A Large-Scale Multi-ancestry Genome-wide Study Accounting for Smoking Behavior Identifies Multiple Significant Loci for Blood Pressure


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 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 loci that were genome-wide significant ($p < 5 \times 10^{-8}$) 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 ($p < 5 \times 10^{-8}$). Of the newly identified loci, ten 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, RPRGRIP1L), telomere maintenance (TNKS, PINX1, AKTIP), and central dopaminergic signaling (MSRA, EBF2).

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 (GWASs).2 Specifically, the common variants

1Division of Biostatistics, Washington University School of Medicine, St. Louis, MO 63110, USA; 2Department of Genetic Epidemiology, University of Regensburg, Regensburg 93051, Germany; 3Cardiovascular Division, Department of Medicine, Washington University, St. Louis, MO 63110, USA; 4Center for Research on Genomics and Global Health, National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA; 5Department of Epidemiology, Human Genetics, and Environmental Sciences, The University of Texas School of Public Health, Houston, TX 77030, USA; 6Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO 63108, USA; 7William Harvey Research Institute, Clinical Pharmacology, Queen Mary University of London, London EC1M 6BQ, UK; 8Genomic Outcomes, Department of Pediatrics, LABioMed at Harbor-UCLA Medical Center, Torrance, CA 90502, USA; 9Department of Epidemiology, University of North Carolina Gillings School of Global Public Health, Chapel Hill, NC 27514, USA; 10Icahn School of Medicine at Mount Sinai, The Charles Bronfman Institute for Personalized Medicine, New York, NY 10029, USA; 11Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore 169856, Singapore; 12Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore 117597, Singapore; 13Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore 117549, Singapore; 14Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands; 15Medical Research Council Human Genetics Unit, Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh EH4 2UX, UK; 16Jackson Heart Study, Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39213, USA; 17Department of Epidemiology and Biostatistics, University of Georgia at Athens College of Public Health, Athens, GA 30602, USA; 18Section of Metabolic Genetics, Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 2100, Denmark; 19Department of Environmental Medicine and Public Health, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 20Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 21Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden 2300RC, the Netherlands; 22Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA; 23Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA; 24Centre de Bioinformatique Biostatistique et Biologie Integrative (C3BI), Institut Pasteur, Paris 75015, France; 25Cardiovascular Health Research Unit, Biostatistics

© 2018 American Society of Human Genetics.
and University of Washington, Seattle, WA 98101, USA; 23Genome Institute of Singapore, Agency for Science Technology and Research, Singapore 138672, Singapore; 24Division of Biostatistical Sciences, Department of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; 25Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA 02114, USA; 26Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 27Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA 70808, USA; 28Icelandic Heart Association, Kopavogur 201, Iceland; 29Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland; 30Health Disparities Research Section, Laboratory of Epidemiology and Population Sciences, National Institute on Aging, NIH, Baltimore, MD 21224, USA; 31Department of Public Health Sciences, Loyola University Chicago, Maywood, IL 60153, USA; 32NIHR Cardiovascular Biomedical Research Unit, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 1BB, UK; 33School of Public Health, Department of Epidemiology, University of Michigan, Ann Arbor, MI 48109, USA; 34Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA; 35Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Yokohama 224-0028, Japan; 36Department of Biostatistics, University of Washington, Seattle, WA 98105, USA; 37Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; 38Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, the Netherlands; 39CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, University of Lille, Lille 59000, France; 40Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117549, Singapore; 41Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden; 42Department of Medical Sciences, University of Trieste, Trieste 34137, Italy; 43Department of Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; 44Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02215, USA; 45Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, Oxfordshire OX3 9DU, UK; 46Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; 47Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, Edinburgh EH9 9JZ, UK; 48Medical 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; 49Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, RS 96020220, Brazil; 50Lab Genetics and Molecular Cardiology, Department of Cardiology, Heart Institute, University of Sao Paulo, Sao Paulo, Brazil; 51Department of Biostatistics and Center for Statistical Genetics, Michigan University, Ann Arbor, MI 48109, USA; 52Department of Clinical Physiology, Faculty of Life Sciences, University of Tromsø, Tromsø 9037, Norway; 53Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland; 54Department of Public Health, University of Helsinki, Helsinki, Finland; 55Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg 85764, Germany; 56Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg 85764, Germany; 57Institute of Environmental Medicine, Karolinska Institutet, Stockholm 17177, Sweden; 58Department of Medical Research, Taichung Veterans General Hospital, Department of Social Work, Taichung, Taiwan; 59Institute of Medical Technology, National Yang-Ming University, Taipei, Taiwan; 60School of Medicine, National Defense Medical Center, Taipei, Taiwan; 61Institute of Medical Technology, National Chung-Hsing University, Taichung 40705, Taiwan; 62Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio 70210, Finland; 63Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo 1628655, Japan; 64Genetic and Molecular Epidemiology Unit, Department of Clinical Sciences, Lund University, Malmö, Skåne 205 02, Sweden; 65Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA; 66Institute for Social Research, Social Research Center for Group Dynamics, University of Michigan, Ann Arbor, MI 48104, USA; 67Interfaculty Institute for Genetics and Functional Genomics, University Medicine and Ernst-Moritz-Arndt University Greifswald, Greifswald 17487, Germany; 68DIZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald 17487, Germany; 69Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN 37203, USA; 70General Internal Medicine, GeneSTAR Research Program, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA; 71Department of Epidemiology and Biostatistics, Imperial College London, London W2 IPG, UK; 72Department of Cardiology, Ealing Hospital, Middlesex UB1 3HW, UK; 73McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 74Departments of Medicine and Epidemiology, Columbia University Medical Center, New York, NY.
initially identified through three collaborative consortia for genome-wide BP genetics in people of European ancestry,\textsuperscript{1,3,4} explain less than 2.5% of the variance in systolic BP (SBP) or diastolic BP (DBP). Recent reports based on larger sample sizes have increased the number of BP-associated variants which together explain about 3.5% of BP variance.\textsuperscript{5–7} In contrast, only six BP loci have been identified by GWASs in African ancestry which explain less than 0.54% of BP variance.\textsuperscript{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 an additional route for discovery of genetic effects on complex traits,\textsuperscript{10} including BP, and may more generally extend our knowledge of the genetic architecture of complex traits.\textsuperscript{11}

Many lifestyle factors including physical activity, tobacco use, alcohol consumption, stress, and dietary factors influence BP.\textsuperscript{12} These lifestyle exposures may also modify the effect of genetic variants on BP. Cigarette smoking is known to influence BP in both acute and chronic fashion, motivating genetic association studies accounting for potential gene-by-smoking interactions. This may help identify 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.\textsuperscript{16} The primary goal of these investigations is to use interactions to identify 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.”\textsuperscript{17} Within this setting, we performed a genome-wide association meta-analysis incorporating gene-smoking interactions (overview shown in Figure 1) to identify SBP- and DBP-associated 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.\textsuperscript{16}

Material and Methods

Overview of Participating Studies

Men and women between the ages of 18 and 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 Note. 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 (\(p < 5 \times 10^{-8}\)) 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 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, resting SBP (mmHg) and DBP (mmHg), were analyzed separately. 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.\textsuperscript{5} Summary statistics are shown in Table 1 (more details in Tables S3 and 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 six 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 had 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 or Affymetrix genotyping arrays. Each study performed imputation to impute genotypes for SNPs, short insertions and deletions (indels), and larger deletions that were not genotyped directly but are available from the 1000 Genomes Project. Information on genotype and imputation for each study is presented in Tables S6 and 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,

\[
E[Y | G, C] = \beta_0 + \beta_S Smk + \beta_C G + \beta_{GE} Smk \times G + \gamma_C C
\]

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

\[
E[Y | C, Smk = 0] = \gamma_0^0 + \gamma_C^0 G + \gamma_C^0 C
\]

\[
E[Y | C, Smk = 1] = \gamma_0^1 + \gamma_C^1 G + \gamma_C^1 C
\]

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) (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. Therefore, we considered only the first “joint” model in stage 2.

Each study in stage 1 performed GWAS analysis within each ancestry and provided (1) the estimated genetic main effect \(\hat{\beta}_G\), estimated interaction effect \(\hat{\beta}_{GE}\), and a robust estimate of the corresponding covariance matrix under the joint model; and (2) estimates of the stratum-specific effects \(\hat{\gamma}_{G,1}\) and \(\hat{\gamma}_{C,1}\) and robust estimates of their standard errors (SE) under the stratified model. Each study in stage 2 provided estimates of the genetic main effect \(\hat{\beta}_G\), the interaction effect \(\hat{\beta}_{GE}\), and robust estimates of the corresponding covariance matrix under the joint model at 4,459 select variants. Robust estimates of covariance matrices and SEs were used to
We have examined QQ plots and genomic control inflation factors to identify problems with population substructures or relatedness. We performed two QC levels: “study-level” and “meta-level.”

We performed extensive QC using the R package EasyQC for all cohort-specific GWAS results. In stage 1, each cohort provided 12 GWAS result files (2 BPs × 2 smoking exposures × 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 GWASs 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. 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 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.

### Quality Control

Study investigators participating in this study have ample experience in main-effect-based GWASs 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 for all cohort-specific GWAS results. In stage 1, each cohort provided 12 GWAS result files (2 BPs × 2 smoking exposures × 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 GWASs 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. 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 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.

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 * MAF0 * N0 for the unexposed group (with MAF0 and sample size N0 for E = 0 stratum) and MAC1 = 2 * MAFe1 * N1E 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 the Supplemental Note. 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.23

**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-square distribution with 1 df under \( H_0: \beta_{GE} = 0 \) for the 1 df test or a chi-square distribution with 2 df under \( H_0: \beta_{GI} = \beta_{GE} = 0 \), 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.29 for the 2 df test, both using METAL.27 In the stratified model, we performed meta-analysis using the approach of Randall et al.29 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.28 More details are described elsewhere.27

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 × 2 smoking exposures × 4 (2 tests in the joint model, 2 stratified groups in the stratified model) × 5 ancestries (4 ancestry-specific and 1 trans-ancestry to combine ancestry-specific results). In this stage, genomic control correction31 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 (the number of variants varied across the ancestry). Similarly, we considered all index variants representing previously unreported loci as “novel” variants. Similarly, we considered all index variants representing previously unreported 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 four 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 GTEx34 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 more than 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 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 ($P_{\text{EUR,SBP}} < 5 \times 10^{-8}$), (2) DBP in Europeans ($P_{\text{EUR,DBP}} < 5 \times 10^{-8}$), (3) SBP in trans-ancestry analysis ($P_{\text{Trans.SBP}} < 5 \times 10^{-8}$), or (4) DBP in trans-ancestry analysis ($P_{\text{Trans,DBP}} < 5 \times 10^{-8}$). 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, 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 GWASs. 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. 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 Acumeta 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 (Figure 1). For the 2-step approach, we performed ancestry-specific meta-analysis in each of five 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 ($p < 5 \times 10^{-8}$) 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. 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 stages 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 least 1 Mb away from any previously

The American Journal of Human Genetics 102, 375–400, March 1, 2018
Table 2. Newly Identified Loci that Are Significant in Stage 1 and Formally Replicated in Stage 2

<table>
<thead>
<tr>
<th>Locusa</th>
<th>Nearest Genesb</th>
<th>rsID</th>
<th>Chr:Posc</th>
<th>EA</th>
<th>EAF</th>
<th>Stage</th>
<th>Genetic Main Effect Estd</th>
<th>Genetic Main Effect SEd</th>
<th>Interaction Effect Estd</th>
<th>Interaction Effect SEd</th>
<th>2 df Jointp Valuee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MTHFR;CLCN6*;NPPA</td>
<td>rs202071545</td>
<td>1:11878161</td>
<td>d</td>
<td>0.945</td>
<td>ALL.SBP</td>
<td>1.24</td>
<td>0.28</td>
<td>−0.16</td>
<td>0.38</td>
<td>3.77 × 10⁻⁸</td>
</tr>
<tr>
<td>2</td>
<td>CLCN6;NPPA;NPPB*</td>
<td>rs3753581</td>
<td>1:11920189</td>
<td>a</td>
<td>0.327</td>
<td>ALL.SBP</td>
<td>−0.63</td>
<td>0.09</td>
<td>0.16</td>
<td>0.21</td>
<td>4.34 × 10⁻¹²</td>
</tr>
<tr>
<td>3</td>
<td>NPPA;NPPB</td>
<td>rs72640287</td>
<td>1:11965792</td>
<td>t</td>
<td>0.039</td>
<td>EUR.SBP</td>
<td>−2.05</td>
<td>0.43</td>
<td>−0.04</td>
<td>0.59</td>
<td>1.59 × 10⁻¹⁰</td>
</tr>
<tr>
<td>4</td>
<td>WNT2B*</td>
<td>rs351364</td>
<td>1:113045061</td>
<td>a</td>
<td>0.297</td>
<td>ALL.SBP</td>
<td>−0.60</td>
<td>0.10</td>
<td>0.53</td>
<td>0.22</td>
<td>1.67 × 10⁻⁸</td>
</tr>
<tr>
<td>5f</td>
<td>CEP170;SDCCAG8*;AKT3</td>
<td>rs3897821</td>
<td>1:243420388</td>
<td>a</td>
<td>0.705</td>
<td>ALL.DBP</td>
<td>−0.35</td>
<td>0.06</td>
<td>0.20</td>
<td>0.13</td>
<td>2.49 × 10⁻⁹</td>
</tr>
<tr>
<td>6f</td>
<td>FER1L5*</td>
<td>rs7599598</td>
<td>2:97351840</td>
<td>a</td>
<td>0.564</td>
<td>EUR.DBP</td>
<td>−0.30</td>
<td>0.06</td>
<td>−0.15</td>
<td>0.14</td>
<td>5.93 × 10⁻⁸</td>
</tr>
<tr>
<td>7</td>
<td>SLC4A7*</td>
<td>rs13063291</td>
<td>3:27446285</td>
<td>a</td>
<td>0.204</td>
<td>ALL.DBP</td>
<td>0.33</td>
<td>0.08</td>
<td>0.03</td>
<td>0.12</td>
<td>4.00 × 10⁻⁸</td>
</tr>
<tr>
<td>8f</td>
<td>PRAG1;MFHAS1</td>
<td>rs7823056</td>
<td>8:8382705</td>
<td>a</td>
<td>0.397</td>
<td>EUR.SBP</td>
<td>−0.56</td>
<td>0.10</td>
<td>−0.02</td>
<td>0.22</td>
<td>1.54 × 10⁻⁸</td>
</tr>
<tr>
<td>9f</td>
<td>PPP1R3B;TNKS</td>
<td>rs62493780</td>
<td>8:9151051</td>
<td>t</td>
<td>0.238</td>
<td>EUR.SBP</td>
<td>0.89</td>
<td>0.18</td>
<td>−0.19</td>
<td>0.25</td>
<td>3.47 × 10⁻⁸</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>Locus</th>
<th>Nearest Genes</th>
<th>rsID</th>
<th>Chr:Pos</th>
<th>EA</th>
<th>EAF</th>
<th>Ancestry and Trait</th>
<th>Stage</th>
<th>Genetic Main Effect Est</th>
<th>Genetic Main Effect SE</th>
<th>Interaction Effect Est</th>
<th>Interaction Effect SE</th>
<th>2 df Joint p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>MIR124-1*;MSRA</td>
<td>rs13271489</td>
<td>8:9803712</td>
<td>t</td>
<td>0.478</td>
<td>EUR.SBP</td>
<td>1</td>
<td>0.55</td>
<td>0.10</td>
<td>-0.13</td>
<td>0.14</td>
<td>9.35 × 10^{-16}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.44</td>
<td>0.05</td>
<td>-0.13</td>
<td>0.14</td>
<td>6.37 × 10^{-9}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1+2</td>
<td>0.46</td>
<td>0.05</td>
<td>-0.08</td>
<td>0.12</td>
<td>*4.56 × 10^{-23}</td>
</tr>
<tr>
<td>11</td>
<td>TNNI2;LSP1*;TNNT3</td>
<td>rs7483477</td>
<td>11:1920255</td>
<td>t</td>
<td>0.75</td>
<td>ALL.SBP</td>
<td>1</td>
<td>-0.65</td>
<td>0.11</td>
<td>0.17</td>
<td>0.27</td>
<td>2.25 × 10^{-6}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>-0.36</td>
<td>0.05</td>
<td>0.08</td>
<td>0.12</td>
<td>1.77 × 10^{-13}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1+2</td>
<td>-0.40</td>
<td>0.04</td>
<td>0.09</td>
<td>0.11</td>
<td>*2.12 × 10^{-20}</td>
</tr>
<tr>
<td>12</td>
<td>POC1B;ATP2B1</td>
<td>rs7313874</td>
<td>12:89965049</td>
<td>t</td>
<td>0.325</td>
<td>ALL.SBP</td>
<td>1</td>
<td>-0.64</td>
<td>0.11</td>
<td>0.01</td>
<td>0.15</td>
<td>1.85 × 10^{-14}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>-0.48</td>
<td>0.06</td>
<td>-0.23</td>
<td>0.09</td>
<td>1.07 × 10^{-29}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1+2</td>
<td>-0.52</td>
<td>0.05</td>
<td>-0.17</td>
<td>0.08</td>
<td>*2.49 × 10^{-54}</td>
</tr>
<tr>
<td>13</td>
<td>ATP2B1*</td>
<td>rs111337717</td>
<td>12:90037506</td>
<td>t</td>
<td>0.943</td>
<td>ALL.SBP</td>
<td>1</td>
<td>1.27</td>
<td>0.33</td>
<td>0.66</td>
<td>0.46</td>
<td>9.23 × 10^{-11}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1.09</td>
<td>0.15</td>
<td>-0.26</td>
<td>0.22</td>
<td>2.86 × 10^{-18}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1+2</td>
<td>1.13</td>
<td>0.13</td>
<td>-0.07</td>
<td>0.20</td>
<td>*1.27 × 10^{-27}</td>
</tr>
<tr>
<td>14</td>
<td>PTPN11</td>
<td>rs7974266</td>
<td>12:113007602</td>
<td>t</td>
<td>0.513</td>
<td>ALL.DBP</td>
<td>1</td>
<td>0.19</td>
<td>0.13</td>
<td>0.57</td>
<td>0.18</td>
<td>3.58 × 10^{-8}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.23</td>
<td>0.06</td>
<td>0.19</td>
<td>0.09</td>
<td>6.50 × 10^{-12}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1+2</td>
<td>0.22</td>
<td>0.06</td>
<td>0.28</td>
<td>0.08</td>
<td>*5.91 × 10^{-19}</td>
</tr>
<tr>
<td>15</td>
<td>AKTIP;RIPKIP1;FTO*</td>
<td>rs11642015</td>
<td>16:53802494</td>
<td>t</td>
<td>0.334</td>
<td>ALL.SBP</td>
<td>1</td>
<td>0.57</td>
<td>0.09</td>
<td>-0.19</td>
<td>0.21</td>
<td>2.78 × 10^{-9}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.29</td>
<td>0.05</td>
<td>0.08</td>
<td>0.11</td>
<td>6.74 × 10^{-13}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1+2</td>
<td>0.35</td>
<td>0.04</td>
<td>0.03</td>
<td>0.10</td>
<td>*9.91 × 10^{-21}</td>
</tr>
</tbody>
</table>

Each locus is genome-wide significant (p < 5 × 10^{-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. Abbreviations: 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).

*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.

bGene names were obtained using variant effect predictor (VEP) from Ensembl. Genes with intragenic index variants are indicated with an asterisk (*).

cPositions are based on build 37.

dEffect is in mmHg unit.

*The most significant p value (between 1 df interaction test and 2 df joint test) are indicated with an asterisk (*).

fThese loci indicate “completely novel” loci, at least 1 Mb away from any of known BP loci.
published BP variants, and we term them “completely novel.” 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 two independent signals identified in European ancestry ($p = 4.1 \times 10^{-41}$), Asian ($p = 1.5 \times 10^{-13}$), and trans-ancestry ($p = 2.5 \times 10^{-54}$) analyses. Near another well-known BP locus, MTHFR-NPPB-CLCN6, we identified three additional independent signals (with p values as small as $4.3 \times 10^{-34}$ at index variants, spanning 196 kb [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 ($p < 5 \times 10^{-5}$) 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.\textsuperscript{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 ($p < 1 \times 10^{-5}$) in stage 1 analyses but became significant in the combined stages 1 and 2 meta-analysis (Tables 3, 4, and 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 ($p = 9.4 \times 10^{-7}$) and identified through the combined analysis in European ancestry ($p = 6.0 \times 10^{-28}$), Asian ($p = 1.2 \times 10^{-10}$), and trans-ancestry ($p = 3.6 \times 10^{-41}$, 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 $p = 5.8 \times 10^{-10}$ in stage 1; $p = 0.03$ in stage 2; $p = 4.2 \times 10^{-12}$ in stages 1–2). Six loci were suggestive ($p < 1 \times 10^{-6}$) 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 $p = 8.7 \times 10^{-7}$ in stage 1; $p = 0.00047$ in stage 2; $p = 1.8 \times 10^{-10}$ in stages 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\textsuperscript{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 $P_{SBP} = 4.8 \times 10^{-85}$; $P_{DBP} = 5.5 \times 10^{-57}$) and SH2B3-ATXN2 (rs3184504 on chromosome 12, trans-ancestry $P_{SBP} = 3.2 \times 10^{-36}$; $P_{DBP} = 6.0 \times 10^{-67}$).

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, 3, 4, and 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 $p < 5 \times 10^{-8}$). 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 (1 df interaction $p < 0.05/158$). 27 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 WGH5). 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
Figure 2. Forest Plots and LocusZoom Plots for Three Newly Identified Loci
(A and 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 and 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 and F) Variant rs11642015 is intronic to the well-known obesity/diabetes locus FTO. In addition, AKTIP in this locus has role in telomere maintenance.
Loci selected from Table 2.
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 and 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) that 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
vertebrates were 6 variants as identified via GERP and 5 variants via SiPhy. RegulomeDB assigned class 1f well-known BP locus MTHFR target. Of these, rs12741980 (locus 2, Table 4) is near the regulatory elements and is linked to expression of a gene (Table S12), each of which likely affects the binding of regulatory proteins (strong evidence for enhancer function) for 2 variants each.

Main Interaction 2 df Joint Trait

<table>
<thead>
<tr>
<th>Effect 1 df Interaction 2 df Joint Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>P value</td>
</tr>
<tr>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Genetic Main</td>
</tr>
</tbody>
</table>

Each locus is genome-wide significant ($p < 5 \times 10^{-8}$) 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. Abbreviations: 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.

*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.

**Gene names were obtained using variant effect predictor (VEP) from Ensembl. Genes with intragenic index variants are indicated with an asterisk (*).

***Positions are based on build 37.

**Effect is in MmHg unit.

**The most significant p value (between 1 df interaction test and 2 df joint test) was set in bold.

These loci indicate “completely novel” loci, at least 1 Mb away from any of known BP loci.

Table 4. Additional Significant Loci from the Combined Analyses of Stages 1 and 2 in European Ancestry

<table>
<thead>
<tr>
<th>Locus</th>
<th>Nearest Genes</th>
<th>rsID</th>
<th>Chr:Pos</th>
<th>EA</th>
<th>EAF</th>
<th>Effect</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MTHFR*;CLCN6</td>
<td>rs6541006</td>
<td>1:11857526</td>
<td>a</td>
<td>0.071</td>
<td>-0.85</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>KCNC3;DYNC2LI1</td>
<td>rs73923009</td>
<td>2:43141074</td>
<td>a</td>
<td>0.099</td>
<td>-0.36</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>SLC17A1-4;HFE</td>
<td>rs7753826</td>
<td>6:26042239</td>
<td>a</td>
<td>0.189</td>
<td>0.36</td>
<td>-0.06</td>
</tr>
<tr>
<td>4</td>
<td>SLC44A4;EHMT2*;STK19;CYP21A2;TNNB</td>
<td>rs2243873</td>
<td>6:31863433</td>
<td>a</td>
<td>0.556</td>
<td>0.45</td>
<td>-0.19</td>
</tr>
<tr>
<td>5</td>
<td>SLC44A4;EHMT2*;HLA-DQB2*;STK19;CYP21A2;TNNB</td>
<td>rs2071550</td>
<td>6:32370940</td>
<td>a</td>
<td>0.307</td>
<td>0.32</td>
<td>-0.22</td>
</tr>
<tr>
<td>6</td>
<td>TNKS;MSRA</td>
<td>rs4841235</td>
<td>8:9683358</td>
<td>a</td>
<td>0.426</td>
<td>0.37</td>
<td>-0.1</td>
</tr>
<tr>
<td>7</td>
<td>SOX7*;PINX1</td>
<td>rs6995692</td>
<td>8:10587008</td>
<td>c</td>
<td>0.563</td>
<td>-0.44</td>
<td>0.31</td>
</tr>
<tr>
<td>8</td>
<td>ADARB2*</td>
<td>rs150155092</td>
<td>10:1769881</td>
<td>d</td>
<td>0.013</td>
<td>4.76</td>
<td>-18.32</td>
</tr>
<tr>
<td>9</td>
<td>KAT5;RNASEH2C</td>
<td>rs72941051</td>
<td>11:65478893</td>
<td>t</td>
<td>0.074</td>
<td>-0.39</td>
<td>0.07</td>
</tr>
<tr>
<td>10</td>
<td>FAM19A2*;AVPR1A</td>
<td>rs17713040</td>
<td>12:6246714</td>
<td>t</td>
<td>0.977</td>
<td>0.24</td>
<td>0.31</td>
</tr>
<tr>
<td>11</td>
<td>FAM109A;SH2B3*;ATXN2</td>
<td>rs4375492</td>
<td>12:111835990</td>
<td>a</td>
<td>0.794</td>
<td>0.35</td>
<td>0.03</td>
</tr>
<tr>
<td>12</td>
<td>MPI;COX5A;SCAMP5</td>
<td>rs12050494</td>
<td>15:75620896</td>
<td>a</td>
<td>0.316</td>
<td>0.32</td>
<td>0.06</td>
</tr>
<tr>
<td>13</td>
<td>NAA38*;KCNA8;VAMP2</td>
<td>rs74439044</td>
<td>17:7781019</td>
<td>t</td>
<td>0.903</td>
<td>-0.36</td>
<td>-0.14</td>
</tr>
</tbody>
</table>

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 and RegulomeDB. 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 and 5 variants via SiPhy. 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-ASI1, 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 ADRBI, an adrenergic receptor that mediates effects of the hormone epinephrine and the neurotransmitter norepinephrine, although it is about 80 kb upstream of this locus. Of note, our results identified this known BP locus (rs2782980, $p = 1.1 \times 10^{-21}$ and rs1801253, $p = 1.3 \times 10^{-22}$, 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 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 on chromosome 8) are cis-eQTLs for PRAG1, which is expressed in multiple tissues including the cerebral and thyroid. The most abundant evidence of cis-eQTL association (with 44 eQTL hits from multiple studies) was observed for rs2243873, an intronic variant of EHMT2; it is predicted to regulate expression of many genes including HLA-C, HLA-B, and HLA-DBR1 across multiple tissues.
Table 5. Additional Significant Loci from the Combined Analyses of Stages 1 and 2 in African Ancestry

<table>
<thead>
<tr>
<th>Locus</th>
<th>Nearest Genes</th>
<th>rsID</th>
<th>ChrPos</th>
<th>EA</th>
<th>EAF</th>
<th>Effect</th>
<th>p Value</th>
<th>Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A2AP1*</td>
<td>rs12135881</td>
<td>1:4781922</td>
<td>c</td>
<td>0.988</td>
<td>-2.05</td>
<td>16.94</td>
<td>SBP</td>
</tr>
<tr>
<td>2</td>
<td>FABP3;SERINC2;TINAGL1</td>
<td>rs11809589</td>
<td>1:31907118</td>
<td>a</td>
<td>0.012</td>
<td>-1.11</td>
<td>18.04</td>
<td>SBP</td>
</tr>
<tr>
<td>3</td>
<td>LOC10128219</td>
<td>rs182662555</td>
<td>1:96289336</td>
<td>t</td>
<td>0.988</td>
<td>6.15</td>
<td>-4.45</td>
<td>SBP</td>
</tr>
<tr>
<td>4</td>
<td>PXDN;MYT11*</td>
<td>rs75247762</td>
<td>2:1893133</td>
<td>t</td>
<td>0.014</td>
<td>-2.37</td>
<td>-12.93</td>
<td>DBP</td>
</tr>
<tr>
<td>5</td>
<td>ASR3;ELENC1;GPR75</td>
<td>rs115234772</td>
<td>2:53650295</td>
<td>a</td>
<td>0.987</td>
<td>-0.1</td>
<td>8.5</td>
<td>DBP</td>
</tr>
<tr>
<td>6</td>
<td>SEATAD2;SLC1A4</td>
<td>rs145162854</td>
<td>2:65104447</td>
<td>a</td>
<td>0.015</td>
<td>-3.17</td>
<td>-2.61</td>
<td>SBP</td>
</tr>
<tr>
<td>7</td>
<td>ACOXL*</td>
<td>rs116008367</td>
<td>2:111807546</td>
<td>c</td>
<td>0.014</td>
<td>-0.86</td>
<td>-5.35</td>
<td>DBP</td>
</tr>
<tr>
<td>8</td>
<td>KCNE4;SCG2</td>
<td>rs10166552</td>
<td>2:224036537</td>
<td>t</td>
<td>0.016</td>
<td>-0.15</td>
<td>-10.83</td>
<td>SBP</td>
</tr>
<tr>
<td>9</td>
<td>TRY1;MCM2</td>
<td>rs139963642</td>
<td>3:127314188</td>
<td>t</td>
<td>0.013</td>
<td>-6.35</td>
<td>1.23</td>
<td>DBP</td>
</tr>
<tr>
<td>10</td>
<td>PCDH7</td>
<td>rs11931572</td>
<td>4:30086104</td>
<td>a</td>
<td>0.968</td>
<td>-0.45</td>
<td>3.28</td>
<td>DBP</td>
</tr>
<tr>
<td>11</td>
<td>SPRY1;LINC01091*</td>
<td>rs62319742</td>
<td>4:124581262</td>
<td>a</td>
<td>0.014</td>
<td>1.98</td>
<td>-10.98</td>
<td>DBP</td>
</tr>
<tr>
<td>12</td>
<td>HSD17B4</td>
<td>rs140543491</td>
<td>5:11923601</td>
<td>a</td>
<td>0.017</td>
<td>-3</td>
<td>16.29</td>
<td>SBP</td>
</tr>
<tr>
<td>13</td>
<td>OFCC1</td>
<td>rs148387718</td>
<td>6:9446000</td>
<td>t</td>
<td>0.014</td>
<td>0.59</td>
<td>-7.84</td>
<td>DBP</td>
</tr>
<tr>
<td>14</td>
<td>NEDD9;LOC105374928*</td>
<td>rs9348895</td>
<td>6:11496048</td>
<td>a</td>
<td>0.586</td>
<td>0.11</td>
<td>1.21</td>
<td>SBP</td>
</tr>
<tr>
<td>15</td>
<td>MYO6;MPG1*</td>
<td>rs58806982</td>
<td>6:76688006</td>
<td>t</td>
<td>0.01</td>
<td>-11.24</td>
<td>14.92</td>
<td>DBP</td>
</tr>
<tr>
<td>16</td>
<td>TAR1;SLC2A12</td>
<td>rs76987554</td>
<td>6:134080555</td>
<td>t</td>
<td>0.062</td>
<td>-1.57</td>
<td>0</td>
<td>DBP</td>
</tr>
<tr>
<td>17</td>
<td>ARID1B*</td>
<td>rs112140754</td>
<td>6:157245233</td>
<td>t</td>
<td>0.988</td>
<td>0.97</td>
<td>7.6</td>
<td>SBP</td>
</tr>
<tr>
<td>18</td>
<td>BZW2*</td>
<td>rs116196735</td>
<td>7:16710605</td>
<td>a</td>
<td>0.018</td>
<td>-2.88</td>
<td>-13.75</td>
<td>SBP</td>
</tr>
<tr>
<td>19</td>
<td>MED30;EXT1</td>
<td>rs74701635</td>
<td>8:118758316</td>
<td>t</td>
<td>0.016</td>
<td>3.79</td>
<td>-19.2</td>
<td>SBP</td>
</tr>
<tr>
<td>20</td>
<td>ADAMTS11;MR3152</td>
<td>rs146250839</td>
<td>9:18189778</td>
<td>a</td>
<td>0.976</td>
<td>0.35</td>
<td>2.79</td>
<td>DBP</td>
</tr>
<tr>
<td>21</td>
<td>SPIN1;SPR3;SHC3;CKS2</td>
<td>rs192642798</td>
<td>9:91503938</td>
<td>a</td>
<td>0.012</td>
<td>-8.38</td>
<td>3.95</td>
<td>SBP</td>
</tr>
<tr>
<td>22</td>
<td>FZD8</td>
<td>rs76726877</td>
<td>10:36331497</td>
<td>t</td>
<td>0.015</td>
<td>-1.55</td>
<td>-9.14</td>
<td>DBP</td>
</tr>
<tr>
<td>23</td>
<td>SFRP5;CRTA1*</td>
<td>rs11599481</td>
<td>10:9640463</td>
<td>t</td>
<td>0.058</td>
<td>-0.9</td>
<td>-3.33</td>
<td>DBP</td>
</tr>
<tr>
<td>24</td>
<td>TSPAN18;PRDM11;SYT13</td>
<td>rs148772934</td>
<td>11:45005681</td>
<td>t</td>
<td>0.986</td>
<td>-0.57</td>
<td>11.66</td>
<td>DBP</td>
</tr>
<tr>
<td>25</td>
<td>SLC15A3;Cd6;LOC105369325*;CD5</td>
<td>rs11601370</td>
<td>11:60834043</td>
<td>t</td>
<td>0.976</td>
<td>1.34</td>
<td>6.63</td>
<td>SBP</td>
</tr>
<tr>
<td>26</td>
<td>LOC101928944</td>
<td>rs74601585</td>
<td>11:80140007</td>
<td>t</td>
<td>0.017</td>
<td>-3.93</td>
<td>-2.58</td>
<td>DBP</td>
</tr>
<tr>
<td>27</td>
<td>LOC105369408</td>
<td>rs78103586</td>
<td>11:133893928</td>
<td>a</td>
<td>0.029</td>
<td>-1.65</td>
<td>5.27</td>
<td>DBP</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>Locus</th>
<th>Nearest Genes</th>
<th>rsID</th>
<th>Chr:Pos</th>
<th>EA</th>
<th>EAF</th>
<th>Effect</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>28f</td>
<td>PLEKHG7;EEA1;LOC643339*</td>
<td>rs61935525</td>
<td>12:93645481</td>
<td>c</td>
<td>0.985</td>
<td>1.26</td>
<td>10.15</td>
</tr>
<tr>
<td>29f</td>
<td>DICER1;CLMN</td>
<td>rs187852559</td>
<td>14:95794914</td>
<td>a</td>
<td>0.013</td>
<td>−1.67</td>
<td>−4.93</td>
</tr>
<tr>
<td>30f</td>
<td>SETD3;CCNK</td>
<td>rs1257310</td>
<td>14:99810427</td>
<td>a</td>
<td>0.784</td>
<td>1.03</td>
<td>0.98</td>
</tr>
<tr>
<td>31</td>
<td>GPR139;GP2;UMOD;PDILT</td>
<td>rs148753653</td>
<td>16:20230175</td>
<td>a</td>
<td>0.981</td>
<td>5.25</td>
<td>−9.4</td>
</tr>
<tr>
<td>32f</td>
<td>LOC339166*;WSCD1</td>
<td>rs138973557</td>
<td>17:5699720</td>
<td>t</td>
<td>0.903</td>
<td>0.36</td>
<td>2.09</td>
</tr>
<tr>
<td>33f</td>
<td>DYM;LIPG;ACAA2;MYO5B</td>
<td>rs9965695</td>
<td>18:47261614</td>
<td>t</td>
<td>0.982</td>
<td>0.29</td>
<td>13.32</td>
</tr>
<tr>
<td>34f</td>
<td>ZNF98*</td>
<td>rs10405764</td>
<td>19:22598479</td>
<td>t</td>
<td>0.017</td>
<td>0.91</td>
<td>−19.1</td>
</tr>
<tr>
<td>35f</td>
<td>BMP7</td>
<td>rs115893283</td>
<td>20:55401465</td>
<td>t</td>
<td>0.042</td>
<td>0.9</td>
<td>−9.08</td>
</tr>
</tbody>
</table>

Each locus is genome-wide significant (p < 5 × 10⁻⁸) 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. Abbreviations: 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.

*Each locus was determined through LD-based clumping, using ± 1 Mb around index variants, followed by LD threshold of r² > 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.

*Gene names were obtained using variant effect predictor (VEP) from Ensembl. Genes with intragenic index variants are indicated with an asterisk (*).

*Positions are based on build 37.

*Effect is in mmHg unit.

*The most significant p value (between 1 df interaction test and 2 df joint test) is indicated with an asterisk (*).

*These loci indicate “completely novel” loci, at least 1 Mb away from any of known BP loci.
The majority of the available data on tissue expression are 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 more than 5,000 individuals of the Framingham Heart Study. 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 more than 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 separately by the four combinations of ancestry (EUR versus trans-ancestry) and BP trait (DBP versus SBP; Material and 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. The genes near our findings were enriched by GeneGO disease class “chronic kidney failure” (p = 9.2 × 10^-9). These same genes were also included in the much larger network representing the GeneGO disease class “fibrosis” (p = 3.39 × 10^-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%), AT2B1 (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%), AT2B1 (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%), AT2B1 (0.37%), MTHFR (0.2%), SH2B3 (0.16%), TNFSF12 (0.09%), and APSB1 (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, $P_{\text{Binomial chance alone}} = 2.8 \times 10^{-6}$), the observed count is higher than that expected by chance alone (Table S27). For example, we observed 9 of 37 SNPs associated with stroke-related traits, 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. 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 effect was influenced by adjustment and/or stratification by smoking status (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, 3, and 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. 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.

Telomere Maintenance

Since telomere length shortens with successive cell divisions, it has been proposed as a reflection of biologic age. Several genes with significant association with BP have roles in telomere maintenance including TNKS, PINX1, AKTIP (Tables 2, 3, and 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 and has additionally been associated with cardiovascular disease (CVD) risk and the inflammatory biomarker, C-reactive protein. PINX1 has been previously associated with CVD, carotid artery intima-media thickness, and serum triglyceride levels, and has also been associated with obesity- and diabetes-related traits (Tables S23 and S24). AKTIP has been previously associated with stroke-related traits in other studies (Table S22).

Central Dopaminergic Signaling

Dopaminergic signaling in the kidney is known to modulate the secretion of renin and other key regulators of salt-water balance. There is evidence that central dopamine signaling also modulates BP via mechanisms that are independent of changes in sodium excretion. Early stages of Parkinson disease, a neurodegenerative disorder characterized by the loss of dopamine-secreting neurons, is characterized by autonomic dysfunction and BP dysregulation. The current study, genes involved in central dopaminergic 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 D2 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, BCA1-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). BCA1-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); a potential causal variant in a BCA1 regulatory domain has been identified. 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 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 × 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 × 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 cells62 but has not been previously identified, even though other components of RAAS including NPPA, NPPB, and SLC17A1-4 have been found in previous GWASs 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,56 LRP667 and PPP1R3B69 have been associated with serum low-density lipoprotein levels and the metabolic syndrome. MSRA70 and SERTAD271 (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 FTO73,74 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 RPRGIP1L,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 and 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,9 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 three validated BP loci,96 one of which overlaps with our newly identified loci.

35 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 (Figures S10 and 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 whether any of our loci newly identified through the 2 df 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 GCxT) 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 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. 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 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 Supplemental Notes, 17 figures, and 28 tables and can be found with this article online at https://doi.org/10.1016/j.ajhg.2018.01.015.

Conflicts of Interest
The authors declare no competing financial interests except for the following. B.M.P 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; O.H.F received grants from Metagenics (on women’s health and epigenetics) and from Nestle (on child health); L.J.B. 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; P.S. has received research awards from Pfizer Inc., is a consultant for Mundipharma Co. (Cambridge, UK), is a 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 has a 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); P.W.F. 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; M.A.N.’s participation is supported by a consulting contract between Data Tecnica Internation and the National Institute on Aging (NIH, Bethesda, MD, USA), and he also consults for Illumina, Inc., the Michael J. Fox Foundation, and University of California Health-care among others; and M.J.C. is Chief Scientist for Genomics England, a UK government company.

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, RO1HL118305. A Career Development Award (K25HL121091), also from the NHLBI, enabled Y.J.S. to lead this project. Full set of study-specific funding sources and acknowledgments appear in the Supplemental Note.

Received: October 10, 2017
Accepted: January 18, 2018
Published: February 15, 2018

Web Resources
DEPICT, https://data.broadinstitute.org/mpg/depict
GSEA, http://software.broadinstitute.org/gsea/msigdb/annotate.jsp
GTEx Portal, https://www.gtexportal.org/home/
Literature Lab, http://www.acumenta.com
LocusZoom, http://locuszoom.sph.umich.edu/locuszoom/
METAL, http://genome.sph.umich.edu/wiki/METAL_Documentation
RegulomeDB, http://RegulomeDB.org/
Roadmap, http://www.roaddmapepigenomics.org/

References


