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

Cristian Pattaro, European Academy of Bozen/Bolzano (EURAC)
Idris Guessous, Emory University

Journal Title: Nature Communications
Volume: Volume 7
Publisher: Nature Publishing Group: Nature Communications | 2016-01-01, Pages 10023-10023
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1038/ncomms10023
Permanent URL: https://pid.emory.edu/ark:/25593/rms3s

Final published version: http://dx.doi.org/10.1038/ncomms10023

Copyright information:
© 2016, Nature Publishing Group, a division of Macmillan Publishers Limited. This is an Open Access work distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

Accessed March 1, 2019 3:29 AM EST
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, 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 genetic 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), and estimated glomerular filtration rate (eGFR), a measure of the kidney’s filtration ability that is used to diagnose and stage CKD. 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 genetic loci associated with kidney function traits and provide extensive locus characterization and bioinformatics analyses, further delineating the physiologic basis of kidney function.

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 \(< 5.0 \times 10^{-8}\) (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 \(< 5.0 \times 10^{-6}\), \( q\)-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, \( P = 2.5 \times 10^{-8}\)), and six loci (NFKB1, UNXC, 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 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 \( P = 5.0 \times 10^{-8}\) 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, LR2P, IGFBP5, SKIL, UNXC, 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.
(Supplementary Fig. 7), suggesting that our findings can likely be generalized beyond the European ancestry 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-analysis, 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 the gene 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, SDCAG8, 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), 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 in 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 reconstructed 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 DNA hypomethylation regions 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 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.

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 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 values from 3.1 × 10^-3 to 7.9 × 10^-6, multiple testing threshold z = 5.6 × 10^-3).

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 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).

**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 within the kidney. The kidney is a highly vascular and

![Figure 3](image-url)

**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 permuted test P values (see Methods). (b) Enrichment of eGFR-associated SNPs in DHS according to kidney action value threshold. SNPs from the eGFR discovery genome-wide scan meeting a series of P value thresholds in the range $10^{-4}$, $10^{-6}$ 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 indicated with coloured lines had nominally significant enrichment (* indicate particular tissues, including kidney and urinary tract. Shown are permutation test P values from Fishers’ exact tests for 2 $	imes$ 2 tables are reported (significance level = $5.6 \times 10^{-3}$, 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, DPEPI, 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).
metabolically active organ that receives 20% of all cardiac output, contains an extensive endothelium-lined capillary network, and is sensitive to ischaemic 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)\textsuperscript{27}, low birth weight, nephron endowment and early growth and early-life nutrition\textsuperscript{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 genomci 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\textsuperscript{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 \times (\text{serum cystatin C})^{-1.19} (ref. 31). eGFRcrea and eGFRcys values

---

**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\textsuperscript{48}. 
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 independent loci through LD pruning based on an estimated frequency (MAF) of 0.01. Genome-wide association studies were included in the regression and family-based studies appropriately accounted for relatedness.

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 group, N = 118,448 and 39 for the diabetes group, N = 11,522). GWAS of CKD were comprised by 43 studies, for a total sample size of 117,165, including 12,385 CKD cases. GWAS of eGFRcrea were comprised by 16 studies for a total sample size of 32,834. All GWAS files underwent quality control using the Genotype+Phenotype (Supplementary Table 1) and genetic principal components were included in the regression and family-based studies appropriately accounted for relatedness.

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 Park (OGP), Talana, whose GWAS was included in the discovery analysis, SAPHIR and SKIPOGH studies (Supplementary Table 20). Summarizing all in silico and de novo replication studies (Supplementary Table 1), replication data for eGFRcrea were contributed by 14 studies (total sample size = 42,166), which also contributed eGFRcrea results from non-diabetes (13 studies, N = 36,433) and diabetes (14 studies, N = 4,935). Thirteen studies contributed replication data on CKD (N = 33,972; 4,245 CKD cases; studies with <50 CKD cases were excluded) and five on eGFRcrea (N = 14,930).

Association between eGFRcrea, CKD and eGFRcrea and each of the 48 SNPs in the replication studies was assessed using the same analysis protocol detailed for the discovery studies above. Quality control of the replication files was performed with the same software as described above.

We performed a combined fixed-effect meta-analysis of the double-GC corrected results from the discovery meta-analysis and the replication studies, based on inverse-variance weighting. The total sample size in the combined analysis of eGFRcrea was estimated to be 98,486, from 72 to 96 for the CKD-associated SNPs, and was equal to 59% for the eGFRcrea-associated SNP (Supplementary Table 3).

Associations stratified by diabetes and hypertension status. For all the 24 novel and 29 known SNPs, the difference between the SNP effect on eGFRcrea in the diabetes versus the non-diabetes groups was assessed by means of a two-sample t-test for correlated data at a significance level of 0.05. We used the following two-sample t-test for correlated data:

\[ t = \frac{(b_{\text{DM}} - b_{\text{nonDM}})}{\sqrt{\frac{s.e.}{\text{DM}} + \frac{s.e.}{\text{nonDM}}} - \frac{2}{s}(p_{\text{DM}} \times \text{var(SNP)}_{\text{DM}} + p_{\text{nonDM}} \times \text{var(SNP)}_{\text{nonDM}})} \]

where \( b_{\text{DM}} \) and \( b_{\text{nonDM}} \) represent the SNP effects on log(eGFRcrea) in the two groups, i.e. the standard error of the estimate and \( p \) 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 eGFRcrea. For a large sample size, as in our case, this provided a standard normal distribution.

A similar analysis was performed to compare results in subjects with and without hypertension, based on results from our previous work. The correlation between the two strata was 0.01.

Proportion of phenotypic variance explained. The percent of phenotypic variance explained by novel and known loci was estimated as \( R^2_{\text{var}} \), where \( R^2_{\text{var}} = \text{var(SNP)}_{\text{DM}} / \text{var(SNP)}_{\text{nonDM}} \) is the coefficient of determination for each of the 53 individual SNPs associated with eGFRcrea uncovered to date (24 novel and 29 known ones), \( b \) is the estimated effect of the \( i \)th SNP on \( y \), \( y \) corresponds to the sex- and age-adjusted residuals of the logarithm of eGFRcrea and \( \text{var(SNP)} \) is the variance explained by novel and known loci associated with eGFRcrea. For a large sample size, as in our case, this provided a standard normal distribution.

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 CARDIOGRAM16, 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.

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 consortium11.

African ancestry meta-analysis. We performed integrated 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 > 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 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 9.9 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 (log₁₀BF) ≥6 (ref. 41).

Gene Relationships Across Implicated Loci (GRAIL). To prioritize the gene(s) most likely to give rise to association signals in a given region, the software GRAIL was used. The index SNP of all previously known kidney function associated loci, as well as the novel SNPs identified here was used as input, using the CEU HapMap (hg18 assembly) and the functional data resource textovear 2009_03, 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 alias rsIDs and proxies (r²>0.8) for our index SNPs using SNP software across 4 HapMap builds. SNPs and aliases were searched for primary SNPs and LD proxies against a collected database of expression SNP (eSNP) results. The collected eSNP results met criteria for statistical thresholds for association with gene transcript levels in their respective original analyses (for references see Supplementary Table 11). A correlation of selected eSNPs to the best eSNPs per transcript per expression quantitative trait loci (eQTL) data set were assessed by pairwise LD. All results are reported in Supplementary Table 11.

DEPICT analysis. In this work, we first used PLINK to identify independently associated SNPs using all SNPs with eGFRcra association P values < 10⁻⁵ (HapMap release 27 CEU data); LD r² threshold = 0.01; physical kb threshold = 1,000. We then used the DEPICT method to construct associated regions by mapping to genes independently associated SNPs if they overlapped or resided within 50 kb of a given gene. 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 help with the interpretation especially of intergenic variants identified in association studies. 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. Chromatin immune-precipitation sequencing (ChIP-seq) data from human kidney samples were generated by NIH Roadmap Epigenomics Mapping Consortium. Briefly, proximal tubule cells derived from an adult human kidney were collected and fixed with 1% formaldehyde. Subsequently, ChIP-seq was conducted using whole-cell extract from adult human tissue as the input (GSM621638) and assessing the following chromatin marks: H3K36me3 (GSM621634), H3K4me1 (GSM670025), H3K4me3 (GSM621648), H3K9ac (GSM772811) and H3K9me3 (GSM621651). The MACS version 1.41 (model-based analysis of ChIP-seq) peak finding algorithm was used to identify regions of ChIP-seq enrichment. 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 v0.06e for chromatin annotation.

For comparison, the same genomic coordinates were investigated in chromatin annotation maps of regions with SNPs or with other different cell lines from the ENCODE Project: umbilical vein endothelial cells (HUVEC), mammary epithelial cells (HMEC), normal epidermal keratinocytes (NHK), B-lymphoblastoid cells (GM12878), erythrocye 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 present over or within regions was different from that of the other nine tissues by means of a Fisher’s exact test for 2 x 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 x 10⁻⁴.

Functional characterization of new loci. Replicated regional gene priorities were generated for functional studies using the following criteria: (1) GRAIL identification of a gene in each region of P value <0.05 or DEPICT, FDR <0.05; (2) an eGFRcra 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, EU) were maintained according to established Harvard Medical School Institutional Animal Care and Use Committee protocols. 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.

References


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


168Center for Statistical Genetics, Department of Biostatistics, University of Iowa, School of Public Health, Ann Arbor, Michigan 48103, USA. 169Department of Nutrition, University of North Carolina, Chapel Hill, North Carolina 27599, USA. 170Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK. 171Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. 172Department of Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. 173Center for Human Genetic Research, Harvard Medical School, Boston, Massachusetts 02114, USA. 174Department of Geriatric Rehabilitation Unit, Azienda Sanitaria Firenze (ASF), 50125 Florence, Italy. 175Department of Genématique, Université de Lausanne, Lausanne 1015, Switzerland. 176Pathology and Laboratory Medicine, University of Western Australia, 6009 Crawley, Western Australia, Australia. 177Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. 178Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle 98195, Washington, USA. 179National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. 180Department of Medicine III, Medical Faculty Carl Gustav Carus at the Technical University of Dresden, Dresden 01307, Germany. 181Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Heidelberglaan 100, Utrecht 3508 GA, The Netherlands. 182Leibniz-Institute for Arteriosclerosis Research, Department of Molecular Genetics of Cardiovascular Disease, University of Münster, Münster, Germany. 183University Hospital Münster, International Medicine D, Münster, Germany. 184Center for Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester LE3 9QP, UK. 185Clinical Pharmacology Unit, University of Cambridge, Addenbrookes Hospital, Hills Road, Cambridge CB2 2QQ, UK. 186Department of Health Sciences, University of Leicester, University Rd, Leicester LE1 7RH, UK. 187Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. 188Clinical 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. 189Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK. 190Centre for Cellular and Molecular Biology (CCMB), Council of Scientific and Industrial Research (CSIR), Uppal Road, Hyderabad 500 007, India. 191University of Maryland, School of Medicine, Baltimore, Maryland 21201, USA. 192School of Science and Engineering, University of Ballarat, Ballarat 3353, Australia. 193International Centre for Circulatory Health, National Heart & Lung Institute, Imperial College, London, UK. 194Department of Genome Science, National Institute of Health, Seoul, Korea. 195Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford OX3 7LF, UK. 196University of Dundee, Ninewells Hospital & Medical School, Dundee DD1 9SY, UK. 197Centre for Cardiovascular Genetics, University College London, London WC1E 6JF, UK. 198Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Crawley, Western Australia, Australia. 199Department of Preventive Medicine and Epidemiology, Loyola University Medical School, Maywood, Illinois, USA. 200Tuscany Regional Health Agency, Florence, Italy. 201Department of Internal Medicine B, Ernst-Moritz-Arndt-University Greifswald, Greifswald 17487, Germany. 202Occupational and Environmental Medicine, Department of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg 40530, Sweden. 203MRC Centre for Causal Analyses in Translational Epidemiology, School of Social & Community Medicine, University of Bristol, Bristol BS8 2BN, UK. 204BHF Glasgow Cardiovascular Research Centre, University of Glasgow, 126 University Place, Glasgow G12 8TA, UK. 205Georgia Prevention Institute, Department of Pediatrics, Medical College of Georgia, 30912 Augusta, Georgia, USA. 206Cardiovascular Epidemiology and Genetics, Institut Municipal d'Investigacio Medica, Barcelona Biomedical Research Park, 88 Doctor Aiguader, Barcelona 08003, Spain. 207Medizinische Klinik II, Universität zu Lübeck, 23562 Lübeck, Germany. 208Department of Statistics, Pontificia Universidad Católica de Chile, Viña Mackenna, Santiago 84860, Chile. 209Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. 210Department of Clinical Sciences, Lund University, SE-205 02 Malmö, Sweden. 211Tropical Medicine Research Institute, University of the West Indies, Mona, Kingston, Jamaica. 212Department of Medicine, University of Mississippi Medical Center, 2500 North State St, Jackson, Mississippi 39216, USA. 213Genetics of Complex Traits, Peninsula Medical School, University of Exeter, EXI 2LU Exeter, UK. 214Division of Epidemiology, Biometry, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. 215Laboratory of Epidemiology, Diabetes and Endocrinology Research Unit, Harvard Medical School, SE-205 02 Malmö, Sweden. 216Human Genetics Foundation (HUGEF), Via de la Sante´ et de la Recherche Me´dicale, U1125 Institut National de la Recherche Agronomique, Universite ´ Paris 13, F-93017 Bobigny, France. 217Laboratory of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. 218Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan Medical Center, Ann Arbor 48108, Michigan, USA. 219Department of Clinical Sciences, Lund University, SE-205 02 Malmö, Sweden. 220Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles 90048 California, USA. 221Division of Community Health Sciences, St George's University of London, London SW17 0RE, UK. 222Atherosclerosis Research Unit, Department of Medicine, Karolinska Institute, Stockholm, Sweden. 223Department of Public Health and Ageing, University of Washington, Seattle, Washington 98195, USA. 224Laboratory of Epidemiology, Biometry, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. 225Centre National de Ge´notypage, Commissariat de L'Energie Atomique, Institut de Ge ´nomique, 91000 Evry, France. 226Department of Epidemiology, University of Washington, Seattle, Washington, 98195, USA. 227Epidemiology Public Health, UCL, London WC1E 6BT, UK. 228William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. 229Cardiovascular Genetics, University of Utah, School of Medicine, Salt Lake City, 84132 Utah, USA. 230HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger 7600, Norway.
Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital, Boston, 02114 Massachusetts, USA. 340Center for Human Genetic Research, Massachusetts General Hospital, Boston, 02114 Massachusetts, USA. 341Division of Research, Kaiser Permanente, Oakland, 94611 California, USA. 342Institute for Human Genetics, University of California, San Francisco, San Francisco, 94143 California, USA. 343Department of Cardiovascular Medicine, Cleveland Clinic, 7255 Old Oak Blvd, Cleveland, Ohio 44130, USA. 344Medizinische Klinik und Poliklinik, Johannes-Gutenberg Universität Mainz, Universitätssmediizin, 55122 Mainz, Germany. 345Institut für Klinische Chemie und Laboratoriumsmedizin, Johannes-Gutenberg Universität Mainz, Universitätssmediizin, 55122 Mainz, Germany. 346INSERM UMR 937, Pierre and Marie Curie University (UPMC, Paris 6) and Medical School, 75005 Paris, France. 347Boston University, School of Public Health, Boston, 02118 Massachusetts, USA. 348University of Minnesota School of Public Health, Division of Epidemiology and Community Health, School of Public Health (A.R.F.), Minneapolis, 55455 Minnesota, USA. 349University of Washington, Department of Internal Medicine, Seattle, 98195-6420 Washington, USA. 350University of Texas, School of Public Health, Houston, 77030 Texas, USA. 351Department of Medicine, Landspitali University Hospital, Reykjavik 101, Iceland. 352Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany. 353Division of Endocrinology and Diabetes, Graduate School of Molecular Endocrinology and Diabetes, University of Ulm, 89069 Ulm, Germany. 354Division of Endocrinology, Department of Medicine, Medical University of Graz, 8010 Graz, Austria. 355Synlab Center of Laboratory Diagnostics Heidelberg, 69037 Heidelberg, Germany. 356Division of Clinical Chemistry, Department of Medicine, Albert Ludwigs University, 79085 Freiburg, Germany. 357Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University Graz, 8010 Graz, Austria. 358Cardiology Group Frankfurt-Sachsenhausen, 60594 Frankfurt, Germany. 359The Center for Applied Genomics, Children’s Hospital of Philadelphia, 19104 Philadelphia, Pennsylvania, USA. 360Cardiovascular Research Institute, Medstar Health Research Institute, Washington Hospital Center, Washington, DC 20010, USA. 361Genetics Division and Drug Discovery, GlaxoSmithKline, King of Prussia, Pennsylvania 19406, USA. 362The Institute for Translational Medicine and Therapeutics, School of Medicine, University of Pennsylvania, Philadelphia, 19104-5158 Pennsylvania, USA. 363Department of Cardiovascular Surgery, University of Leicester, Leicester LE1 7RH, UK. 364Division of Cardiovascular and Diabetes Research, Multidisciplinary Cardiovascular Research Centre, Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds LS2 9JT, UK. 365LIGHT Research Institute, Faculty of Medicine and Health, University of Leeds, Leeds LS2 9JT, UK.

**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**


**ECHO Gen Consortium**