington, USA. 224 National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. 225 Centre National de Ge´notypage, Commissariat de L’Energie Atomique, Institut de Ge ´nomique, 91000 Evry, France. 226 William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. 227 Cardiovascular Genetics, University of Utah, School of Medicine, Salt Lake City, 84132 Utah, USA. 228 HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger 7600, Norway. 229 Cardiovascular Genetics, University of Utah, School of Medicine, Salt Lake City, 84132 Utah, USA. 230 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle 98195, Washington, USA. 231 Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger 7600, Norway. 232 Cardiovascular Genetics, University of Utah, School of Medicine, Salt Lake City, 84132 Utah, USA.
Charge-Heart Failure Group
Laura R. Loehr, Wayne D. Rosamond, Emelia Benjamin, Talin Haritunians, David Couper, Joanne Murabito, Ying A. Wang, Bruno H. Stricker, Patricia P. Chang, James T. Willerson

EchoGen Consortium

Nature Communications | 7:10023 | DOI: 10.1038/ncomms10023 | www.nature.com/naturecommunications