A large-scale multi-ancestry genome-wide study accounting for smoking behavior identifies multiple genome-wide significant loci for systolic and diastolic blood pressure

(The full list of authors and affiliations appears at the end)

Corresponding authors:

Dr. Yun J Sung (yunju@wustl.edu)

Dr. Daniel I Chasman (dchasman@research.bwh.harvard.edu)

Dr. Yun J. Sung
Division of Biostatistics
Washington University School of Medicine
660 South Euclid Avenue, Campus Box 8067
St. Louis, MO 63110-1093
Phone: (314) 362-0053; Fax: (314) 362-2693

Dr. Daniel I. Chasman
Division of Preventive Medicine
Brigham and Women's Hospital
Boston, MA 02215
Phone: 617-278-0821;
Fax: 617-525-4751

Abstract

Genome-wide association analysis advanced understanding of blood pressure (BP), a major risk factor for vascular conditions such as coronary heart disease and stroke. Accounting for smoking behavior may help identify novel BP loci and extend our knowledge of its genetic architecture. We performed genome-wide association meta-analyses of systolic and diastolic BP incorporating gene-smoking interactions in 610,091 individuals. Stage 1 analysis examined ~18.8 million SNPs and small insertion/deletion variants in 129,913 individuals from four ancestries (European, African, Asian, and Hispanic) with follow-up analysis of promising variants in 480,178 additional individuals from five ancestries. We identified 15 new loci that were genome-wide significant ($P < 5\times10^{-8}$) in Stage 1 and formally replicated in Stage 2. A combined Stage 1 and 2 meta-analysis identified 66 additional genome-wide significant loci (13, 35, and 18 loci in European, African and trans-ancestry, respectively). A total of 56 known BP loci were also identified by our results ($P < 5\times10^{-8}$). Of the newly identified loci, 10 showed significant interaction with smoking status, but none of them were replicated in Stage 2. Several loci were identified in African ancestry, highlighting the importance of genetic studies in diverse populations. The identified loci show strong evidence for regulatory features and support shared pathophysiology with cardiometabolic and addiction traits. They also highlight a role in BP regulation for biological candidates such as modulators of vascular structure and function (*CDKN1B*, *BCAR1-CFDP1*, *PXDN*, *EEA1*), ciliopathies (*SDCCAG8, RPGRIP1L*), telomere maintenance (*TNKS*, *PINX1*, *AKTIP*), and central dopaminergic signaling (*MSRA*, *EBF2*).

Words: 245 (250 words max in AJHG)

Introduction

The management of blood pressure (BP) is a major public health priority with implications for the prevention of coronary heart disease, heart failure, stroke, and other vascular conditions. BP is partly under genetic control with moderately high heritability (30-60%),[1] although only a small fraction of the heritability has been explained by variants identified through genome-wide association studies (GWAS).[2] Specifically, the common variants initially identified through three collaborative consortia for genome-wide BP genetics in people of European ancestry[1; 3; 4] explain less than 2.5% of the variance in systolic BP (SBP) or diastolic BP (DBP).[4] Recent reports based on larger sample sizes have increased the number of BP associated variants which together explain about 3.5% of BP variance.[5-7] In contrast, only six BP loci have been identified by GWAS in African ancestry which explain less than 0.54% of BP variance.[8; 9] A focus on main effects to the exclusion of interactions in these studies may have limited the discovery of a full complement of genetic influences on BP. In particular, incorporating interactions between genetic variants and environmental exposures (GxE) represents a novel additional route for discovery of genetic effects on complex traits,[10] including BP, and may more generally extend our knowledge of the genetic architecture of complex traits.[11]

Many lifestyle factors including physical activity, tobacco use, alcohol consumption, stress, and dietary factors influence BP.[12] These lifestyle exposures may also modify the effect of genetic variants on BP. Cigarette smoking is known to influence BP in both acute[13] and chronic[14; 15] fashion, motivating genetic association studies accounting for potential gene-by-smoking interactions. This may help identify novel BP loci, and such BP loci driven by GxE interactions may reveal new biological insights and mechanisms that can be explored for treatment or prevention of hypertension.

The recently established Gene-Lifestyle Interactions Working Group within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium has designed a series of multi-ancestry genome-wide interaction projects focused on assessing the impact of interactions with multiple lifestyle factors on the genetics of cardiovascular traits.[16] The primary goal of these investigations is to use interactions to identify novel trait loci that act

synergistically with lifestyle factors. Large scale interaction studies like this one represent "an important milestone on the path toward a far more complete understanding of the origins of cardiovascular disease and a better understanding of how to manage it".[17] Within this setting, we performed a genome-wide association meta-analysis incorporating gene-smoking interactions (overview shown in **Figure 1**) to identify novel SBP and DBP loci and understand the modulating role of cigarette smoking in the genetic architecture of BP. Here we report our findings based on a total of 610,091 individuals from five ancestry groups which provide adequate power for discovery.[16]

Methods

Overview of participating studies

Men and women between the ages of 18-80 years from five self-reported ancestry groups are represented in this study: European (EUR), African (AFR), Asian (ASN), Hispanic (HIS), and Brazilian admixed (BRA). These participating studies are described in the **Supplemental Notes**. Each study obtained informed consent from participants and approval from the appropriate institutional review boards. Although the participating studies are based on different study designs and populations, all of them have data on BP, smoking, and genotypes across the genome (data imputed using the 1000 Genomes reference panel in most cohorts). In total, this study involves two stages comprising 610,091 individuals.

A total of 48 cohorts participated in Stage 1 and performed genome-wide interaction analyses (**Table S1**). This stage included 80,552 EUR, 27,118 AFR, 13,438 ASN, and 8,805 HIS for an overall total of 129,913 individuals. A total of 76 cohorts participated in Stage 2 and performed analyses of 4,459 variants that were identified in Stage 1 as either genome-wide significant ($P < 5\times10^{-8}$) or suggestive ($P < 10^{-6}$) for any of the BP-smoking combinations for either 1 DF or 2 DF tests (**Table S2**). This stage included 305,513 EUR, 7,786 AFR, 148,932 ASN, 13,533 HIS, and 4,414 Brazilian Admixed (BRA) individuals to a total of 480,178 individuals in Stage 2. Since discoveries to date are largely from EUR populations, we optimized the chances of novel discovery in non-EUR populations (especially in AFR) by recruiting most of the available non-EUR cohorts into Stage 1.

Phenotypes and Lifestyle Variables

The two BP traits were analyzed separately: resting SBP (mmHg) and DBP (mmHg). For individuals taking any anti-hypertensive (BP lowering) medications, their SBP and DBP values were first adjusted for medication effects by adding 15 mmHg to SBP and adding 10 mmHg to DBP.[3] Summary statistics are shown in **Table 1** (more details in **Tables S3-S4)**. These medication-adjusted BP variables were approximately normally distributed, as shown in **Table S5** and **Figure S1**. In addition, to reduce the influence of possible outliers, winsorizing has been applied for each BP value that was more than 6 standard deviations away from the mean.

The participating cohorts have varying levels of information on smoking, some with a simple binary variable and others (such as UK Biobank) with more precise data. We considered two dichotomized smoking variables 'current smoking' status (CurSmk) and 'ever smoking' status (EverSmk), as they were the most widely available information (**Table 1**). Current smoking status was coded as 1 if the subject smoked regularly in past year (and as 0 for non-current smokers which includes both never and former smokers). Ever smoking status was coded as 1 if the subject smoked at least 100 cigarettes during his/her lifetime (and as 0 for the never-smokers). Smoking status was assessed at the time of the BP measurements. When subjects have the multiple smoking measures that were inconsistent, they were excluded from analysis. Subjects with missing data for BP, the smoking variable, or any covariates were excluded from analysis.

Genotype Data

Genotyping was performed using Illumina (San Diego, CA, USA) or Affymetrix (Santa Clara, CA, USA) genotyping arrays. Each study performed imputation to impute genotypes for single nucleotide polymorphisms (SNPs), short insertions and deletions (indels), and larger deletions that were not genotyped directly but are available from the 1000 Genomes Project.[18] Information on genotype and imputation for each study is presented in **Tables S6-S7**. For imputation, most studies used the 1000 Genomes Project Phase I Integrated Release Version 3 Haplotypes (2010-11 data freeze, 2012-03-14 haplotypes), which contain haplotypes of 1,092 individuals of all ethnic backgrounds.

Cohort-specific GWAS Analysis

For SBP and DBP separately, each study performed association analyses accounting for two smoking exposure variables, current smoking (CurSmk) and ever smoking (EverSmk). In Stage 1, we considered two models to account for gene-smoking interactions. For the first 'joint' model, a regression model including both genetic main and GxE interaction effects

$$\mathbb{E}[Y|G,C] = \beta_0 + \beta_E Smk + \beta_G G + \beta_{GE} Smk * G + \boldsymbol{\beta_C C}$$

was applied to the entire sample. For the second 'stratified' model, analyses of the genetic main-effect regression models

$$\mathbb{E}[Y|C, Smk = 0] = \gamma_0^{(0)} + \gamma_G^{(0)} G + \boldsymbol{\gamma_C^{(0)} C}$$

$$\mathbb{E}[Y|C, Smk = 1] = \gamma_0^{(1)} + \gamma_G^{(1)} G + \boldsymbol{\gamma_C^{(1)} C}$$

were applied separately to the Smk = 0 unexposed group and to the Smk = 1 exposed group (smokers). *Y* is the medication-adjusted BP value, Smk is the smoking variable (with 0/1 coding for the absence/presence of the smoking exposure), *G* is the dosage of the imputed genetic variant coded additively (from 0 to 2), and **C** is the vector of all other covariates, which include age, sex, field center (for multi-center studies), and principal component (PC)s (to account for population stratification and admixture). No additional cohort-specific covariates were included. Our previous work showed that the two (joint and stratified) models provided highly similar inference.[19] Therefore, we considered only the first 'joint' model in Stage 2.

Each study in Stage 1 performed GWAS analysis within each ancestry and provided a) the estimated genetic main effect β_G, estimated interaction effect β_{GE} and a robust estimate of the corresponding covariance matrix under the joint model; and b) estimates of the stratum-specific effects $\gamma_G^{(0)}, \gamma_G^{(1)}$ and robust estimates of their standard errors (SE) under the stratified model. Each study in Stage 2 provided estimates of the genetic main effect β_G, the interaction effect β_{GE} and robust estimates of the corresponding covariance matrix under the joint model at 6,370 select variants. Robust estimates of covariance matrices and SEs were used to safeguard against both mis-specification of the mean model and violation of the assumption of constant BP variance across smoking groups (heteroscedasticity).[20; 21] Association analysis was performed using various software (**Tables S6-S7**). To obtain robust estimates of covariance matrices and robust SEs, studies of unrelated subjects used either the R package sandwich[22] or ProbABEL.[23] To

account for relatedness in families, family studies used the generalized estimating equations (GEE) approach, treating each family as a cluster, or the linear mixed effect model approach with a random polygenic component (for which the covariance matrix depends on the kinship matrix).

Quality control

Study investigators participating in this study have ample experience in main-effect based GWAS for multiple phenotypes and are very familiar with validated approaches for quality control (QC) of phenotype, genotype, and imputed data. For example, cohort level analyses used PCs as covariates to deal with population structure; family studies used suitable software packages to deal with relatedness (Table S6). Overlap among some of the participating cohorts is a potential possibility. However, when there was known overlap of samples across cohorts, one of the cohorts used a non-overlapping sub-sample for their analysis.

We performed extensive QC using the R package EasyQC[24] for all cohort-specific GWAS results. In Stage 1, each cohort provided 12 GWAS result files (2 BPs x 2 Smoking exposures x 3 analyses, 1 for model 1 and 2 for model 2) for each ancestry group. Each GWAS result file included approximately 8-15 million high-quality variants (depending on ancestry), as cohorts applied a preliminary filter on their imputed data excluding variants with minor allele frequency (MAF) < 1% or imputation quality measure < 0.1. We performed two QC levels: "study-level" and "meta-level". To identify problems with population substructures or relatedness, we have examined QQ plots and genomic control inflation factors (lambdas) on a study-by-study level (to identify study-specific issues) as well as on the meta-analysis result (to identify cross-study issues). Because GWAS were performed within each ancestry, the "study-level" QC also carefully checked the provided allele frequencies against the retrospective ancestry-specific 1000 genomes reference panel. Finally, marker names were harmonized to ensure consistencies across cohorts. In addition, we contrasted results from the joint model and stratified models in Stage 1 cohorts, as explained elsewhere.[19] The "meta-level" QC reviewed result files of a specific analysis (e.g., SBP-CurSmk-Model1) across all cohorts: this included 1) visually comparing summary statistics (mean, median, standard deviation, inter-quartile range, minimum, maximum) on all effect estimates standard errors (SEs) and p-values; and 2) examining SE-N and QQ plots

to reveal issues with trait transformation[24] or other analytical problems. Any problems found during QC steps, including major differences from the ancestry-specific reference panel and any inflation of lambdas within studies, were communicated and resolved with the individual cohorts. Similar QC steps were applied to cohort-specific results in Stage 2. More detailed information about the QC steps, including major QC problems encountered and how they were resolved, are described elsewhere.[16]

The most crucial filter during the meta-analysis was approximate DF = min (MAC0, MAC1) * imputation quality measure; this is based on the minor allele count (MAC) in each stratum (MAC0 and MAC1) and imputation quality measure, where MAC0 = 2 * MAF_{E0} * N_{E0} for the unexposed group (with MAF_{E0} and sample size N_{E0} for E=0 stratum) and MAC1 = 2 * MAF_{E1} * N_{E1} for the exposed group. In meta-analysis, to exclude unstable cohort-specific results that reflect small sample size, low MAF, or low imputation quality measures, variants were excluded if approximate DF < 20. This filtering threshold was decided after considering various thresholds and examining the resulting QQ and Manhattan plots. More details are provided in Supplemental Notes. Variants were further excluded if imputation quality measure < 0.5. This value of 0.5 was used regardless of the software used for imputation, because imputation quality measures are shown to be similar across imputation software.[25]

Meta-analysis

After conducting extensive quality control and selecting high-quality variants, approximately 18.8 million SNPs and small insertion and deletion (indels) variants were included in the meta-analysis (the number of variants varied across the ancestry groups). We performed meta-analysis using both models in Stage 1 and using the joint model in Stage 2. For both stages, we performed meta-analysis using the 1 degree of freedom (DF) test of interaction effect and 2 DF tests of testing both SNP main and interaction effects. Wald test statistics approximately follow either a chi-squared distribution with 1 DF under H_0: $\beta_{GE} = 0$ for the 1 DF test or a chi-squared distribution with 2 DF under H_0: $\beta_G = \beta_{GE} = 0$, for the 2 DF test. In the joint model, inverse-variance weighted meta-analysis was performed for the 1 DF test and the joint meta-analysis of Manning et al[26] for the 2 DF test, both using METAL.[27] In the stratified model, we performed meta-analysis using the approach of Randall et al[28] for the 1 DF test and the approach of Aschard

et al[29] for the 2 DF test. Both tests in the stratified model were computed using the R package EasyStrata.[30] More details are described elsewhere.[19]

Ancestry-specific meta-analyses using inverse-variance weighting were performed to combine cohort-specific results within each ancestry. The ancestry-specific results were then combined through meta-analysis to obtain evidence of "trans-ancestry" association. In Stage 1, 80 separate genome-wide meta-analyses were performed: 2 BPs x 2 smoking exposures x 4 (2 tests in the joint model, 2 stratified groups in the stratified model) x 5 ancestries (4 ancestry-specific and one trans-ancestry to combine ancestry-specific results). In this stage, genomic control correction[31] was applied twice, first for cohort-specific GWAS results if their genomic control lambda value was greater than 1, and again after the meta-analysis results. Variants were excluded if they were represented by valid data in fewer than 5,000 samples and 3 cohorts. Variants that were genome-wide significant ($P < 5\times10^{-8}$) or suggestive ($P < 1\times10^{-6}$) in any of Stage 1 analyses were pursued for Stage 2 analysis. In Stage 2, 48 separate meta-analyses were performed using the joint model: 2 BPs x 2 smoking exposures x 2 (2 tests; 1 DF and 2 DF tests) x 6 ancestries (5 ancestry-specific and one trans-ancestry to combine ancestry-specific results). Genomic control correction was not applied to the replication statistics as association analysis was performed only at select variants. Similarly, 48 separate meta-analyses were performed to combine Stages 1 and 2 results.

Genome-wide Significant Variants

If a variant reached genome-wide significance ($P < 5\times10^{-8}$) through any of these 48 combined association meta-analyses (which are not independent), then the variant was considered as genome-wide significant. To identify a set of independent (index) variants through ancestry-specific and trans-ancestry analysis, we performed the linkage disequilibrium (LD)-based clumping procedure using PLINK[32] and EasyStrata.[30] A locus is defined through LD-based clumping that uses both physical distance (± 1 Mb) and LD threshold of $r^2 > 0.1$. Since valid methods do not exist for conditional analysis involving interactions across multi-ancestry studies, we relied on a relatively more stringent LD threshold ($r^2 > 0.1$) for identifying "independent" loci. As LD reference, ancestry-specific 1000 Genomes Project data were used for ancestry-specific results and the entire cosmopolitan dataset was used for trans-ancestry results. False discovery rate (FDR) q-values were computed using the R function p.adjust using the step-up method by Benjamini and Hochberg.[33]

BP variance explained

Since variants weakly correlated with index variants ($0.1 \leq r^2 \leq 0.2$) can contribute to the percent variance, for the purposes of calculating percent variance, we carried out clumping using slightly less conservative LD threshold ($r^2 > 0.2$ instead of > 0.1). The percent of variance explained in SBP and DBP by all previously known (158) and newly identified variants (132 using LD threshold of >0.2 for clumping) was evaluated in several studies from multiple ancestries (see **Table S8**). BP variants previously identified in any ancestry were considered as 'known' variants. Similarly, we considered all index variants representing newly identified loci as novel for this purpose regardless of which ancestry they were identified in; separate interaction terms were included for newly identified variants. Known and newly identified variants (combined from all ancestries) were used in assessing the percent variance.

Percent variance was calculated using standard regression models. Four nested models were considered. The first model included the smoking variables and standard covariates (age, sex, PCs, etc.); the second model included those covariates and all known variants; the third model contained all those previous variables and all newly identified variants (excluding any interaction terms); finally, the fourth model contained all those (covariates, known, and novel) plus the interaction terms. Each of SBP and DBP was regressed on the relevant predictors in each of the 4 models. The r^2 values obtained from the regressions were used as measures of the percent variance explained by the respective models. Through sequential subtraction of appropriate r^2 values, we determined the "additional" percent variance explained by a given set of variants. For studies with $N < 20{,}000$, we used a stepwise regression procedure with significance tests for inclusion of one variant at a time and for backward elimination of redundant variants.

Functional inference

Variant Effect Predictor (VEP) from Ensembl was used to obtain the gene name for each locus. For the variants whose gene names were not identified by VEP, NCBI SNP database was used to obtain the closest gene. We applied several computational strategies to infer biological functions associated with our newly identified loci. We used HaploReg, RegulomeDB, and GTEx[34] to obtain annotations of the noncoding genome, chromatin state and protein binding annotation from the Roadmap Epigenomics and ENCODE projects, sequence conservation across

mammals, and the effect of SNPs on expression from eQTL studies. To further assess putative functionality for the new loci, we searched for *cis* associations between new variants and gene transcripts using previously published eQTL analyses, which includes the GTEx.[34]

Further eQTL evidence was queried using the eQTL database of Joehanes et al.[35] for transcripts associated in both *cis* and *trans* in over 5,000 individuals from the Framingham Heart Study, with genome-wide false discovery rate (FDR) < 0.05. Two Gene-Set Enrichment Analysis (GSEA) queries were then performed on December 23, 2016 to determine the enrichment of biological processes and disease pathways of the resulting transcripts. Prior to the queries, duplicated gene names and genes with provisional names (such as LOCXXX) were first removed. Then, for each transcript probe associated with more than one gene name, only the first gene name was taken. This process yielded 127 gene names for the GSEA query. For querying biological processes, option C5:BP was selected on the GSEA website. For querying disease pathway, option C2:CP was selected. Both GSEA queries were set at FDR < 0.05 threshold to guard against multiple comparison errors.

Pathway and gene set enrichment analysis

We conducted four separate DEPICT analyses based on the following criteria that were applied to our combined association meta-analysis results. We utilized variants showing genome-wide significant joint effect association with 1) SBP in Europeans ($P_{EUR.SBP} < 5\times10^{-8}$), 2) DBP in Europeans ($P_{EUR.DBP} < 5\times10^{-8}$), 3) SBP in trans-ancestry analysis ($P_{Trans.SBP} < 5\times10^{-8}$) or 4) DBP in trans-ancestry analysis ($P_{Trans.DBP} < 5\times10^{-8}$). For each combination, DEPICT first performed the following steps to obtain the input of the prioritization and enrichment analyses: non-overlapping regions lists of independent variants were obtained using 500 kb flanking regions and LD $r^2 > 0.1$ using the 1000 Genomes data,[18] resulting variants were merged with overlapping genes ($r^2 > 0.5$ with a functional coding variant within the gene or *cis*-acting regulatory variant), and the major histocompatibility complex region on chromosome 6 (base position 25,000,000 - 35,000,000) was excluded.

DEPICT prioritized genes at the associated loci based on their functional similarity. Functional similarity of genes across associated loci was quantified by computing a gene score that was

adjusted for bias through confounders such as gene length. Experiment-wide FDR for the gene prioritization was obtained by repeating the scoring step 50 times based on lead variants from 500 pre-compiled null GWAS. For the gene-set enrichment analyses, DEPICT utilized a total of 14,461 pre-compiled reconstituted gene sets comprising 737 Reactome database pathways, 2,473 phenotypic gene sets (derived from the Mouse Genetics Initiative), 184 Kyoto Encyclopedia of Genes and Genomes (KEGG) database pathways, 5,083 Gene Ontology database terms, and 5,984 protein molecular pathways (derived from protein-protein interactions). For the tissue and cell type enrichment analyses, DEPICT tested whether genes harboring associated loci are enriched for expression in any of the 209 MeSH annotations for 37,427 microarrays of the Affymetrix U133 Plus 2.0 Array platform.

To further identify connected gene sets and pathways implicated by our findings, we performed GeneGO analysis and text data mining using Literature Lab.[36] GeneGO (known also as MetaCore) evaluates p-values for pathways by mapping a list of target genes to each pathway and comparing those that arise by chance using a hypergeometric distribution formula. GeneGO implements a correction of p-values using a false discovery rate. Literature Lab of Acumenta evaluates co-occurrences in the publication records of a list of genes and biological and biochemical terms. The analysis compares the gene input set against the average of 1,000 randomly generated similar size sets, providing a spectrum of statistically significant associations. Our Literature Lab analysis included the use of 17,261,987 PubMed abstracts, out of which 10,091,778 abstracts include one or more human genes.

Results

Study overview

We performed the traditional 2-step approach with discovery in Stage 1 followed by formal replication in Stage 2. Because this study was not optimally designed for replications in non-EUR (especially in AFR) ancestry, to identify additional loci, we performed combined analysis of stages 1 and 2 to maximize power for discovery[37] (**Figure 1**). For the 2-step approach, we performed ancestry-specific meta-analysis in each of 5 ancestries and trans-ancestry analysis in Stage 2. We checked whether each of the genome-wide significant loci in Stage 1 was replicated

in Stage 2 using Bonferroni-adjusted significance level (0.05/74, see details below). For the combined analysis, we performed ancestry-specific meta-analysis combining both Stages 1 and 2 (discovery and follow-up) in each of 5 ancestries; these ancestry-specific meta-analyses results were then combined to perform trans-ancestry analysis at 4,459 variants using a total of up to 610,091 individuals.

Two-step approach of discovery followed by replication

Of the 4,459 significant or suggestive variants selected from Stage 1 meta-analyses, 3,222 were replicated in Stage 2 with $P < 0.05/4,459$ (to an aggregate replication rate of 72.3%). Of the 1,993 variants that were genome-wide significant ($P < 5\times10^{-8}$) in Stage 1 analysis, 1,836 were replicated in Stage 2 with $P < 0.05/1,993$ to a replication rate of 92.1%. These 1,993 genome-wide significant variants in Stage 1 belong to 114 independent loci . Of the 114 loci, 40 loci (consisting of 1,644 variants) contain previously published BP variants.[1; 3-7] Of the remaining 74 newly identified loci (consisting of 349 variants), 15 loci were formally replicated in Stage 2 using Bonferroni-adjusted significance level ($P < 0.05/74$) (**Table 2**); all 15 novel loci were replicated even when using the more conservative adjustment threshold $P < 0.05/349$. In addition, 25 more of the remaining 59 loci were nominally replicated ($P < 0.05$) in one or more of the analyses in Stage 2 ($P < 0.05$), and 27 more showed the same direction of effect in Stages 1 and 2. For 7 loci, no additional data were available in Stage 2 and, therefore, it was not possible to check for replication. For the 15 formally replicated loci, estimates of the genetic main effects were all consistent between Stage 1 and 2; estimates of SNP-smoking interaction effects were not statistically significant (forest plots; **Figure S3**). All of the 15 replicated loci were genome-wide significant in European ancestry. Furthermore, 10 loci also had supporting evidence from non-European ancestry, resulting in stronger statistical significance from trans-ancestry analysis (**Figure S3**, **Table 2**). Quantile-quantile (QQ) plots for the genome-wide Stage 1 meta-analysis are shown in **Figure S2**.

Of the 15 formally replicated loci, six loci (indicated by f in Table 2) are completely novel, at least 1MB away from any previously published BP variants. Three of them (near *PRAG1*, *MIR124-1*, and *FTO*) show compelling biological relevance (see below) and eQTL evidence (**Figure 2**). The locus zoom plots of all newly identified loci identified in this paper are shown in

Figure S4. The remaining 9 loci are novel signals (which meet our definition of a locus) near but not in LD ($r^2 < 0.1$) with known BP loci. For example, near the well-known BP locus *ATP2B1* on chromosome 12, there were 2 independent signals identified in European ($P = 4.1\times10^{-41}$), Asian ($P = 1.5\times10^{-13}$), and trans-ancestry ($P = 2.5\times10^{-54}$) analyses. Near another well-known BP locus *MTHFR-NPPB-CLCN6*, we identified 3 additional independent signals (with p-values as small as 4.3×10^{-34} at index variants, spanning 196kb [from 11,827,796 to 12,023,500] on chromosome 1).

Combined analysis of Stages 1 and 2

Combined meta-analysis of Stages 1 and 2 identified a total of 82 additional independent loci ($P < 5 \times 10^{-8}$) not identified by the 2-step approach. Association statistics for all genome-wide significant variants in the combined meta-analysis are provided in **Table S9**. Manhattan plots of the combined meta-analysis for each BP trait using the 1 DF interaction and 2 DF joint tests are shown in **Figures S5-S8**. Summary Manhattan plots for SBP and DBP with the minimum p-values across all analyses are shown in **Figure S9**. QQ plots are shown in **Figure S10**.

Of these 82 additional loci identified through combined analysis, 16 loci contain previously published BP variants.[1; 3-7] All of the remaining 66 loci had a low false discovery rate (FDR *q* value < 0.1 for all 66 loci and < 0.01 for 60 of the loci, **Table S10**). Of these 66 loci, 18 and 13 loci were identified through trans-ancestry (**Table 3**) and European-ancestry (**Table 4**), respectively. Except for one locus, they were suggestive ($P < 1\times10^{-6}$) in Stage 1 analyses but became significant in the combined Stages 1 and 2 meta-analysis (**Tables 3-5**). The strength of the combined analysis was exemplified by a locus in *HOTTIP* on chromosome 7 (locus 4 in **Table 3**), which were suggestive in Stage 1 analysis ($P = 9.4\times10^{-7}$) and identified through the combined analysis in European ($P = 6.0\times10^{-29}$), Asian ($P = 1.2\times10^{-10}$), and trans-ancestry ($P = 3.6\times10^{-41}$, see **Figure S3**). Genome-wide significant loci from trans-ancestry analysis did not show strong evidence of heterogeneity across ancestry groups.

Of the 66 identified loci, 35 were found through African-ancestry only (**Table 5**). These loci were mostly low-frequency with MAF between 1% and 5% (**Table 5**). Of these 35 loci, 4 were genome-wide significant in Stage 1 African ancestry and stayed significant in the combined

analysis (although not formally replicated in Stage 2). One such locus was near *BMP7* on chromosome 20 (with $P = 5.8\times10^{-10}$ in Stage 1; P = 0.03 in Stage 2; $P = 4.2\times10^{-12}$ in Stage 1+2). Six loci were suggestive ($P < 1\times10^{-6}$) in Stage 1 analyses but became significant in the combined Stages 1 and 2 meta-analysis. One such locus was near *WSCD1* on chromosome 17 (with $P = 8.7\times10^{-7}$ in Stage 1; P = 0.00047 in Stage 2; $P = 1.8\times10^{-10}$ in Stage 1+2). The remaining 25 loci were genome-wide significant in Stage 1 African ancestry but not represented in Stage 2 African ancestry due to limited sample sizes and low MAF. Furthermore, 15 loci were African-specific loci; they had MAF < 1% in the other ancestry groups and were filtered out by the individual studies (by design), and therefore results are unavailable for further analysis. In the non-AFR ancestry results, genome-wide significant variants at newly identified loci were mostly common (with MAF ≥ 5%) and had similar MAF distributions as those at known loci (**Figure S10**).

Known BP loci

At most of the 56 known BP loci[1; 3-7] identified in the two-step or combined analyses, the lead variant identified by our analyses was the same as the one previously published (**Table S11**); European-, Asian-, and trans-ancestry results identified 48, 14, and 50 of these variants, respectively. In the remaining loci, our results identified a variant in the same locus as the known BP variant. The most significant results were observed at well-known BP loci: *ATP2B1* (rs17249754 on chromosome 12, trans-ancestry $P_{SBP} = 4.8\times10^{-85}$; $P_{DBP} = 5.5\times10^{-57}$) and *SH2B3-ATXN2* (rs3184504 on chromosome 12, trans-ancestry $P_{SBP} = 3.2\times10^{-36}$; $P_{DBP} = 6.0\times10^{-67}$).

The role of interactions

Interaction effects contributed in varying degrees to the evidence of association for the 81 newly reported genome-wide significant loci (**Tables 2-5**). The genetic effects of these new index variants (each index variant representing a locus with the smallest p-value) were different in smokers and non-smokers, thus highlighting the potentially important role of interactions (**Figure 3**). Among the 81 index variants, 10 variants showed genome-wide significant interactions with smoking exposure status (1 DF interaction $P < 5\times10^{-8}$). All 10 of these variants, most of which were identified in African ancestry, show larger effects on BP in smokers (**Figure 3**). However, none of the interactions were replicated in stage 2. In addition, of the 158 previously reported BP variants, two (rs3752728 in *PDE3A* and rs3184504 in *SH2B3-ATXN2*)

show significant evidence of interactions with smoking using Bonferroni correction (1DF interaction $P < 0.05/158$). Twenty seven additional variants show nominal evidence of interaction (with $P < 0.05$).

To minimize spurious results, we winsorized extreme BP values and used robust standard errors in cohort-specific analyses. Moreover, since non-normality and unequal BP variances among smokers and non-smokers can lead to false positives, we examined these characteristics in three large studies (ARIC, UK Biobank, and WGHS). The distributions look very similar in exposed and unexposed groups (histograms in **Figure S1**). The variances across strata are also very similar (**Table S5**). Moreover, on average across all Stage 1 cohorts, skewness is 0.64 for SBP and 0.36 for DBP; kurtosis is 3.52 for SBP and 3.32 for DBP (**Table S3**). There do not seem to be substantial deviations from normality although moderate deviations exist. , Therefore, it is less likely that the interaction effects at these 10 newly identified loci are spurious.

BP variance explained

In several large cohorts, we calculated the percent of BP variance explained by various loci across four ancestries (**Table S8**). The variance explained by the 158 previously known loci ranges from 1.1% (in HIS) to 3.2% (in EUR) for SBP and ranges from 1.6% (in ASN & HIS) to 3.4% (in AFR) for DBP. The additional variance explained by the newly identified loci and their interactions ranges from 0.6% (in EUR) to 2.6% (in AFR) for SBP and ranges from 0.3% (in ASN) to 3.2% (in AFR) for DBP. The percent variance explained is ideally calculated in large individual studies which did not participate in our analysis in Stage 1 or 2. However, having recruited most of the studies available to us into Stage 1 or 2 (for maximizing power), we had to use some of the same studies for this purpose and therefore some of the variance estimates may be somewhat inflated. In an independent EUR study (Airwave study, N=14,002) which did not participate in Stage 1 or 2, known variants explained 1.6% of variance in SBP and DBP, and newly identified variants and their interactions explained 1.2% variance in SBP and 1.3% variance in DBP (**Table S8**). These variances are within the ranges noted, lending credibility to the results from other studies. Note that both known and newly identified variants (with their interactions) explain some of the BP variance across ancestry groups.

Functional annotation and eQTL evidence

For all 81 index variants representing the newly identified loci, we obtained functional annotations using HaploReg[38] and RegulomeDB.[39] There were 2 coding variants (1 missense and 1 synonymous). Of the remaining non-coding variants (29 intronic and 52 intergenic), 17 are located in promoter histone marks, 53 in enhancer histone marks, 29 in DNase I marks, and 10 altered the binding sites of regulatory proteins (**Table S12**). Conserved among vertebrates were 6 variants as identified via GERP[40] and 5 variants via SiPhy.[41] RegulomeDB assigned class 1f (strong evidence for enhancer function) for 2 variants (**Table S12**), each of which likely affects the binding of regulatory elements and is linked to expression of a gene target. Of these, rs12741980 (locus 2, **Table 4**) is near the well-known BP locus *MTHFR-NPPB-CLCN6* and a *cis*-acting expression quantitative trait locus (eQTL) for *NPPA-AS1*, which is expressed in multiple tissues, including thyroid and whole blood. Also, newly identified variant rs180940 (locus 10, **Table 4**), with RegulomeDB score of 1f, is a *cis*-eQTL for the known locus *ADRB1*, an adrenergic receptor that mediates effects of the hormone epinephrine and the neurotransmitter norepinephrine,[42] although it is about 80 kb upstream from this locus. Of note, our results identified this known BP locus (rs2782980, $P = 1.1\times10^{-21}$ and rs1801253, $P = 1.3\times10^{-22}$, in **Table S11**).

Among the 81 newly identified index variants, *cis*-eQTL evidence was available for 39 variants with varying degrees of association with expression probes (**Table S12**). In particular, 21 of them were identified by GTEx[34] as *cis*-eQTLs across various tissues (**Table S13**). However, most of them are for *cis*-eQTLs that differ from their nearest assigned genes. For example, an intronic variant in *WNT2B* (rs351364) is a *cis*-eQTL for *RHOC*, which serves as a microtubule-dependent signal that is required for the myosin contractile ring formation during cell cycle cytokinesis. Additionally, 11 variants (including rs7823056 in **Figure 2**) on chromosome 8 are *cis*-eQTLs for *PRAG1*, which is expressed in multiple tissues including the cerebellum and thyroid. The most abundant evidence of *cis*-eQTL association (with 44 eQTL hits from multiple studies) was observed for rs2243873, a intronic variant of *EHMT2*; it is predicted to regulate expression of many genes including *HLA-C*, *HLA-B* and *HLA-DRB1* across multiple tissues.

The majority of the available data on tissue expression is derived from studies with a breadth of tissue types but with small sample sizes that limit the statistical power to detect association. A more in-depth but single-tissue functional annotation, reporting both *cis*- and *trans*-acting elements, was recently performed using microarray-based gene and exon expression levels in whole blood from over 5,000 individuals of the Framingham Heart Study.[35] In this database, a total of 170 variant-transcript pairs (representing 36 variants) were significant at false discovery rate (FDR) < 0.05 (**Table S14**). There were 113 pairs for *cis*-eQTL, 3 pairs for *trans*-eQTL, and 54 pairs for long-range *cis*-eQTL where the variant is located over 1 Mb away from the target transcript on the same chromosome. Among 36 variants, 9 variants were eQTLs for more than 5 gene transcripts. For example, the 4 SNPs with the most significant eQTL evidence were rs2243873 (described in the previous paragraph), rs2071550, rs7823056, and rs13271489 (Locus 8 in **Table 2** and **Figure 2**) associated with 29, 12, 11, and 10 transcripts, respectively.

Pathway and gene set enrichment analysis

In order to distinguish between functional properties of loci with SBP compared to DBP effects, as well as between European-specific and trans-ancestry mechanisms, we conducted gene prioritization, gene set enrichment and tissue enrichment analyses using DEPICT[43] separately by the four combinations of ancestry (EUR vs trans-ancestry) and BP trait (DBP vs SBP, **Methods**, **Tables S15-S20**). DEPICT significantly prioritized genes (FDR < 5%) at 12 European DBP loci, 26 European SBP loci, 34 trans-ancestry DBP loci and 27 trans-ancestry SBP loci (**Tables S15-S19**). In 43 cases, the prioritized gene for a specific locus differed from the nearest gene of the lead variant. Our DEPICT gene-set enrichment analyses highlighted a role for the identified variants in the cardiovascular system – predominantly affecting blood vessel biology (FDR < 0.05 for a total of 134 gene-sets across the four analyses, **Table S20**).

To identify connected gene sets and pathways implicated by our findings, we performed GeneGO analysis and text data mining using Literature Lab.[36] The genes near our findings were enriched by GeneGO disease class “Chronic Kidney Failure” ($P = 9.2\times10^{-6}$). These same genes were also included in the much larger network representing the GeneGO disease class “Fibrosis” ($P = 3.39\times10^{-7}$) suggesting that genetic contribution of chronic kidney disease to BP is likely mediated by fibrosis. With Literature Lab, for the “Diseases” medical subject heading (MeSH),

Hypertension was strongly enriched (P = 0.0011), with contributions from *ACE* (93.4%), *MTHFR* (2.12%), *ATP2B1* (1.18%), *NPPB* (0.54%), *SH2B3* (0.43%), and *SLC4A7* (0.13%). For the "Physiology" MeSH, blood pressure and cardiovascular physiological phenomena were enriched. Blood pressure (P = 0.0026) had contributions from *ACE* (96.77%), *ATP2B1* (1.16%), *NPPB* (0.6%), *MTHFR* (0.46%), *SH2B3* (0.46%), and *FTO* (0.3%); Cardiovascular physiological phenomena (P = 0.0056) had contributions from *ACE* (97.89%), *NPPB* (1%), *ATP2B1* (0.37%), *MTHFR* (0.2%), *SH2B3* (0.16%), *TNFSF12* (0.09%), and *AP5B1* (0.05%).

Associations of BP loci with cardiometabolic traits

To test association of all 81 newly identified BP-associated index variants with other cardiometabolic traits, we obtained lookup results for coronary artery disease (CAD), stroke, and other cardiometabolic traits related to adiposity, diabetes, and renal function (**Tables S21-S27**). We found that several of our newly identified index variants corroborate those previously associated with these cardiometabolic traits. To quantify this, we counted the number of variants that show association with p-value < 0.05 (highlighted in red). In the vast majority of cases (39 out of 47, $P_{Binomial}=2.8 \times 10^{-6}$), the observed count is higher than that expected by chance alone (**Table S27**). For example, we observed 9 and 14 such associations with CAD and myocardial infarction, respectively, where the expected count is 2.6 for both traits. This is consistent with the known association of increased BP with CAD mortality, independent of other risk factors.[44] Likewise, overlapping signals with other cardiometabolic traits, including those related to adiposity, diabetes, and renal function, support the notion that these traits share a common pathophysiology. For many of the obesity-related trait associations found in the GIANT Consortium, the genetic effects was influenced by adjustment and/or stratification by smoking status[45] (**Table S26**).

We also found corroborating evidence for some well-known loci associated with the renin-angiotensin-aldosterone system (RAAS), including *NPPA*, *NPPB*, and *SLC17A1-4* (**Tables 2-4**).[4] Variants in and near these loci have also been associated with CAD-related traits (*NPPA*/*NPPB*; **Table S21**), stroke (*NPPA*/*NPPB*, and *SLC17A1-4*; **Table S22**), obesity-related traits (*NPPA*/*NPPB*, and *SLC17A1-4*; **Table S23**), and diabetes-related traits (*SLC17A1-4*; **Table S24**)

The confluence of these data provide further evidence of the biologic relevance of these loci to BP regulation and the shared pathophysiology among cardiometabolic traits.

Biological relevance of newly identified variants associated with BP

Ciliopathies: Cilia are cellular protuberances found in several tissues including the kidney and brain that serve several purposes including cellular structure, growth, mobility, secretion, and environmental response. New BP candidate genes *SDCCAG8* (locus zoom plot in **Figure 2**), *RPGRIP1L*, and *TMEM231* encode products that play critical roles in the structure and function of primary cilia including microtubules, basal bodies, and centrosomes. Mutations in these genes can lead to nephronophthisis-related ciliopathy, a monogenic cause of end-stage renal disease. *DPYSL2*, which encodes a microtubule assembly protein, has also been implicated in polycystic kidney disease.[46] Cilia also contain actin fibers with motor proteins (dynein and kinesin) responsible for the transport of mitochondria and other cargo. *DYNC2LI1* is another dynein-associated protein associated with BP; dynein proteins co-localize in the kidney with the water channel aquaporin-2.[47]

Telomere Maintenance: Since telomere length shortens with successive cell divisions, it has been proposed as a reflection of biologic age.[48] Several genes with significant association with BP have roles in telomere maintenance including *TNKS*, *PINX1*, *AKTIP* (**Tables 2-4**), and *TERF2IP*. *TNKS*, which is in a locus previously associated with stroke-, obesity-, and diabetes-related traits in other studies (**Tables S22-S24**), plays a role in the insulin-stimulated translocation of GLUT4 (glucose transporter) to the plasma membrane,[49] and has additionally been associated with cardiovascular disease (CVD) risk and the inflammatory biomarker, C-reactive protein.[50] *PINX1* has been previously associated with CVD,[51] carotid artery intima-media thickness,[52] and serum triglyceride levels,[53] and has also been associated with obesity- and diabetes-related traits (**Tables S23-S24**). *AKTIP* has been previously associated with stroke-related traits in other studies (**Table S22**). Of note, the association at *TNKS*, *PINX1*, and *AKTIP* with multiple adiposity traits in the GIANT Consortium were strengthened by adjustment for smoking status (**Table S26**). *TERF2IP* has also been associated with stroke risk[50] and coronary artery disease traits (**Tables S21-S22**).

Central Dopaminergic Signaling: Dopaminergic signaling in the kidney is known to modulate the secretion of renin[54] and other key regulators of salt-water balance.[55] There is evidence that central dopamine signaling also modulates BP via mechanisms that are independent of changes in sodium excretion.[56] Early stages of Parkinson's disease, a neurodegenerative disorder characterized by the loss of dopamine-secreting neurons, is characterized by autonomic dysfunction and BP dysregulation.[57] In the current study, genes involved in central dopamine signaling were associated with BP, including *MSRA* and *EBF2*, which promote the survival and development of dopaminergic neurons, and *GPR19*, a G-protein coupled receptor for the dopamine D_2 receptor. *MSRA* has been previously associated with body mass index after adjustment with smoking status in the GIANT Consortium (**Table S26**) and *GPR19* with renal function (**Table S25**) in the COGENT-Kidney Consortium.

Modulators of vascular structure and function: *CDKN1B*, *BCAR1-CFDP1*, *PXDN*, and *EEA1* are involved in pathways that contribute to angiotensin II-induced vascular hypertrophy. Notably, the association of *PXDN* and *EEA1* with BP is limited to AFR. *CDKN1B* has been previously associated with renal function (**Table S25**). *BCAR1-CFDP1* has furthermore been identified as a genome-wide significant locus for carotid artery intima-media thickness and coronary artery disease risk (also **Table S21**);[58] a potential causal variant in a *BCAR1* regulatory domain has been identified.[59] *KCNG3* and *KCNE4* are subunit modifiers of voltage-gated potassium channels expressed in vascular smooth muscle cells; activation of these channels leads to vasodilation. *AVPR1A*, which was associated with BP in AFR only, is a receptor for the vasoconstrictor vasopressin; murine knock-out models are hypotensive with impaired baroreceptor reflexes.[60]

Discussion

This is a large-scale multi-ancestry study to systematically use GxE interactions for identifying novel trait loci and for evaluating the role of GxE interactions in cardiovascular traits. In Stage 1, we performed a genome-wide analysis of gene-smoking interactions in 129,913 individuals across four ancestry groups using 1000 Genomes-imputed data, with follow up analysis in Stage 2 of a small set of promising variants in 480,178 additional individuals across five ancestry

groups. We identified 40 known BP loci at genome-wide significance level ($P < 5 \times 10^{-8}$) in Stage 1 as well as 15 novel loci that are genome-wide significant in Stage 1 and replicated in Stage 2 using Bonferroni correction. A combined meta-analysis of Stages 1 and 2 results yielded 16 additional known BP loci and 66 additional genome-wide significant loci ($P < 5 \times 10^{-8}$); 13, 35, and 18 loci were identified in European, African and trans-ancestry, respectively. These 66 additional loci were validated with low false discovery rate (FDR *q* value < 0.1) (e.g., see Nelson et al.[61]).

Identification of novel loci in this GxE analysis demonstrates the importance of incorporating environmental exposures in association discovery. Our newly identified loci including interactions with smoking collectively explained up to 1.7% additional variance in BP (beyond that explained by known BP variants) in several European cohorts. Furthermore, it may be particularly striking that our analyses also identified *VAMP2*, a component of the renin-angiotensin-aldosterone system (RAAS), as a likely mediator of hypertension. *VAMP2* modulates cAMP-stimulated renin release by renal juxtaglomerular cells[62] but has not been previously identified, even though other components of RAAS including *NPPA*, *NPPB*, and *SLC17A1*-4 have been found in previous GWAS and, indeed, among the 56 known BP loci identified in our study[4; 63-65]

Several of our newly identified BP loci show evidence for shared pathophysiology with cardiometabolic traits. This is encouraging as hypertension is a frequent comorbidity of a variety of cardiometabolic traits, including dyslipidemia, type 2 diabetes, obesity and other disorders of substrate metabolism and storage. *XKR6-MIR598* and *MFHAS1* have been associated with serum triglyceride levels.[66] *LRP6*[67; 68] and *PPP1R3B*[69] have been associated with serum low-density lipoprotein levels and the metabolic syndrome. *MSRA*[70] and *SERTAD2*[71] (associated in AFR) have been associated with obesity-related traits and adipocyte function, and *PPP1R3B* has been associated with steatohepatitis.[72] We also identified the well-known obesity/diabetes locus *FTO*[73; 74] as a newly identified BP locus (**Figure 2**). In addition to a recent discovery of the effect of an *FTO* variant on *IRX3* and *IRX5*,[75] variants in intron 1 of *FTO* have been identified that regulate the expression of nearby *RPGRIP1L*,[74] shown to modulate leptin receptor trafficking and signaling in the hypothalamus.[76] Variants in and near *XKR6-MIR598*, *MFHAS1*, *MSRA*, and

FTO have been associated with obesity- and diabetes-related traits in other studies (**Tables S23-S24**). Among other variants in genes related to cardiometabolic traits, *VAMP2* plays a role in the trafficking of the GLUT4 glucose receptor to the adipocyte plasma membrane.[77] Finally, we identified a SNP (in AFR) in *FABP3*, a gene known to regulate mitochondrial β-oxidation.[78] Studies have shown that serum *FABP3* transcript and protein levels are elevated in animal models and humans with hypertension compared with normotensive controls.[79; 80] Consistent with a recent paper,[6] our findings provide additional BP variants overlapping with metabolic trait loci.

Some of the newly identified BP loci have been previously reported as suggestive (but not genome-wide significant) for smoking and other addiction traits. Among our newly identified loci, *FTO*, *DPYSL2-ADRA1A*, *AJAP1*, and *SERINC2* have shown suggestive evidence of association with smoking-related traits,[81; 82] illicit drug use,[83] and alcohol consumption and dependence.[84; 85] In addition, dopaminergic signaling has been implicated in addictive behaviors.[86] Moreover, located in an intron of *TNFSF12* (tumor necrosis factor superfamily member), our newly identified variant rs9899183 has many compelling regulatory features supporting its candidacy (**Table S12**); it resides in a region characterized by promoter histone marks in 23 tissues, in enhancer histone marks in 7 tissues, and by DNAse marks in 12 tissues. This variant is also identified as an eQTL for genes *TNFSF12*, *CHRNB1* and *SAT2*; *CHRNB1* (1 nicotinic acetylcholine receptor subunit) may also contribute to nicotine dependence.[87]

BP regulation critically involves both central and peripheral regulation via neuroendocrine and hormonal regulation in a complex integrated system that includes the brain, kidneys, adrenal glands, and vasculature. In addition to validating loci known for their involvement in the RAAS system, natriuretic peptide signaling, solute channels, and adrenergic and cholinergic receptor signaling (among others), we identified variants in or near new biological candidates for BP regulation. For example, several of our newly identified loci identified genes that have been previously implicated in monogenic causes of ciliopathy (nephronophthisis-related ciliopathy), a cause of end-stage renal disease in children and young adults.[88; 89] This condition is a genetically heterogeneous autosomal recessive disease, and heterozygote siblings and other adults with incompletely penetrant versions of this disease may have variable degrees of hypertension, renal

insufficiency, obesity, and diabetes.[90] Newly identified loci also include genes involved in dopaminergic signaling which may act both centrally and in the kidney to modulate BP regulation. Still other newly identified loci reside in or near genes involved in telomere maintenance.

Of the 81 newly identified loci, 10 show genome-wide significant interactions although none were replicated in Stage 2. Nine were identified with current smoking status. The ever smoking status is more heterogeneous since the effect of (former) smoking on BP decays over time from cessation.[91] It is therefore not surprising that the analyses with the more homogeneous current smoking (CurSmk) status yielded larger (and more robust) effects on BP than did analyses using ever smoker (EverSmk) status. Although the joint 2 df test succeeded in identifying 71 of the 81 newly identified loci, the precise role of interaction is unclear. It is sobering to note that, although gene-smoking interactions may have helped identify a reasonably large number of the newly identified loci, the sample size we used here for genome-wide analysis in Stage 1 appears inadequate for identifying a large number of interaction effects (should they exist) through the 1 DF interaction test alone. This may be because, if the pathobiology of BP involves large numbers of interactions, the majority of the interaction effects are likely (relatively) small enough whose identification requires the 2 DF joint test and/or require much larger sample sizes for identifying them through the 1 DF interaction test. Moreover, smoking is only one of many lifestyle attributes that may have interaction effects on BP.[12] It is possible that some interactions we report here are driven by other lifestyle factors that may be correlated with smoking. A follow up study (such as Young et al[92] and Tyrrell et al[93]) that jointly examines multiple lifestyle factors can shed light on further understanding of interaction effects on BP.

Several large consortia-based BP GWAS papers have been published in recent years, dramatically increasing the number of BP loci. We treated 158 as known BP loci, which included the 71 loci that were reported by three recent papers.[5-7] Of the 56 known BP loci we identified, 8 overlap with these newly identified 71 loci. Hoffmann et al[94] reported 75 novel loci (and 241 additional loci not validated) based on > 300,000 individuals. The use of repeated measurements, beside the large sample size, appears to be responsible for the large number of novel loci discovered. Their study demonstrates the power of large sample sizes and repeated

measurements. Warren et al[95] reported 107 validated loci. As shown in **Table S28** in detail, nine of our newly identified loci include variants reported by these two papers.[94; 95] Based on African ancestry, Liang et al reported 3 validated BP loci,[96] one of which overlaps with our newly identified loci.

Thirty-five loci were identified in African ancestry meta-analyses. As previous discoveries of BP loci were mostly in European ancestry, some using very large sample sizes, it may be harder to detect newly identified signals in European ancestry in our study. There are also more opportunities to identify lower frequency variants in African ancestry meta-analysis because there are more of these variants in this genetically more diverse population. However, because of the highly limited sample sizes available for African ancestry in Stage 2, genome-wide significant loci in Stage 1 African ancestry could not be formally replicated in Stage 2. Nevertheless, there is evidence supporting the validity of many of the African-specific newly identified loci: African-specific QQ plots were very similar with and without the known BP loci (**Figure S10; Figure S12**). Genomic control values are all close to 1, and the top signals are away from the expected null line in the QQ plots, suggesting that these may be real associations. Forest plots at the African-specific loci (**Figure S13**) were not heterogeneous across cohorts. For most loci, there exists at least one non-African ancestry showing effects in the same direction as those in African ancestry. They may also relate at least in part to unique smoking behaviors or BP regulation or both in African ancestry. However, these African-specific loci require further validation.

There are several limitations in this large-scale multi-ancestry genome-wide investigation incorporating gene-smoking interactions. First, main effect only analysis without regard to smoking was not performed, and this limits our ability to resolve if any of our loci newly identified through the 2DF joint test could be found without smoking or gene-smoking interaction in the model. Second, although the strategy of clumping with a stringent LD threshold ($r^2 > 0.1$) in addition to large physical distance threshold (± 1 Mb) is reasonable for inferring independent loci, conditional analysis of summary statistics from interaction analysis (similar to GCTA) would be more rigorous; however, such methods do not exist currently. Third, the relatively smaller stage 2 sample sizes available in African and Hispanic ancestries limit our

ability to formally replicate the loci that were newly identified in Stage 1 in those ancestries (including the 10 interactions). Fourth, power for novel discovery using interactions may be limited even in this reasonably large sample size. Fifth, if there is a G-C correlation, a potential confounding of GxE with interaction between covariate and smoking exposure (CxE) may exist, which can inflate Type I error of the GxE interaction test;[97; 98] using a stratified model may help overcome such confounding. Sixth, our use of the fixed effect meta-analysis for trans-ancestry analysis may have limited the power in the presence of heterogeneous effects across ancestries; however, specialized trans-ancestry methods for GxE interactions do not exist. Seventh, subjects were grouped into each ancestry based on self-reported information instead of genetically computed ancestry. Finally, the use of multiple hypothesis tests, multiple phenotypes and exposures, and multiple ancestries may contribute to inflation at some level. Striking a balance between false positives and false negatives, especially in the context of interactions, remains a challenge.

In summary, our study identified a total of 137 genome-wide significant loci; 56 known loci, 15 new loci identified in Stage 1 and formally replicated in Stage 2, and 66 additional BP loci identified through the combined analysis of Stages 1 and 2 and validated through low FDR. Our ability to identify this many loci is likely due to four factors: focus on gene-smoking interactions, consideration of multiple ancestries, the large aggregate sample sizes available, and the densely-imputed data using the recent 1000 Genomes Project reference panel in Stage 1 analysis. The 10 newly identified loci with significant interactions showed larger effects on BP in smokers. Thirty-five (35) loci were identified only in African ancestry, highlighting the importance of pursuing genetic studies in diverse populations. In addition to evidence for shared pathophysiology with cardiometabolic traits, smoking and other addiction traits, our results provide compelling evidence for biological candidates for BP regulation such as modulators of vascular structure and function, ciliopathies, telomere maintenance, and central dopaminergic signaling. Our findings demonstrate how the interplay between genes and environment can help identify novel loci, open up new avenues for investigation about BP homeostasis, and highlight the promise of gene-lifestyle interactions for more in-depth genetic and environmental dissection of BP and other complex traits.

Supplemental Data

Supplemental Data include 5 Supplemental Notes, 17 figures, and 28 tables, all of which can be found in online.

URLs

DEPICT, https://data.broadinstitute.org/mpg/depict/; dbSNP, ncbi.nlm.nih.gov/snp/; GeneGo, http://lsresearch.thomsonreuters.com/; GSEA, http://software.broadinstitute.org/gsea/msigdb/annotate.jsp; GTEx, gtextportal.org; Literature Lab, http://www.acumenta.com/acumenta/overview/index.php; LocusZoom, http://locuszoom.sph.umich.edu/; HaploReg, http://www.broadinstitute.org/mammals/haploreg/haploreg_v3.php; National Human Genome Research Institute (NHGRI) GWAS catalog, http://www.genome.gov/gwastudies/; NCBI Entrez gene, ncbi.nlm.nih.gov/gene/; METAL, http://genome.sph.umich.edu/wiki/METAL_Documentation; RegulomeDB, http://regulomedb.org/; Roadmap Epigenomics, http://www.roadmapepigenomics.org/.

Data Availability

All summary results ("data") will be used for pleiotropy and pathway analyses as part of the N.I.H. grant that supports our GxE investigations. All data results on which this manuscript is based will be made available on dbGaP (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000930.v5.p1) within 6 months after completing the pleiotropy and pathway analyses.

Acknowledgments

The various Gene-Lifestyle Interaction projects, including this one, are largely supported by a grant from the U.S. National Heart, Lung, and Blood Institute (NHLBI), the National Institutes of Health, R01HL118305. A Career Development Award (K25HL121091), also from the

NHLBI, enabled Dr. Sung to lead this project. Full set of study-specific funding sources and acknowledgments appear in the **Supplemental Notes**.

References

1. Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J., Dehghan, A., Glazer, N.L., Morrison, A.C., Johnson, A.D., Aspelund, T., et al. (2009). Genome-wide association study of blood pressure and hypertension. Nature genetics 41, 677-687.
2. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature 461, 747-753.
3. Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M.D., Bochud, M., Coin, L., Najjar, S.S., Zhao, J.H., Heath, S.C., Eyheramendy, S., et al. (2009). Genome-wide association study identifies eight loci associated with blood pressure. Nature genetics 41, 666-676.
4. International Consortium for Blood Pressure Genome-Wide Association, S., Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C., et al. (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 478, 103-109.
5. Ehret, G.B., Ferreira, T., Chasman, D.I., Jackson, A.U., Schmidt, E.M., Johnson, T., Thorleifsson, G., Luan, J., Donnelly, L.A., Kanoni, S., et al. (2016). The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. Nat Genet 48, 1171-1184.
6. Liu, C., Kraja, A.T., Smith, J.A., Brody, J.A., Franceschini, N., Bis, J.C., Rice, K., Morrison, A.C., Lu, Y., Weiss, S., et al. (2016). Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. Nature genetics 48, 1162-1170.
7. Surendran, P., Drenos, F., Young, R., Warren, H., Cook, J.P., Manning, A.K., Grarup, N., Sim, X., Barnes, D.R., Witkowska, K., et al. (2016). Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. Nature genetics 48, 1151-1161.
8. Franceschini, N., Fox, E., Zhang, Z., Edwards, T.L., Nalls, M.A., Sung, Y.J., Tayo, B.O., Sun, Y.V., Gottesman, O., Adeyemo, A., et al. (2013). Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. American journal of human genetics 93, 545-554.
9. Zhu, X., Feng, T., Tayo, B.O., Liang, J., Young, J.H., Franceschini, N., Smith, J.A., Yanek, L.R., Sun, Y.V., Edwards, T.L., et al. (2015). Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. American journal of human genetics 96, 21-36.
10. Manning, A.K., Hivert, M.F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.T., Bielak, L.F., Prokopenko, I., et al. (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nature genetics 44, 659-669.
11. Hunter, D.J. (2005). Gene-environment interactions in human diseases. Nature reviews Genetics 6, 287-298.

12. Go, A.S., Mozaffarian, D., Roger, V.L., Benjamin, E.J., Berry, J.D., Blaha, M.J., Dai, S., Ford, E.S., Fox, C.S., Franco, S., et al. (2013). Heart Disease and Stroke Statistics--2014 Update: A Report From the American Heart Association. Circulation.
13. Mann, S.J., James, G.D., Wang, R.S., and Pickering, T.G. (1991). Elevation of ambulatory systolic blood pressure in hypertensive smokers. A case-control study. JAMA : the journal of the American Medical Association 265, 2226-2228.
14. Primatesta, P., Falaschetti, E., Gupta, S., Marmot, M.G., and Poulter, N.R. (2001). Association between smoking and blood pressure: evidence from the health survey for England. Hypertension 37, 187-193.
15. Green, M.S., Jucha, E., and Luz, Y. (1986). Blood pressure in smokers and nonsmokers: epidemiologic findings. Am Heart J 111, 932-940.
16. Rao, D.C., Sung, Y.J., Winkler, T.W., Schwander, K., Borecki, I., Cupples, L.A., Gauderman, W.J., Rice, K., Munroe, P.B., and Psaty, B. (2017). A Multi-ancestry study of gene-lifestyle interactions for cardiovascular traits in 610,475 individuals from 124 cohorts: Design and rationale. Circulation Cardiovascular genetics 10, e001649.
17. Kirk, E.P. (2017). Genes, Environment, and the Heart: Putting the Pieces Together. Circulation Cardiovascular genetics 10.
18. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56-65.
19. Sung, Y.J., Winkler, T.W., Manning, A.K., Aschard, H., Gudnason, V., Harris, T.B., Smith, A.V., Boerwinkle, E., Brown, M.R., Morrison, A.C., et al. (2016). An Empirical Comparison of Joint and Stratified Frameworks for Studying G x E Interactions: Systolic Blood Pressure and Smoking in the CHARGE Gene-Lifestyle Interactions Working Group. Genetic epidemiology 40, 404-415.
20. Tchetgen Tchetgen, E.J., and Kraft, P. (2011). On the robustness of tests of genetic associations incorporating gene-environment interaction when the environmental exposure is misspecified. Epidemiology 22, 257-261.
21. Voorman, A., Lumley, T., McKnight, B., and Rice, K. (2011). Behavior of QQ-plots and genomic control in studies of gene-environment interaction. PloS one 6, e19416.
22. Zeileis, A. (2006). Object-oriented computation of sandwich estimators. J Stat Softw 16.
23. Aulchenko, Y.S., Struchalin, M.V., and van Duijn, C.M. (2010). ProbABEL package for genome-wide association analysis of imputed data. BMC bioinformatics 11, 134.
24. Winkler, T.W., Day, F.R., Croteau-Chonka, D.C., Wood, A.R., Locke, A.E., Magi, R., Ferreira, T., Fall, T., Graff, M., Justice, A.E., et al. (2014). Quality control and conduct of genome-wide association meta-analyses. Nature protocols 9, 1192-1212.
25. Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. Nature reviews Genetics 11, 499-511.
26. Manning, A.K., LaValley, M., Liu, C.T., Rice, K., An, P., Liu, Y., Miljkovic, I., Rasmussen-Torvik, L., Harris, T.B., Province, M.A., et al. (2011). Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP x environment regression coefficients. Genetic epidemiology 35, 11-18.
27. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190-2191.

28. Randall, J.C., Winkler, T.W., Kutalik, Z., Berndt, S.I., Jackson, A.U., Monda, K.L., Kilpelainen, T.O., Esko, T., Magi, R., Li, S., et al. (2013). Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. PLoS genetics 9, e1003500.
29. Aschard, H., Hancock, D.B., London, S.J., and Kraft, P. (2010). Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. Human heredity 70, 292-300.
30. Winkler, T.W., Kutalik, Z., Gorski, M., Lottaz, C., Kronenberg, F., and Heid, I.M. (2015). EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. Bioinformatics 31, 259-261.
31. Devlin, B., and Roeder, K. (1999). Genomic control for association studies. Biometrics 55, 997-1004.
32. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics 81, 559-575.
33. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met 57, 289-300.
34. Consortium, G.T. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648-660.
35. Joehanes, R., Zhang, X., Huan, T., Yao, C., Ying, S.X., Nguyen, Q.T., Demirkale, C.Y., Feolo, M.L., Sharopova, N.R., Sturcke, A., et al. (2017). Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. Genome biology 18, 16.
36. Febbo, P.G., Mulligan, M.G., Slonina, D.A., Stegmaier, K., Di Vizio, D., Martinez, P.R., Loda, M., and Taylor, S.C. (2007). Literature Lab: a method of automated literature interrogation to infer biology from microarray analysis. BMC genomics 8, 461.
37. Skol, A.D., Scott, L.J., Abecasis, G.R., and Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nature genetics 38, 209-213.
38. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 40, D930-934.
39. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome research 22, 1790-1797.
40. Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol 6, e1001025.
41. Garber, M., Guttman, M., Clamp, M., Zody, M.C., Friedman, N., and Xie, X. (2009). Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics 25, i54-62.
42. Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.B., Stephens, M., Gilad, Y., and Pritchard, J.K. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464, 768-772.

43. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. Nature communications 6, 5890.
44. Lewington, S., Clarke, R., Qizilbash, N., Peto, R., Collins, R., and Prospective Studies Collaboration. (2002). Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 360, 1903-1913.
45. Justice, A., Winkler, T., Feitosa, M., Graff, M., Fisher, V.A., ..., Loos, R., Kilpelainen, T.O., Liu, T., Borecki, I., et al. (In press). Genome-wide meta-analysis of 241,258 adults accounting for smoking behavior identified novel loci for obesity traits. Nature communications.
46. Husson, H., Moreno, S., Smith, L.A., Smith, M.M., Russo, R.J., Pitstick, R., Sergeev, M., Ledbetter, S.R., Bukanov, N.O., Lane, M., et al. (2016). Reduction of ciliary length through pharmacologic or genetic inhibition of CDK5 attenuates polycystic kidney disease in a model of nephronophthisis. Hum Mol Genet.
47. Marples, D., Schroer, T.A., Ahrens, N., Taylor, A., Knepper, M.A., and Nielsen, S. (1998). Dynein and dynactin colocalize with AQP2 water channels in intracellular vesicles from kidney collecting duct. Am J Physiol 274, F384-394.
48. Butt, H.Z., Atturu, G., London, N.J., Sayers, R.D., and Bown, M.J. (2010). Telomere length dynamics in vascular disease: a review. Eur J Vasc Endovasc Surg 40, 17-26.
49. Guo, H.L., Zhang, C., Liu, Q., Li, Q., Lian, G., Wu, D., Li, X., Zhang, W., Shen, Y., Ye, Z., et al. (2012). The Axin/TNKS complex interacts with KIF3A and is required for insulin-stimulated GLUT4 translocation. Cell Res 22, 1246-1257.
50. Zee, R.Y., Ridker, P.M., and Chasman, D.I. (2011). Genetic variants in eleven telomere-associated genes and the risk of incident cardio/cerebrovascular disease: The Women's Genome Health Study. Clin Chim Acta 412, 199-202.
51. Hemerich, D., van der Laan, S.W., Tragante, V., den Ruijter, H.M., de Borst, G.J., Pasterkamp, G., de Bakker, P.I., and Asselbergs, F.W. (2015). Impact of carotid atherosclerosis loci on cardiovascular events. Atherosclerosis 243, 466-468.
52. Bis, J.C., Kavousi, M., Franceschini, N., Isaacs, A., Abecasis, G.R., Schminke, U., Post, W.S., Smith, A.V., Cupples, L.A., Markus, H.S., et al. (2011). Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. Nature genetics 43, 940-947.
53. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466, 707-713.
54. Imbs, J.L., Schmidt, M., and Schwartz, J. (1975). Effect of dopamine on renin secretion in the anesthetized dog. Eur J Pharmacol 33, 151-157.
55. Boone, M., and Deen, P.M. (2008). Physiology and pathophysiology of the vasopressin-regulated renal water reabsorption. Pflugers Arch 456, 1005-1024.
56. Li, X.X., Bek, M., Asico, L.D., Yang, Z., Grandy, D.K., Goldstein, D.S., Rubinstein, M., Eisner, G.M., and Jose, P.A. (2001). Adrenergic and endothelin B receptor-dependent hypertension in dopamine receptor type-2 knockout mice. Hypertension 38, 303-308.

57. Sakata, M., Sei, H., Toida, K., Fujihara, H., Urushihara, R., and Morita, Y. (2002). Mesolimbic dopaminergic system is involved in diurnal blood pressure regulation. Brain Res 928, 194-201.
58. Gertow, K., Sennblad, B., Strawbridge, R.J., Ohrvik, J., Zabaneh, D., Shah, S., Veglia, F., Fava, C., Kavousi, M., McLachlan, S., et al. (2012). Identification of the BCAR1-CFDP1-TMEM170A locus as a determinant of carotid intima-media thickness and coronary artery disease risk. Circ Cardiovasc Genet 5, 656-665.
59. Boardman-Pretty, F., Smith, A.J., Cooper, J., Palmen, J., Folkersen, L., Hamsten, A., Catapano, A.L., Melander, O., Price, J.F., Kumari, M., et al. (2015). Functional Analysis of a Carotid Intima-Media Thickness Locus Implicates BCAR1 and Suggests a Causal Variant. Circ Cardiovasc Genet 8, 696-706.
60. Koshimizu, T.A., Nasa, Y., Tanoue, A., Oikawa, R., Kawahara, Y., Kiyono, Y., Adachi, T., Tanaka, T., Kuwaki, T., Mori, T., et al. (2006). V1a vasopressin receptors maintain normal blood pressure by regulating circulating blood volume and baroreflex sensitivity. Proc Natl Acad Sci U S A 103, 7807-7812.
61. Nelson, C.P., Goel, A., Butterworth, A.S., Kanoni, S., Webb, T.R., Marouli, E., Zeng, L., Ntalla, I., Lai, F.Y., Hopewell, J.C., et al. (2017). Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nature genetics 49, 1385-1391.
62. Yu, K.Y., Wang, Y.P., Wang, L.H., Jian, Y., Zhao, X.D., Chen, J.W., Murao, K., Zhu, W., Dong, L., Wang, G.Q., et al. (2014). Mitochondrial KATP channel involvement in angiotensin II-induced autophagy in vascular smooth muscle cells. Basic Res Cardiol 109, 416.
63. Del Greco, M.F., Pattaro, C., Luchner, A., Pichler, I., Winkler, T., Hicks, A.A., Fuchsberger, C., Franke, A., Melville, S.A., Peters, A., et al. (2011). Genome-wide association analysis and fine mapping of NT-proBNP level provide novel insight into the role of the MTHFR-CLCN6-NPPA-NPPB gene cluster. Hum Mol Genet 20, 1660-1671.
64. Ganesh, S.K., Tragante, V., Guo, W., Guo, Y., Lanktree, M.B., Smith, E.N., Johnson, T., Castillo, B.A., Barnard, J., Baumert, J., et al. (2013). Loci influencing blood pressure identified using a cardiovascular gene-centric array. Human molecular genetics 22, 1663-1678.
65. Johnson, T., Gaunt, T.R., Newhouse, S.J., Padmanabhan, S., Tomaszewski, M., Kumari, M., Morris, R.W., Tzoulaki, I., O'Brien, E.T., Poulter, N.R., et al. (2011). Blood pressure loci identified with a gene-centric array. American journal of human genetics 89, 688-700.
66. Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., et al. (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. Nature genetics 41, 56-65.
67. Tomaszewski, M., Charchar, F.J., Barnes, T., Gawron-Kiszka, M., Sedkowska, A., Podolecka, E., Kowalczyk, J., Rathbone, W., Kalarus, Z., Grzeszczak, W., et al. (2009). A common variant in low-density lipoprotein receptor-related protein 6 gene (LRP6) is associated with LDL-cholesterol. Arterioscler Thromb Vasc Biol 29, 1316-1321.
68. Singh, R., Smith, E., Fathzadeh, M., Liu, W., Go, G.W., Subrahmanyan, L., Faramarzi, S., McKenna, W., and Mani, A. (2013). Rare nonconservative LRP6 mutations are associated with metabolic syndrome. Hum Mutat 34, 1221-1225.
69. Waterworth, D.M., Ricketts, S.L., Song, K., Chen, L., Zhao, J.H., Ripatti, S., Aulchenko, Y.S., Zhang, W., Yuan, X., Lim, N., et al. (2010). Genetic variants influencing

circulating lipid levels and risk of coronary artery disease. Arterioscler Thromb Vasc Biol 30, 2264-2276.
70. Reilly, D., Hao, K., Jensen, M.K., Girman, C.J., and Rimm, E.B. (2013). Use of systems biology approaches to analysis of genome-wide association studies of myocardial infarction and blood cholesterol in the nurses' health study and health professionals' follow-up study. PLoS One 8, e85369.
71. Liew, C.W., Boucher, J., Cheong, J.K., Vernochet, C., Koh, H.J., Mallol, C., Townsend, K., Langin, D., Kawamori, D., Hu, J., et al. (2013). Ablation of TRIP-Br2, a regulator of fat lipolysis, thermogenesis and oxidative metabolism, prevents diet-induced obesity and insulin resistance. Nat Med 19, 217-226.
72. Speliotes, E.K., Yerges-Armstrong, L.M., Wu, J., Hernaez, R., Kim, L.J., Palmer, C.D., Gudnason, V., Eiriksdottir, G., Garcia, M.E., Launer, L.J., et al. (2011). Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS Genet 7, e1001324.
73. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W., et al. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 316, 889-894.
74. Stratigopoulos, G., Martin Carli, J.F., O'Day, D.R., Wang, L., Leduc, C.A., Lanzano, P., Chung, W.K., Rosenbaum, M., Egli, D., Doherty, D.A., et al. (2014). Hypomorphism for RPGRIP1L, a ciliary gene vicinal to the FTO locus, causes increased adiposity in mice. Cell Metab 19, 767-779.
75. Claussnitzer, M., Dankel, S.N., Kim, K.H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puviindran, V., et al. (2015). FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. The New England journal of medicine 373, 895-907.
76. Stratigopoulos, G., LeDuc, C.A., Cremona, M.L., Chung, W.K., and Leibel, R.L. (2011). Cut-like homeobox 1 (CUX1) regulates expression of the fat mass and obesity-associated and retinitis pigmentosa GTPase regulator-interacting protein-1-like (RPGRIP1L) genes and coordinates leptin receptor signaling. J Biol Chem 286, 2155-2170.
77. Kawaguchi, T., Tamori, Y., Kanda, H., Yoshikawa, M., Tateya, S., Nishino, N., and Kasuga, M. (2010). The t-SNAREs syntaxin4 and SNAP23 but not v-SNARE VAMP2 are indispensable to tether GLUT4 vesicles at the plasma membrane in adipocyte. Biochem Biophys Res Commun 391, 1336-1341.
78. Pelsers, M.M., Hermens, W.T., and Glatz, J.F. (2005). Fatty acid-binding proteins as plasma markers of tissue injury. Clin Chim Acta 352, 15-35.
79. Stoynev, N., Dimova, I., Rukova, B., Hadjidekova, S., Nikolova, D., Toncheva, D., and Tankova, T. (2014). Gene expression in peripheral blood of patients with hypertension and patients with type 2 diabetes. J Cardiovasc Med (Hagerstown) 15, 702-709.
80. Setsuta, K., Seino, Y., and Mizuno, K. (2014). Heart-type fatty acid-binding protein is a novel prognostic marker in patients with essential hypertension. Int J Cardiol 176, 1323-1325.
81. Swan, G.E., Hops, H., Wilhelmsen, K.C., Lessov-Schlaggar, C.N., Cheng, L.S., Hudmon, K.S., Amos, C.I., Feiler, H.S., Ring, H.Z., Andrews, J.A., et al. (2006). A genome-wide screen for nicotine dependence susceptibility loci. Am J Med Genet B Neuropsychiatr Genet 141B, 354-360.

82. Bierut, L.J., Madden, P.A., Breslau, N., Johnson, E.O., Hatsukami, D., Pomerleau, O.F., Swan, G.E., Rutter, J., Bertelsen, S., Fox, L., et al. (2007). Novel genes identified in a high-density genome wide association study for nicotine dependence. Hum Mol Genet 16, 24-35.
83. Levran, O., Peles, E., Randesi, M., Correa da Rosa, J., Ott, J., Rotrosen, J., Adelson, M., and Kreek, M.J. (2015). Susceptibility loci for heroin and cocaine addiction in the serotonergic and adrenergic pathways in populations of different ancestry. Pharmacogenomics 16, 1329-1342.
84. Taylor, A., and Wang, K.S. (2014). Association between DPYSL2 gene polymorphisms and alcohol dependence in Caucasian samples. J Neural Transm (Vienna) 121, 105-111.
85. Edwards, A.C., Aliev, F., Bierut, L.J., Bucholz, K.K., Edenberg, H., Hesselbrock, V., Kramer, J., Kuperman, S., Nurnberger, J.I., Jr., Schuckit, M.A., et al. (2012). Genome-wide association study of comorbid depressive syndrome and alcohol dependence. Psychiatr Genet 22, 31-41.
86. Nutt, D.J., Lingford-Hughes, A., Erritzoe, D., and Stokes, P.R. (2015). The dopamine theory of addiction: 40 years of highs and lows. Nat Rev Neurosci 16, 305-312.
87. Saccone, N.L., Schwantes-An, T.H., Wang, J.C., Grucza, R.A., Breslau, N., Hatsukami, D., Johnson, E.O., Rice, J.P., Goate, A.M., and Bierut, L.J. (2010). Multiple cholinergic nicotinic receptor genes affect nicotine dependence risk in African and European Americans. Genes Brain Behav 9, 741-750.
88. Simms, R.J., Hynes, A.M., Eley, L., and Sayer, J.A. (2011). Nephronophthisis: a genetically diverse ciliopathy. Int J Nephrol 2011, 527137.
89. Schueler, M., Halbritter, J., Phelps, I.G., Braun, D.A., Otto, E.A., Porath, J.D., Gee, H.Y., Shendure, J., O'Roak, B.J., Lawson, J.A., et al. (2016). Large-scale targeted sequencing comparison highlights extreme genetic heterogeneity in nephronophthisis-related ciliopathies. J Med Genet 53, 208-214.
90. Croft, J.B., and Swift, M. (1990). Obesity, hypertension, and renal disease in relatives of Bardet-Biedl syndrome sibs. Am J Med Genet 36, 37-42.
91. Ambatipudi, S., Cuenin, C., Hernandez-Vargas, H., Ghantous, A., Le Calvez-Kelm, F., Kaaks, R., Barrdahl, M., Boeing, H., Aleksandrova, K., Trichopoulou, A., et al. (2016). Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics 8, 599-618.
92. Young, A.I., Wauthier, F., and Donnelly, P. (2016). Multiple novel gene-by-environment interactions modify the effect of FTO variants on body mass index. Nature communications 7, 12724.
93. Tyrrell, J., Wood, A.R., Ames, R.M., Yaghootkar, H., Beaumont, R.N., Jones, S.E., Tuke, M.A., Ruth, K.S., Freathy, R.M., Davey Smith, G., et al. (2017). Gene-obesogenic environment interactions in the UK Biobank study. International journal of epidemiology 46, 559-575.
94. Hoffmann, T.J., Ehret, G.B., Nandakumar, P., Ranatunga, D., Schaefer, C., Kwok, P.Y., Iribarren, C., Chakravarti, A., and Risch, N. (2016). Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. Nature genetics.
95. Warren, H. (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nature genetics.

96. Liang, J., Le, T.H., Edwards, D.R.V., Tayo, B.O., Gaulton, K.J., Smith, J.A., Lu, Y., Jensen, R.A., Chen, G., Yanek, L.R., et al. (2017). Single-trait and multi-trait genome-wide association analyses identify novel loci for blood pressure in African-ancestry populations. PLoS genetics 13, e1006728.
97. Keller, M.C. (2014). Gene x environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. Biological psychiatry 75, 18-24.
98. Shi, G., and Nehorai, A. (2017). Robustness of meta-analyses in finding gene x environment interactions. PloS one 12, e0171446.

Competing Financial Interests

The authors declare no competing financial interests except for the following. Bruce M. Psaty serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson; Oscar H. Franco Received grants from Metagenics (on women's health and epigenetics) and from Nestle (on child health); Laura J. Bierut is listed as an inventor on Issued U.S. Patent 8,080,371,“Markers for Addiction” covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction; Peter Sever has received research awards from Pfizer Inc; Consultant for Mundipharma Co. (Cambridge, UK); Patent holder with Biocompatibles UK Ltd. (Franham, Surrey, UK) (Title: Treatment of eye diseases using encapsulated cells encoding and secreting neuroprotective factor and / or anti-angiogenic factor; Patent number: 20120263794), and Patent application with University of Heidelberg (Heidelberg, Germany) (Title: Agents for use in the therapeutic or prophylactic treatment of myopia or hyperopia; Europäische Patentanmeldung 15 000 771.4); Paul W. Franks has been a paid consultant for Eli Lilly and Sanofi Aventis and has received research support from several pharmaceutical companies as part of a European Union Innovative Medicines Initiative (IMI) project; and Mark J Caulfield is Chief Scientist for Genomics England, a UK government company.

Figure legends

Figure 1. Study design and overall workflow. Stage 1 analysis identified 74 significant novel loci of which 15 were replicated in Stage 2. Replication in Stage 2 was hampered by limited sample sizes for African and Hispanic ancestries. Combined analysis leverages the full power of Stages 1 and 2, identifying 66 additional BP loci missed by the 2-step approach which were validated by FDR. Association analyses were performed for each of SBP and DBP, accounting for two smoking exposure variables, 'current smoking' status (CurSmk) and 'ever smoking' status (EverSmk). For each ancestry, cohort-specific results were combined to perform the 1 degree of freedom (DF) test of the interaction effect and the 2 DF joint test of genetic main and interaction effects.

Figure 2. Forest plots and LocusZoom plots for 3 newly identified loci (selected from **Table 2**)

A-B: Variant rs7823056 and 10 additional variants on chromosome 8 are an eQTL for *PRAG1*, which is expressed in multiple tissues including the cerebellum and thyroid.
C-D: Variant rs13271489 is a *cis*-eQTL for *MSRA* and predicted to modify enhancers in brain cells. *MSRA* has been shown to be associated with obesity-related traits and adipocyte function; it also promotes the survival and development of dopaminergic neurons .
E-F: Variant rs11642015 is intronic to the well-known obesity/diabetes locus *FTO*. In addition, *AKTIP* in this locus has role in telomere maintenance .

Figure 3. Scatter-plots of smoking-specific genetic effect sizes for BP traits at the 15 newly identified and 66 putative index variants listed in **Tables 2-5**. The red points show variants with 1 DF interaction $P < 5\times10^{-8}$ (1=rs12135881; 2=rs115234772; 3=rs62319742; 4=rs148387718; 5=rs74701635; 6=rs150155092; 7=rs148772934; 8=rs138973557; 9=rs115893283; 10=rs148753653). The blue points show variants with 1 DF interaction $P < 1\times10^{-5}$ (11=rs11809589; 12=rs10166552; 13=rs11931572; 14=rs9348895; 15=rs76726877; 16=rs61935525; 17=rs9965695; 18=rs10405764).

Tables

Table 1. Basic characteristics of cohorts in Stages 1 and 2 in each ancestry

	Current Smoker		Former Smoker		Never Smoker		% Male	% HT	% HT Meds	Age		SBP		DBP	
Stage 1	N	%	N	%	N	%				Mean	SD	Mean	SD	Mean	SD
EUR	14,607	18.1	28,409	35.3	37,535	46.6	32.6	38.2	25.4	54.63	8.0	129.31	19.2	77.29	11.2
AFR	5,545	21.5	7,185	27.8	13,121	50.8	26.5	55.9	39.5	54.49	9.1	136.39	22.8	81.75	12.8
ASN	2,465	18.3	1,677	12.5	9,296	69.2	51.2	46.9	27.0	55.42	9.7	137.29	21.5	79.41	11.1
HIS	1,068	12.1	2,160	24.5	5,577	63.3	24.9	43.5	13.3	55.50	11.0	130.50	22.0	76.95	11.8
Stage 1 Total	*23,685*	*18.4*	*39,431*	*30.7*	*65,529*	*50.9*	*32.8*	*43.1*	*27.7*	*54.74*	*8.6*	*131.69*	*20.4*	*78.42*	*11.6*
Stage 2	N	%	N	%	N	%	% Male	% HT	% HT Meds	Mean	SD	Mean	SD	Mean	SD
EUR	48,198	17.0	89,597	31.6	145,914	51.4	47.8	44.8	25.0	55.91	8.6	139.02	20.4	83.76	11.5
AFR	1,971	29.8	1,579	23.8	3,075	46.4	40.9	54.3	42.8	53.66	10.2	137.00	21.6	83.32	12.8
ASN	29,485	19.8	40,850	27.4	78,597	52.8	54.9	50.3	33.1	60.76	12.3	134.92	20.2	80.01	12.3
HIS	2,739	20.3	2,559	18.9	8,231	60.8	41.0	26.9	16.3	45.86	13.8	124.08	20.0	75.09	11.9
BRZ	998	22.6	514	11.6	2,902	65.8	48.0	15.5	6.3	27.78	3.2	119.91	16.0	74.68	11.5
Stage 2 Total	*83,391*	*18.2*	*135,099*	*29.6*	*238,719*	*52.2*	*49.7*	*45.9*	*27.4*	*56.84*	*9.9*	*137.12*	*20.3*	*82.26*	*11.8*
TOTAL	**107,076**	**18.3**	**174,530**	**29.8**	**304,248**	**51.9**	**46.1**	**45.3**	**27.4**	**56.40**	**9.6**	**135.96**	**20.3**	**81.44**	**11.7**

The cell entries for the covariates and BP traits correspond to sample-size weighted averages across all cohorts in each category.

Table 2. Newly identified loci that are significant in Stage 1 and formally replicated in Stage 2.

Locus[a]	Nearest Genes[b]	rsID	Chr:Pos[c]	EA	EAF	Ancestry & Trait	Stage	Genetic Main Effect Est[d]	Genetic Main Effect SE[d]	Interaction Effect Est[d]	Interaction Effect SE[d]	2 DF Joint P-value[e]
1	*MTHFR;**CLCN6**;NPPA*	rs202071545	1:11878161	d	0.945	ALL.SBP	1	1.24	0.28	-0.16	0.38	3.77×10^{-8}
							2	0.88	0.17	0.01	0.25	7.44×10^{-12}
							1+2	0.99	0.14	-0.04	0.20	$\mathbf{9.39\times10^{-20}}$
2	*CLCN6;NPPA;**NPPB***	rs3753581	1:11920189	a	0.327	ALL.SBP	1	-0.63	0.09	0.16	0.21	4.34×10^{-12}
							2	-0.43	0.05	0.00	0.11	5.52×10^{-23}
							1+2	-0.48	0.04	0.04	0.10	$\mathbf{1.31\times10^{-34}}$
3	*NPPA;NPPB*	rs72640287	1:11965792	t	0.039	EUR.SBP	1	-2.05	0.43	-0.04	0.59	1.59×10^{-10}
							2	-0.86	0.19	-0.33	0.28	8.19×10^{-13}
							1+2	-1.06	0.18	-0.31	0.25	$\mathbf{2.79\times10^{-21}}$
4	***WNT2B***	rs351364	1:113045061	a	0.297	ALL.SBP	1	-0.60	0.10	0.53	0.22	1.67×10^{-8}
							2	-0.42	0.05	0.14	0.11	5.38×10^{-19}
							1+2	-0.45	0.04	0.22	0.10	$\mathbf{1.20\times10^{-26}}$
5[f]	*CEP170;**SDCCAG8**;AKT3*	rs3897821	1:243420388	a	0.705	ALL.DBP	1	-0.35	0.06	0.20	0.13	2.49×10^{-9}
							2	-0.20	0.03	0.00	0.07	1.51×10^{-12}
							1+2	-0.23	0.03	0.05	0.06	$\mathbf{1.67\times10^{-20}}$
6[f]	***FER1L5***	rs7599598	2:97351840	a	0.564	EUR.DBP	1	-0.30	0.06	-0.15	0.14	5.93×10^{-8}
							2	-0.16	0.03	0.02	0.08	4.10×10^{-7}
							1+2	-0.19	0.03	-0.03	0.07	$\mathbf{4.25\times10^{-13}}$
7	***SLC4A7***	rs13063291	3:27446285	a	0.204	ALL.DBP	1	0.33	0.08	0.03	0.12	4.00×10^{-8}
							2	0.20	0.04	-0.14	0.06	3.75×10^{-6}
							1+2	0.23	0.04	-0.09	0.05	$\mathbf{1.67\times10^{-11}}$
8[f]	*PRAG1;MFHAS1*	rs7823056	8:8382705	a	0.397	EUR.SBP	1	-0.56	0.10	-0.02	0.22	1.54×10^{-8}
							2	-0.42	0.05	0.16	0.13	1.55×10^{-14}
							1+2	-0.45	0.05	0.10	0.11	$\mathbf{3.01\times10^{-22}}$
9[f]	*PPP1R3B;TNKS*	rs62493780	8:9151051	t	0.238	EUR.SBP	1	0.89	0.18	-0.19	0.25	3.47×10^{-8}

							2	0.46	0.09	-0.27	0.13	2.37×10^{-7}
							1+2	0.54	0.08	-0.24	0.12	**2.95×10^{-13}**
10[f]	***MIR124-1;****MSRA*	rs13271489	8:9803712	t	0.478	EUR.SBP	1	0.55	0.10	0.02	0.22	6.37×10^{-8}
							2	0.44	0.05	-0.13	0.14	9.35×10^{-16}
							1+2	0.46	0.05	-0.08	0.12	**4.56×10^{-23}**
11	*TNNI2;****LSP1****;TNNT3*	rs7483477	11:1920255	t	0.75	ALL.SBP	1	-0.65	0.11	0.17	0.27	2.25×10^{-8}
							2	-0.36	0.05	0.08	0.12	1.77×10^{-13}
							1+2	-0.40	0.04	0.09	0.11	**2.12×10^{-20}**
12	*POC1B;ATP2B1*	rs7313874	12:89965049	t	0.325	ALL.SBP	1	-0.64	0.11	0.01	0.15	1.85×10^{-14}
							2	-0.48	0.06	-0.23	0.09	1.07×10^{-39}
							1+2	-0.52	0.05	-0.17	0.08	**2.49×10^{-54}**
13	***ATP2B1***	rs111337717	12:90037506	t	0.943	ALL.SBP	1	1.27	0.33	0.60	0.46	9.23×10^{-11}
							2	1.09	0.15	-0.26	0.22	2.86×10^{-18}
							1+2	1.13	0.13	-0.07	0.20	**1.27×10^{-27}**
14	*PTPN11*	rs7974266	12:113007602	t	0.513	ALL.DBP	1	0.19	0.13	0.57	0.18	3.58×10^{-8}
							2	0.23	0.06	0.19	0.09	6.50×10^{-12}
							1+2	0.22	0.06	0.28	0.08	**5.91×10^{-19}**
15[f]	*AKTIP;RPGRIP1L;****FTO***	rs11642015	16:53802494	t	0.334	ALL.SBP	1	0.57	0.09	-0.19	0.21	2.78×10^{-9}
							2	0.29	0.05	0.08	0.11	6.74×10^{-13}
							1+2	0.35	0.04	0.03	0.10	**9.91×10^{-21}**

Each locus is genome-wide significant ($P < 5\times10^{-8}$) in Stage 1 and formally replicated in Stage 2 using Bonferroni-adjusted significance level ($P < 0.05/74$). Forest plots and LocusZoom plots are shown in Figures S3 and S4, respectively.

[a]Each locus was determined through LD-based clumping, using ± 1 Mb around index variants, followed by LD threshold of $r^2 > 0.1$; ancestry-specific LDs from 1000 Genomes Project were used when clumping within each ancestry and the entire cosmopolitan data were used for trans-ancestry clumping.

[b]Gene names were obtained using variant effect predictor (VEP) from Ensembl. Genes with intragenic index variants bolded.

[c]Positions are based on build 37.

[d]Effect is in mmHg unit.

[e]The most significant p-value (between 1 DF interaction test and 2 DF joint test) was set in bold.

[f]These loci indicate completely novel loci, at least 1Mb away from any of known BP loci.

BP: blood pressure. SBP: systolic BP. DBP: diastolic BP. EA: effect allele; EAF: effect allele frequency; 2 DF Joint P: P-value of the joint test with 2 degrees of freedom of genetic main and interaction effects. 1 DF Interaction P: P-value of the interaction test with 1 degree of freedom; EUR: European ancestry. ALL: trans-ancestry (i.e., combining all ancestry groups through meta-analysis).

Table 3. Additional significant loci from the combined analyses of Stages 1 and 2 in trans-ancestry

Locus[a]	Nearest Genes[b]	rsID	Chr:Pos[c]	EA	EAF	Effect[d]		P value[e]		Trait
						Genetic Main	Interaction	1 DF Interaction	2 DF Joint	
1	*NPPA;NPPB*	rs12741980	1:11939593	a	0.943	0.68	0.02	0.852	**3.04×10^{-14}**	SBP
2[f]	***RSRC1***	rs201851995	3:157837508	d	0.648	-0.6	0.38	0.0016	**4.65×10^{-12}**	SBP
3[f]	*INPP4B;GAB1*	rs78763922	4:144054552	d	0.303	0.34	0.05	0.5067	**4.03×10^{-13}**	SBP
4	***HOTTIP***	rs2023843	7:27243221	t	0.837	0.7	-0.2	0.1634	**3.69×10^{-41}**	SBP
5[f]	***MFHAS1;****ERI1;PPP1R3B*	rs201133964	8:8607849	d	0.174	-0.52	-0.16	0.4366	**1.24×10^{-9}**	SBP
6[f]	*PPP1R3B;TNKS*	rs35904419	8:9376810	d	0.816	-0.19	-0.15	0.1761	**1.34×10^{-8}**	DBP
7	***FAM167A-AS1****;FAM167A;BLK*	rs4841531	8:11293390	t	0.161	-0.31	0.03	0.7825	**1.32×10^{-8}**	SBP
8[f]	*EBF2;****LOC105379336;*** *PPP2R2A;DPYSL2;ADRA1A*	rs58429174	8:26011922	t	0.262	-0.12	-0.14	0.026	**2.60×10^{-9}**	DBP
9	*ADRB1*	rs180940	10:115722411	a	0.391	-0.19	0.06	0.1514	**5.00×10^{-12}**	DBP
10	*AP5B1;OVOL1*	rs201316070	11:65548558	d	0.061	-0.6	-0.23	0.462	**1.54×10^{-9}**	SBP
11[f]	*LRP6;GPR19;****APOLD1****;GPRC5A*	rs72656645	12:12881055	a	0.7	0.36	-0.13	0.064	**4.49×10^{-15}**	SBP
12	*SLCO1C1;SLCO1B3; SLCO1B7;SLCO1B1*	rs73073686	12:20354507	a	0.231	-0.24	-0.07	0.2553	**1.68×10^{-18}**	DBP
13	*ATP2B1*	rs10858948	12:90478651	a	0.578	-0.18	0	0.6992	**4.74×10^{-15}**	DBP
14	*MED13L*	rs11067762	12:116198214	a	0.176	-0.24	-0.05	0.1951	**5.30×10^{-18}**	DBP
15	*CYP1A1-2;ULK3;****SCAMP2;****MPI*	rs10628234	15:75211142	d	0.3	0.32	-0.22	0.0253	**1.57×10^{-24}**	DBP
16[f]	*LDHD;****CFDP1;****TMEM231; TERF2IP*	rs4888411	16:75443183	a	0.56	0.26	0.12	0.0467	**1.19×10^{-18}**	SBP
17[f]	*SLC2A4;KCTD11;****TNFSF12;*** *TNFSF13;ATP1B2*	rs9899183	17:7452977	t	0.742	-0.35	0.07	0.6683	**1.24×10^{-12}**	SBP
18[f]	***ACE***	rs4968782	17:61548476	a	0.616	-0.2	0.08	0.2179	**3.30×10^{-16}**	DBP

Each locus is genome-wide significant ($P < 5\times10^{-8}$) in the combined analyses of Stages 1 and 2 and had FDR *q* value < 0.1. Forest plots and LocusZoom plots are shown in Figures S3 and S4, respectively.

[a]Each locus was determined through LD-based clumping, using ± 1 Mb around index variants, followed by LD threshold of $r^2 > 0.1$; ancestry-specific LDs from 1000 Genomes Project were used when clumping within each ancestry and the entire cosmopolitan data were used for trans-ancestry clumping.

[b]Gene names were obtained using variant effect predictor (VEP) from Ensembl. Genes with intragenic index variants bolded.

[c]Positions are based on build 37.

[d]Effect is in mmHg unit.

[e]The most significant p-value (between 1 DF interaction test and 2 DF joint test) was set in bold.

[f]These loci indicate completely novel loci, at least 1Mb away from any of known BP loci.

BP: blood pressure. SBP: systolic BP. DBP: diastolic BP. EA: effect allele; EAF: effect allele frequency; 2 DF Joint P: P-value of the joint test with 2 degrees of freedom of genetic main and interaction effects. 1 DF Interaction P: P-value of the interaction test with 1 degree of freedom.

Table 4. Additional significant loci from the combined analyses of Stages 1 and 2 in European-ancestry

Locus[a]	Nearest Genes[b]	rsID	Chr:Pos[c]	EA	EAF	Effect[d]		P value[e]		Trait
						Genetic Main	Interaction	1 DF Interaction	2 DF Joint	
1	***MTHFR;****CLCN6*	rs6541006	1:11857526	a	0.071	-0.85	0	0.6454	**3.17×10^{-19}**	SBP
2[f]	*KCNG3;DYNC2LI1*	rs73923009	2:43141074	a	0.099	-0.36	0.07	0.6165	**1.21×10^{-14}**	DBP
3	*SLC17A1-4;HFE*	rs7753826	6:26042239	a	0.189	0.36	-0.05	0.4371	**1.72×10^{-25}**	DBP
4	*SLC44A4;****EHMT2****; STK19;CYP21A2;TNXB*	rs2243873	6:31863433	a	0.556	0.45	-0.19	0.0472	**3.33×10^{-14}**	SBP
5	*SLC44A4;EHMT2;* ***HLA-DQB2;****STK19;CYP21A2;TNXB*	rs2071550	6:32730940	a	0.307	0.29	-0.22	0.0003	**1.17×10^{-9}**	DBP
6[f]	*TNKS;MSRA*	rs4841235	8:9683358	a	0.426	0.37	-0.1	0.7078	**4.78×10^{-15}**	SBP
7	***SOX7****;PINX1*	rs6995692	8:10587008	c	0.563	-0.44	0.31	0.0102	**4.11×10^{-19}**	SBP
8[f]	***ADARB2***	rs150155092	10:1769881	d	0.013	4.76	-18.32	**7.43×10^{-9}**	1.94×10^{-8}	SBP
9	*KAT5;RNASEH2C*	rs72941051	11:65478893	t	0.074	-0.39	0.07	0.3701	**1.75×10^{-11}**	DBP
10[f]	***FAM19A2;****AVPR1A*	rs17713040	12:62467714	t	0.977	0.24	0.31	0.7633	**3.44×10^{-8}**	DBP
11	*FAM109A;****SH2B3;****ATXN2*	rs4375492	12:111835990	a	0.794	0.35	0.03	0.8187	**1.03×10^{-26}**	DBP
12	*MPI;COX5A;SCAMP5*	rs12050494	15:75260896	a	0.316	0.32	-0.06	0.525	**3.01×10^{-27}**	DBP
13[f]	***NAA38;****KCNAB3;VAMP2*	rs74439044	17:7781019	t	0.903	-0.36	-0.14	0.1507	**2.43×10^{-21}**	DBP

Each locus is genome-wide significant ($P < 5\times10^{-8}$) in the combined analyses of Stages 1 and 2 and had FDR *q* value < 0.1. Forest plots and LocusZoom plots are shown in Figures S3 and S4, respectively.

[a]Each locus was determined through LD-based clumping, using ± 1 Mb around index variants, followed by LD threshold of $r^2 > 0.1$; ancestry-specific LDs from 1000 Genomes Project were used when clumping within each ancestry and the entire cosmopolitan data were used for trans-ancestry clumping.

[b]Gene names were obtained using variant effect predictor (VEP) from Ensembl. Genes with intragenic index variants bolded.

[c]Positions are based on build 37.

[d]Effect is in mmHg unit.

[e]The most significant p-value (between 1 DF interaction test and 2 DF joint test) was set in bold.

[f]These loci indicate completely novel loci, at least 1Mb away from any of known BP loci.

BP: blood pressure. SBP: systolic BP. DBP: diastolic BP. EA: effect allele; EAF: effect allele frequency; 2 DF Joint P: P-value of the joint test with 2 degrees of freedom of genetic main and interaction effects. 1 DF Interaction P: P-value of the interaction test with 1 degree of freedom.

Table 5. Additional significant loci from the combined analyses of Stages 1 and 2 in African-ancestry

Locus[a]	Nearest Genes[b]	rsID	Chr:Pos[c]	EA	EAF	Effect[d]		P value[e]		Trait
						Genetic Main	Interaction	1 DF Interaction	2 DF Joint	
1[f]	***AJAP1***	rs12135881	1:4781922	c	0.988	-2.05	16.94	2.06×10^{-8}	**3.09×10^{-9}**	SBP
2[f]	*FABP3;SERINC2;TINAGL1*	rs11809589	1:31970118	a	0.012	-1.11	-18.04	1.54×10^{-7}	**7.71×10^{-10}**	SBP
3[f]	*LOC101928219*	rs182662555	1:96289336	t	0.988	6.15	-4.45	0.00201	**1.79×10^{-8}**	DBP
4[f]	*PXDN;**MYT1L***	rs75247762	2:1893133	t	0.014	-2.37	-12.93	1.45×10^{-5}	**1.17×10^{-9}**	SBP
5[f]	*ASB3;ERLEC1;GPR75*	rs115234772	2:53650295	a	0.987	-0.1	8.5	2.13×10^{-9}	**1.07×10^{-11}**	DBP
6[f]	*SERTAD2;SLC1A4*	rs145162854	2:65104447	a	0.015	-3.17	-2.61	0.171	**6.63×10^{-9}**	SBP
7[f]	***ACOXL***	rs116008367	2:111807546	c	0.014	-0.86	-5.35	5.00×10^{-5}	**3.09×10^{-8}**	DBP
8[f]	*KCNE4;SCG2*	rs10166552	2:224036537	t	0.016	-0.15	-10.83	4.28×10^{-6}	**1.52×10^{-9}**	SBP
9[f]	***TPRA1;**MCM2*	rs139963642	3:127314188	t	0.013	-6.35	1.23	0.6742	**1.55×10^{-8}**	DBP
10[f]	*PCDH7*	rs11931572	4:30086104	a	0.968	-0.45	3.28	2.71×10^{-6}	**2.91×10^{-8}**	DBP
11[f]	*SPRY1;**LINC01091***	rs62319742	4:124581262	a	0.014	1.98	-10.98	**3.43×10^{-8}**	4.09×10^{-8}	DBP
12[f]	*HSD17B4*	rs140543491	5:118923601	a	0.017	-3	-16.29	1.24×10^{-5}	**5.34×10^{-9}**	SBP
13[f]	*OFCC1*	rs148387718	6:9446000	t	0.014	0.59	-7.84	2.70×10^{-8}	**1.77×10^{-11}**	DBP
14[f]	*NEDD9;**LOC105374928***	rs9348895	6:11496048	a	0.586	0.11	1.21	6.15×10^{-6}	**1.71×10^{-8}**	DBP
15[f]	*MYO6;**IMPG1***	rs58806982	6:76688806	t	0.01	-11.24	14.92	1.47×10^{-5}	**4.57×10^{-8}**	SBP
16[f]	***TARID;**SLC2A12*	rs76987554	6:134080855	t	0.062	-1.57	0	0.6676	**1.63×10^{-8}**	SBP
17[f]	***ARID1B***	rs112140754	6:157245233	t	0.988	0.97	7.6	0.00104	**2.44×10^{-8}**	DBP
18[f]	***BZW2***	rs116196735	7:16710605	a	0.018	-2.88	-13.75	0.00037	**6.98×10^{-10}**	SBP
19[f]	*MED30;EXT1*	rs74701635	8:118758316	t	0.016	3.79	-19.2	2.38×10^{-9}	**2.13×10^{-9}**	SBP
20[f]	***ADAMTSL1;**MIR3152*	rs146250839	9:18189778	a	0.976	0.35	2.79	0.00029	**4.36×10^{-8}**	DBP
21[f]	*SPIN1;S1PR3;SHC3;CKS2*	rs192642798	9:91503987	a	0.012	-8.38	3.95	0.346	**4.23×10^{-9}**	SBP
22[f]	*FZD8*	rs76726877	10:36313497	t	0.015	-1.55	-9.14	4.17×10^{-6}	**4.47×10^{-10}**	DBP
23[f]	*SFRP5;**CRTAC1***	rs11599481	10:99640463	t	0.058	-0.9	-3.33	1.38×10^{-5}	**4.55×10^{-11}**	SBP
24[f]	*TSPAN18;PRDM11;SYT13*	rs148772934	11:45005681	t	0.986	-0.57	11.66	1.00×10^{-8}	**1.20×10^{-9}**	DBP
25	*SLC15A3;CD6;* ***LOC105369325****; CD5*	rs11601370	11:60834043	t	0.976	1.34	6.63	0.00867	**3.01×10^{-9}**	SBP
26[f]	*LOC101928944*	rs74601585	11:80140007	t	0.017	-3.93	-2.58	0.2715	**8.06×10^{-9}**	SBP

27[f]	*LOC105369408*	rs78103586	11:133893928	a	0.029	-1.65	-5.27	0.00163	**2.26×10^{-9}**	DBP
28[f]	*PLEKHG7;EEA1;**LOC643339***	rs61935525	12:93645481	c	0.985	1.26	10.15	3.28×10^{-7}	**3.28×10^{-11}**	DBP
29[f]	*DICER1;CLMN*	rs187852559	14:95794914	a	0.013	-1.67	-4.93	0.0246	**8.74×10^{-10}**	DBP
30[f]	*SETD3;CCNK*	rs1257310	14:99810427	a	0.784	1.03	0.98	0.1335	**1.67×10^{-8}**	SBP
31	*GPR139;GP2;UMOD;PDILT*	rs148753653	16:20230175	a	0.981	5.25	-9.4	**1.89×10^{-8}**	6.30×10^{-8}	DBP
32[f]	***LOC339166;****WSCD1*	rs138973557	17:5699720	t	0.903	0.36	2.09	2.12×10^{-8}	**1.81×10^{-10}**	DBP
33[f]	*DYM;LIPG;ACAA2;MYO5B*	rs9965695	18:47261614	t	0.982	0.29	13.32	8.36×10^{-6}	**1.63×10^{-8}**	SBP
34[f]	***ZNF98***	rs10405764	19:22598479	t	0.017	0.91	-19.1	2.13×10^{-7}	**4.30×10^{-8}**	SBP
35[f]	*BMP7*	rs115893283	20:55404165	t	0.042	0.9	-9.05	2.53×10^{-8}	**4.24×10^{-12}**	SBP

Each locus is genome-wide significant ($P < 5\times10^{-8}$) in the combined analyses of Stages 1 and 2 and had FDR *q* value < 0.1. Forest plots and LocusZoom plots are shown in Figures S3 and S4, respectively.

[a]Each locus was determined through LD-based clumping, using ± 1 Mb around index variants, followed by LD threshold of $r^2 > 0.1$; ancestry-specific LDs from 1000 Genomes Project were used when clumping within each ancestry and the entire cosmopolitan data were used for trans-ancestry clumping.

[b]Gene names were obtained using variant effect predictor (VEP) from Ensembl. Genes with intragenic index variants bolded.

[c]Positions are based on build 37.

[d]Effect is in mmHg unit.

[e]The most significant p-value (between 1 DF interaction test and 2 DF joint test) was set in bold.

[f]These loci indicate completely novel loci, at least 1Mb away from any of known BP loci.

BP: blood pressure. SBP: systolic BP. DBP: diastolic BP. EA: effect allele; EAF: effect allele frequency; 2 DF Joint P: P-value of the joint test with 2 degrees of freedom of genetic main and interaction effects. 1 DF Interaction P: P-value of the interaction test with 1 degree of freedom.

List of all authors & affiliations

Yun J. Sung[1,219], Thomas W. Winkler[2,219], Lisa de las Fuentes[3,219], Amy R. Bentley[4,219], Michael R. Brown[5,219], Aldi T. Kraja[6,219], Karen Schwander[1,219], Ioanna Ntalla[7,219], Xiuqing Guo[8], Nora Franceschini[9], Lu Yingchang[10], Ching-Yu Cheng[11,12,13], Xueling Sim[14], Dina Vojinovic[15], Jonathan Marten[16], Solomon K. Musani[17], Changwei Li[18], Mary F. Feitosa[6], Tuomas O. Kilpeläinen[19,20], Melissa A. Richard[21], Raymond Noordam[22], Stella Aslibekyan[23], Hugues Aschard[24,25], Traci M. Bartz[26], Rajkumar Dorajoo[27], Yongmei Liu[28], Alisa K. Manning[29,30], Tuomo Rankinen[31], Albert Vernon Smith[32,33], Salman M. Tajuddin[34], Bamidele O. Tayo[35], Helen R. Warren[7,36], Wei Zhao[37], Yanhua Zhou[38], Nana Matoba[39], Tamar Sofer[40], Maris Alver[41], Marzyeh Amini[42], Mathilde Boissel[43], Jin Fang Chai[44], Xu Chen[45], Jasmin Divers[46], Ilaria Gandin[47], Chuan Gao[48], Franco Giulianini[49], Anuj Goel[50,51], Sarah E. Harris[52,53], Fernando Pires Hartwig[54], Andrea R. V. R. Horimoto[55], Fang-Chi Hsu[46], Anne U. Jackson[56], Mika Kähönen[57], [58], Anuradhani Kasturiratne[59], Brigitte Kühnel[60,61], Karin Leander[62], Wen-Jane Lee[63], Keng-Hung Lin[64], Jian'an Luan[65], Colin A. McKenzie[66], He Meian[67], Christopher P. Nelson[68,69], Rainer Rauramaa[70], Nicole Schupf[71], Robert A. Scott[65], Wayne H.H. Sheu[72,73,74,75], Alena Stančáková[76], Fumihiko Takeuchi[77], Peter J. van der Most[42], Tibor V. Varga[78], Heming Wang[79], Yajuan Wang[79], Erin B. Ware[80,37], Stefan Weiss[81, 82], Wanqing Wen[83], Lisa R. Yanek[84], Weihua Zhang[85,86], Jing Hua Zhao[65], Saima Afaq[85], Tamuno Alfred[10], Najaf Amin[15], Dan Arking[87], Tin Aung[11,12,13], R Graham Barr[88], Lawrence F. Bielak[37], Eric Boerwinkle[89, 90], Erwin P. Bottinger[10], Peter S. Braund[68, 69], Jennifer A. Brody[91], Ulrich Broeckel[92], Claudia P. Cabrera[7, 36], Brian Cade[93], Yu Caizheng[67], Archie Campbell[94], Mickaël Canouil[43], Aravinda Chakravarti[95], The CHARGE Neurology Working Group, Sabanayagam Charumathi[11,12], Ganesh Chauhan[96], Kaare Christensen[97], The COGENT-Kidney Consortium, Francis S. Collins[98], John M. Connell[99], Renée de Mutsert[100], H. Janaka de Silva[101], Stephanie Debette[102,103,104], Marcus Dörr[105,82], Qing Duan[106], Charles B. Eaton[107], Georg Ehret[95,108], Evangelos Evangelou[85,109], Jessica D. Faul[110], Virginia A. Fisher[38], Nita G. Forouhi[65], Oscar H. Franco[15], Yechiel Friedlander[111], He Gao[85,112], The GIANT Consortium, Bruna Gigante[62], Misa Graff[9], C. Charles Gu[1], Dongfeng Gu[113], Preeti Gupta[11], Saskia P. Hagenaars[52,114], Tamara B. Harris[115], Jiang He[116,117], Sami Heikkinen[118], Chew-Kiat Heng[119,120], Makoto Hirata[121], Albert Hofman[15], Barbara V. Howard[122, 123], Steven Hunt[124, 125], Marguerite R. Irvin[23], Yucheng Jia[8], Roby Joehanes[126, 127], Anne E. Justice[9], Tomohiro Katsuya[128, 129], Joel Kaufman[130], Nicola D. Kerrison[65], Chiea Chuen Khor[27, 131], Woon-Puay Koh[44, 132], Heikki A. Koistinen[133, 134], Pirjo Komulainen[70], Charles Kooperberg[135], Jose E. Krieger[55], Michiaki Kubo[136], Johanna Kuusisto[137], Carl D. Langefeld[46], Claudia Langenberg[65], Lenore J. Launer[115], Benjamin C. Lehne[85], Cora E. Lewis[138], Yize Li[1], Lifelines cohort study[139], Sing Hui Lim[11], Shiow Lin[6], Ching-Ti Liu[38], Jianjun Liu[14, 27], Jingmin Liu[140], Kiang Liu[141], Yeheng Liu[8], Marie Loh[85, 142], Kurt K. Lohman[28], Jirong Long[83], Tin Louie[40], Reedik Mägi[41], Anubha Mahajan[51], Cocca Massimiliano[47], Thomas Meitinger[143, 144], Andres Metspalu[41], Lili Milani[41], Yukihide Momozawa[145], Andrew P. Morris[51, 146], Thomas H. Mosley, Jr.[147], Peter Munson[148], Alison D. Murray[149], Mike A. Nalls[150, 151], Ubaydah Nasri[8], Jill M. Norris[152], Kari North[153], Adesola Ogunniyi[154], Sandosh Padmanabhan[155], Walter R. Palmas[156], Nicholette D. Palmer[157], James S. Pankow[158], Nancy L. Pedersen[45], Annette Peters[61, 159], Patricia A. Peyser[37], Ozren Polasek[160], Olli T. Raitakari[161, 162], Frida Renström[78, 163], Treva K. Rice[1], Paul M. Ridker[49], Antonietta Robino[164], Jennifer G. Robinson[165], Lynda M. Rose[49], Igor Rudan[166], Babatunde L. Salako[154], Kevin Sandow[8], Carsten O. Schmidt[82, 167], Pamela J. Schreiner[168], William R. Scott[85, 169], Sudha Seshadri[127, 170], Peter Sever[171], Colleen M. Sitlani[91], Jennifer A.

Smith[37], Harold Snieder[42], John M. Starr[52, 172], Konstantin Strauch[173, 174], Hua Tang[175], Kent D. Taylor[8], Yik Ying Teo[14,176,177,27,178], Yih Chung Tham[11], André G Uitterlinden[179], Melanie Waldenberger[60, 61], Lihua Wang[6], Ya X. Wang[180], Wen Bin Wei[181], Christine Williams[6], Gregory Wilson[182], Mary K. Wojczynski[6], Jie Yao[8], Jian-Min Yuan[183, 184], Alan B. Zonderman[185], Diane M. Becker[84], Michael Boehnke[56], Donald W. Bowden[157], John C. Chambers[85, 86], Yii-Der Ida Chen[8], Ulf de Faire[62], Ian J. Deary[52, 114], Tõnu Esko[41, 186], Martin Farrall[50, 51], Terrence Forrester[66], Paul W. Franks[78, 187, 188], Barry I. Freedman[189], Philippe Froguel[43, 190], Paolo Gasparini[47, 191], Christian Gieger[60, 192], Bernardo Lessa Horta[54], Yi-Jen Hung[193], Jost B. Jonas[181, 194], Norihiro Kato[77], Jaspal S. Kooner[86, 169], Markku Laakso[137], Terho Lehtimäki[195, 196], Kae-Woei Liang[73, 197, 198], Patrik KE Magnusson[45], Anne B. Newman[183], Albertine J. Oldehinkel[199], Alexandre C. Pereira[55, 200], Susan Redline[93], Rainer Rettig[82, 201], Nilesh J. Samani[68, 69], James Scott[169], Xiao-Ou Shu[83], Pim Van der Harst[202], Lynne E. Wagenknecht[203], Nicholas J. Wareham[65], Hugh Watkins[50, 51], David R. Weir[110], Ananda R. Wickremasinghe[59], Tangchun Wu[67], Wei Zheng[83], Yoichiro Kamatani[39], Cathy C. Laurie[40], Claude Bouchard[31], Richard S. Cooper[35], Michele K. Evans[34], Vilmundur Gudnason[32, 33], Sharon L.R. Kardia[37], Stephen B. Kritchevsky[204], Daniel Levy[127, 205], Jeff R. O'Connell[206, 207], Bruce M. Psaty[208, 209], Rob M. van Dam[44, 210], Mario Sims[17], Donna K. Arnett[211], Dennis O. Mook-Kanamori[100, 212], Tanika N. Kelly[116], Ervin R. Fox[213], Caroline Hayward[16], Myriam Fornage[21], Charles N. Rotimi[4], Michael A. Province[6], Cornelia M. van Duijn[15], E. Shyong Tai[210,14,132], Tien Yin Wong[11,12,13], Ruth J.F. Loos[10,214], Alex P. Reiner[135], Jerome I. Rotter[215,8], Xiaofeng Zhu[79], Laura J. Bierut[216], W. James Gauderman[217], Mark J. Caulfield[7, 36,220], Paul Elliott[85,112,220], Kenneth Rice[40,220], Patricia B. Munroe[7,218,220], Alanna C. Morrison[5,220], L. Adrienne Cupples[38, 127,220], Dabeeru C. Rao[1,220], Daniel I. Chasman[49,220]

1. Division of Biostatistics, Washington University School of Medicine, St. Louis, MO 63110, USA.
2. Department of Genetic Epidemiology, University of Regensburg, Regensburg, 93051, Germany.
3. Cardiovascular Division, Department of Medicine, Washington University, St. Louis, MO 63110, USA.
4. Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA.
5. Department of Epidemiology, Human Genetics, and Environmental Sciences, The University of Texas School of Public Health, Houston, TX 77030, USA.
6. Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO 63108, USA.
7. William Harvey Research Institute, Clinical Pharmacology, Queen Mary University of London, London, London EC1M 6BQ, UK.
8. Genomic Outcomes, Department of Pediatrics, LABioMed at Harbor-UCLA Medical Center, Torrance, CA 90502, USA.
9. Department of Epidemiology, University of North Carolina Gilling School of Global Public Health, Chapel Hill, NC 27514, USA.
10. Icahn School of Medicine at Mount Sinai, The Charles Bronfman Institute for Personalized Medicine, New York, NY 10029, USA.

11. Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore 169856, Singapore.
12. Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore 169857, Singapore.
13. Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore 117597, Singapore.
14. Saw Swee Hock School of Public Health, National University Health System and National University of Singapore, Singapore, Singapore 117549, Singapore.
15. Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands.
16. Medical Research Council Human Genetics Unit, Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.
17. Jackson Heart Study, Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39213, USA.
18. Department of Epidemiology and Biostatistics, University of Giorgia at Athens College of Public Health, Athens, GA 30602, USA.
19. Section of Metabolic Genetics, Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, DK-2100, Denmark.
20. Department of Environmental Medicine and Public Health, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
21. Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA.
22. Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden 2300RC, The Netherlands.
23. Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA.
24. Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA.
25. Centre de Bioinformatique Biostatistique et Biologie Integrative (C3BI), Institut Pasteur, Paris 75015, France.
26. Cardiovascular Health Research Unit, Biostatistics and Medicine, University of Washington, Seattle, WA 98101, USA.
27. Genome Institute of Singapore, Agency for Science Technology and Research, Singapore 138672, Singapore.
28. Public Health Sciences, Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA.
29. Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA 02114, USA.
30. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.
31. Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA 70808, USA.
32. Icelandic Heart Association, Kopavogur 201, Iceland.
33. Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland.

34. Health Disparities Research Section, Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA.
35. Department of Public Health Sciences, Loyola University Chicago, Maywood, IL 60153, USA.
36. NIHR Cardiovascular Biomedical Research Unit, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK.
37. School of Public Health, Department of Epidemiology, University of Michigan, Ann Arbor, MI 48109, USA.
38. Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA.
39. Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Yokohama 230-0045, Japan.
40. Department of Biostatistics, University of Washington, Seattle, WA 98105, USA.
41. Estonian Genome Center, University of Tartu, Tartu 51010, Estonia.
42. Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, The Netherlands.
43. CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, University of Lille, Lille 59000, France.
44. Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117549, Singapore.
45. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Stockholm 17177, Sweden.
46. Division of Biostatistical Sciences, Department of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA.
47. Department of Medical Sciences, University of Trieste, Trieste 34137, Italy.
48. Department of Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA.
49. Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02215, USA.
50. Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, Oxfordshire OX3 9DU, UK.
51. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.
52. Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, Edinburgh EH8 9JZ, UK.
53. Medical Genetics Section, University of Edinburgh Centre for Genomic and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, The University of Edinburgh, Edinburgh EH4 2XU, UK.
54. Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, RS 96020220, Brazil.
55. Lab Genetics and Molecular Cardiology, Department of Cardiology, Heart Institute, University of Sao Paulo, Sao Paulo, Brazil.
56. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA.

57. Department of Clinical Physiology, Faculty of Medicine and Lifes Sciences, University of Tampere, Tampere 33014, Finland.
58. Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland.
59. Department of Public Health, University of Kelaniya, Ragama, Sri Lanka.
60. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg 85764, Germany.
61. Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg 85764, Germany.
62. Institute of Environmental Medicine, Karolinska Institutet, Stockholm 17177, Sweden.
63. Department of Medical Research, Taichung Veterans General Hospital, Department of Social Work, Tunghai University, Taichung 40705, Taiwan.
64. Department of Opthalmology, Taichung Veterans General Hospital, Taichung 40705, Taiwan.
65. MRC Epidemiology Unit, University of Cambridge, Cambridge CB2 0QQ, UK.
66. Tropical Metabolism Research Unit, Tropical Medicine Research Institute, University of the West Indies, Mona JMAAW15, Jamaica.
67. Department of Occupational and Environmental Health and State Key Laboratory of Environmental Health for Incubating, School of Public Health, Tongji Medical College Huazhong University of Science and Technology, Wuhan, China.
68. Department of Cardiovascular Sciences, University of Leicester, Leicester LE3 9QP, UK.
69. NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester LE3 9QP, UK.
70. Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio 70100, Finland.
71. Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY 10032, USA.
72. Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung 40705, Taiwan.
73. School of Medicine, National Yang-ming University, Taipei, Taiwan.
74. School of Medicine, National Defense Medical Center, Taipei, Taiwan.
75. Institute of Medical Technology, National Chung-Hsing University, Taichung 40705, Taiwan.
76. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio 70210, Finland.
77. Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo 1628655, Japan.
78. Genetic and Molecular Epidemiology Unit, Department of Clinical Sciences, Lund University, Malmö, Skåne SE-205 02, Sweden.
79. Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA.
80. Institute for Social Research, Research Center for Group Dynamics, University of Michigan, Ann Arbor, MI 48104, USA School of Public Health, Department of Epidemiology, University of Michigan, Ann Arbor, MI 48109, USA.

81. Interfaculty Institute for Genetics and Functional genomics, University Medicine and Ernst-Moritz Arndt University Greifswald, Greifswald 17487, Germany.
82. DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald 17475, Germany.
83. Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN 37203, USA.
84. General Internal Medicine, GeneSTAR Research Program, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA.
85. Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK.
86. Department of Cardiology, Ealing Hospital, Middlesex UB1 3HW, UK.
87. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.
88. Departments of Medicine and Epidemiology, Columbia University Medical Center, New York, NY 10032, USA.
89. Human Genetics Center, The University of Texas School of Public Health, Houston, TX 77030, USA.
90. Institute of Molecular Medicine, The University of Texas Health Science Center, Houston, TX 77030, USA.
91. Cardiovascular Health Research Unit, Medicine, University of Washington, Seattle, WA 98101, USA.
92. Section of Genomic Pediatrics, Department of Pediatrics, Medicine and Physiology, Medical College of Wisconsin, Milwaukee, WI 53226, USA.
93. Sleep Medicine and Circadian Disorders, Brigham and Women's Hospital, Boston, MA 02115, USA.
94. Centre for Genomic & Experimental Medicine, Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.
95. Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.
96. Centre for Brain Research, Indian Institute of Schience, Bangalore 560012, India.
97. The Danish Aging Research Center, Institute of Public Health, University of Southern Denmark, Odense, Denmark.
98. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA.
99. Ninewells Hospital & Medical School, University of Dundee, Dundee, DD1 9SY, UK.
100. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden 2300RC, The Netherlands.
101. Department of Medicine, University of Kelaniya, Ragama, Sri Lanka.
102. Inserm U1219 Neuroepidemiology, University of Bordeaux, Bordeaux, France.
103. Department of Neurology, University Hospital, Bordeaux, France.
104. Boston University School of Medicine, Boston, MA 02118, USA.
105. Department of Internal Medicine B, University Medicine Greifswald, Greifswald 17475, Germany.
106. Department of Genetics, University of North Carolina, Chapel Hill, NC 27514, USA.
107. Department of Family Medicine and Epidemiology, Alpert Medical School of Brown University, Providence, RI 02860, USA.

108. Division of Cardiology, Department of Specialties of Medicine, Geneva University Hospital, Geneva, foreign 1211, Switzerland.
109. Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina 45110, Greece.
110. Institute for Social Research, Survey Research Center, University of Michigan, Ann Arbor, MI 48104, USA.
111. Braun School of Public Health, Hebrew University-Hadassah Medical Center, Jerusalem 91120, Israel.
112. MRC-PHE Centre for Environment and Health, Imperial College London, London, UK.
113. Department of Epidemiology, State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.
114. Department of Psychology, The University of Edinburgh, Edinburgh EH8 9JZ, UK.
115. Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA.
116. Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA 70112, USA.
117. Department of Medicine, Tulane University School of Medicine, New Orleans, LA 70112, USA.
118. University of Eastern Finland, Institute of Biomedicine, Kuopio 70211, Finland.
119. Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore.
120. Khoo Teck Puat – National University Children's Medical Institute, National University Health System, Singapore 119228, Singapore.
121. Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, The University of Tokyo, Minato-ku 108-8639, Japan.
122. MedStar Health Research Institute, Hyattsville, MD 20782, USA.
123. Center for Clinical and Translational Sciences and Department of Medicine, Georgetown-Howard Universities, Washington, DC 20057, USA.
124. Cardiovascular Genetics, Department of Internal Medicine, University of Utah, Salt Lake City, UT 84108, USA.
125. Weill Cornell Medicine in Qatar, Doha, Qatar.
126. Hebrew SeniorLife, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02131, USA.
127. Framingham Heart Study, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20982, USA.
128. Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Suita 5650871, Japan.
129. Department of Geriatric and General Medicine, Osaka University Graduate School of Medicine, Suita 5650871, Japan.
130. Epidemiology, Department of Occupational and Environmental Medicine Program, University of Washington, Seattle, WA 98105, USA.
131. Department of Biochemistry, National University of Singapore, Singapore, Singapore 117596, Singapore.
132. Duke-NUS Medical School, Singapore 169857, Singapore.

133. Department of Health, National Institute for Health and Welfare, Helsinki 00271, Finland.
134. Department of Medicine and Abdominal Center: Endocrinology, University of Helsinki and Helsinki University Central Hospital, Helsinki 00029, Finland.
135. Fred Hutchinson Cancer Research Center, University of Washington School of Public Health, Seattle, WA 98109, USA.
136. Center for Integrative Medical Sciences, RIKEN, Yokohama 230-0045, Japan.
137. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio 70210, Finland.
138. Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35205, USA.
139. Lifelines cohort study, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, The Netherlands.
140. WHI CCC, Fred Hutchinson Cancer Research Center, Seattle, WA 98115, USA.
141. Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA.
142. Translational Laboratory in Genetic Medicine, Agency for Science, Technology and Research 138648, Singapore.
143. Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg 85764, Germany.
144. Institute of Human Genetics, Technische Universität München, Munich Germany, Technische Universität München, Munich 80333, Germany.
145. Laboratory for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama 230-0045, Japan.
146. Department of Biostatistics, University of Liverpool, Liverpool L69 3GL, UK.
147. Geriatrics, Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA.
148. Mathematical and Statistical Computing Laboratory, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892, USA.
149. The Institute of Medical Sciences, Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen AB25 2ZD, UK.
150. Data Tecnica International, Glen Echo, MD 20812, USA.
151. Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD 20892, USA.
152. Department of Epidemiology, Colorado School of Public Health, Aurora, CO 80045, USA.
153. Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27599, USA.
154. Department of Medicine, University of Ibadan, Ibadan, Nigeria.
155. Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow G12 8TA, UK.
156. Internal Medicine, Department of Medicine, Columbia University, New York, NY 10032, USA.
157. Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA.
158. Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN 55454, USA.

159. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Neuherberg 85764, Germany.
160. Faculty of Medicine, University of Split, Split, Croatia.
161. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20521, Finland.
162. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland.
163. Department of Biobank Research, Umeå University, Umeå, Västerbotten SE-901 87, Sweden.
164. Medical Genetics, IRCCS Burlo Garofolo 34137, Italy.
165. Department of Epidemiology and Medicine, University of Iowa, Iowa City, IA 52242, USA.
166. Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH8 9AG, UK.
167. Institute for Community Medicine, University Medicine Greifswald, Greifswald 17475, Germany.
168. Epidemiology & Community Health, University of Minnesota, Minneapolis, MN 55454, USA.
169. National Heart and Lung Institute, Imperial College London, London W12 0NN, UK.
170. Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA.
171. International Centre for Circulatory Health, Imperial College London, London, London W2 1PG, UK.
172. Alzheimer Scotland Dementia Research Centre, The University of Edinburgh, Edinburgh EH8 9JZ, UK.
173. Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg 85764, Germany.
174. Chair of Genetic Epidemiology, IBE, Faculty of Medicine, Ludwig-Maximilians-Universität, Munich 81377, Germany.
175. Department of Genetics, Stanford University, Stanford, CA 94305, USA.
176. Life Sciences Institute, National University of Singapore, Singapore, Singapore 117456, Singapore.
177. NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore, Singapore 117456, Singapore.
178. Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore 117546, Singapore.
179. Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands.
180. Beijing Institute of Ophthalmology, Beijing Ophthalmology and Visual Science Key Lab, Beijing Tongren Eye Center, Capital Medical University, Beijing, China 100730, China.
181. Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China 100730, China.
182. Jackson Heart Study, Department of Public Health, Jackson State University, Jackson, MS 39213, USA.
183. Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA.

184. Division of Cancer Control and Population Sciences, University of Pittsburgh Cancer Institute, Pittsburgh, PA 15232, USA.
185. Behavioral Epidemiology Section, Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA.
186. Broad Institute of the Massachusetts Institute of Technology and Harvard University, Boston, MA 02142, USA.
187. Harvard T. H. Chan School of Public Health, Department of Nutrition, Harvard University, Boston, MA 02115, USA.
188. Department of Public Health & Clinical Medicine, Umeå University, Umeå, Västerbotten SE-901 85, Sweden.
189. Division of Nephrology, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA.
190. Department of Genomics of Common Disease, Imperial College London, London W12 0NN, UK.
191. Division Experimental Genetics, Sidra, Doha 26999, Qatar.
192. German Center for Diabetes Research (DZD e.V.), Neuherberg 85764, Germany.
193. Endocrinology and Metabolism, Tri-Service General Hospital, National Defense Medical Center, Taipei City, Taipei 11490, Taiwan.
194. Department of Ophthalmology, Medical Faculty Mannheim, University Heidelberg, Mannheim, Germany 68167, Germany.
195. Department of Clinical Chemistry, Fimlab Laboratories, Tampere 33520, Finland.
196. Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Lifes Sciences, University of Tampere, Tampere 33014, Finland.
197. Cardiovascular Center, Taichung Veterans General Hospital.
198. Department of Medicine, China Medical University, Taichung 40705, Taiwan.
199. Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, The Netherlands.
200. Genetics, Harvard Medical School, Boston, MA 02115, USA.
201. Institute of Physiology, University Medicine Greifswald, Greifswald 17495, Germany.
202. Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, The Netherlands.
203. Department of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA.
204. Sticht Center for Health Aging and Alzheimer's Prevention, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA.
205. Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA.
206. Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD 21201, USA.
207. Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA.
208. Cardiovascular Health Research Unit, Epidemiology, Medicine and Health Services, University of Washington, Seattle, WA 98101, USA.
209. Kaiser Permanente Washington, Health Research Institute, Seattle, WA 98101, USA.

210. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore.
211. Dean's Office, University of Kentucky College of Public Health, Lexington, KY 40536, USA.
212. Department of Public Health and Primary Care, Leiden University Medical Center, Leiden 2300RC, The Netherlands.
213. Cardiology, Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA.
214. Icahn School of Medicine at Mount Sinai, The Mindich Child Health and Development Institute, New York, NY 10029, USA.
215. Genomic Outcomes, Department of Medicine, LABioMed at Harbor-UCLA Medical Center, Torrance, CA 90502, USA.
216. Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110, USA.
217. Division of Biostatistics, Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90032, USA.
218. NIHR Barts Cardiovascular Biomedical Research Unit, Queen Mary University of London, London, London EC1M 6BQ, UK.
219. These authors contributed equally to this work.
220. These authors jointly directed this work.