African ancestry genome-wide association study of blood pressure and hypertension identifies 25 novel loci through predicted gene expression

Jacklyn Hellwege (jacklyn.hellwege@vumc.org)
Vanderbilt University Medical Center https://orcid.org/0000-0001-7479-0920

Jingjing Liang
Case Western Reserve University

Ayush Giri
Vanderbilt University Medical Center

Jacob Keaton
NHGRI

Xiao Sun
Tulane University School of Public Health and Tropical Medicine

Eric Torstenson
Vanderbilt University Medical Center

Megan Shuey
Vanderbilt University Medical Center

Otis Wilson
Vanderbilt University Medical Center

Helen Warren
William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London https://orcid.org/0000-0003-0511-1386

Donna Arnett
University of Kentucky College of Public Health https://orcid.org/0000-0003-2219-657X

Joshua Bis
Cardiovascular Health Research Unit https://orcid.org/0000-0002-3409-1110

Brian Cade
Brigham & Women's Hospital

Adolfo Correa
University of Mississippi Medical Center https://orcid.org/0000-0002-9501-600X

Paul de Vries
Human Genetics Center

Myriam Fornage
University of Texas Health Science Center at Houston https://orcid.org/0000-0003-0677-8158

Xiuqing Guo
The Institute for Translational Genomics and Population Sciences https://orcid.org/0000-0002-5264-5068
Marguerite Irvin
University of Alabama at Birmingham School of Public Health
Sharon Kardia
University of Michigan
Eimear Kenny
Icahn School of Medicine at Mount Sinai
Charles Kooperberg
Fred Hutchinson Cancer Research Center https://orcid.org/0000-0002-7986-8560
Ruth Loos
Icahn School of Medicine at Mount Sinai https://orcid.org/0000-0002-8532-5087
JoAnn Manson
Harvard Medical School
Alanna Morrison
The University of Texas Health Science Center at Houston https://orcid.org/0000-0001-6381-4296
Lawrence Phillips
Atlanta VA Medical Center https://orcid.org/0000-0002-6542-8046
Bruce Psaty
Cardiovascular Health Research Unit https://orcid.org/0000-0002-7278-2190
Jennifer Smith
University of Michigan https://orcid.org/0000-0002-3575-5468
Laura Raffield
University of North Carolina https://orcid.org/0000-0002-7892-193X
Susan Redline
Harvard University
Jerome Rotter
The Institute for Translational Genomics and Population Sciences
Dhananjay Vaidya
Johns Hopkins University School of Medicine https://orcid.org/0000-0002-7164-1601
Lisa Yanek
Johns Hopkins University School of Medicine https://orcid.org/0000-0001-7117-1075
Nora Franceschini
University of North Carolina at Chapel Hill
Patricia Munroe
Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ https://orcid.org/0000-0002-4176-2947

Peter Wilson
Atlanta VA Medical Center
Tanika Kelly
Tulane University School of Public Health and Tropical Medicine
Christopher O'Donnell
U.S. Department of Veterans Affairs https://orcid.org/0000-0002-2667-8624
Yan Sun
Emory University https://orcid.org/0000-0002-2838-1824
Adriana Hung
Division of Nephrology and Hypertension, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
Xiaofeng Zhu
Case Western Reserve University https://orcid.org/0000-0003-0037-411X
Todd Edwards
Vanderbilt University Medical Center

Article

Keywords:

Posted Date: March 9th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1392991/v1

African ancestry genome-wide association study of blood pressure and hypertension identifies 25 novel loci through predicted gene expression

Jacklyn N. Helllwege[1,2], Jingjing Liang[3], Ayush Giri[2,4], Jacob M. Keaton[5,6], Xiao Sun[7], Eric S. Torstenson[2,8], Megan M. Shuey[1], Otis Wilson[2,9], Helen R. Warren[10,11], Donna K. Arnett[12], Joshua C. Bis[13], Brian E. Cade[14,15], Adolfo Correa[16], Paul S. de Vries[17], Myriam Fornage[17,18], Xiuqing Guo[19], Marguerite R. Irvin[20], Sharon L.R. Kardia[21], Eimear E. Kenny[22,23,24], Charles Kooperberg[25], Ruth J.F. Loos[26,27], JoAnn E. Manson[28], Alanna C. Morrison[17], Lawrence S. Phillips[29,30], Bruce M. Psaty[13], Jennifer A. Smith[21], Laura M. Raffield[31], Susan Redline[15,32] , Jerome I. Rotter[19], Dhananjay Vaidya[33], Lisa R. Yanek[33], Nora Franceschini[34], Patricia B. Munroe[10,11], Peter W.F. Wilson[30,35], Tanika N. Kelly[7], Christopher J. O'Donnell[36,37], Yan V. Sun[30,38], Adriana M. Hung[2,9†] , Xiaofeng Zhu[3†], Todd L. Edwards[2,5†Δ], on Behalf of the VA Million Veteran Program, the COGENT Consortium, and with contributions from the TOPMed Blood Pressure working group.

† Indicates that these authors share the senior author position

Δ Corresponding author

Correspondence should be addressed to:
Todd L. Edwards
Office 615, Suite 600
2525 West End Ave.
Vanderbilt University Medical Center
Nashville, TN 37204
todd.l.edwards@vumc.org

1. Division of Genetic Medicine, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA.
2. Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, Tennessee, USA.
3. Department of Population and Quantitative and Health Sciences, Case Western Reserve University, Cleveland, OH.

4. Division of Quantitative Sciences, Department of Obstetrics & Gynecology, Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, Tennessee, USA.
5. Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA.
6. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA.
7. Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA.
8. Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.
9. Division of Nephrology and Hypertension, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA.
10. Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.
11. NIHR Barts Cardiovascular Biomedical Research Centre, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.
12. College of Public Health, University of Kentucky, Lexington, KY.
13. Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, WA, USA.
14. Division of Sleep and Circadian Disorders Brigham and Women's Hospital, Boston, MA.
15. Division of Sleep Medicine, Harvard Medical School, Boston, MA.
16. Department of Population Health Science, University of Mississippi Medical Center, Jackson, MS.
17. Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX.

18. Institute of Molecular Medicine McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX.
19. The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA.
20. Epidemiology, University of Alabama at Birmingham, Birmingham, AL.
21. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI.
22. Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY.
23. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.
24. Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.
25. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center.
26. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.
27. Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
28. Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.
29. Division of Endocrinology and Metabolism, Department of Medicine, Emory University School of Medicine, Atlanta, GA.
30. Atlanta VA Health Care System, Decatur, GA.
31. Department of Genetics, The University of North Carolina at Chapel Hill, Chapel Hill, NC.
32. Department of Medicine, Brigham and Women's Hospital, Boston, MA.
33. Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

34. Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA.
35. Emory Clinical Cardiovascular Research Institute, Atlanta, Georgia, USA.
36. VA Boston Health Care System, Boston, Massachusetts, USA.
37. Department of Medicine, Harvard Medical School, Boston, MA.
38. Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA.

ABSTRACT

African Americans (AAs) experience a high burden of hypertension but have been underrepresented in genetic studies of blood pressure. We performed common and rare variant genome-wide association studies (GWAS) of systolic (SBP) and diastolic (DBP) blood pressure, pulse pressure and hypertension in 95,457 AAs from the Million Veteran Program and Continental Origins and Genetic Epidemiology Network (COGENT) consortium with replication in up to 41,014 African-ancestry participants. GWAS meta-analysis confirmed 26 previously reported loci including 15 with SBP, four with DBP, 10 with pulse pressure and five with hypertension. Predicted gene expression analysis identified 25 genes newly associated with blood pressure. We provide validated African-ancestry polygenic scores for SBP and DBP that are strongly associated with hypertension. This work provides evidence for similar genetic architectures of hypertension and blood pressure across racial/ethnic groups and demonstrates the utility of predicted expression analysis in identifying novel genes beyond GWAS alone.

Introduction

Decades of scientific evidence implicate elevated blood pressure in the etiology of cardiovascular disease (CVD), including coronary artery disease, peripheral arterial disease, and stroke, as well as renal and ocular damage. Elevated blood pressure accounts for at least 13% of annual deaths worldwide.[1,2] Risk of death from ischemic heart disease and stroke increases linearly with systolic blood pressure (SBP) and diastolic blood pressure (DBP) elevations greater than 115 mmHg and 75 mmHg, respectively.[3] In the United States, 34% and 44% of adults of European and African ancestry, respectively, have hypertension, which rank among the highest prevalence rates of hypertension among ancestrally and geographically defined world populations.[4,5] Hypertension observed in African Americans is generally more severe, develops earlier, and leads to increased rates of cardiovascular disease (CVD) compared to European Americans[6]. Specific observations include a two-fold increase in risk of stroke and three-fold increase in risk of death among African Americans compared to European Americans.[7,8] Additionally, the progression from pre-hypertension to hypertension is accelerated in African American adults by an average of one year,[9] and African Americans with pre-hypertension also have a higher burden of CVDs than European Americans.[10] The higher rates of CVD among African Americans compared to European Americans are explained to some extent by higher burdens of obesity, diabetes (T2D), and other CVD risk factors, in addition to hypertension.[11] However, the adjusted cumulative incidence of hypertension is still higher in African Americans than European Americans across levels of body mass index (BMI, kg/m^2).[12]

The canonical risk factors for hypertension traits have been identified primarily in European- and Asian-ancestry participants and include physical inactivity, diet, obesity, smoking, high alcohol consumption, and hypertension-increasing alleles.[13-16] Recent efforts to determine reasons for disparities between African- and European-ancestry individuals in the prevalence of hypertension and related adverse sequelae have included assessment of the impact of socioeconomic status, discrimination, and systemic racism.[17-25] However, exposure rates and strength of effects of well-recognized risk factors, together with socioeconomic factors, do not fully explain the associated disparities.

Large-scale genome-wide association studies (GWAS) have reported 3,800 independent SNP signals associated with blood pressure traits (SBP, DBP, and pulse pressure), establishing that the genetic architecture of blood pressure is complex with many genetic determinants of modest effect.[15,26-47] The few GWAS, that have been published in African ancestry populations,[26,28,29,38,48], have relatively small sample sizes and have identified a small number of novel loci.

In this study, we have combined evidence for association from the Million Veteran Program (MVP) and Continental Origins and Genetic Epidemiology Network (COGENT) consortium to study common and rare genetic determinants of BP in 95,457 African-ancestry participants (Figure 1). Additionally, we evaluated replication in up to 41,014 African-ancestry participants from five cohorts. Meta-analysis in discovery and replication samples detected 26 significantly associated loci, of which 15 loci associated with SBP, four loci with DBP, ten loci associated with pulse pressure, and five loci with hypertension. Additionally, we evaluated associations with genetically predicted gene expression (GPGE) using S-PrediXcan[49] and combined association evidence across traits using CPASSOC.[50] We further developed and validated African-specific polygenic scores for blood pressure that demonstrate substantial associations with blood pressure values between top and bottom deciles of the polygenic scores in independent data.

RESULTS

This study compiled discovery resources totaling 95,457 participants of African ancestry with blood pressure measurements and genome-wide genotype data (Supplementary Table 1). Generally, although MVP is predominantly male, other cohorts are more balanced (replication consisted of 43% male), resulting in discovery and replication combined near 60% male. Mean age ranged across studies from 44-61 years, while mean BMI was similar across studies, ranging from 29-31.

Single trait GWAS Results

In discovery-stage meta-analyses, 22 loci were identified at genome-wide significance for SBP (Supplementary Table 2a), and 20 significant loci were detected for DBP (Supplementary Table 2b). There were eight significant loci identified for pulse pressure and seven for HTN in discovery analyses (Supplementary Tables 2c and 2d,

respectively). SNP-based heritability of blood pressure phenotypes in discovery analysis as estimated with LD Score Regression was 0.065 (SBP), 0.067 (DBP), 0.081 (pulse pressure) and 0.059 (HTN). LD intercepts indicated that inflation was not due to population stratification (LD intercepts of 1.035, 1.027, 0.98, and 1.001 for SBP, DBP, pulse pressure, and HTN, respectively).

Replication of suggestive ($p<1x10^{-6}$) common variants was performed in up to 27,301 additional African-ancestry participants from MVP, UK Biobank, and BioVU (Supplementary Table 1). In total, 112 common sentinel SNPs were evaluated for replication across all four phenotypes. We report a tiered system of replication where tier 1 represents the most stringently replicated variants ($p<5x10^{-8}$ in discovery, $p<0.05/N$ SNPs in replication, and $p<5x10^{-8}$ in meta-analysis with consistent direction), tier 2 replication is defined by $p<5x10^{-8}$ in discovery, $p<0.05$ in replication, and $p<5x10^{-8}$ in meta-analysis with consistent direction, and tier 3 represents SNPs which were suggestive in discovery but with addition of replication became statistically significant after meta-analysis ($p<1x10^{-6}$ in discovery, $p<0.05$ in replication, and $p<5x10^{-8}$ in meta-analysis with consistent direction). All tiers of SNPs represent statistically significant results following replication and meta-analysis. Replication results are summarized in Supplementary Table 2a-d.

Among 41 variants (p<1e-6) carried forward from discovery analysis for replication with SBP, 15 replicated significantly: two in tier 1, nine in tier 2, and four in tier 3, with the strongest association at rs3821845 in *CACNA1D* on chromosome 3 (p-value = $9.94x10^{-19}$, effect allele frequency [EAF] = 0.24; Table 1, Supplementary Table 2a). Of 29 variants evaluated in replication for DBP, four replicated significantly: two in tier 1 and two in tier 2 (Supplementary Table 2b). A different sentinel SNP, rs9821489, in *CACNA1D* ($r^2 = 0.272$ with rs3821845) was the strongest association for DBP as well (p-value=$1.13x10^{-18}$; Supplementary Table 2b). Ten of 33 variants from pulse pressure discovery analyses replicated significantly: five in tier 1, one in tier 2, and four in tier 3 (Supplementary Table 2c). The SNP rs7006531 near *CDH17* was the strongest association for pulse pressure (p-value = $4.56x10^{-32}$, EAF = 0.14; Supplementary Table 2c). For HTN, of 24 candidate variants evaluated, five replicated significantly, two in tier 1, two in tier 2, and one in tier 3 (Supplementary Table 2d). An intronic SNP

rs17035646 in *CASZ1* was the lead signal for HTN (p-value = 4.20×10^{-16}, EAF = 0.18; Supplementary Table 2d).

We used a joint conditional analysis implemented in GCTA to assess novelty of these results against 3800 previously identified SNPs associated with blood pressure (Supplementary Table 3). This conditional analysis of discovery results identified seven secondary signals (six unique SNPs) across the four phenotypes, however, none of these SNPs replicated (Supplementary Table 4).

Pathway Analysis

We used DEPICT to identify significantly enriched tissues (Supplementary Tables 5a-d) and pathways (Supplementary Tables 6a-c) for any of the four phenotypes using all discovery sentinel variants with $p<10^{-6}$. For pulse pressure, several cell types were implicated in the tissue enrichment analysis (Mesenchymal Stem Cells, Osteoblasts, Connective Tissue Cells, Stromal Cells, and Fibroblasts) using a false discovery rate of 5%. For the pathway gene set enrichment, 145 pathways were significant only for DBP. The top result was XY body, followed by EDC4 subnetwork and UBE2L3 subnetwork.

Multi-trait GWAS Results

Combination of phenotype results in a multi-trait approach using CPASSOC[50] identified a large number of significant variants. Specifically, 9 loci that did not attain genome-wide significance in discovery analysis of any trait individually (Supplementary Table 7). Of these, 7 were independent of any of the 3,800 previously identified sentinel lead or secondary blood pressure SNPs (Supplementary Table 3; $r^2<0.1$). Eight of the nine sentinel SNPs were identified using the model assuming heterogeneity between phenotypes, while one SNP was significant in the model assuming homogeneity between phenotypes: rs17429177 on chromosome 7 between *EVX1* and *HIBADH*, which was in LD with a known locus. Only one SNP was directly confirmed by multi-trait analysis of replication results, rs143320468 near the previously identified locus *CDH17*. Although not replicated in multi-trait analysis, rs10898966 near *PPME1* attained genome-wide significance in meta-analysis of the discovery and replication phases for pulse pressure (Supplementary Table 2c). Another three were not sentinel candidates

for single-trait common SNP variation, but upon investigation into multi-trait results were found to also surpass genome-wide significance with a single trait in the combined discovery and replication results (Supplementary Table 7).

Rare Variants Analysis

Discovery analysis of variants with minor allele frequency less than 1% was performed in MVP alone, as COGENT data for low frequency SNPs was unavailable. A total of three low frequency variants (MAF<1%, effective minor allele count≥20) were suggestively ($p<10^{-6}$) associated with SBP, seven with DBP, 19 with pulse pressure, and 11 with HTN (Supplementary Figure 1a-d; Supplementary Table 8). All suggestive rare variants ($p<10^{-6}$) were evaluated for replication in up to 41,014 participants from four cohorts of recent African-ancestry – those included in common variant replication with the addition of TOPMed. TOPMed samples were included only in rare variant analyses due to extensive sample overlap with COGENT. Among the 53 suggestive variants from MVP, 40 were available in at least one replication cohort. Four SNPs had p-values less than 0.05 though none met Bonferroni correction for the number of tests performed. Intergenic SNP rs149762703 (near *LINC02301*) was suggestively associated with HTN in replication ($p_{replication}=0.037$) and meta-analysis of discovery and replication indicated consistent directions of association, though not quite attaining genome-wide significance ($p_{meta}=2.68x10^{-7}$; risk allele [A] Odds Ratio = 2.09 [95% Confidence interval 1.58-2.76], EAF=0.0033).

Genetically-predicted gene expression

Discovery and Replication meta-analysis results for common variants were used to evaluate the associations between BP traits and GPGE levels across five cardiovascular-related tissues (whole blood, heart-left ventricle, heart-atrial appendage, aorta, and tibial artery) from Gene Tissue Expression Project (GTEx) v7 using an African-ancestry covariance matrix developed in independent samples to account for population LD and S-PrediXcan software. Colocalization was confirmed at a majority of results. We detected significant ($p<3.5x10^{-6}$) GPGE associations for 18 gene-tissue pairs with SBP, 19 with DBP, 49 for pulse pressure, and three with hypertension (Supplementary Tables 9a-d, Figure 3a-d). These findings represent 59 distinct genes, of which 52 also significantly colocalized. A total of 25 have not been definitively linked

to blood pressure in the literature (Table 1). Nine genes were significantly associated with two or more traits. Predicted increased expression of *HOXA2* in Tibial Artery was significantly associated with decreases in all three BP phenotypes, with the largest effects seen with SBP (-0.56 mm Hg per standard deviation of *HOXA2* transcript level; Figure 4). Predicted increased *SH3PXD2A* expression in heart-left ventricle tissue was associated with both decreased SBP and DBP (Figure 5). Additional associations across multiple phenotypes were detected with predicted *CCDC71L* in aorta with increased SBP and pulse pressure, predicted *BCAR1* in tibial artery (increased SBP and pulse pressure), aorta and whole blood (pulse pressure), *ULK4* in tibial artery (DBP) and whole blood (DBP and pulse pressure) (increased DBP, decreased pulse pressure), *CTSW* in left ventricle (decreased SBP and pulse pressure), *PRSS45* in aorta (increased DBP and risk of HTN) , *RP11-259G18.3* in tibial artery (decreased DBP and SBP), and *SNX32* in left ventricle (increased SBP and pulse pressure).

Polygenic Scores of Blood Pressure

Polygenic scores were developed for DBP and SBP, derived from the discovery African-ancestry summary statistics (Supplementary Tables 10a and b). The African-ancestry-derived scores had optimal p-value thresholds of 0.1 for both SBP and DBP as determined in the BioVU replication set (Supplementary Figure 2). The PRS explained 1.1 and 0.9 percent of the variance of SBP and DBP, respectively, in BioVU. Validation in UKB indicated higher percentages of variance explained (1.5 and 1.3 percent, for SBP and DBP respectively; Supplementary Table 11). In parallel, we evaluated the polygenic score catalog[51] scores for SBP and DBP[52] which are derived from European-ancestry genome-wide significant results[47]. The European-derived polygenic scores were also significantly associated with blood pressure in BioVU African-ancestry individuals, with effect estimates per standard deviation of PRS being consistent for DBP scores (3.41 mmHg [p-value = 2.4×10^{-7}] and 3.42 mmHg [p-value = 9.31×10^{-7}] for African- and European-derived scores respectively). The African-derived SBP score however had a much larger effect (8.07 mmHg [p-value = 9.5×10^{-14}]) than the European-derived score (4.15 mmHg [p-value = 4.4×10^{-16}]), though both were highly significant.

Comparison of top to bottom deciles for African-ancestry derived polygenic scores in UKB African-ancestry individuals represented an increase of 7.66 mmHg for SBP and 4.35 mmHg for DBP (Supplementary Figure 3). Deciles of each score also showed increasing association with hypertension status, with the top decile relative to the bottom decile having odds ratios of 4.46 (95% Confidence Interval [CI]: 3.04-6.55; p-value = 1.99×10^{-14}) and 3.31 (95% CI 2.26-4.86; p-value = 9.28×10^{-10}), for SBP and DBP scores, respectively, when modeled separately adjusting for age, age squared, sex, BMI, and proportions of ancestry (Figure 6; Supplementary Table 12). The joint effect of both African-ancestry derived scores (SBP and DBP PRS) in the same model had a stronger relationship than each score individually, with a top-relative-to-bottom-decile odds ratio of 5.73 (95% CI 5.29-6.16; Figure 6, Supplementary Table 12). Performance of the model including both scores and basic covariates trained in BioVU and applied in UKB revealed an area under the receiver-operator curve (AUC) of 0.720 (95% confidence interval: 0.708-0.731; Supplementary Figure 4), an increase of 0.01 from the covariate-only model (AUC = 0.710).

DISCUSSION

The results of this analysis provide additional insight into the genetic and biological architecture of blood pressure traits in African ancestry populations. This population remains relatively understudied despite experiencing disproportionate health impacts of conditions related to high blood pressure. We leveraged the GWAS results to observe evidence of associations between expression of 25 previously unreported genes and blood pressure traits. Many of these genes exist in or are nearby regions that were mapped to blood pressure traits in prior studies but were not annotated in those reports or in the GWAS catalog. Additionally, we provide validated polygenic scores for predicting blood pressure traits, with base, target, and validation steps performed in recent African ancestry populations.

In the GWAS stage, we observed and replicated 35 sentinel loci. Of these, only one SNP was conditionally independent compared with prior GWAS sentinel SNPs. Previous GWAS of blood pressure have approached sample sizes of 1 million participants and identified thousands of associated signals. The estimated chip-based

heritability of BP traits from these analyses ranged between six and eight percent which is substantially less than the heritability estimated by family studies and larger SNP-based studies. Additionally, the substantial difference between LDSC intercepts and λ_{GC} indicates that blood pressure traits are likely very polygenic in African ancestry populations. Together, these facts suggest that the genetic architecture of blood pressure traits is highly complex and dominated by subtle effects in recent African ancestry populations as has been observed in European and East Asian populations.

We leveraged the GWAS results to identify genes that are involved in blood pressure traits. We used a combination of S-PrediXcan to detect GPGE associations and estimate effects and COLOC to mitigate LD contamination, where the causal variants for expression and trait association are in LD but are distinct loci. We detected and colocalized 25 genes, pseudogenes, and non-coding RNAs that were previously unreported in the GWAS literature (Table 1).

The 25 unreported gene regions including the genes and 50kb flanking sequences contain previously reported associations with several non-blood pressure traits that may offer some insight into mechanisms of influence on blood pressure. We observed relationships between genes associated with blood pressure and prior findings from studies of both psychiatric and physical traits. Twelve of these genes are at or nearby previous associations with various blood cell and composition traits. *TRIP4*, *ALS2CL*, *RP11-259G18.3*/*CCDC36*, *SMIM4*, *IFT52*, *HSD17B6 TPD52L1*, *HEMK1, HOXA2*, *SH3PXD2A*, and *SDCCAG3* are all in genomic regions with statistically significant associations with one or more blood cell labs or blood protein levels[53-59]. The *RP11-890B15.3*, *PPM1M, HSD17B6*, *HOXA2* and *HEMK1* regions have been associated with BMI, adiposity measures, or waist-hip ratio adjusted for BMI[60-63]. *TIGD7* and *HSD17B6* are nearby associations with bone mineral density[64].

Several blood pressure genes are nearby loci associated with various psychiatric traits. The *MBTPS1* gene region has been associated with schizophrenia, while *RP11-3B7.1*/*CCDC36*, and *PPM1M* have been associated with neurotic behaviors[65-68]. *RP11-3B7.1*/*CCDC36,* and *KDELR3* were also nearby associations with sleep quality or duration[69-71]. Additionally, *CCDC36* and *SH3PXD2A* are in regions associated with educational attainment[72-74]. The alcohol dehydrogenase 1a (*ADH1A*) gene region has

been associated with alcohol dependency[75], and another gene in this pathway *ALDH2* has been associated with blood pressure traits and has been under strong recent positive selection in East Asian populations[15]. The *TIGD7, HSD17B6,* and *AC005022.1* regions have also been associated with smoking behaviors[76].

The *RP11-819C21.1* region is nearby an association with use of beta blockers[77], which is a class of drug sometimes used in combination with ACE inhibitors to control blood pressure in African ancestry patients with chronic kidney disease or myocardial infarction[78]. *AC005022.1* is also nearby a renal function sentinel SNP[79]. Additionally, *SH3PXD2A* is nearby a sentinel locus for atrial fibrillation[80] and the *KDELR3* region is associated with Electrocardiogram T-peak to T-end[81], suggesting a role for heart rhythm and contractility in African ancestry blood pressure traits. *SH3PXD2A* is also within loci associated with intraocular pressure, cerebral small vessel disease, and white matter hyperintensities, all closely related to blood pressure levels[82-84]. *MBTPS1* has also been implicated in stroke[85].

The *HOXA2* and previously reported *HOXA7* genes are located in a gene cluster with several homeobox genes. This region was first mapped to blood pressure traits by the COGENT consortium in Franceschini et al[28], and later observed in a transethnic study including African American individuals by Giri et al[29] and in a Japanese population by Takeuchi et al[86]. These studies identified this region, but substantial uncertainty remained about what the causal gene(s) were for blood pressure. In this study, we report evidence that this association is driven at least in part by regulatory effects on *HOXA2* and *HOXA7* expression in African ancestry individuals (Figure 4).

Another aspect of this work is the development of externally validated polygenic risk scores to predict blood pressure levels and hypertension prevalence. The polygenic score catalog[51] does not currently contain scores for SBP or DBP derived from African ancestry populations, although genetic risk scores based on sentinel SNPs do exist for European-derived populations[52]. These scores have a wide range of potential applications, including predicting hypertension and hypertension sequelae, evaluations of genetic predisposition for increasing blood pressure to other traits in populations such as pediatric that may not be accessible for large-scale genomic interrogation[87], facilitating Mendelian randomization studies[88], and phenome-wide association studies[29].

In our comparison of these African ancestry-derived PRS and European GRS[52] from the polygenic score catalog (PGS000301 and PGS000302) in African ancestry individuals, we observed that the African-ancestry derived scores explained nearly three-fold as much variance (1.1% and 0.9% for SBP and DBP, respectively compared to 0.35% and 0.32%) as the PGS catalog scores. Additionally, we observed substantial and clinically relevant changes in average blood pressures between top and bottom deciles of the SBP and DBP PRSs. The odds ratio for hypertension for the scores separately as well as jointly were also very significant compared with effects of other classical hypertension risk factors. A predictive model of hypertension including the PRSs and only basic covariates has an AUC of 0.717. These results demonstrate that the PRSs are robustly associated with the target traits, are substantive predictors of blood pressure levels and hypertension risk when compared with other important risk factors, and that even with a modest percentage of explained variance have the potential to be a useful variable in research and clinical prediction.

Our study of rare variants failed to detect any replicated associations between blood pressure traits and variants with an allele frequency less than 0.01. There were limitations to this study, such as the use of older imputation references from the 1000 Genomes project instead of TOPMed in the GWAS, and the mixture of GWAS and sequencing data in the replication stage. Despite these limitations, the data should have been sufficient to detect large effects at variants close to 0.01 frequency, with a sample size of about 100,000 across stages for most variants. We had at least 95% power to detect an effect size of 5mm Hg for SBP at a frequency of at least 0.003, 6mm Hg at a frequency of 0.002, and 9mm Hg at a frequency of 0.001 (Quanto v1.2.4). Therefore, we believe it is unlikely that very large effect alleles exist in this frequency range in the recent African ancestry population, although we acknowledge genomic coverage at these allele frequencies is likely incomplete. This suggests that intermediate-frequency variants between 0.01 and 0.001 do not substantively explain the racial disparities in blood pressure and hypertension traits such as has been reported in large studies of other populations, and that larger studies with better genomic coverage will be required to interrogate the influence of alleles with frequencies less than 0.001. African populations have more rare genetic variation than populations with other geographic

origins[89], and so it is possible that studies in these populations will produce more significant associations and insights into important genes from rare variants than similar size studies conducted elsewhere.

In our analyses we have evaluated associations between common genetic variants, rare variants, GPGE and blood pressure traits using the largest sample of African Americans to date. These analyses provided new insights into biological factors and the effect sizes at rare variants that contribute to hypertension risk in recent African ancestry populations. We also provide externally validated African ancestry PRSs for researchers to use in subsequent research and predictive models. These advances, as well as providing some resolution of the association in the HOX gene cluster, represent progress in blood pressure genetics for African ancestry populations. The results of this study also support further efforts to expand sample sizes of African ancestry blood pressure GWAS.

METHODS

Discovery Cohorts

The Million Veteran Program

The Million Veteran's Program (MVP) is a large cohort of fully consented participants who were recruited from the patient populations of 63 Veteran's Administration (VA) medical facilities[90]. Recruitment began in 2011 and is conducted in-person, which is initiated by an invitation letter and completed by answering baseline and lifestyle questionnaires, providing a blood sample, and providing access to medical records, and giving permission for re-contact. Consent to participate is provided after counseling by research staff and mailing of informational materials. All documents and protocols have been approved by the VA Central Institutional Review Board. Blood samples are collected by phlebotomists and banked at the VA Central Biorepository in Boston, MA. Genotyping was conducted using a customized Affymetrix Axiom Biobank Array chip with additional content added to provide coverage of African and Hispanic haplotypes, as well as markers for common diseases in the VA population. Researchers are provided with de-identified versions of these data, and do not have the ability or authorization to link these details with a participants' identity.

MVP Genotype QC

Blood samples drawn from consenting MVP participants were shipped to a central biorepository in Boston, Massachusetts, where DNA was extracted and shipped to two external genotyping centers for genotyping on an Affymetrix Axiom Biobank array designed specifically for the MVP. The MVP genomics working group applied standard quality control and genotype calling algorithms to the data in batches using the Affymetrix Power Tools Suite (v1.18). Standard quality control pipelines were used to exclude duplicate samples, samples with more heterozygosity than expected, or discordance genetically inferred sex versus self-report. We also excluded related individuals (halfway between 2nd and 3rd degree relatives or closer) as measured by KING software[91].

We excluded: duplicate samples, samples with more heterozygosity than expected, an excess (>2.5%) of missing genotype calls, or discordance between genetically inferred sex and phenotypic gender. In addition, one individual from each pair of related individuals (as measured by KING software[91]) were removed. Prior to imputation, variants that were poorly called or that deviated from their expected allele frequency based on reference data from the 1000 Genomes Project[92] were excluded. After pre-phasing using EAGLE v2[93], genotypes from the 1000 Genomes Project[92] phase 3, version 5 reference panel were imputed into Million Veteran Program (MVP) participants via Minimac3 software[94]. Principal component analysis was performed using FlashPCA[95], to generate top 10 genetic principal components explaining the greatest variability.

Race/ethnicity

Information on race (non-Hispanic blacks, Asians, and Native Americans) and ethnicity (Hispanic: Yes or No) were obtained based on self-report through centralized VA data collection methods using standardized survey forms, or through the use of information from corporate data warehouse (CDW), or Observational Medical Outcomes Partnership (OMOP) data, when information from self-report survey was missing. Race and ethnicity categories were then merged to form the following administratively assigned race/ethnicity variables: non-Hispanic whites, non-Hispanic blacks, non-Hispanic Asians, non-Hispanic Native Americans and Hispanics. Individuals for whom

race and ethnicity could not be confidently assigned due to conflicting records and/or missing data, race/ethnicity category was set to unknown. Prior to analysis QC, there were 15,710 with unknown status for race/ethnicity. For these individuals, we used a K-means clustering approach in R following the McQueen algorithm with top 10 genetic principal components as input variables. In order to obtain the most reliable cluster designations for the missing data, the k-means approach was applied to the maximum available samples: the 1000 Genomes reference populations and all individuals for whom PCs were available regardless of whether race/ethnicity designations were unknown. K-clusters were optimized by testing values K=2 through K=10. K = 4 was ultimately chosen as the most optimal value, as visual examination of these most closely corresponded to non-Hispanic whites (N=5,265), non-Hispanic blacks (N=4,671), Asians (N= 3,936) and Hispanics (N= 1,838). Only non-Hispanic blacks were included in this analysis.

MVP BP Phenotypes

We selected adults (age ≥ 18) and used the earliest median eligible non-Emergency Department outpatient measured SBP in the EHR, and also used the corresponding DBP from this measure. Measures are ineligible if they occur at or after an ICD-9 code from the groups 585, 405, or 428. If pain scores were available, we censored BP measures taken during encounters when a pain score ≥ 5 was recorded, because severe pain can elevate BP[96,97]. For measures taken while a patient was on an antihypertensive medication we added 15 mmHg to SBP and 10 mmHg to DBP[40,98].

MVP Analysis

For the MVP GWAS we performed linear regression association tests with additive models for untransformed BP traits, after adjusting for medication use. We adjusted linear regression models analyzing SNP associations for age at BP measure, age^2, sex, BMI measured within 1 year of BP measure, and top 10 genetic principal components in analyses. All primary analyses for the MVP were conducted by strata of administratively assigned race/ethnicity or by their empirically designated clusters. All regression based analyses were conducted in SNPTEST-v2.5.4-beta[99]. Inference was limited to genotyped and imputed variants with SNPTEST Info scores of 0.4 or higher, with Hardy

Weinberg equilibrium p-value > $5x10^{-8}$ for common variant analysis (minor allele frequency > 0.1).

Continental Origins and Genetic Epidemiology Network (COGENT) consortium

COGENT consists of 19 studies (n = 29,378 subjects) with GWAS level data which has been previously reported[28,38]. Each study followed protocols for phenotype harmonization. Each cohort was genotyped on either Affymetrix or Illumina genotyping platforms. Pre-imputation quality criteria were applied, and individuals with discordant self-reported gender and genetic gender were excluded. Imputation was performed using the software MACH-ADMIX, MACH-minimac or IMPUTE2[100] using the Phase 1 integrated (March 2012 release) multi-ethnic reference panel from the 1000G Consortium[92]. Autosomal chromosome SNP associations for SBP, DBP, and pulse pressure were assessed by linear regression for unrelated data or by the generalized linear mixed-effects model for family data, under the assumption of an additive genetic model. Analysis of hypertension used logistic models. All models were adjusted for age, age^2, sex, and body mass index. Up to ten principal components were included, as needed as covariates in the regression models, to control population stratification[101,102]. We used standardized pre-meta-analysis QC criteria for all 19 discovery studies[103]. At the SNP level, we excluded variants with 1) imputation quality $r^2 < 0.3$ in MACH or <0.4 in IMPUTE2; 2) the number of informative individuals ($2\times MAF\times N\times r^2$) ≤ 30; 3) an effect allele frequency (EAF) difference larger than 0.3 in comparison with the mixture of 80% YRI and 20% CEU of 1000G; and 4) the absolute regression coefficient ≥ 10. SNPs that passed the QC were carried forward for inverse variance weighted meta-analyses, implemented in METAL[104]

Meta-analysis of discovery datasets

Inverse-variance weighted fixed-effects meta-analysis of common variants across MVP subsets and summary statistics from COGENT was performed using the METAL software[104]. Genomic inflation factors were calculated, and λ_{GC} for the discovery from MVP were 1.195, and 1.053 for SBP and DBP, respectively, 1.303 and 1.315, respectively, from COGENT, and 1.275 and 1.140, respectively, in the overall discovery analysis.

Selection of SNPs for Replication

We considered for follow-up sentinel SNPs with meta-analysis p-value < $1x10^{-6}$ for any BP traits. *In silico* replication summary statistics were requested for 4,578 SNPs from five studies of blood pressure phenotypes. Rare variants (MAF<0.01) were extracted from MVP-only discovery summary statistics as COGENT did not analyze low-frequency variants. Due to sample overlap between the COGENT contributing consortia and those contributing to the TOPMed dataset, TOPMed results were only contributed to replication meta-analyses of rare variants. Discovery and Replication data were combined using fixed-effects inverse-variance weighted meta-analysis implemented in METAL[104].

MVP Phase 3

MVP Phase 3 data was generated in the same manner as described above for an additional 10,392 samples of recent African ancestry.

BioVU

The BioVU DNA Repository is a deidentified database of electronic health records (EHR) that are linked to patient DNA samples at Vanderbilt University Medical Center. A detailed description of the database and how it is maintained has been published elsewhere[105]. BioVU participant DNA samples were genotyped on a custom Illumina Multi-Ethnic Genotyping Array (MEGA-ex; Illumina Inc., San Diego, CA, USA). Quality control included excluding samples or variants with missingness rates above 2%. Samples were also excluded if consent had been revoked, sample was duplicated, or failed sex concordance checks. Imputation was performed on the Michigan Imputation Server (MIS) v1.2.4[94] using Minimac4 and the Haplotype Reference Consortium (HRC) panel v1.1[106].

Among BioVU participants, we selected unrelated self-reported Black adults (age ≥ 18) and used the earliest median eligible non-Emergency Department outpatient measured SBP in the EHR, and the corresponding DBP. Measures were considered ineligible if they occurred at or after an ICD-9/10 billing code from the groups 585/N18, 405/I15, or 428/I50. For measures taken while a patient was on an antihypertensive medication we added 15 mmHg to SBP and 10 mmHg to DBP. Electronic health records were used to define HTN status through presence of an HTN ICD-9 code, treatment with an

antihypertensive drug, or having two SBP measures >140 mmHg and/or two DBP measures >90 mmHg.

We performed linear regression association tests with additive models for untransformed BP traits, after adjusting for medication use. We adjusted linear or logistic regression models analyzing SNP associations for age at BP measure, age^2, bmi, sex, and the top 10 genetic principal components using SNPTEST-v2.5.4-beta[99]. Inferences were limited to genotyped and imputed variants with imputation info scores of 0.4 or higher, Hardy-Weinberg equilibrium p-values >$5x10^{-8}$, and minor allele frequencies >0.01. Blood pressure and hypertension phenotype data were available from 9,268 self-reported black participants with MEGA genotyping.

UK Biobank

SNPs were imputed centrally by UKB using a reference panel that merged the UK10K and 1000 Genomes Phase 3 panel as well as the Haplotype Reference Consortium (HRC) panel. For the current analysis, only SNPs imputed from the HRC panel were considered. The mean SBP and DBP values were determined from available blood pressure measurements, and when only one blood pressure measurement was available, we used this single value. We adjusted for medication use by adding 15 and 10 mm Hg to SBP and DBP, respectively. We performed genome-wide analysis separately in reported “Black” and “Caribbean Black” UKB participants using linear regression for untransformed BP traits after adjusting for medication use and logistic regression for hypertension status. For each dataset, we adjusted regression models analyzing additively modeled SNP associations for age at BP measure, age^2, BMI, sex, and the top 10 genetic principal components using SNPTEST-v2.5.4-beta. Inferences were limited to genotyped and imputed variants with imputation info scores of 0.4 or higher, Hardy-Weinberg equilibrium p-values >$5x10^{-8}$, and minor allele frequencies >0.01.

TOPMed

The TOPMed consortium study consists of diverse, well-phenotyped epidemiologic cohorts with whole genome sequencing (WGS) data. Details of TOPMed and analysis

of blood pressure are described elsewhere[107-109]. Briefly, single variants with minor allele count (MAC) ≥10 were tested for association with SBP, DBP, pulse pressure, and hypertension in ancestry-stratified analyses of 13,713 participants from 11 studies from the TOPMed WGS program freeze 6: Atherosclerosis Risk in Communities (ARIC) Study, The Cardiovascular Health Study (CHS), The Cleveland Family Study (CFS), Coronary Artery Risk Development in Young Adults (CARDIA), Genetic Epidemiology Network of Arteriopathy (GENOA), Genetic Studies of Atherosclerosis Risk (GeneSTAR), Hypertension Genetic Epidemiology Network (HyperGEN), Jackson Heart Study (JHS), Multi-ethnic Study of Atherosclerosis (MESA), Women's Health Initiative (WHI), and The Mount Sinai BioMe Biobank (BioMe).

Classifying results by evidence for association

For results that reached statistical significance of p-value ≤ $5x10^{-8}$ at any stage of the analysis, and that had consistent direction of effect between discovery and replication stages, we established three tiers of evidence that are annotated in results tables:

1) Genome-wide significance in the discovery stage, and Bonferroni-corrected significance in replication.
2) Genome-wide significance in the discovery stage, and p-value ≤ 0.05 in the replication stage.
3) All other associations reaching genome-wide significance across stages with replication p-value ≤ 0.05.

Conditional Analysis

For conditional analysis of common variants we used two parallel approaches implemented in the Genome-wide Complex Traits Analysis (GCTA) software[110], : (i) genome-wide joint conditional analysis; and (ii) locus-specific conditional analysis.

(i) Genome-wide joint conditional analysis

Conditional analysis was conducted within GCTA software, using the *–cojo* method, which performs iterative conditional and joint analysis simultaneously with stepwise model selection[111]. The summary statistics from the GWAS discovery meta-analysis of MVP and COGENT were used as the input

summary data (separately by trait), and the imputed (info score ≥0.4), hard-called BioVU AA genetic data (N = 19,726) was used as the reference genotype-level data, in PLINK format. Combination of these two input data files restricted the GCTA analysis to the imputed SNPs in common to the GWAS discovery meta-analysis (which was itself restricted to MAF > 1%). Within the BioVU genetic data, LD was calculated between all pairwise SNPs. A p-value cut-off of 5×10^{-8} was used as the selection threshold within GCTA, and the collinearity threshold was set at the default value of 0.9, so that SNPs are not selected if the multiple regression with the current SNPs in the model has $R^2 \geq 0.9$. After combining results across all 22 chromosomes, each trait-specific analysis resulted in a distinct set of jointly independent significant signals. We then merged together genome-wide results across all four BP traits to exclude signals duplicated across traits. For SNPs in LD ($r^2 \geq 0.1$), we selected the most significant SNP with the minimum p-value across all BP traits from the GCTA joint model. Hence all final SNPs are pairwise-LD-independent.

(ii) Locus-specific conditional analysis

Within each of the Tier 1 or 2 loci (Supplementary Tables 2a-d), we searched separately by trait for any potential secondary signals, which are independently associated in addition to the sentinel SNP. Tier 3 was excluded due to not attaining genome-wide significance in the discovery analysis, which was the basis for performing the conditional analyses. Each conditional analysis was performed across all imputed SNPs with MAF ≥ 1% within the 1Mb locus region centered ±500kb around the sentinel SNP, conditioning on the sentinel SNP.

We evaluated independence of our results compared to 3,800 previously reported BP sentinel SNPs from lead or secondary loci. For loci containing only one SNP, the 1Mb locus region centered ±500kb around the SNP was used for analysis. For loci containing multiple identified SNPs, the interval

was wider than 1Mb, with the locus region starting 500kb downstream from the first SNP and ending 500kb upstream from the last SNP. For known loci containing only one sentinel SNP, conditional analysis was performed on all imputed SNPs with MAF ≥ 1% within the 1Mb region, conditioning on the single published SNP within the locus, testing for association of all three BP traits. Conditional analysis was performed within wider locus regions in order to condition jointly on all sentinel SNPs within the locus. If any pairs of SNPs at a locus were in high LD ($r^2 \geq 0.9$) beyond the collinearity cut-off, the most significant SNP with the minimum P-value across all BP traits from the GWAS discovery meta-analysis was selected.

All locus-specific conditional analyses used the "--cojo-cond" command in GCTA, with the list of sentinel or published SNPs being input as the conditional SNP-list. As for the genome-wide approach, the trait-specific GWAS discovery meta-analysis results were used as the input summary data, and the BioVU EA imputed genetic data was used as the reference PLINK dataset. The output provides the conditional analysis results of all SNPs within the locus region after conditioning on the sentinel or published SNPs. These results are then filtered to obtain a list of potential secondary SNPs which are both significant and independent according to the following criteria:

(a) $P < 5x10^{-8}$ from original GWAS discovery primary meta-analysis, so the SNP is significantly associated with BP itself, at genome-wide significance level

(b) $Pc < 5x10^{-8}$ from the conditional analysis, so that the SNP is also significantly associated with BP after conditioning on the sentinel / published SNPs

(c) -log10(p) / -log10(p_cond) < 1.5, i.e. there is less than a 1.5 fold difference between the GWAS P-value and the conditional *P*-value of the SNP, implying that conditioning on the sentinel / published SNPs has had little impact on the association of the potential secondary SNP, and hence it is statistically independent

All significant independent SNPs meeting the above criteria, from all loci across all chromosomes were combined together into one list. This is a longer list than from approach (i), as it contains all possible secondary SNPs, rather than only one lead SNP per independent signal, and many of the SNPs corresponding to the same signal will be in LD with each other.

The outputs from the two different approaches were then combined together to identify those SNPs which are genome-wide significant in the discovery dataset and jointly independent on a genome-wide level, as well as residing within an existing BP locus (either novel or known). For robustness, a secondary signal was only claimed if the SNP is validated from both approaches.

Enrichment and Pathway Analyses

Enrichment analyses in DEPICT[112] were performed by using trait-specific GWAS significant sentinel SNPs from known and novel loci from final meta-analysis as input. DEPICT is based on predefined phenotypic gene sets from multiple databases and Affymetrix HGU133a2.0 expression microarray data from more than >37k subjects to build highly-expressed gene sets for Medical Subject Heading (MeSH) tissue and cell type annotations. Output includes a p-value for enrichment and a yes/no indicator of whether the FDR q-value is <0.05. Tissue level and gene-set enrichment features with FDR <5% are considered.

Cross-phenotype association analysis

We applied the CPASSOC software to combine association evidence of SBP and DBP. CPASSOC provides two statistics, S_{Hom} and S_{Het}, as previously described [50]. S_{Hom} is similar to the fixed effect meta-analysis method [104] but accounts for the correlation of summary statistics of the multi-traits and for overlapping or related samples among the cohorts. S_{Hom} uses the trait sample size as the weight, so that it is possible to combine traits with different measure scales. S_{Het} is an extension of S_{Hom}, and it can increase the statistical power over S_{Hom} when a variant affect only a subset of traits. The distribution of S_{Het} under the null hypothesis was obtained through an estimated beta distribution. To calculate the statistics, S_{Hom} and S_{Het}, and to account for the correlation among the

traits, a correlation matrix is required. In this study, we used the correlation matrix calculated from the residuals of the four traits after adjustments for covariates and principal components.

S-PrediXcan Analysis

Genetically predicted gene expression was evaluated for the common variant subset with S-PrediXcan[49], a gene-level approach that estimates the genetically determined component of gene expression in a given tissue and tests it for association with SNP-level summary statistics. The covariance matrix used to account for SNP-SNP relationships (linkage disequilibrium) was constructed using genotypes from 1000 Genomes African ancestry samples. We utilized all four BP discovery+replication meta-analysis results for common variants and 5 cardiovascular disease-related tissues from GTEx v7[113] for this analysis. (Supplementary Table 9a-d).

Polygenic Risk Scores

Polygenic scores were constructed separately for DBP and SBP using discovery meta-analysis results from this manuscript. PRS-CS[114] was used to shrink effect estimates instead of pruning based on linkage disequilibrium, and PLINKv1.9[115] was used to compute weighted sum PRS at p-value thresholds from 0.5 to $5x10^{-8}$ in BioVU participants. Linear regressions were performed in R for SBP or DBP as a function of each PRS, age, age^2, BMI, sex, and 10 principal components separately in 9,268 BioVU participants of African ancestry to evaluate variance explained by the PRS and model fit (r^2). Upon selecting the best performing PRS by maximum variance explained by the PRS for each phenotype (Supplementary Tables 10a and b, scores were validated in UK Biobank (N = 7,641). Associations between PRS decile and hypertension status were evaluated through logistic regressions performed separately in both BioVU and UK Biobank data using R and adjusting for age, age^2, BMI, sex, and proportions of genetic ancestry determined using ADMIXTURE with K=6 on an identical 100000 SNPs with 1000 Genomes reference populations as described previously[116]. Predictive performance of logistic models for hypertension both with and without PRS were trained

in BioVU using 10-fold cross-validation to estimate weightings, and applied in UKB for evaluation of the area under the receiver operator curve (AUROC).

AUTHOR CONTRIBUTIONS

Conceptualization: NF, TLE, AMH, CJO, YVS, XZ; Data curation: JCB, BEC, AC, NF, AG, XG, JNH, AMH, MRI, SLRK, EEK, JL, LSP, SR, MMS, JAS, YVS, OW, LRY; Formal analysis: JNH, NF, AG, JMK, JL; Funding acquisition: DKA, AC, TLE, NF, SLRK, AMH, LSP, SR, JIR, XZ; Investigation: DKA, XG, SLRK, JNH, JIR, MMS; Methodology: DKA, HRM, YVS; Project administration: DKA, AC, TLE, AMH, EEK, BMP, JIR, JAS, LRY; Resources: JCB, BEC, AC, MF, NF, XG, MRI, SLRK, EEK, CK, JL, JEM, PBM, LSP, BMP, LMR, SR, JIR, MMS, JAS, XS, DV, HRW, LRY; Software: EST; Supervision: JNH, AMH, XZ TLE; Validation: DKA, PSdV, NF, AG, MRI, JNH, TNK, ACM, XS, JCB, BEC, AC, MF, NF, XG, MRI, SLRK, JMK, EEK, CK, JL, JEM, PBM, LSP, BMP, LMR, SR, JIR, MMS, JAS, XS, DV, HRW, LRY; Visualization: JNH, AG; Writing-original draft: JNH, TLE; Writing- review and editing: DKA, JCB, BEC, AC, PSdV, TLE, NF, AG, XG, MRI, JNH, SLRK, TNK, EEK, CK, RJFL, JEM, ACM, PBM, CJO, LSP, BMP, LMR, JIR, JAS, YVS, HRW, PFWF, LRY, XZ.

DECLARATION OF INTERESTS

The authors declare no conflicts of interest other than the following:
CJO is currently employed by Novartis Institutes for Biomedical Research (unrelated to this project) and remains credentialed as a without compensation researcher with the Veterans Administration. LSP declares that there is no duality of interest associated with this manuscript. With regard to potential conflicts of interest, LSP has served on Scientific Advisory Boards for Janssen, and has or had research support from Merck, Pfizer, Eli Lilly, Novo Nordisk, Sanofi, PhaseBio, Roche, Abbvie, Vascular Pharmaceuticals, Janssen, Glaxo SmithKline, and the Cystic Fibrosis Foundation. LSP is also a cofounder and Officer and Board member and stockholder for a company, Diasyst, Inc., which markets software aimed to help improve diabetes management. BMP serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. EEK received speaker honorarium at Regeneron and Illumina.

LMR is a consultant for the TOPMed Administrative Coordinating Center (through Westat).

ACKNOWLEDGEMENTS

This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by award MVP VA Award #I01BX003360 (AMH). JNH is supported by K12HD04348 (PI K.E. Hartmann). BioVU is supported by institutional funding, the 1S10RR025141-01 instrumentation award, and by the Vanderbilt CTSA grant UL1TR000445 from NCATS/NIH. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN, supported in part by an S10 instrumentation award (1S10OD023680-01). The authors also acknowledge the many contributors to the COGENT and TOPMed consortia. This publication does not represent the vision of the Department of Veterans Affairs or the United States Government. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services. The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and genera lprogram coordination were provided by the TOPMed Administrative Coordinating Center (R01HL-120393; U01HL-120393; contractHHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. Individual contributing study acknowledgements are included in

supplementary materials. We also acknowledge support from the CHARGE infrastructure grant (HL105756).

Additional funding was provided in part by the following awards: I01-BX004821 (PWFW), R01-NS070941 (MPI Taylor/Crusto), HG011052 (XZ) , HL086694 NHLBI (PI Chakravarti), HHSN2682013000001I / N01-HC-65233 (JC), R01 DK117445 R01-MD012765, R21-HL140385, R21 HL123677, R01HL120393, 1R01HG010297, R01 HL105756, and X01 HL134588. MMS's contribution was supported by DK108444-01A1. LSP is supported in part by VA awards CSP #2008, I01 CX001899, I01 CX001737, and HSR&D IIR 07-138, NIH awards R21 DK099716, R18 DK066204, R21 AI156161, U01 DK091958, U01 DK098246, UL1 TR002378, R01 HL113338, R35 HL135818; R01 HL43680, and a Cystic Fibrosis Foundation award PHILLI12A0. The project described was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant KL2TR002490 (LMR).

ACCESSION NUMBERS

MVP genome-wide summary statistics are available at dbGAP phs001672.v6.p1.

REFERENCES

1. Lawes, C.M., Vander Hoorn, S., Rodgers, A. & International Society of Hypertension. Global burden of blood-pressure-related disease, 2001. *Lancet* **371**, 1513-8 (2008).
2. Forouzanfar, M.H. *et al.* Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990-2015. *JAMA* **317**, 165-182 (2017).
3. Lewington, S. *et al.* Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* **360**, 1903-13 (2002).
4. Roger, V.L. *et al.* Executive summary: heart disease and stroke statistics--2012 update: a report from the American Heart Association. *Circulation* **125**, 188-97 (2012).
5. Whelton, P.K. *et al.* 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol* **71**, e127-e248 (2018).
6. Egan, B.M., Zhao, Y. & Axon, R.N. US trends in prevalence, awareness, treatment, and control of hypertension, 1988-2008. *JAMA* **303**, 2043-50 (2010).
7. Lloyd-Jones, D. *et al.* Heart disease and stroke statistics--2010 update: a report from the American Heart Association. *Circulation* **121**, e46-e215 (2010).
8. Rosamond, W. *et al.* Heart disease and stroke statistics--2007 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* **115**, e69-171 (2007).
9. Selassie, A. *et al.* Progression is accelerated from prehypertension to hypertension in blacks. *Hypertension* **58**, 579-87 (2011).
10. Kshirsagar, A.V., Carpenter, M., Bang, H., Wyatt, S.B. & Colindres, R.E. Blood pressure usually considered normal is associated with an elevated risk of cardiovascular disease. *Am J Med* **119**, 133-41 (2006).
11. Berry, J.D. *et al.* Lifetime risks of cardiovascular disease. *N Engl J Med* **366**, 321-9 (2012).
12. Stevens, J., Truesdale, K.P., Katz, E.G. & Cai, J. Impact of body mass index on incident hypertension and diabetes in Chinese Asians, American Whites, and American Blacks: the People's Republic of China Study and the Atherosclerosis Risk in Communities Study. *Am J Epidemiol* **167**, 1365-74 (2008).
13. Forman, J.P., Stampfer, M.J. & Curhan, G.C. Diet and lifestyle risk factors associated with incident hypertension in women. *JAMA* **302**, 401-11 (2009).
14. Ehret, G.B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
15. Kato, N. *et al.* Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet* **43**, 531-8 (2011).
16. Jones, D.W. Hypertension in East Asia. *Am J Hypertens* **8**, 111s-114s (1995).

17. Dragano, N. *et al.* Neighbourhood socioeconomic status and cardiovascular risk factors: a multilevel analysis of nine cities in the Czech Republic and Germany. *BMC Public Health* **7**, 255 (2007).
18. Grotto, I., Huerta, M., Grossman, E. & Sharabi, Y. Relative impact of socioeconomic status on blood pressure lessons from a large-scale survey of young adults. *Am J Hypertens* **20**, 1140-5 (2007).
19. Hoang, V.M., Byass, P., Dao, L.H., Nguyen, T.K. & Wall, S. Risk factors for chronic disease among rural Vietnamese adults and the association of these factors with sociodemographic variables: findings from the WHO STEPS survey in rural Vietnam, 2005. *Prev Chronic Dis* **4**, A22 (2007).
20. Le, C., Chongsuvivatwong, V. & Geater, A. Contextual socioeconomic determinants of cardiovascular risk factors in rural south-west China: a multilevel analysis. *BMC Public Health* **7**, 72 (2007).
21. Pickett, K.E. & Wilkinson, R.G. People like us: ethnic group density effects on health. *Ethn Health* **13**, 321-34 (2008).
22. Wilkinson, R.G. & Pickett, K.E. Income inequality and socioeconomic gradients in mortality. *Am J Public Health* **98**, 699-704 (2008).
23. Jen, K.L., Brogan, K., Washington, O.G., Flack, J.M. & Artinian, N.T. Poor nutrient intake and high obese rate in an urban African American population with hypertension. *J Am Coll Nutr* **26**, 57-65 (2007).
24. Blakely, T., Tobias, M. & Atkinson, J. Inequalities in mortality during and after restructuring of the New Zealand economy: repeated cohort studies. *BMJ* **336**, 371-5 (2008).
25. Forde, A.T. *et al.* Discrimination and Hypertension Risk Among African Americans in the Jackson Heart Study. *Hypertension* **76**, 715-723 (2020).
26. Adeyemo, A. *et al.* A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet* **5**, e1000564 (2009).
27. Ehret, G.B. *et al.* The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat Genet* **48**, 1171-1184 (2016).
28. Franceschini, N. *et al.* Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. *Am J Hum Genet* **93**, 545-54 (2013).
29. Giri, A. *et al.* Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat Genet* **51**, 51-62 (2019).
30. He, J. *et al.* Genome-wide association study identifies 8 novel loci associated with blood pressure responses to interventions in Han Chinese. *Circ Cardiovasc Genet* **6**, 598-607 (2013).
31. Hoffmann, T.J. *et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet* **49**, 54-64 (2017).
32. International Consortium for Blood Pressure Genome-Wide Association Studies *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
33. Kato, N. *et al.* Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat Genet* **47**, 1282-93 (2015).

34. Kelly, T.N. *et al.* Genome-wide association study meta-analysis reveals transethnic replication of mean arterial and pulse pressure loci. *Hypertension* **62**, 853-9 (2013).
35. Kraja, A.T. *et al.* New Blood Pressure-Associated Loci Identified in Meta-Analyses of 475 000 Individuals. *Circ Cardiovasc Genet* **10**(2017).
36. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat Genet* **41**, 677-87 (2009).
37. Li, C. *et al.* Genome-Wide Association Study Meta-Analysis of Long-Term Average Blood Pressure in East Asians. *Circ Cardiovasc Genet* **10**, e001527 (2017).
38. Liang, J. *et al.* Single-trait and multi-trait genome-wide association analyses identify novel loci for blood pressure in African-ancestry populations. *PLoS Genet* **13**, e1006728 (2017).
39. Lu, X. *et al.* Genome-wide association study in Chinese identifies novel loci for blood pressure and hypertension. *Hum Mol Genet* **24**, 865-74 (2015).
40. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet* **41**, 666-76 (2009).
41. Sofer, T. *et al.* Genome-Wide Association Study of Blood Pressure Traits by Hispanic/Latino Background: the Hispanic Community Health Study/Study of Latinos. *Sci Rep* **7**, 10348 (2017).
42. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet* **48**, 1151-1161 (2016).
43. Wain, L.V. *et al.* Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension* (2017).
44. Wain, L.V. *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* **43**, 1005-11 (2011).
45. Warren, H.R. *et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet* **49**, 403-415 (2017).
46. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat Genet* **48**, 1162-70 (2016).
47. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* (2018).
48. Fox, E.R. *et al.* Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study. *Hum Mol Genet* **20**, 2273-84 (2011).
49. Barbeira, A.N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* **9**, 1825 (2018).
50. Zhu, X. *et al.* Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am J Hum Genet* **96**, 21-36 (2015).
51. Lambert, S.A. *et al.* The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat Genet* **53**, 420-425 (2021).

52. Xie, T. *et al.* Genetic Risk Scores for Complex Disease Traits in Youth. *Circ Genom Precis Med* **13**, e002775 (2020).
53. Vuckovic, D. *et al.* The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* **182**, 1214-1231 e11 (2020).
54. Chen, M.H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* **182**, 1198-1213 e14 (2020).
55. Astle, W.J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429.e19 (2016).
56. Ahola-Olli, A.V. *et al.* Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. *Am J Hum Genet* **100**, 40-50 (2017).
57. Sun, B.B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).
58. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* **104**, 65-75 (2019).
59. Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769-773 (2018).
60. Winkler, T.W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* **11**, e1005378 (2015).
61. Rask-Andersen, M., Karlsson, T., Ek, W.E. & Johansson, A. Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nat Commun* **10**, 339 (2019).
62. Zhu, Z. *et al.* Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J Allergy Clin Immunol* **145**, 537-549 (2020).
63. Hoffmann, T.J. *et al.* A Large Multiethnic Genome-Wide Association Study of Adult Body Mass Index Identifies Novel Loci. *Genetics* **210**, 499-515 (2018).
64. Morris, J.A. *et al.* An atlas of genetic influences on osteoporosis in humans and mice. *Nat Genet* **51**, 258-266 (2019).
65. Baselmans, B.M.L. *et al.* Multivariate genome-wide analyses of the well-being spectrum. *Nat Genet* **51**, 445-451 (2019).
66. Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat Genet* **50**, 920-927 (2018).
67. Nagel, M., Watanabe, K., Stringer, S., Posthuma, D. & van der Sluis, S. Item-level analyses reveal genetic heterogeneity in neuroticism. *Nat Commun* **9**, 905 (2018).
68. Hill, W.D. *et al.* Genetic contributions to two special factors of neuroticism are associated with affluence, higher intelligence, better health, and longer life. *Mol Psychiatry* **25**, 3034-3052 (2020).
69. Dashti, H.S. *et al.* Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *Nat Commun* **10**, 1100 (2019).
70. Jones, S.E. *et al.* Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nat Commun* **10**, 343 (2019).

71. Lane, J.M. *et al.* Biological and clinical insights from genetics of insomnia symptoms. *Nat Genet* **51**, 387-393 (2019).
72. Demange, P.A. *et al.* Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nat Genet* **53**, 35-44 (2021).
73. Lee, J.J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* **50**, 1112-1121 (2018).
74. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-42 (2016).
75. Zhou, H. *et al.* Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nat Neurosci* **23**, 809-818 (2020).
76. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* **51**, 237-244 (2019).
77. Wu, Y. *et al.* Genome-wide association study of medication-use and associated disease in the UK Biobank. *Nat Commun* **10**, 1891 (2019).
78. Douglas, J.G. *et al.* Management of high blood pressure in African Americans: consensus statement of the Hypertension in African Americans Working Group of the International Society on Hypertension in Blacks. *Arch Intern Med* **163**, 525-41 (2003).
79. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet* **51**, 957-972 (2019).
80. Nielsen, J.B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet* **50**, 1234-1239 (2018).
81. Ramirez, J. *et al.* Common Genetic Variants Modulate the Electrocardiographic Tpeak-to-Tend Interval. *Am J Hum Genet* **106**, 764-778 (2020).
82. Klein, B.E., Klein, R. & Knudtson, M.D. Intraocular pressure and systemic blood pressure: longitudinal perspective: the Beaver Dam Eye Study. *Br J Ophthalmol* **89**, 284-7 (2005).
83. Armstrong, N.J. *et al.* Common Genetic Variation Indicates Separate Causes for Periventricular and Deep White Matter Hyperintensities. *Stroke* **51**, 2111-2121 (2020).
84. Traylor, M. *et al.* Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. *Lancet Neurol* **20**, 351-361 (2021).
85. Theofilatos, K., Korfiati, A., Mavroudi, S., Cowperthwaite, M.C. & Shpak, M. Discovery of stroke-related blood biomarkers from gene expression network models. *BMC Med Genomics* **12**, 118 (2019).
86. Takeuchi, F. *et al.* Interethnic analyses of blood pressure loci in populations of East Asian and European descent. *Nat Commun* **9**, 5052 (2018).
87. Gopel, W. *et al.* Genetic background of high blood pressure is associated with reduced mortality in premature neonates. *Arch Dis Child Fetal Neonatal Ed* **105**, 184-189 (2020).
88. Klarin, D. *et al.* Genetic Architecture of Abdominal Aortic Aneurysm in the Million Veteran Program. *Circulation* **142**, 1633-1646 (2020).

89. Beichman, A.C., Phung, T.N. & Lohmueller, K.E. Comparison of Single Genome and Allele Frequency Data Reveals Discordant Demographic Histories. *G3 (Bethesda)* **7**, 3605-3620 (2017).
90. Gaziano, J.M. *et al.* Million Veteran Program (MVP): A Mega-Biobank to Study Genetic Influences On Health and Disease. *J Clin Epidemiol* (2015).
91. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-73 (2010).
92. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
93. Loh, P.R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448 (2016).
94. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nature genetics* **48**, 1284-1287 (2016).
95. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS One* **9**, e93766 (2014).
96. Maixner, W., Gracely, R.H., Zuniga, J.R., Humphrey, C.B. & Bloodworth, G.R. Cardiovascular and sensory responses to forearm ischemia and dynamic hand exercise. *Am J Physiol* **259**, R1156-63 (1990).
97. Chawla, P.S. & Kochar, M.S. Effect of pain and nonsteroidal analgesics on blood pressure. *Wmj* **98**, 22-5, 29 (1999).
98. Taylor, J.Y. *et al.* A Genome-wide study of blood pressure in African Americans accounting for gene-smoking interaction. *Sci Rep* **6**, 18812 (2016).
99. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
100. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
101. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
102. Zhu, X., Zhang, S., Zhao, H. & Cooper, R.S. Association mapping, using a mixture model for complex traits. *Genet Epidemiol* **23**, 181-96 (2002).
103. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-212 (2014).
104. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
105. Roden, D.M. *et al.* Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* **84**, 362-9 (2008).
106. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
107. Burgess, D.J. The TOPMed genomic resource for human health. *Nat Rev Genet* **22**, 200 (2021).
108. He, K.Y. *et al.* Leveraging linkage evidence to identify low-frequency and rare variants on 16p13 associated with blood pressure using TOPMed whole genome sequencing data. *Hum Genet* **138**, 199-210 (2019).

109. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299 (2021).
110. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
111. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, s1-3 (2012).
112. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
113. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
114. Ge, T., Chen, C.Y., Ni, Y., Feng, Y.A. & Smoller, J.W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* **10**, 1776 (2019).
115. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
116. Keaton, J.M. *et al.* Evidence that geographic variation in genetic ancestry associates with uterine fibroids. *Hum Genet* (2021).

FIGURES

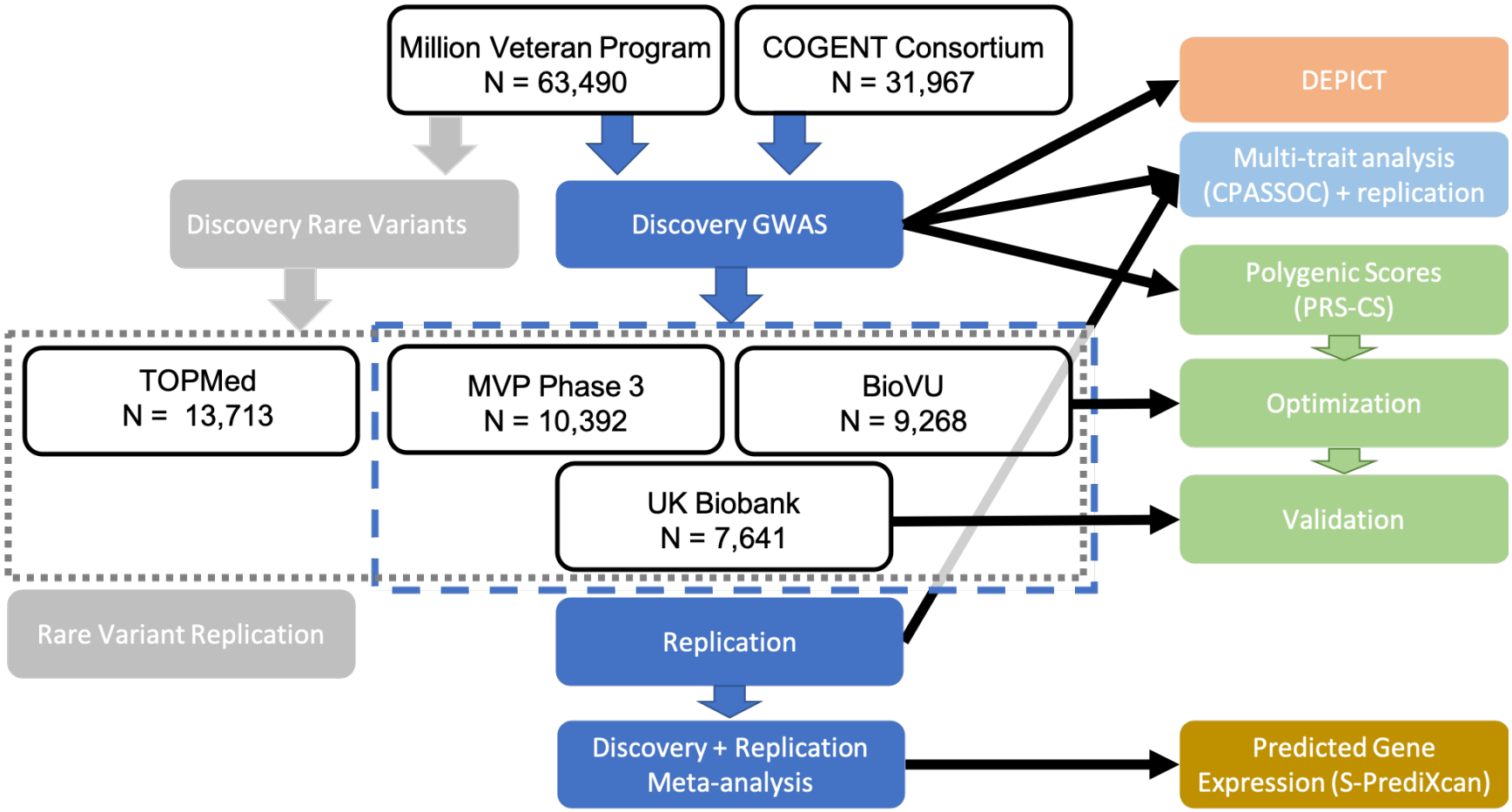


Figure 1. Study resources and workflow diagram

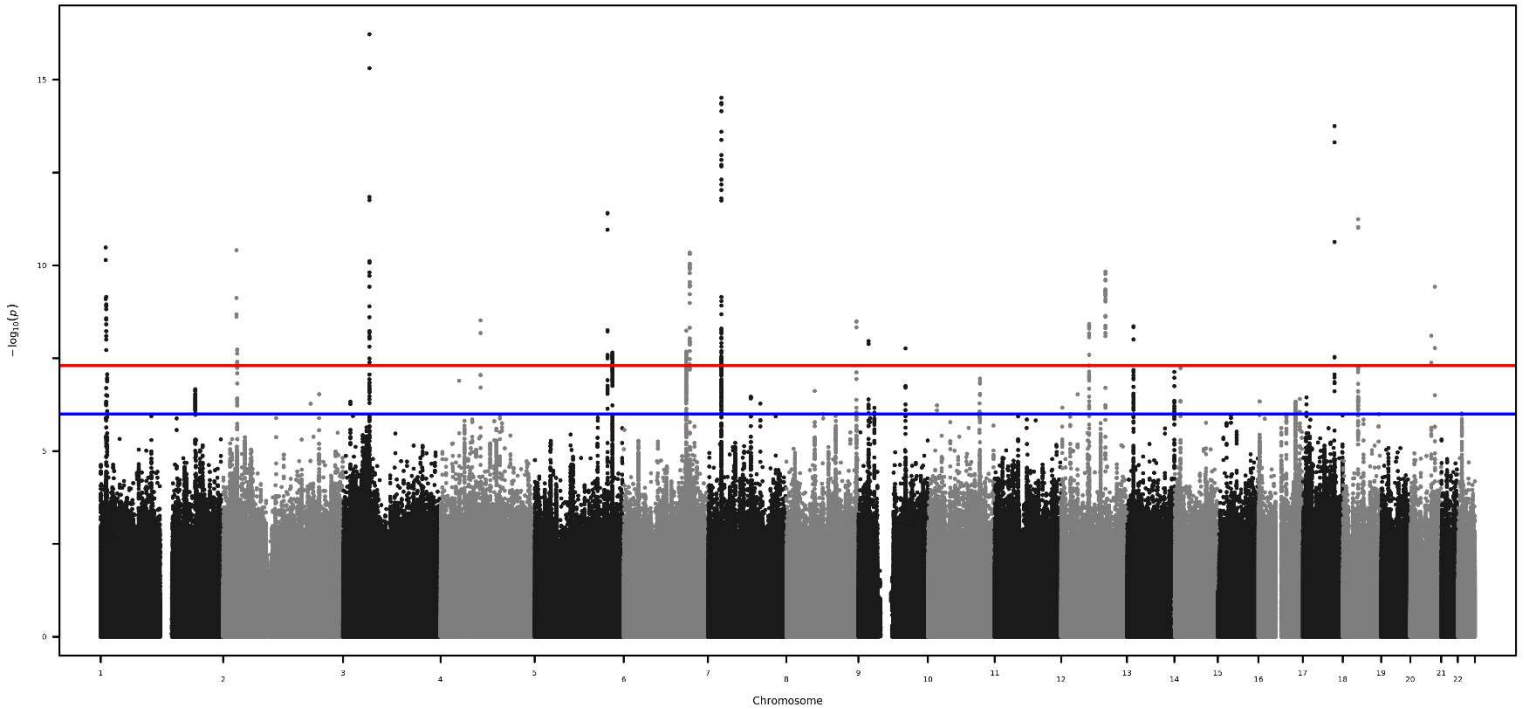


Figure 2a Discovery analysis Manhattan plot for SBP.

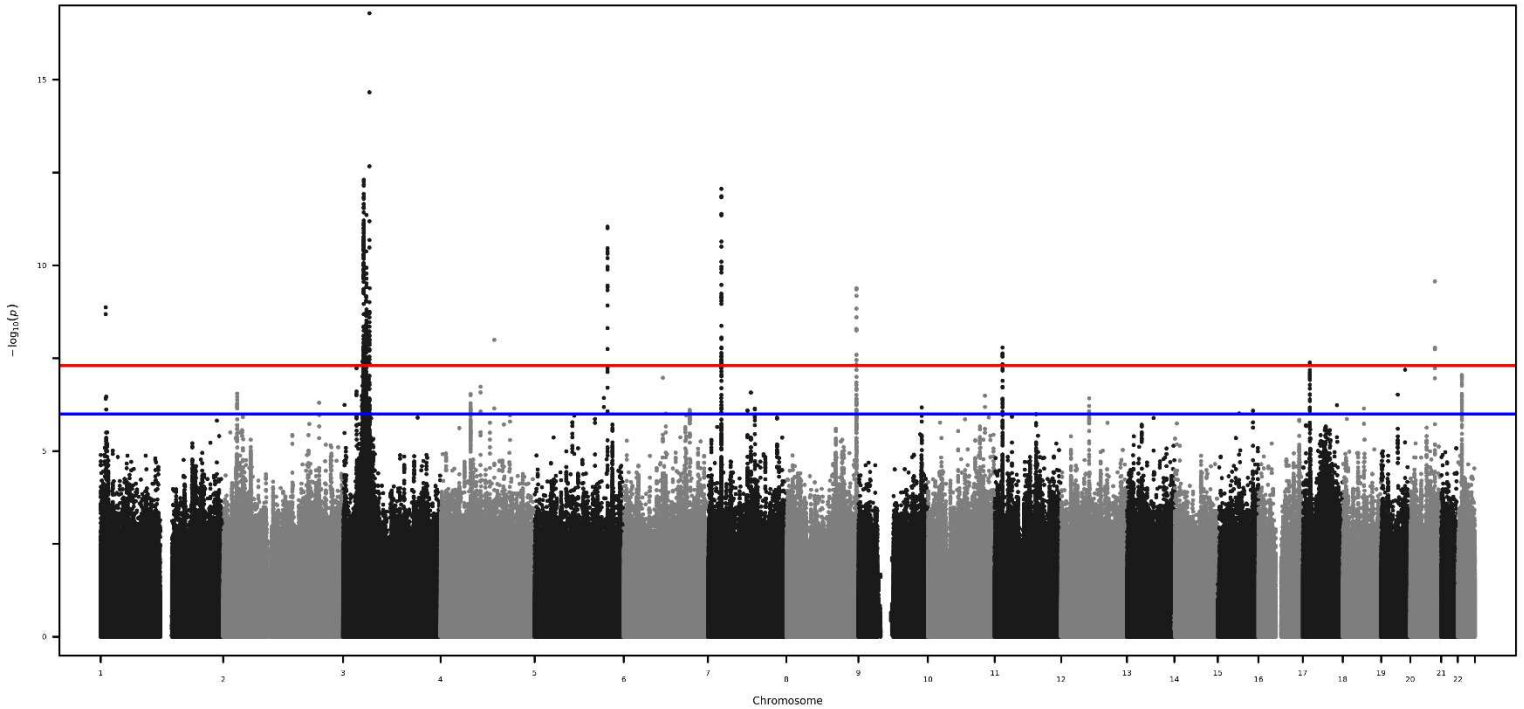


Figure 2b. Discovery analysis Manhattan plot for DBP.

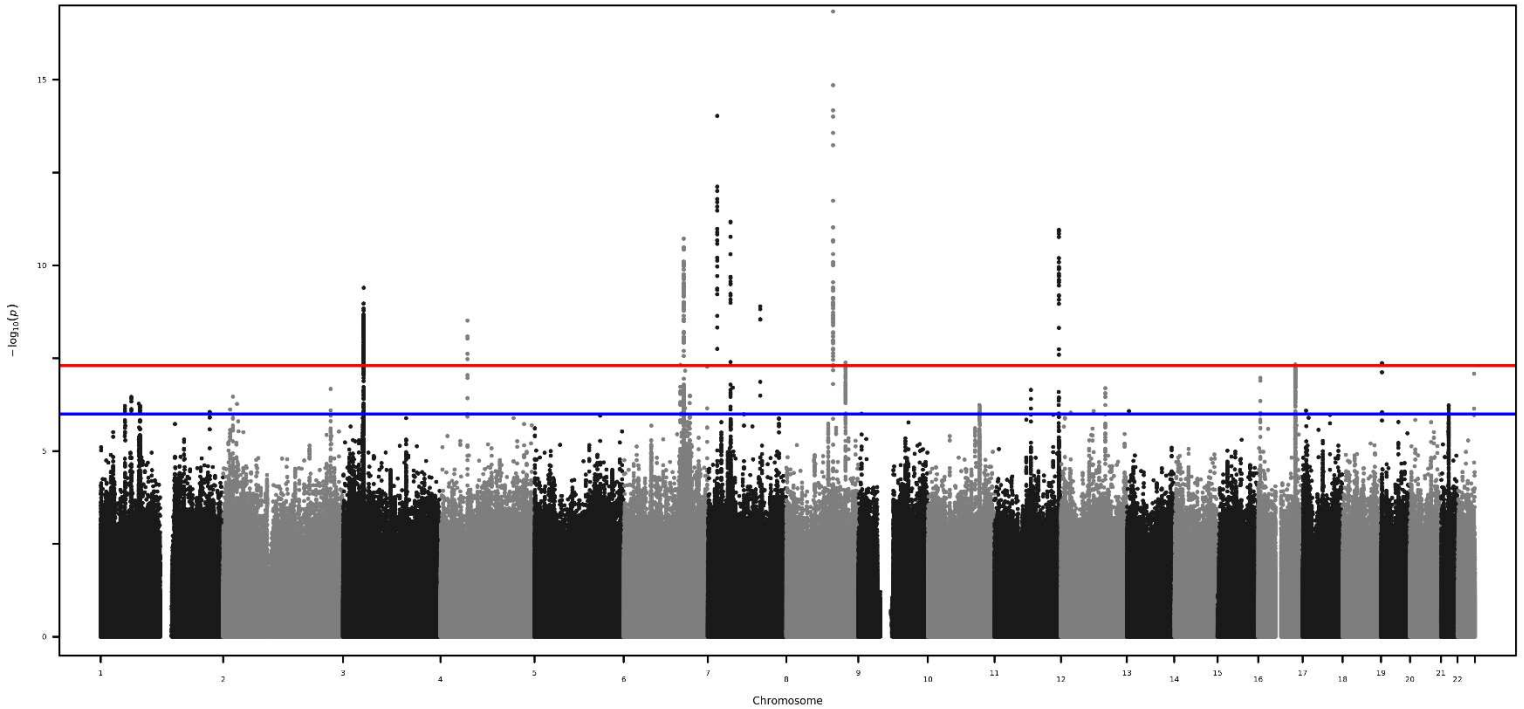


Figure 2c. Discovery analysis Manhattan plot for pulse pressure.

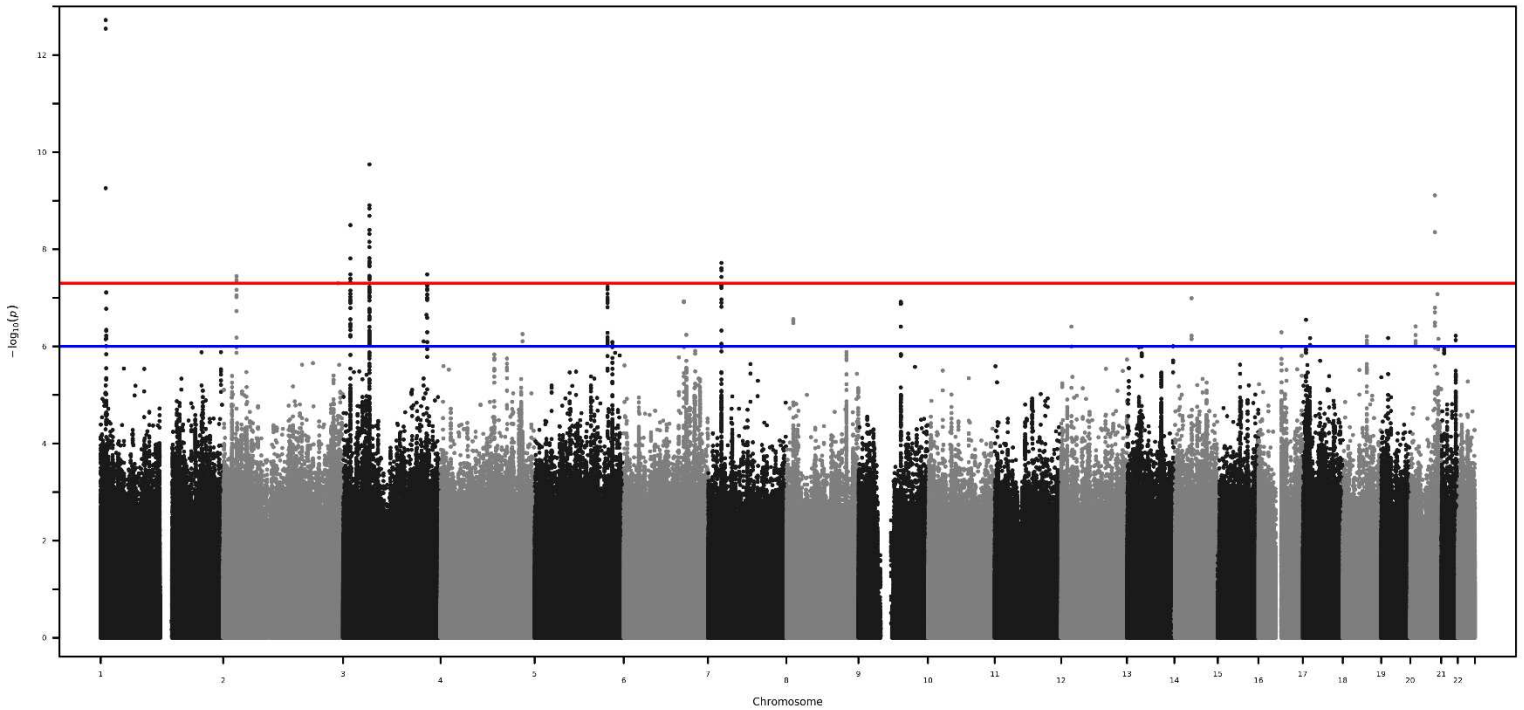


Figure 2d. Discovery analysis Manhattan plots for hypertension status.

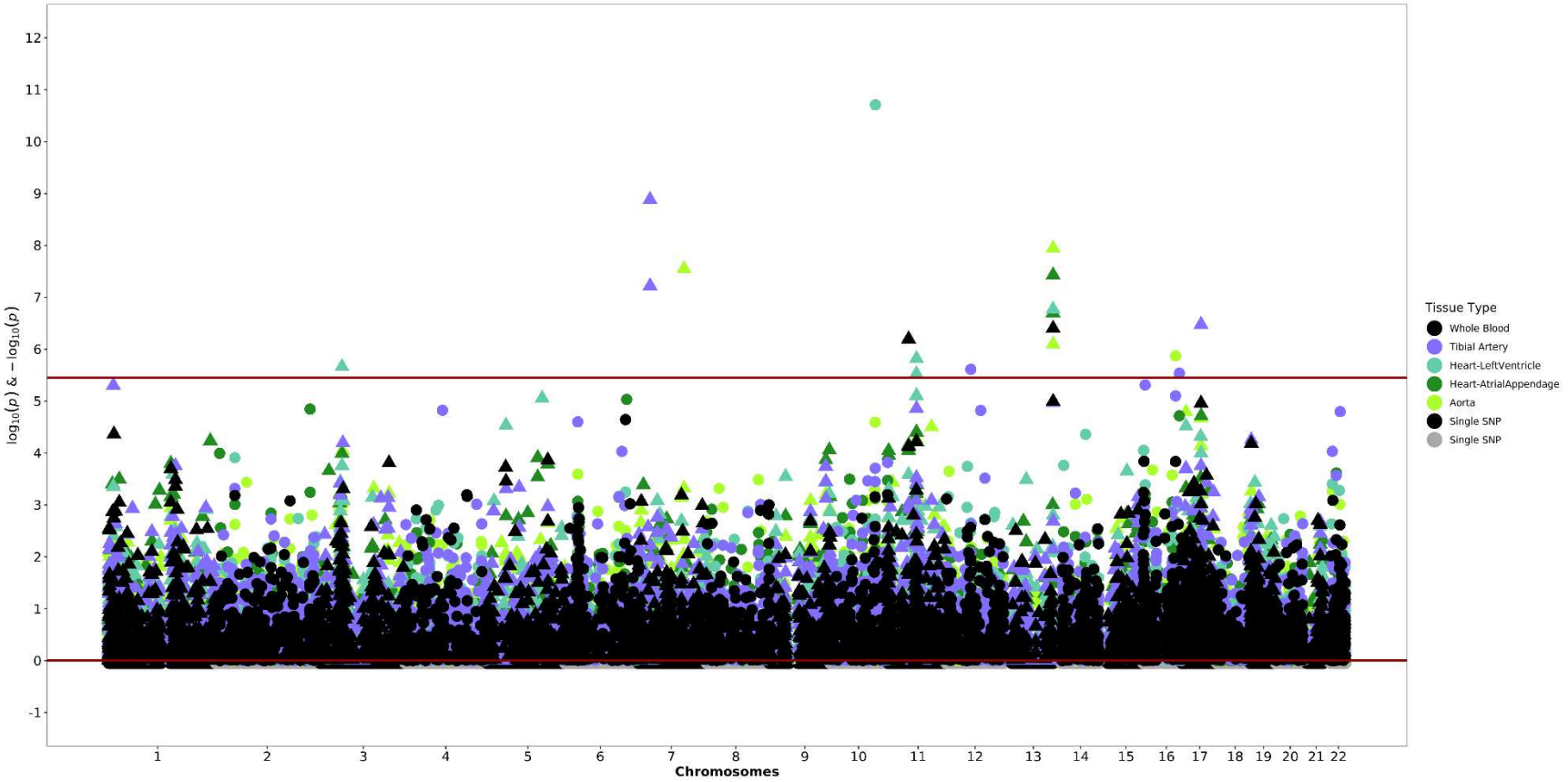


Figure 3a. Genetically-predicted gene expression results for Systolic Blood Pressure from 5 cardiovascular-related tissues from GTEx.

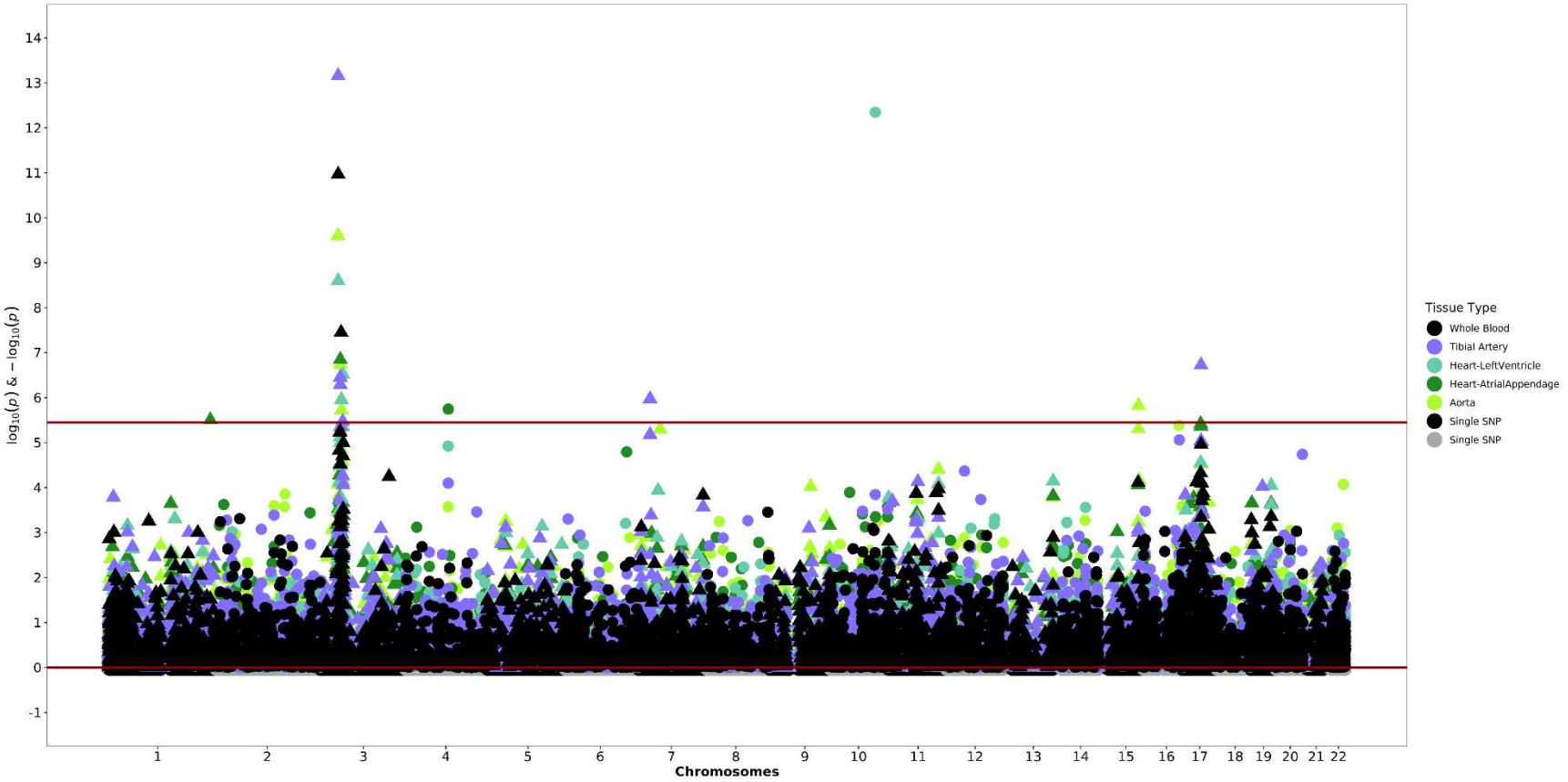


Figure 3b. Genetically-predicted gene expression results for Diastolic Blood Pressure in 5 cardiovascular-related tissues from GTEx.

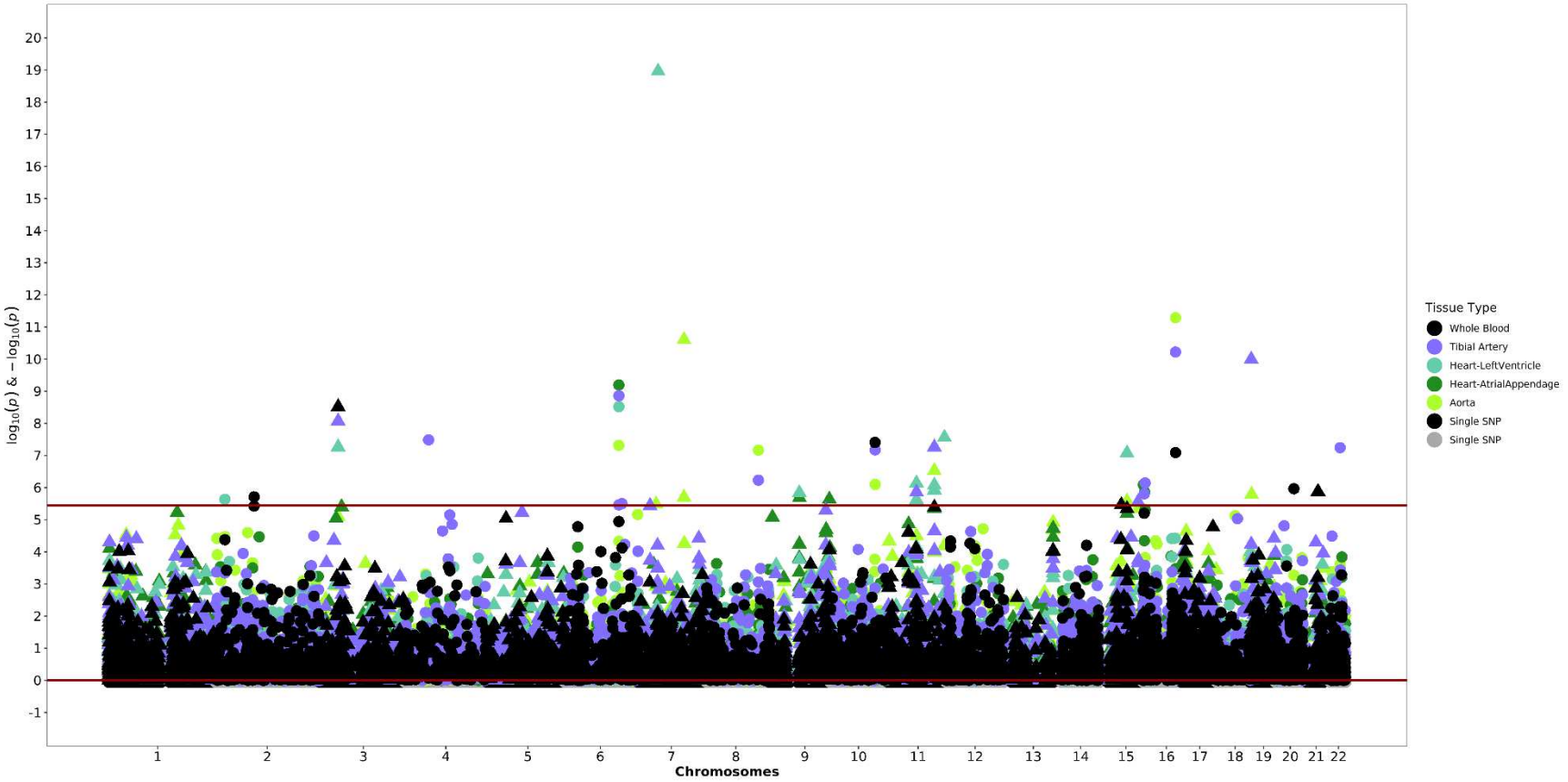


Figure 3c. Genetically-predicted gene expression results for Pulse Pressure from 5 cardiovascular-related GTEx tissues

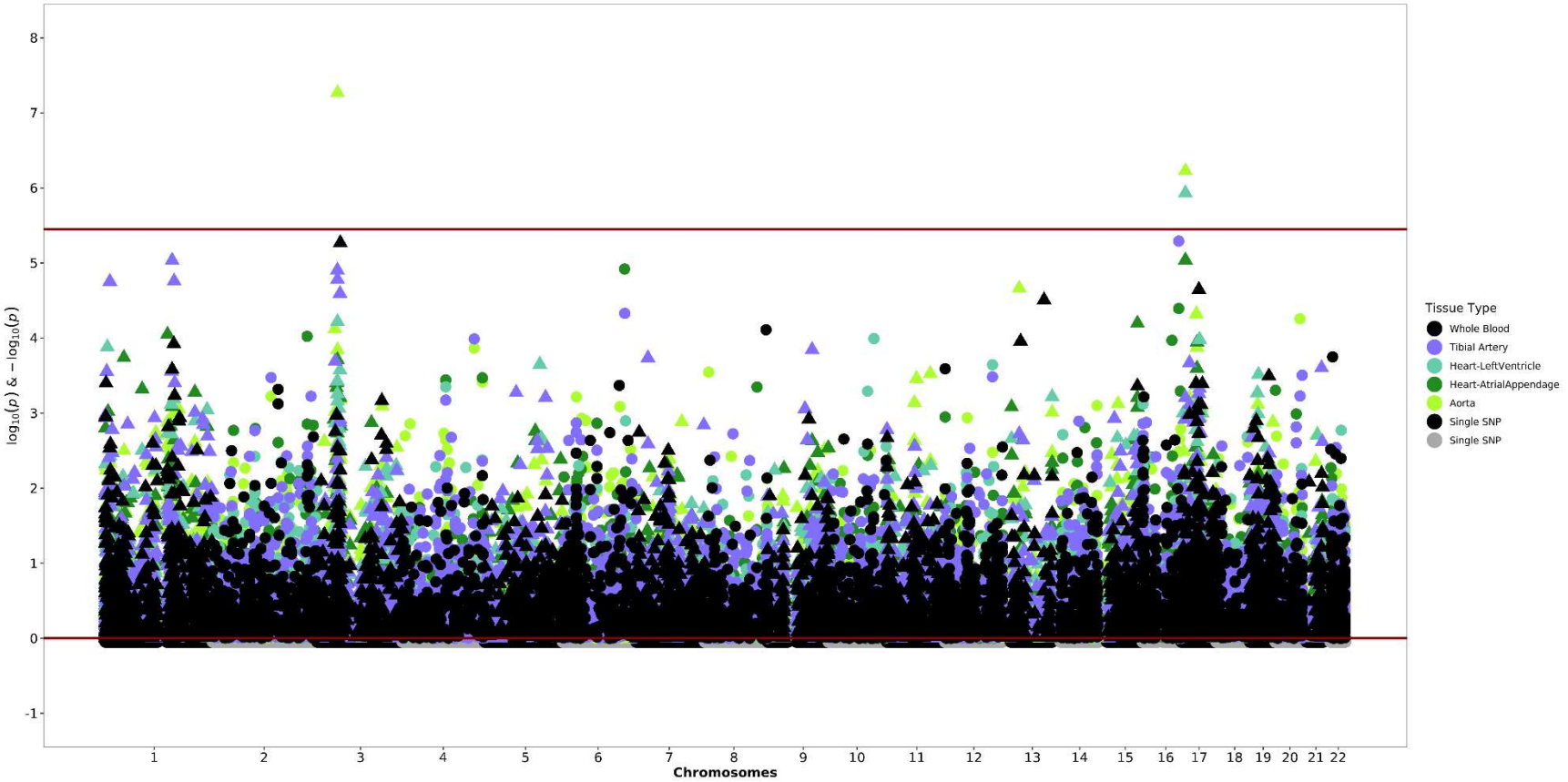


Figure 3d. Genetically-predicted gene expression results for hypertension status from 5 cardiovascular-related GTEx tissues.

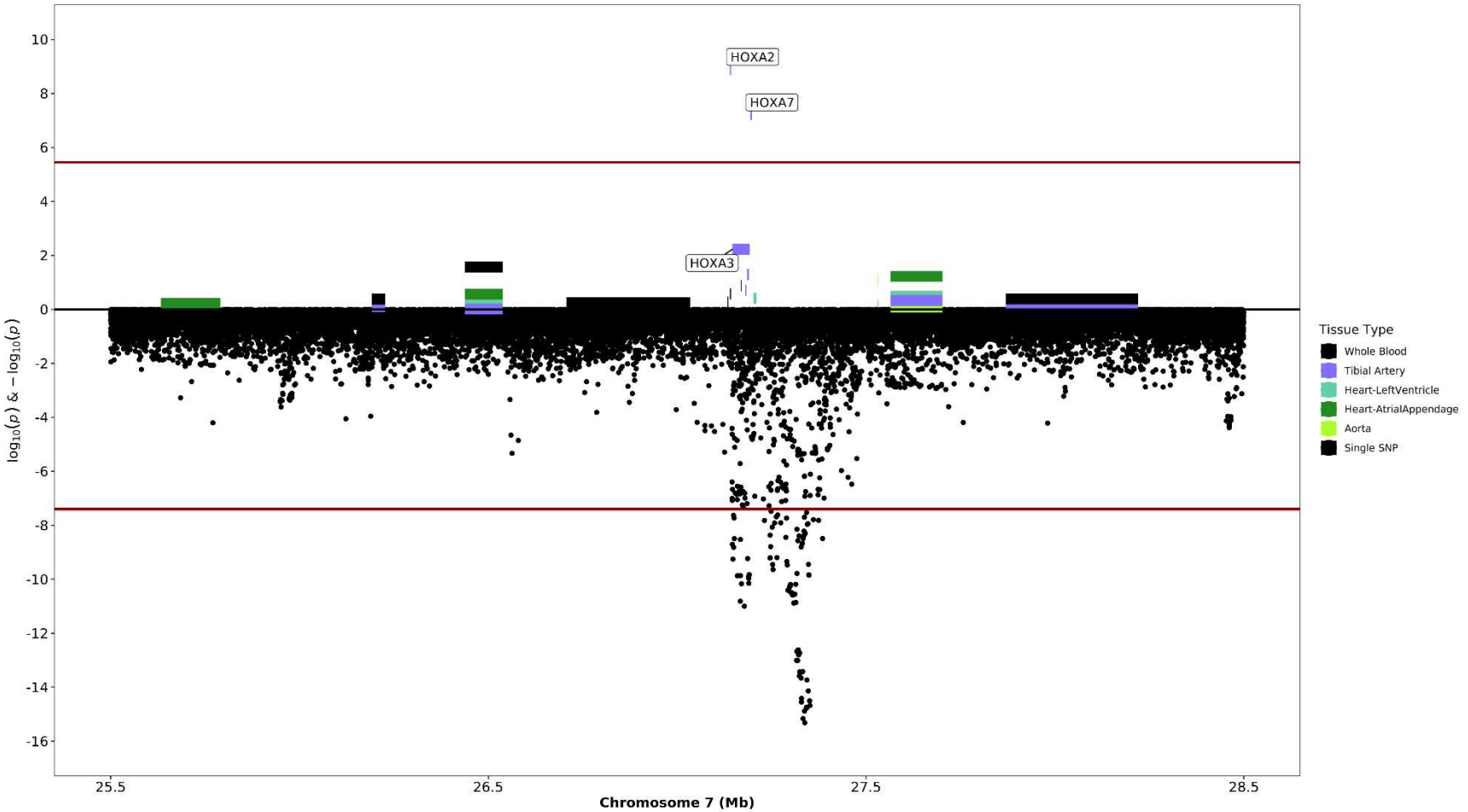


Figure 4. HOX gene region predicted gene expression association with systolic blood pressure.

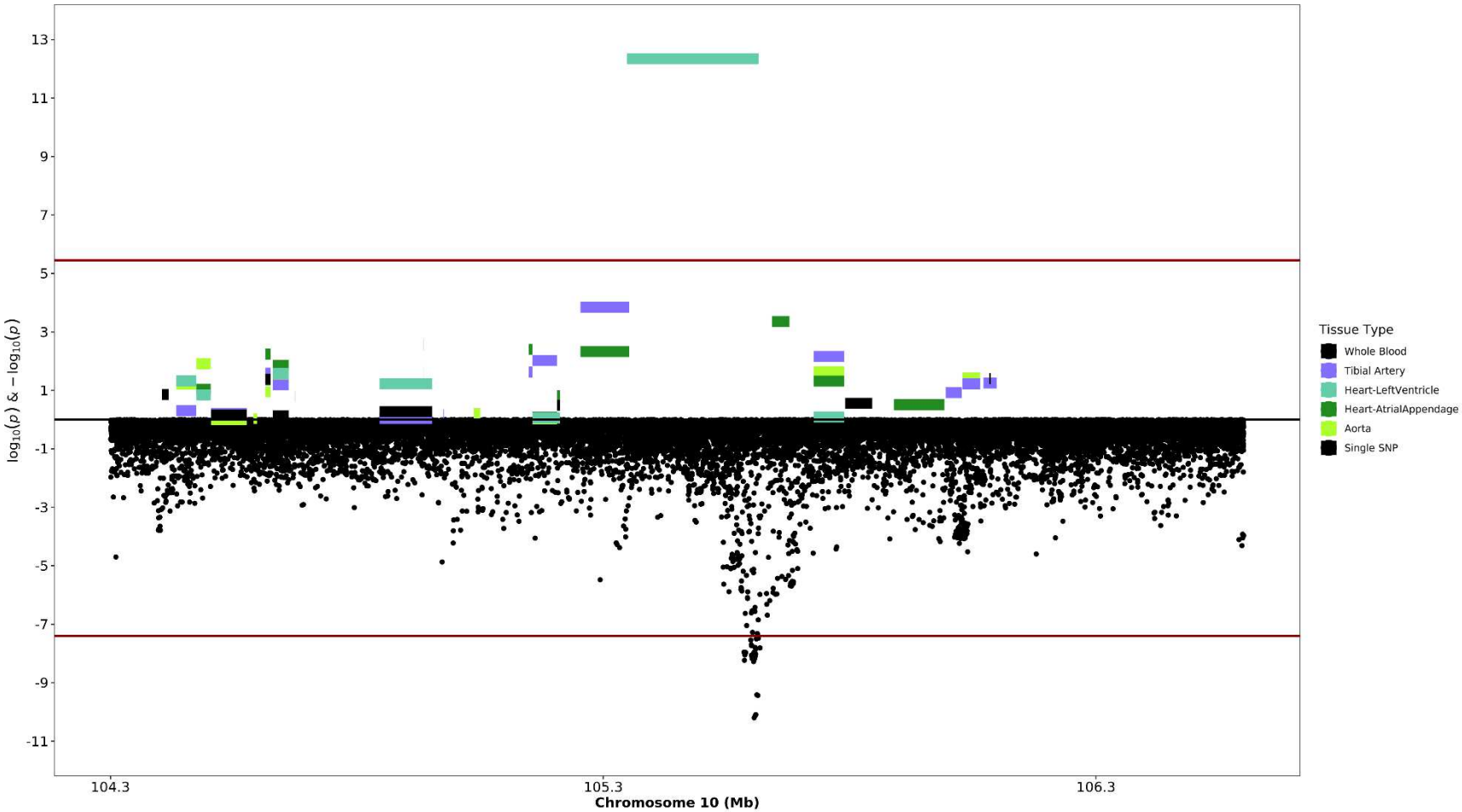


Figure 5. *SH3PXD2A* predicted expression association with diastolic blood pressure.

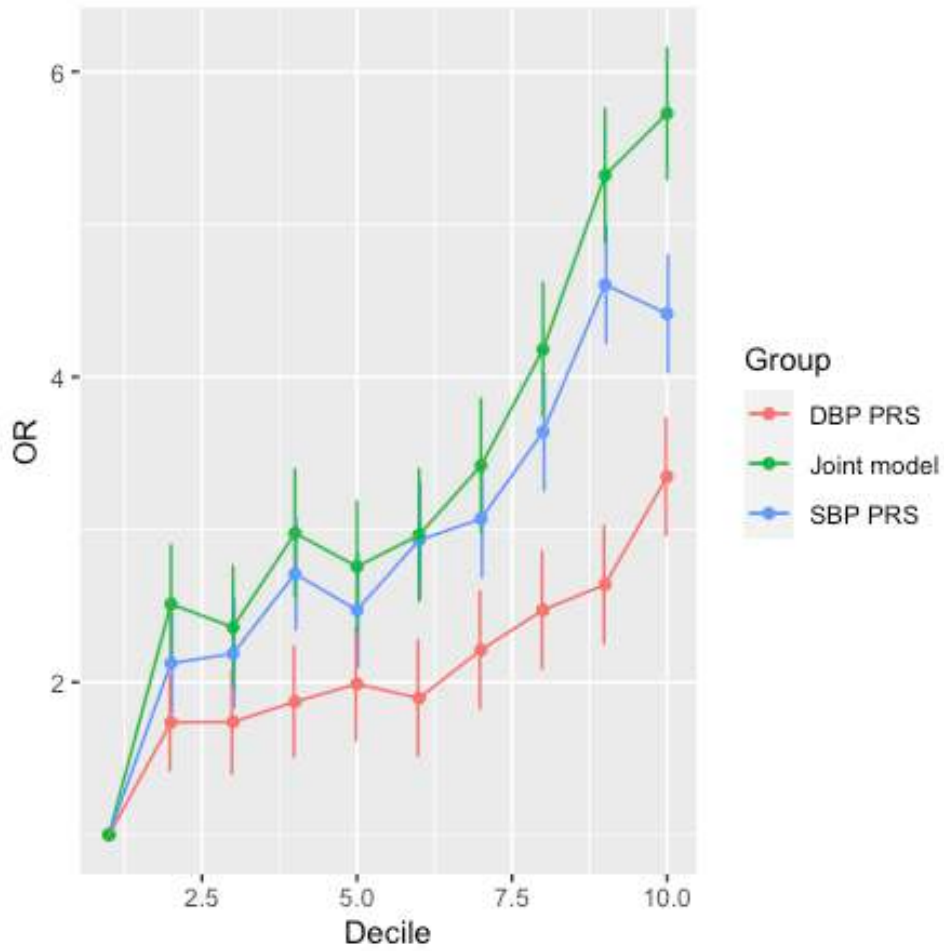
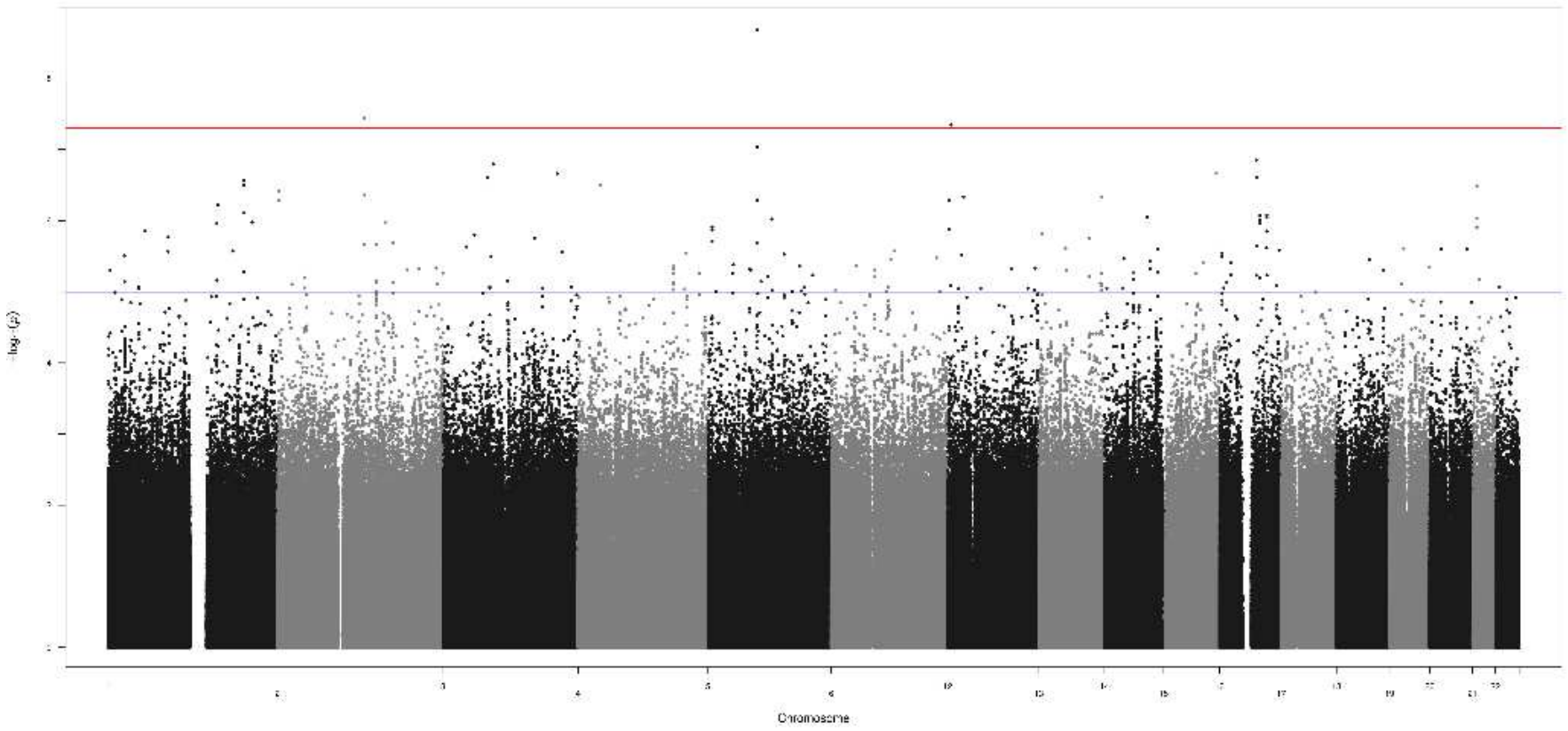


Figure 6. Odds Ratios for hypertension status by decile of optimal PRS in UKB Blacks

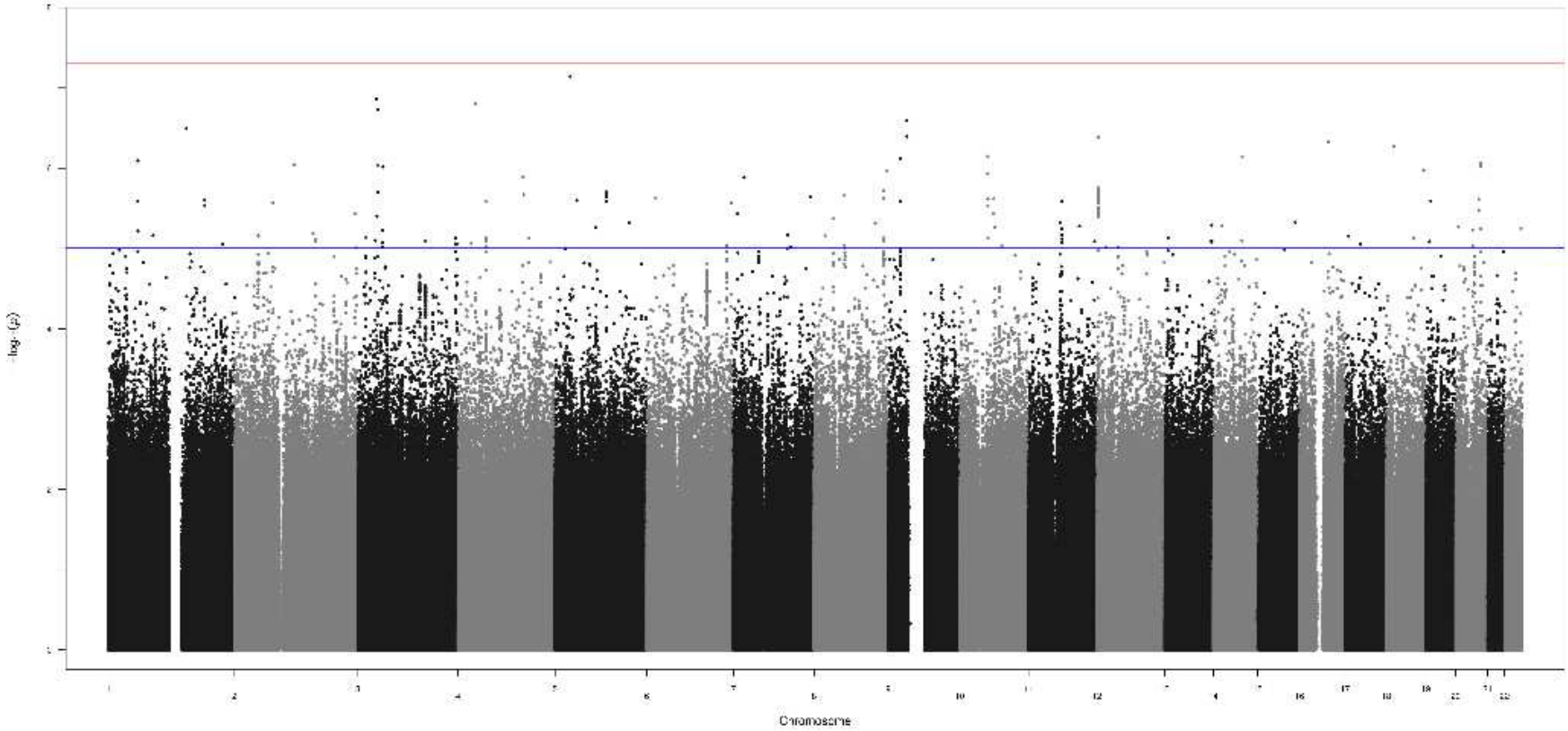
Table 1. Novel genes among S-PrediXcan results with colocalization.

Gene	Chr.	Trait*	S-PrediXcan			COLOC		Association in gene region		
			Best Tissue	Effect	P-value	Lead SNP	Association P-value	Lead SNP	Effect	P-value
LRRC2	3	DBP	Tibial Artery	-1.14	**3.54E-07**	rs11705987	2.168e-06	rs13094615	0.3833	**8.543e-15**
ALS2CL	3	DBP	Heart_Atrial_Appendage	-0.90	**1.41E-07**	rs7633016	1.298e-06			
CCDC36	3	DBP	Heart_Left_Ventricle	-0.75	**1.10E-06**	rs35174559	0.0002476	rs13070214	-0.4045	**3.628e-08**
RP11-3B7.1	3	DBP	Aorta	-0.38	**1.91E-06**	rs11920267	0.0008492			
HEMK1	3	SBP	Heart_Left_Ventricle	-1.23	**2.17E-06**	rs1034405	5.02e-07	rs12629572	-0.4297	**2.003e-07**
PPM1M	3	DBP	Tibial Artery	0.48	**3.37E-06**	rs17052053	1.731e-06	rs2581792	0.4483	**4.808e-08**
SMIM4	3	DBP	Heart_Left_Ventricle	-1.04	**3.03E-07**	rs11710485	2.471e-07	rs9821489	0.4380	**1.133e-18**
ADH1A	4	DBP	Heart_Atrial_Appendage	0.59	**1.80E-06**	rs1789882	2.883e-06	rs2066702	-0.3226	4.009e-07
TPD52L1	6	Pulse pressure	Aorta	-0.46	**4.79E-08**	rs2243390	2.599e-08	rs987166	0.2810	**1.323e-08**
HOXA2	7	*SBP*/DBP/Pulse pressure	Tibial Artery	-0.69	**1.31E-09**	rs739734	**1.35e-10**	rs12535894	-0.7961	**3.83e-15**
AC005022.1	7	Pulse pressure	Aorta	0.94	**3.25E-06**	rs11531505	**1.876e-07**	rs76206723	-0.5112	**8.478e-08**
SDCCAG3	9	Pulse pressure	Heart_Atrial_Appendage	1.04	**2.29E-06**	rs1135314	0.0001277	rs10858100	-0.2396	**3.805e-07**
MARCKSL1P1	10	Pulse pressure	Aorta	-0.69	**7.95E-07**	rs4409766	**1.313e-09**	rs10883797	0.3541	**3.424e-12**
SH3PXD2A	10	*DBP*/SBP	Heart_Left_Ventricle	-0.92	**4.49E-13**	rs4918060	**6.323e-11**	rs4918060	-0.3349	**6.323e-11**
RP11-819C21.1	11	Pulse pressure	Tibial Artery	-1.03	**5.50E-08**	rs1490938	1.31e-06	rs4754196	-0.3027	3.893e-07
RP11-890B15.3	11	Pulse pressure	Heart_Left_Ventricle	-0.74	**2.74E-08**	rs3751039	9.078e-05	rs11222084	-0.3833	**5.338e-12**
HSD17B6	12	SBP	Artery_Tibial	2.17	**2.44E-06**	rs898609	1.245e-10	rs10747770	-0.4708	**1.345e-11**
UPF3A	13	SBP	Heart_Atrial_Appendage	1.21	**1.99E-07**	rs7320104	1.843e-06	rs9590501	-0.4499	**3.37e-11**
TRIP4	15	Pulse pressure	Heart_Left_Ventricle	0.71	**8.44E-08**	rs7165034	4.454e-06	rs1976112	-0.2566	1.33e-07
MIR940	16	Pulse pressure	Tibial Artery	-1.07	**1.56E-06**	rs12599229	4.454e-05	rs9796949	-0.6728	**1.44e-06**
TIGD7	16	Pulse pressure	Heart_Atrial_Appendage	-3.36	**1.38E-06**	rs17684522	4.352e-07	rs1053874	0.4970	**1.591e-09**
STX8	17	HTN	Aorta	-0.43	**5.89E-07**	rs12451858	1.245e-06	rs12452001	-0.1045	4.973e-07
RP11-259G18.3	17	DBP	Tibial Artery	-1.08	**1.86E-07**	rs532193457	0.0001122	rs1819040	-0.6216	1.027e-07
TLE2	19	Pulse pressure	Aorta	-3.81	**1.62E-06**	rs2277739	2.818e-07	rs8102624	0.3981	**1.134e-15**
KDELR3	22	Pulse pressure	Tibial Artery	1.04	**5.71E-08**	rs8141562	8.365e-05	rs138419	0.2656	**1.521e-06**

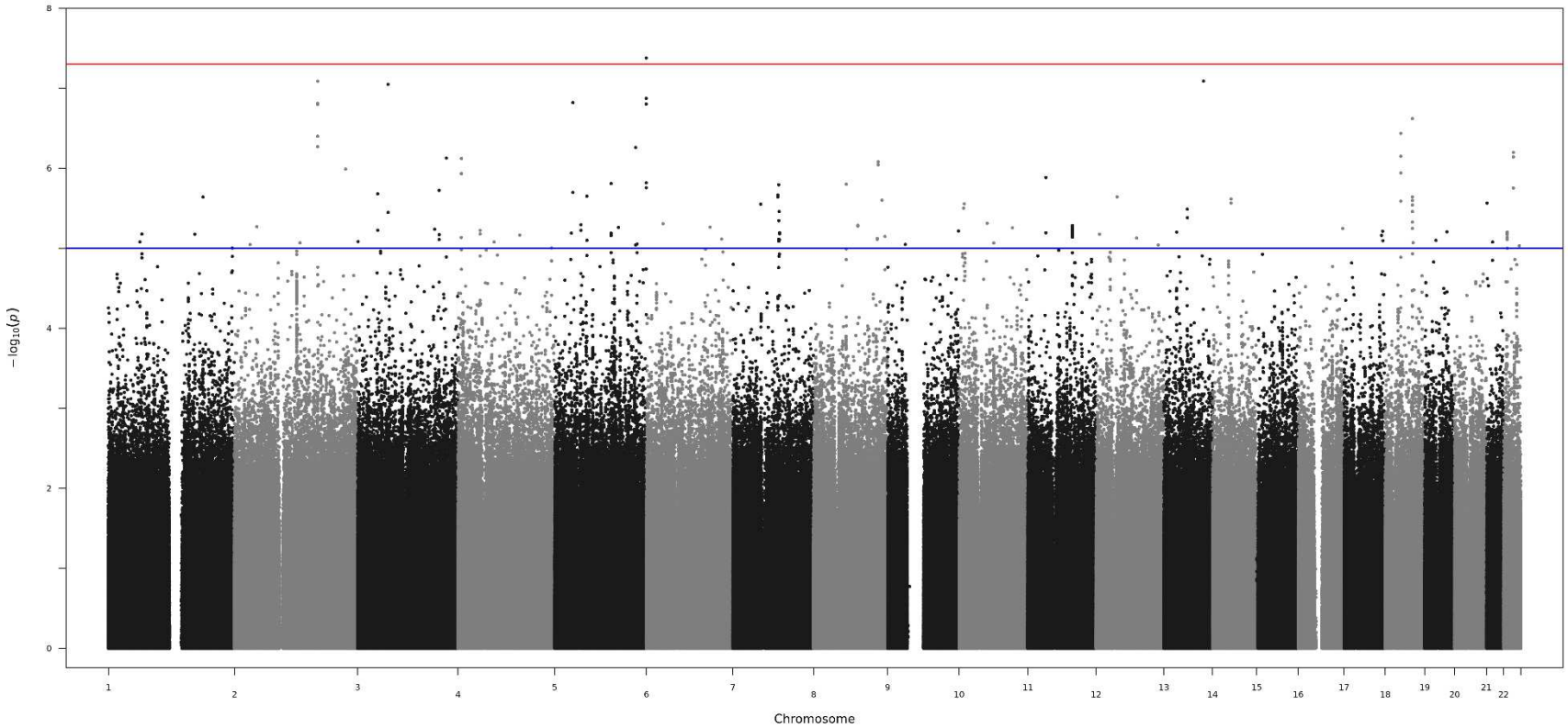
**Italic font indicates strongest associated trait for which results are presented in this table. Boldface indicates significant p-values.*



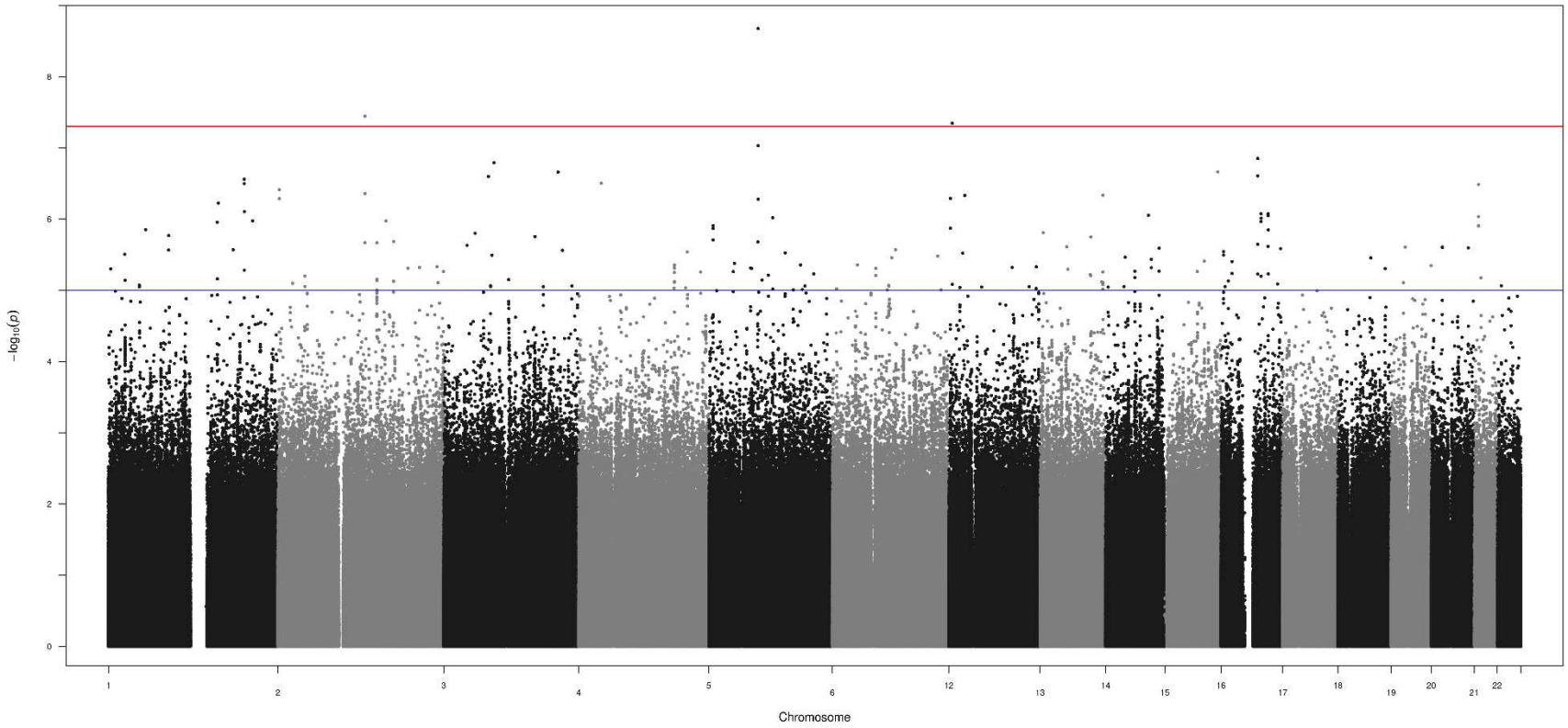
Supplementary Figure 1a. Manhattan plots for rare variant associations in discovery analysis for SBP.



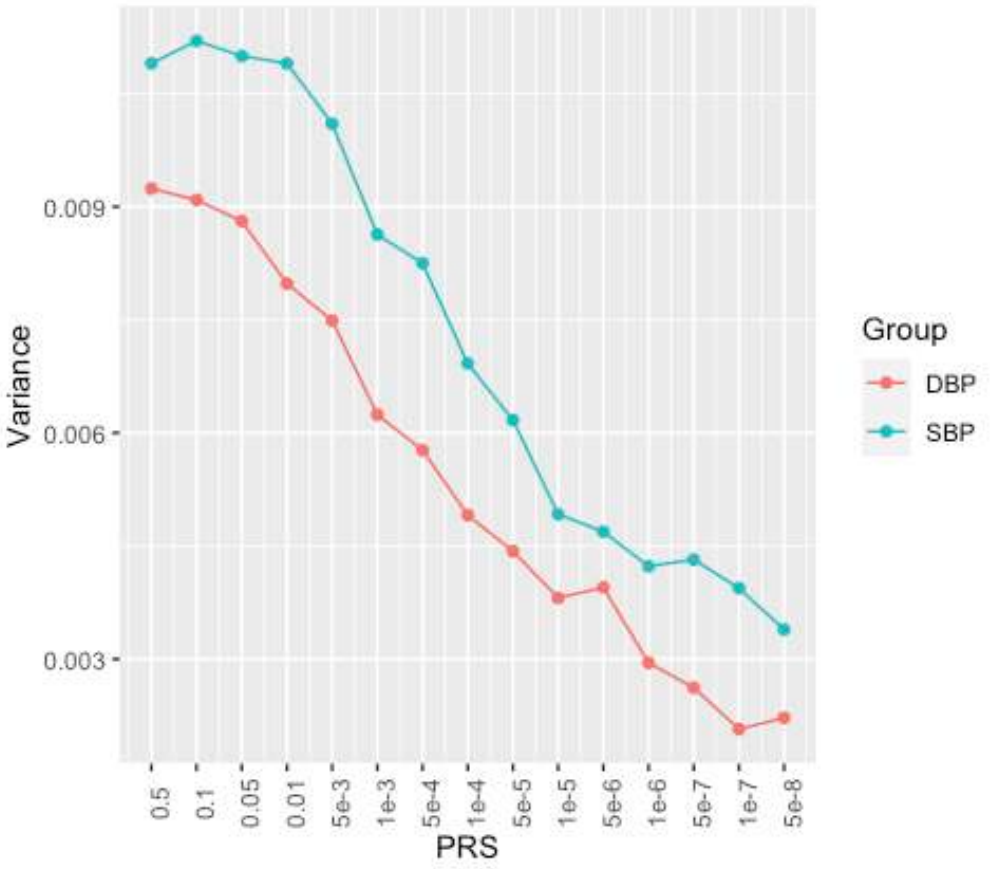
Supplementary Figure 1B. Manhattan plots for rare variant associations in discovery analysis for DBP



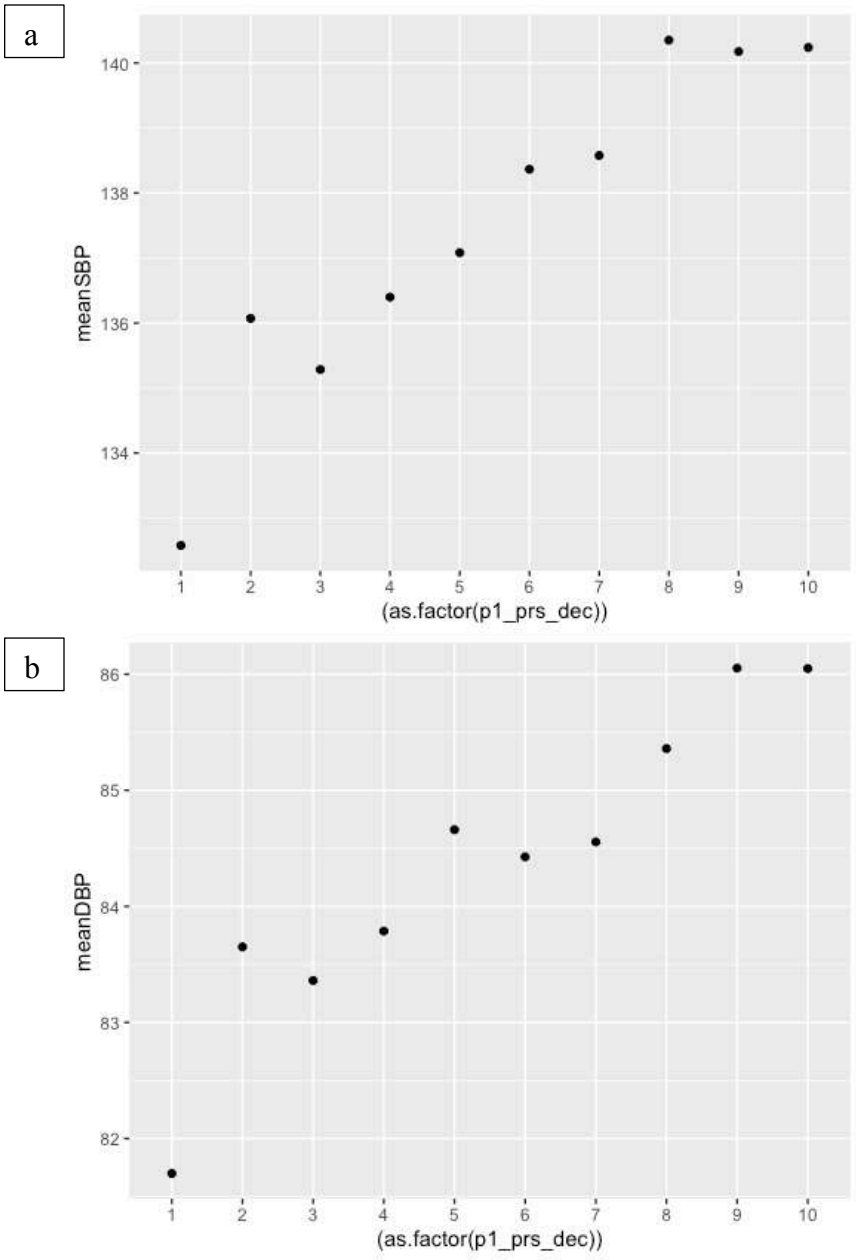
Supplementary Figure 1C. Manhattan plots for rare variant associations in discovery analysis for pulse pressure.



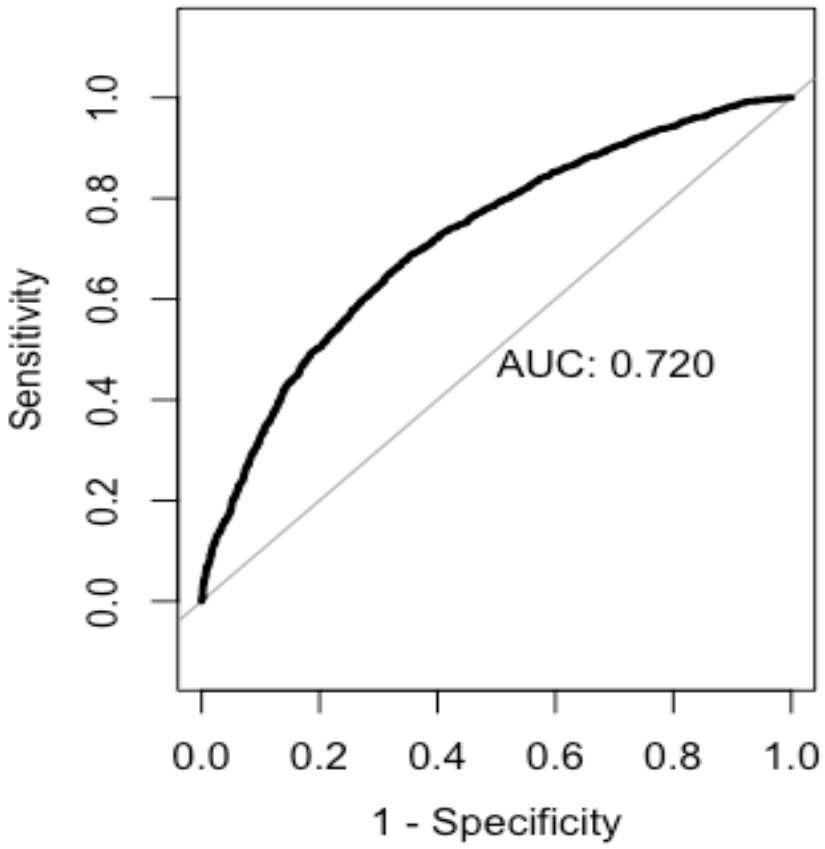
Supplementary Figure 1D. Manhattan plots for rare variant associations in discovery analysis for hypertension status.



Supplementary Figure 2. Optimization of polygenic scores across p-value thresholds for SBP and DBP in BioVU African-ancestry individuals.



Supplementary Figure 3. Relationship between SBP (a) and DBP (b) PRS deciles with measured SBP and DBP, respectively, in UK Biobank.



Supplementary Figure 4. Area under the receiver operator curve for logistic model for hypertension status as a function of both PRS, age, age-squared, sex, BMI, and African ancestry proportions, fit in BioVU and applied to UKB.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AABPSupplementaryTables10282021.xlsx
- SuppTableLegends.docx
- SupplementaryAcknowledgementsv2.docx