# Genetics of kidney traits in worldwide populations: the Continental Origins and Genetic Epidemiology Network (COGENT) Kidney Consortium

*Kidney International* (2020) **98,** 35–41; https://doi.org/10.1016/j.kint.2020.02.036
KEYWORDS: chronic kidney disease; diverse populations; gene expression


Nora Franceschini[1] and Andrew P. Morris[2]
[1]*Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA; and* [2]*Division of Musculoskeletal and Dermatological Sciences, The University of Manchester, Manchester, UK*

**Correspondence:** *Nora Franceschini, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, 123 West Franklin Street, Suite 401, North Carolina 27516, USA. E-mail: noraf@unc.edu*

We describe our collaborative efforts to increase representation of diverse populations in genomic research of kidney phenotypes to fill an unmet need to further understanding of the genetic contribution to chronic kidney disease (CKD) across the globe. These efforts led to the creation of the Continental Origins and Genetic Epidemiology Network Kidney (COGENT-Kidney) Consortium, focused on developing statistical methods for the analysis of genome-wide association studies (GWASs) across diverse populations and their application to renal traits and CKD. Resources generated within the COGENT-Kidney Consortium will provide a framework for future studies of genetic risk of CKD in worldwide populations. Our intent is to foster collaborations for studies of populations that are underrepresented in GWASs, increase awareness of the challenges and opportunities in studying these populations, and promote more genomic research across increasingly diverse populations.

## Scope of the problem: CKD in the US and global populations

The Global Burden of Disease study recently reported rising CKD incidence and prevalence worldwide, and increased deaths and disability due to CKD.[1] US ethnic minorities have a higher burden of CKD compared to those of European ancestry. Globally, CKD has an uneven distribution in its prevalence and causes.[1,2] The limited knowledge of mechanisms underlying CKD development and progression, and of the causes for regional and ethnic variation in CKD, has hampered efforts to prevent the disease globally and locally. It has also prevented the development of therapeutic tools for clinicians to effectively treat the disease.

## Genetic studies in global populations: challenges and opportunities

Genetic susceptibility to disease occurs in the context of lifetime exposures to lifestyle and environmental risk factors, some of which are amenable to intervention. To benefit individuals across the globe, irrespective of race/ethnicity, more studies are needed in populations that have shown a high risk of CKD. There have been relatively few GWAS undertaken across ancestral groups investigating the genetic susceptibility to other common diseases such as hypertension and diabetes.[3] Understanding how genetic variation affects downstream molecular and biological processes provides both greater insight into genomic contributions to human health across the globe and improved ability to apply this knowledge through clinical translation relevant to everyone.

Populations vary in their DNA makeup, including the frequencies of alleles at genetic variants, and the correlation structure between these variants, referred to as linkage disequilibrium (LD; Box 1). Many genetic variants are shared across populations, and the power to identify genomic regions ("loci") containing disease-risk genes will be greatest in the population in which the causal genetic variants are most frequent. Differences in LD structure among populations can help to pinpoint causal variants among multiple variants at a locus through fine-mapping (Box 1). The gain in knowledge and insights into disease mechanisms from gene discovery in one population benefits all groups, even those populations in which the causal variants at a locus are rare or absent.

Challenges in studying populations that are underrepresented in GWAS include the less well-known genetic architecture, the lack of available public databases of genetic variants, and the small number of samples, which can reduce power for gene discovery.[4] Reference panels for African, East Asian, and South Asian ancestry populations, which describe allele frequencies and LD structure, have been greatly improved with increasing availability of large-scale whole-genome sequencing re-

**Box 1 | Genetic definitions**

**GWAS**

An approach used to identify genetic variants associated with traits (e.g., estimated glomerular filtration rate) across the entire genome.

**Trans-ethnic meta-analysis of GWASs**

An approach that combines GWAS data from different population groups for a trait for variants that are shared across populations. The most powerful approaches allow for heterogeneity of effects of variants on the trait across ancestries.

**LD**

Correlation in alleles within individuals across genetic variants mapping to the same genomic region. LD patterns vary across ancestries: shorter-range LD is found in African ancestry populations, owing to their longer population history.

**Fine-mapping**

An approach that attempts to identify the causal variant at a GWAS locus from among the large numbers of all possible variants in the region, using patterns of LD and association with the trait. Identified variants can then be prioritized for downstream experimental studies to understand the role the variant plays in disease pathogenesis.

**Expression quantitative loci:**

Genetic variants that are associated with the expression of genes mapping to the same genomic region. They can be specific to one or a few tissues, or shared ubiquitously across tissues. Causal variants for a disease/trait that are also expression quantitative loci provide insight into the gene through which the trait association is mediated.

**Genetic risk scores**

An approach that aggregates multiple genetic variants associated with a disease or trait identified through GWASs into a risk model for disease prediction.

**Mendelian randomization**

An approach to assess the causal associations of exposures with outcomes. Genetic variants (usually in aggregate) associated with a trait are proxies or "instrumental variables" for an exposure (or risk factor), with the advantage that they are not influenced by the confounding seen in observational studies or by reverse causality (because they are assigned at conception).

**Genotype imputation**

A statistical method to infer genotypes at untyped variants using population-specific reference panels by taking advantage of LD structure between nearby variants. This method provides the probability distribution of possible genotypes for each individual at each untyped variant, based on genotype information at local typed variants, and a measure of the quality of the imputation. Imputation of untyped genotypes allows for combined analyses across different genotyping arrays through meta-analysis.

**Regulatory annotation**

Variants located in noncoding regions can affect gene expression through epigenomic regulation, which could be cell-type and tissue specific. The regulatory annotation of variants is provided by a variety of publicly accessible resources, including the Encyclopedia of DNA Elements (ENCODE) and the Human Epigenome Roadmap.

GWAS, genome-wide association study; LD, linkage disequilibrium

sources,[4] but knowledge of genetic variation for some worldwide populations is still lacking. There has also been a lack of statistical methods appropriate for combining GWAS data that adequately account for these genetic differences across populations. From the perspective of precision medicine, these gaps in knowledge can exacerbate healthcare disparities and prevent effective implementation of genomics in clinical care for all.

## COGENT-Kidney Consortium description and goals

The COGENT-Kidney Consortium was established in 2014 to undertake trans-ethnic meta-

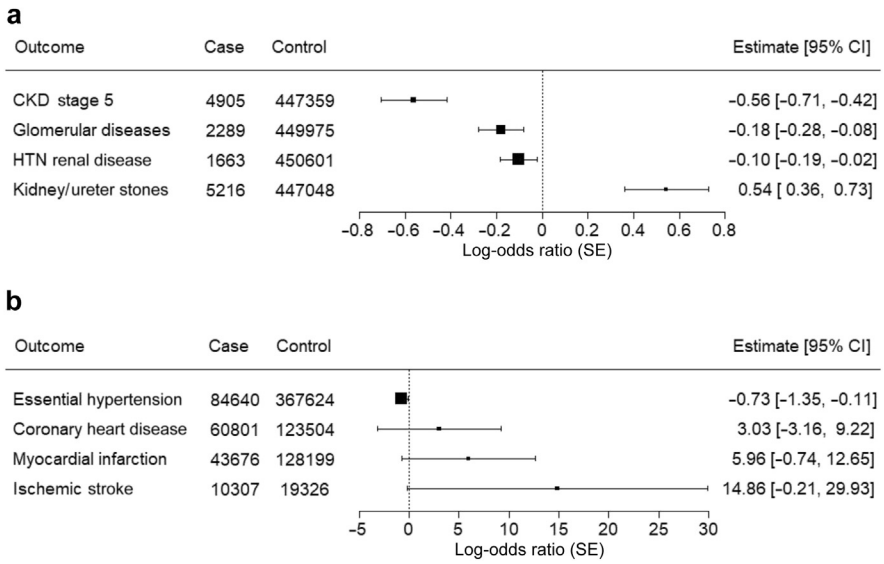


**Figure 1 | Mendelian randomization to assess causal effects of estimated glomerular filtration rate (eGFR) on chronic kidney disease (CKD) and cause-specific kidney disease.** We performed Mendelian randomization to assess the causal association between eGFR and clinical outcomes. We used genetic variants across 94 kidney function loci identified in the trans-ethnic meta-analyses of the Continental Origins and Genetic Epidemiology Network (COGENT)–Kidney Consortium to construct an "instrumental variable" for eGFR. Clinical outcomes were obtained from International Classification of Diseases codes in the UK Biobank, and from the CardiogramplusC4D Consortium and the MEGASTROKE Consortium. The graph shows the Mendelian randomization effect size (log-odds ratio and SE) of eGFR on outcomes for aggregated variants under inverse variance-weighted regression. (**a**) Note that low eGFR is causally associated with CKD stage 5, glomerular disease, and hypertensive (HTN) renal disease, but has an inverse causal association with kidney stones. (**b**) Low eGFR is causally associated with essential hypertension but not with coronary heart disease or ischemic stroke. CI, confidence interval.

analysis of GWASs of kidney traits in diverse populations. The Consortium recruitment targeted GWASs from 4 ancestral groups (African, Hispanic/Latino, European, and East Asian) composed of 71,638 individuals, of whom 67% were of non-European ancestry (Supplementary Table S1).[5] The main goal of the COGENT-Kidney Consortium is to increase representation of populations that have not been well-characterized for genetic risk, particularly ethnic subgroups at high risk of CKD.

The COGENT-Kidney Consortium is an open enrollment and equal partnership among investigators from different studies and it includes expertise in clinical nephrology, population health, human genetics, statistical genetics, bioinformatics, and experimental research. Studies follow standardized protocols for trait definitions and statistical analyses and contribute GWAS results to a centralized repository, after which quality control and meta-analyses are performed. Ancestry-specific meta-analyses are undertaken to maximize the power for discovery of loci that are driven by population-specific variants. Trans-ethnic meta-analysis is also performed, with appropriate modelling of heterogeneity in effects between populations, thereby maximizing power for discovery of associations that are shared across ancestries, and therefore amenable to trans-ethnic fine-mapping. Follow-up analyses include *in silico* and *in vivo* functional investigations and studies of the clinical prediction of variants. We welcome requests to contribute to any of our investigations. The Consortium is modelled on the COGENT Blood Pressure Consortium, which is focused on the genetics of blood pressure and hypertension in individuals of African ancestry.[6]

The COGENT-Kidney Consortium has made a considerable effort to increase the number of Hispanic/Latino GWASs included in genetic studies of CKD, and we are working toward recruitment of other diverse populations. Although sample sizes are not currently (and will not be) large when compared to those in studies of European ancestry, we have already demonstrated important gains in locus discovery and fine-mapping of causal variants for kidney traits, in addition to detection of some population-specific risk variants.[5,7] The Consortium

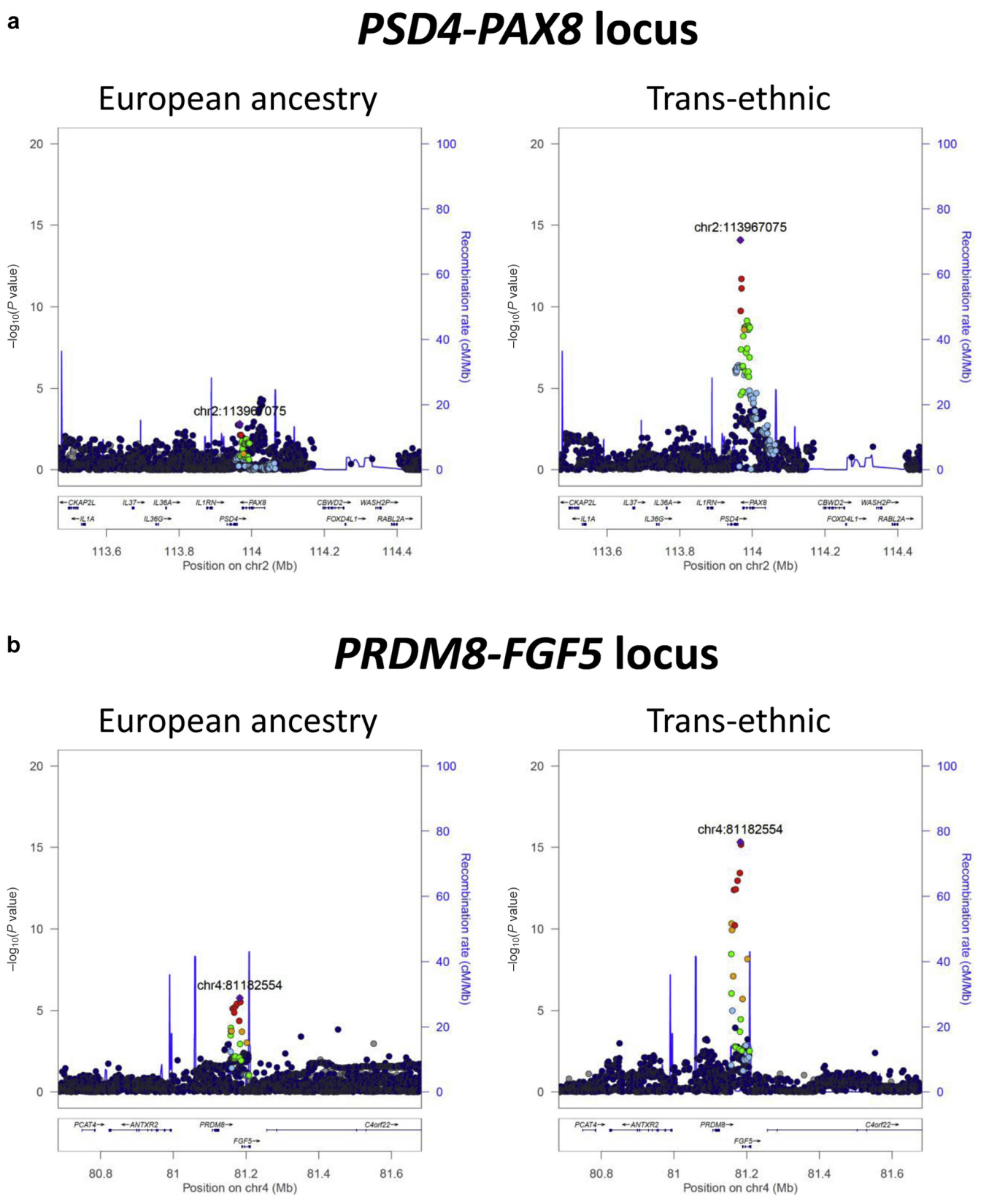


**Figure 2| Examples of improved fine-mapping resolution attained through a trans-ethnic genome-wide association study meta-analysis (312,468 individuals) in the Continental Origins and Genetic Epidemiology Network (COGENT)–Kidney Consortium compared with a meta-analysis restricted to the subset of European-ancestry genome-wide association studies (134,070 individuals) for (a) *PSD4-PAX8* locus and (b) *PRDM8-FGF5* locus.** Each point represents a variant, plotted with its *P* value (on a $-\log_{10}$ scale) as a (continued)

GWAS on kidney traits in African Americans and Hispanics/Latinos is a resource to assess the generalization of findings to other clinical traits, including in studies examining the impact of genetics in hypertension, metabolic abnormalities, and CKD.[8,9] As we expand the samples within each ethnic group, we expect that these data will become a resource for replication of genetic associations identified by multi-ethnic GWASs of kidney traits. We are continuing to develop trans-ethnic methods that account for the genomic differences across ancestries to map variants that are causal to disease. These methods, when integrated with regulatory annotation obtained from relevant cells and tissues[10] (Box 1), reduce the number of variants that are needed for *in vitro* and *in vivo* experimental research to reveal biological mechanisms for CKD.[7]

## What have we learned?

***Trans-ethnic approaches identify novel loci where associations are driven by variants with higher allele frequencies in non-European ancestries.*** Our most recent trans-ethnic GWAS meta-analysis, undertaken in a total of 312,468 individuals of diverse ancestry, has identified 20 novel loci associated with estimated glomerular filtration rate (eGFR) that were replicated in 2 subsequent investigations (Supplementary Table S2). These include at least one locus, mapping to the region encompassing the genes *PMF1* and *BGLAP*, that was not identified in European ancestry GWASs in a much larger sample size. The reported variant at this locus is at higher allele frequency in African ancestry populations than in other ethnic groups, and African American participants were included in our trans-ethnic meta-analysis.

***Causal variants are shared across populations and are amenable to risk prediction.*** Genetic risk scores derived from aggregated GWAS variants can be integrated with traditional risk factors for disease prediction in the clinical setting. However, studies have shown that genetic risk scores identified in European ancestry populations do not generalize to other populations,[11] thereby limiting their clinical utility for disease prediction in diverse populations. Our trans-ethnic GWAS meta-analysis has demonstrated, for the first time, that identified genetic associations with eGFR are shared across ancestries, with mostly homogenous effects in different population groups. Therefore, variants at these loci better represent genetic risk in multi-ethnic populations and can be used for disease prediction, irrespective of ancestry. We have used this genetic risk in Mendelian randomization studies to estimate the causal effects of eGFR on clinical outcomes[7] (Box 1; Figure 1).

***Improved fine-mapping of causal variants at GWAS loci.*** Genetic associations identified in populations of European ancestry typically extend over large genomic regions containing multiple genes, owing to the extensive LD among variants. We have shown that trans-ethnic GWAS meta-analysis approaches that account for differences in LD patterns across populations and the heterogeneity in allelic effects improve fine-mapping resolution, reducing the number of likely causal variants associated with a trait to be queried for regulatory function to one or more nearby genes (Supplementary Figure S1). The improved fine-mapping resolution is driven by both increased sample size (improving power to detect association) and increased ethnic diversity (so that fewer variants are in LD with the causal variant in all populations contributing to the meta-analysis), although it is not possible to disentangle their relative contributions (Figure 2).

***Identified trans-ethnic variants for eGFR are enriched for regulatory annotation in kidney cells and improve understanding of disease biology.*** Our studies have demonstrated significant enrichment of identified trans-ethnic variants associated with eGFR for regulatory sites in kidney-specific cells/tissues (for example, dnase I hypersensitivity sites in multiple kidney cell types, and kidney-specific histone modifications). Our recent trans-ethnic fine-mapping strategy integrated regulatory annotation from kidney cells and tissues to prioritize genomic variants. We further used

**Figure 1** | (continued) function of genomic position (National Center for Biotechnology Information build 37). In each plot, the index variant is represented by the purple symbol. The color coding of all other variants indicates linkage disequilibrium with the index variant in European ancestry haplotypes from the 1000 Genomes Project reference panel: red, $r^2 \geq 0.8$; gold, $0.6 \leq r^2 < 0.8$; green, $0.4 \leq r^2 < 0.6$; cyan, $0.2 \leq r^2 < 0.4$; blue, $r^2 < 0.2$; and grey, $r^2$ unknown. Recombination rates are estimated from Phase II HapMap, and gene annotations are taken from the University of California Santa Cruz genome browser.

kidney-specific gene expression data to link these variants to genes and to map genes to kidney cells in the nephron using single-nucleus RNA sequencing.[7] These studies have provided genes and variants for experimental follow-up studies, for example, genes related to salt sensitivity.[5]

### What is needed?

We still need more information on genetic variation in diverse populations to construct reference panels for the design of efficient GWAS genotyping arrays and for imputation (Box 1), and to uncover population-specific genetic risk variants. The National Human Genome Research Institute Human Heredity and Health in Africa (H3Africa) initiative is generating genomic data in Africans and has several projects related to CKD.[12] The National Heart Lung and Blood Institute Trans-Omic for Precision Medicine (TOPMed) program is performing deep whole-genome sequencing in approximately 150,000 individuals across diverse populations,[2] including African Americans, Hispanics/Latinos, Caribbeans, East Asians, and Pacific Islanders. However, improved representation of global populations at high risk of CKD are needed, such as those of African ancestry and with African admixture, those with Amerindian ancestry, and others for which genetic variation is not well-categorized. This is a primary focus of the COGENT-Kidney Consortium, as these populations can provide invaluable insights into the genetic contribution to disease.

Recent studies have integrated high-dimensional multi-omics and GWASs data to uncover mechanisms by which genotypes influence a trait. However, multi-omics data are still limited in diverse populations. For example, expression quantitative trait loci (Box 1) can help prioritize gene(s) when the associated variant is located in a regulatory region of the genome. The landscape of genetic regulation of gene expression varies across populations, with some genes showing differential expression by ethnicity,[13] but expression quantitative trait loci are, for the most part, available only for populations of European ancestry.

Finally, there is an increasing need for novel statistical methods that integrate omics and GWAS data across diverse populations. Many existing approaches rely on knowledge of LD of the reference population, which may not be available or well-estimated in populations with ancestral admixture. In addition, approaches need to account for differences in allele frequencies and potential heterogeneity in allelic effects across populations when combining data across ancestries.

In summary, studies of diverse populations have not been fully embraced by the research community, for reasons including the small size of available samples, the complexity of genomic LD patterns, the presence of ancestry admixture, and a resistance to adopting methodology that is applicable to populations with or without admixture. We have established the COGENT-Kidney Consortium to overcome some of these challenges by targeting recruitment to diverse ancestries and fostering collaborations in research and methods development appropriate to these populations. By increasing awareness and representation of these populations, we have also shown important gains in translation of some of the loci to biologic mechanisms and in understanding the population impact in diseases associated with low eGFR.

## DISCLOSURE

All the authors declared no competing interests.


## ACKNOWLEDGMENTS

This research is supported by the National Institutes of Health awards R01-DK117445 and R01-MD012765.


## SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

**Table S1.** Sample characteristics of GWASs contributing to the COGENT-Kidney Consortium of diverse ancestries.

**Table S2.** Twenty novel loci for eGFR identified in the COGENT-Kidney Consortium, and replication in subsequent publications of GWASs of eGFR. Note that the *PMF1-BGLAP* was identified only in studies with largely non-European ancestry.

**Figure S1.** Fine-mapping comparison of trans-ethnic GWAS meta-analysis (312,468 individuals) in the COGENT-Kidney Consortium over a meta-analysis restricted to the subset of European ancestry GWAS meta-analysis (134,070 individuals).

## REFERENCES


1. Xie Y, Bowe B, Mokdad AH, et al. Analysis of the Global Burden of Disease study highlights the global, regional, and national trends of chronic kidney disease epidemiology from 1990 to 2016. *Kidney Int*. 2018;94:567–581.
2. Jha V, Garcia-Garcia G, Iseki K, et al. Chronic kidney disease: global dimension and perspectives. *Lancet*. 2013;382:260–272.
3. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538:161–164.
4. Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.

5. Mahajan A, Rodan AR, Le TH, et al. Trans-ethnic fine mapping highlights kidney-function genes linked to salt sensitivity. *Am J Hum Genet*. 2016;99:636–646.
6. Liang J, Le TH, Edwards DRV, et al. Single-trait and multi-trait genome-wide association analyses identify novel loci for blood pressure in African-ancestry populations. *PLoS Genet*. 2017;13:e1006728.
7. Morris AP, Le TH, Wu H, et al. Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nat Commun*. 2019;10:29.
8. Sung YJ, Winkler TW, de Las Fuentes L, et al. A large-scale multi-ancestry genome-wide study accounting for smoking behavior identifies multiple significant loci for blood pressure. *Am J Hum Genet*. 2018;102:375–400.
9. Bentley AR, Sung YJ, Brown MR, et al. Multi-ancestry genome-wide gene-smoking interaction study of 387, 272 individuals identifies new loci associated with serum lipids. *Nat Genet*. 2019;51:636–648.
10. Magi R, Horikoshi M, Sofer T, et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet*. 2017;26:3639–3650.
11. Marquez-Luna C, Loh PR. South Asian Type 2 Diabetes Consortium, et al. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol*. 2017;41:811–823.
12. H3Africa Consortium, Rotimi C, Abayomi A, et al. Research capacity. Enabling the genomic revolution in Africa. *Science*. 2014;344:1346–1348.
13. Stranger BE, Montgomery SB, Dimas AS, et al. Patterns of cis regulatory variation in diverse human populations. *PLoS Genet*. 2012;8:e1002639.