SUPPLEMENTARY NOTE

Trans-ethnic association study of blood pressure determinants in over 750,000 individuals

Ayush Giri[1,2]*, Jacklyn N. Hellwege[2,3]*, Jacob M. Keaton[2,3]*, Jihwan Park[4]*, Chengxiang Qiu[4], Helen R. Warren[5,6], Eric S. Torstenson[2,3], Csaba P. Kovesdy[7], Yan V. Sun[8,9], Otis D. Wilson[2,10], Cassianne Robinson-Cohen[10], Christianne L. Roumie[11,12], Cecilia P. Chung[13], Kelly A. Birdwell[10,14], Scott M. Damrauer[15,16], Scott L. DuVall[17,18], Derek Klarin[19,20,21,22], Kelly Cho[23,24,25], Yu Wang[26], Evangelos Evangelou[27,28], Claudia P. Cabrera[5,6], Louise V. Wain[29,30], Rojesh Shrestha[4], Brian S. Mautz[3], Elvis A. Akwo[10], Muralidharan Sargurupremraj[31], Stéphanie Debette[31,32], Michael Boehnke[33], Laura J. Scott[33], Jian'an Luan[34], Jing-Hua Zhao[34], Sara M. Willems[34], Sébastien Thériault[35,36], Nabi Shah[37,38], Christopher Oldmeadow[39], Peter Almgren[40], Ruifang Li-Gao[41], Niek Verweij[42], Thibaud S. Boutin[43], Massimo Mangino[44,45], Ioanna Ntalla[5], Elena Feofanova[46], Praveen Surendran[47], James P. Cook[48], Savita Karthikeyan[47], Najim Lahrouchi[21,49,50], Chunyu Liu[51], Nuno Sepúlveda[52], Tom G. Richardson[53], Aldi Kraja[54,55,56], Philippe Amouyel[57], Martin Farrall[58], Neil R. Poulter[59], Understanding Society Scientific Group[60], International Consortium for Blood Pressure[60], Blood Pressure-International Consortium of Exome chip studies[60], Markku Laakso[61], Eleftheria Zeggini[62], Peter Sever[63], Robert A. Scott[34], Claudia Langenberg[34], Nicholas J. Wareham[34], David Conen[64], Colin Neil Alexander Palmer[37], John Attia[39,65], Daniel I. Chasman[66], Paul M. Ridker[66], Olle Melander[40], Dennis Owen Mook-Kanamori[41], Pim van der Harst[42], Francesco Cucca[67,68], David Schlessinger[69], Caroline Hayward[43], Tim D. Spector[44], Marjo-Riitta Jarvelin[70,71,72,73,74,] Branwen J. Hennig[75,76,77], Nicholas J. Timpson[53], Wei-Qi Wei[78], Joshua C. Smith[78], Yaomin Xu[26,78], Michael E. Matheny[11,12,26,78], Edward E. Siew[10,12], Cecilia Lindgren[21,79,80], Karl-Heinz Herzig[81,82], George Dedoussis[83], Joshua C. Denny[78], Bruce M. Psaty[84,85,86,87], Joanna M. M. Howson[47], Patricia B. Munroe[5,6], Christopher Newton-Cheh[49], Mark J. Caulfield[5,6], Paul Elliott[70,88,89], J. Michael Gaziano[23,66], John Concato[90,91], Peter W.F. Wilson[92,93], Philip S. Tsao[94,95], Digna R. Velez Edwards[1,2,78], Katalin Susztak[4,96†], Million Veteran Program[60], Christopher J. O'Donnell[66,97†], Adriana M. Hung[2,10†Δ], and Todd L. Edwards[2,3†Δ]

[1]Division of Quantitative Sciences, Department of Obstetrics & Gynecology, Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [2]Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, Tennessee, USA. [3]Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Vanderbilt University, Nashville, Tennessee, USA. [4]Department of Medicine, Renal Electrolyte and Hypertension Division, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [5]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [6]National Institute for Health Research Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, UK. [7]Nephrology Section, Memphis VA Medical Center, Memphis, Tennessee, USA. [8]Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, Georgia, USA. [9]Department of Biomedical Informatics, Emory University School of Medicine,

Atlanta, Georgia, USA. [10]Division of Nephrology and Hypertension, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [11]Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [12]Geriatrics Research Education and Clinical Center, Tennessee Valley Health System, Veteran's Health Administration, Nashville, Tennessee, USA. [13]Division of Rheumatology and Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [14]Division of Nephrology, Department of Medicine, Nashville Veteran Affairs Hospital, Nashville, Tennessee, USA. [15]Department of Surgery, Corporal Michael Crescenz VA Medical Center, Philadelphia, Pennsylvania, USA. [16]Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [17]VA Salt Lake City Health Care System, Salt Lake City, Utah, USA. [18]University of Utah School of Medicine, Salt Lake City, Utah, USA. [19]VA Boston Health Care System, Boston, Massachusetts, USA. [20]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. [21]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [22]Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. [23]Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, Massachusetts, USA. [24]Division of Aging, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. [25]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. [26]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [27]Department of Epidemiology and Biostatistics, Imperial College London, London, UK. [28]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece. [29]Department of Health Sciences, University of Leicester, Leicester, UK. [30]National Institute for Health Research, Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK. [31]University of Bordeaux, Bordeaux Population Health Research Center, INSERM UMR 1219, Bordeaux, France. [32]Department of Neurology, Bordeaux University Hospital, Bordeaux, France. [33]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA. [34]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK. [35]Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada. [36]Department of Molecular Biology, Medical Biochemistry and Pathology, Laval University, Quebec City, Quebec, Canada. [37]Division of Molecular and Clinical Medicine, Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. [38]Department of Pharmacy, COMSATS University Islamabad, Abbottabad, Pakistan. [39]Hunter Medical Research Institute, Newcastle, New South Wales, Australia. [40]Department of Clinical Sciences, Lund University, Malmö, Sweden. [41]Leiden University Medical Center, Leiden, The Netherlands. [42]Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [43]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [44]Department of Twin Research and Genetic Epidemiology, Kings College London, London, UK. [45]NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London, UK. [46]Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, Texas, USA. [47]BHF

Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [48]Department of Biostatistics, University of Liverpool, Liverpool, UK. [49]Cardiovascular Research Center, Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. [50]Amsterdam UMC, University of Amsterdam, Heart Center, Department of Clinical and Experimental Cardiology, Amsterdam Cardiovascular Sciences Amsterdam, The Netherlands. [51]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [52]Immunology and Infection Department, London School of Hygiene & Tropical Medicine, London, UK. [53]MRC Integrative Epidemiology Unit (IEU), Bristol Medical School (Population Health Sciences), University of Bristol, Oakfield House, Oakfield Grove, Bristol, UK. [54]Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA. [55]Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, Missouri, USA. [56]Division of Statistical Genomics, Washington University School of Medicine, St. Louis, Missouri, USA. [57]University of Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167 - RID-AGE - Risk factors and molecular determinants of aging-related diseases, Lille, France. [58]Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, Oxford, UK. [59]International Centre for Circulatory Health, Imperial College London, London, UK. [60]A list of consortium members and affiliations is presented in the Supplementary Note. [61]University of Eastern Finland, School of Medicine, Kuopio, Finland. [62]Wellcome Trust Sanger Institute, Hinxton, UK. [63]National Heart and Lung Institute, Imperial College London, Hammersmith Campus, London, UK. [64]Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada. [65]Faculty of Health, University of Newcastle, Newcastle, New South Wales, Australia. [66]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [67]Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy. [68]Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, Italy. [69]Laboratory of Genetics and Genomics, National Institute on Aging, NIH, Baltimore, Maryland, USA. [70]MRC-PHE Centre for Environment & Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. [71]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland. [72]Biocenter Oulu, University of Oulu, Oulu, Finland. [73]Unit of Primary Health Care, Oulu University Hospital, OYS, Oulu, Finland. [74]Department of Life Sciences, College of Health and Life Sciences, Brunel University London, Uxbridge, Middlesex, UK. [75]Wellcome Trust, London, UK. [76]MRC Unit The Gambia, Atlantic Boulevard, Fajara, Banjul, The Gambia. [77]London School of Hygiene & Tropical Medicine, London, UK. [78]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [79]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [80]Big Data Institute at the Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. [81]Institute of Biomedicine, Biocenter of Oulu, Medical Research Center, Oulu University and Oulu University Hospital, Oulu, Finland. [82]Department of Gastroenterology and Metabolism, Poznan University of Medical Sciences, Poznan, Poland. [83]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, Greece. [84]Departments of Medicine, University of Washington, Seattle, Washington, USA. [85]Departments of Epidemiology, University of Washington, Seattle, Washington, USA. [86]Departments of Health Services,

University of Washington, Seattle, Washington, USA. [87]Kaiser Permanente Washington Health Research Institute, Seattle, Washington, USA. [88]National Institute for Health Research Imperial Biomedical Research Centre, Imperial College Healthcare NHS Trust, Imperial College London, London, UK. [89]UK Dementia Research Institute at Imperial College London, London, UK. [90]Clinical Epidemiology Research Center (CERC), VA Cooperative Studies Program, VA Connecticut Healthcare System, West Haven, Connecticut, USA. [91]Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut, USA. [92]Atlanta VA Medical Center, Atlanta, Georgia, USA. [93]Emory Clinical Cardiovascular Research Institute, Atlanta, Georgia, USA. [94]VA Palo Alto Health Care System, Palo Alto, California, USA. [95]Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California, USA. [96]Department of Genetics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, Pennsylvania, USA. [97]VA Boston Healthcare, Section of Cardiology and Department of Medicine, Boston, Massachusetts, USA.

* Indicates that these authors contributed equally to this work and share the first author position

† Indicates that these authors contributed equally to this work and share the senior author position

Δ Corresponding authors

Correspondence should be addressed to:

Adriana M. Hung

1161 21st Ave. S., MCN S-3223
Nashville, TN 37232
Vanderbilt University Medical Center
adriana.hung@vanderbilt.edu

Todd L. Edwards

Office 615, Suite 600
2525 West End Ave.
Vanderbilt University Medical Center
Nashville, TN 37204
todd.l.edwards@vanderbilt.edu

Supplementary Table Descriptions

Supplementary Table 1. Descriptive summary statistics of MVP participants. Shows mean (SD) for age and body mass index and % for sex, blood pressure-lowering meds (medication) % diabetes, mean (SD) for systolic blood pressure (SBP), diastolic blood pressure (DBP) and pulse pressure. Ethnicities are grouped by self-reported and clustering from PCA and ordered by sample size. Race/ethnicity - race or ethnicity either reported in the EHR or inferred from PCA clustering; N - number of subjects; %Male - percentage of subjects with male sex; %Diabetes - percentage of subjects diagnosed with diabetes in the HER. Age (years) (SD) - mean subject age in years and corresponding standard deviation; BMI (kg/m^2) (SD) - body mass index in kilograms per meter squared and corresponding standard deviation; SBP (mmHg) (SD) - systolic blood pressure in millimeters of mercury and corresponding standard deviation; DBP (mmHg) (SD) - diastolic blood pressure in millimeters of mercury and corresponding standard deviation; pulse pressure (mmHg) (SD) - pulse pressure in millimeters of mercury and corresponding standard deviation; Hypertensive Meds (%) - percentage of subjects prescribed antihypertensive medications in the HER.

Supplementary Table 2a. Association results for all sentinel SBP SNPs in previously reported loci. SNPs are ordered by chromosome and position. rsID - dbSNP accession number; CHR:BP - chromosome, build 37 position; Effect allele - allele corresponding to measured effect; Other allele - allele not corresponding to measured effect; N_{eff} - effective number of subjects in discovery meta-analysis; EAF - effect allele frequency in discovery meta-analysis; Effect - measured effect in discovery meta-analysis; SE - standard error of measured effect in discovery meta-analysis; P-value - association p-value for measured effect in discovery meta-analysis; Nearest Gene - most proximal gene within 250kb of sentinel SNP; Distance - distance in base pairs from sentinel SNP to nearest gene; Location - location of sentinel SNP relative to nearest gene. Maximum effective n=459,670 biologically independent samples. Two-sided Wald test performed to obtain z-scores and resulting p-values. Genome-wide significance threshold ($P<5x10^{-8}$) used to correct for multiple testing.

Supplementary Table 2b. Association results for all sentinel DBP SNPs in previously reported loci. SNPs are ordered by chromosome and position. rsID - dbSNP accession number; CHR:BP - chromosome, build 37 position; Effect allele - allele corresponding to measured effect; Other allele - allele not corresponding to measured effect; N_{eff} - effective number of subjects in discovery meta-analysis; EAF - effect allele frequency in discovery meta-analysis; Effect - measured effect in discovery meta-analysis; SE - standard error of measured effect in discovery meta-analysis; P-value - association p-value for measured effect in discovery meta-analysis; Nearest Gene - most proximal gene within 250kb of sentinel SNP; Distance - distance in base pairs from sentinel SNP to nearest gene; Location - location of sentinel SNP relative to nearest gene. Maximum effective n=459,093 biologically independent samples. Two-sided Wald test performed to obtain z-scores and resulting p-values. Genome-wide significance threshold ($P<5x10^{-8}$) used to correct for multiple testing.

Supplementary Table 2c. Association results for all sentinel pulse pressure SNPs in previously reported loci. SNPs are ordered by chromosome and position. rsID - dbSNP

accession number; CHR:BP - chromosome, build 37 position; Effect allele - allele corresponding to measured effect; Other allele - allele not corresponding to measured effect; N_{eff} - effective number of subjects in discovery meta-analysis; EAF - effect allele frequency in discovery meta-analysis; Effect - measured effect in discovery meta-analysis; SE - standard error of measured effect in discovery meta-analysis; P-value - association p-value for measured effect in discovery meta-analysis; Nearest Gene - most proximal gene within 250kb of sentinel SNP; Distance - distance in base pairs from sentinel SNP to nearest gene; Location - location of sentinel SNP relative to nearest gene. Maximum effective n=459,305 biologically independent samples. Two-sided Wald test performed to obtain z-scores and resulting p-values. Genome-wide significance threshold ($P<5x10^{-8}$) used to correct for multiple testing.

Supplementary Table 3a. Association results for all 128 SBP novel replicated variants from two-stage analysis. SNPs ordered by replication tier, chromosome, and position. rsID - dbSNP accession number; CHR:BP - chromosome, build 37 position; Nearest Gene - most proximal gene within 250kb of sentinel SNP; Distance - base pairs from sentinel SNP to nearest gene; Location - location of sentinel SNP relative to nearest gene; Effect allele - allele corresponding to measured effect; Other allele - allele not corresponding to effect; EAF_{comb} - effect allele frequency in combined discovery and replication meta-analysis; $Effect_{comb}$ - measured effect in combined discovery and replication meta-analysis; SE_{comb} - standard error of measured effect in combined discovery and replication meta-analysis; $P\text{-value}_{comb}$ - association p-value for measured effect in combined discovery and replication meta-analysis; $N_{eff\ comb}$ - effective number of subjects in combined discovery and replication meta-analysis; P_{het} - Cochran's Q test p-value in combined discovery and replication meta-analysis; I^2 - percentage of variation across studies due to statistical heterogeneity; Tier - evidentiary association tier; EAF_{disc} - effect allele frequency in discovery meta-analysis; $Effect_{disc}$ - measured effect in discovery meta-analysis; SE_{disc} - standard error of measured effect in discovery meta-analysis; $P\text{-value}_{disc}$ - association p-value for measured effect in discovery meta-analysis; $Effect_{rep}$ - measured effect in replication meta-analysis; SE_{rep} - standard error of measured effect in replication meta-analysis; $P\text{-value}_{rep}$ - association p-value for measured effect in replication meta-analysis; N - number of subjects in ethnicity-specific discovery meta-analysis. Maximum effective n=760,226 biologically independent samples. Two-sided Wald test performed to obtain z-scores and p-values. Genome-wide significance threshold ($P<5x10^{-8}$) used to correct for multiple testing.

Supplementary Table 3b. Association results for DBP novel replicated sentinel variants from two-stage analysis. SNPs ordered by replication tier, chromosome, and position. rsID - dbSNP accession number; CHR:BP - chromosome, build 37 position; Nearest Gene - most proximal gene within 250kb of sentinel SNP; Distance - base pairs from sentinel SNP to nearest gene; Location - location of sentinel SNP relative to nearest gene; Effect allele - allele corresponding to measured effect; Other allele - allele not corresponding to effect; EAF_{comb} - effect allele frequency in combined discovery and replication meta-analysis; $Effect_{comb}$ - measured effect in combined discovery and replication meta-analysis; SE_{comb} - standard error of measured effect in combined discovery and replication meta-analysis; $P\text{-value}_{comb}$ - association p-value for measured effect in combined discovery and replication meta-analysis; $N_{eff\ comb}$ - effective number of subjects in combined discovery and replication meta-analysis; P_{het} -

Cochran's Q test p-value in combined discovery and replication meta-analysis; I^2 - percentage of variation across studies due to statistical heterogeneity; Tier - evidentiary association tier; EAF_{disc} - effect allele frequency in discovery meta-analysis; $Effect_{disc}$ - measured effect in discovery meta-analysis; SE_{disc} - standard error of measured effect in discovery meta-analysis; $P\text{-value}_{disc}$ - association p-value for measured effect in discovery meta-analysis; $Effect_{rep}$ - measured effect in replication meta-analysis; SE_{rep} - standard error of measured effect in replication meta-analysis; $P\text{-value}_{rep}$ - association p-value for measured effect in replication meta-analysis; N - number of subjects in ethnicity-specific discovery meta-analysis. Maximum effective n=767,920 biologically independent samples. Two-sided Wald test performed to obtain z-scores and p-values. Genome-wide significance threshold ($P<5x10^{-8}$) used to correct for multiple testing.

Supplementary Table 3c. Association results for all 126 pulse pressure novel replicated sentinel variants from two-stage analysis. SNPs ordered by replication tier, chromosome, and position. rsID - dbSNP accession number; CHR:BP - chromosome, build 37 position; Nearest Gene - most proximal gene within 250kb of sentinel SNP; Distance - base pairs from sentinel SNP to nearest gene; Location - location of sentinel SNP relative to nearest gene; Effect allele - allele corresponding to measured effect; Other allele - allele not corresponding to effect; EAF_{comb} - effect allele frequency in combined discovery and replication meta-analysis; $Effect_{comb}$ - measured effect in combined discovery and replication meta-analysis; SE_{comb} - standard error of measured effect in combined discovery and replication meta-analysis; $P\text{-value}_{comb}$ - association p-value for measured effect in combined discovery and replication meta-analysis; $N_{eff\ comb}$ - effective number of subjects in combined discovery and replication meta-analysis; P_{het} - Cochran's Q test p-value in combined discovery and replication meta-analysis; I^2 - percentage of variation across studies due to statistical heterogeneity; Tier - evidentiary association tier; EAF_{disc} - effect allele frequency in discovery meta-analysis; $Effect_{disc}$ - measured effect in discovery meta-analysis; SE_{disc} - standard error of measured effect in discovery meta-analysis; $P\text{-value}_{disc}$ - association p-value for measured effect in discovery meta-analysis; $Effect_{rep}$ - measured effect in replication meta-analysis; SE_{rep} - standard error of measured effect in replication meta-analysis; $P\text{-value}_{rep}$ - association p-value for measured effect in replication meta-analysis; N - number of subjects in ethnicity-specific discovery meta-analysis. Maximum effective n=759,768 biologically independent samples. Two-sided Wald test performed to obtain z-scores and p-values. Genome-wide significance threshold ($P<5x10^{-8}$) used to correct for multiple testing.

Supplementary Table 4. Conditional association results for all jointly conditional SNPs. SNPs are ordered by chromosome, position. rsID - dbSNP accession number; CHR:BP - chromosome, build 37 position; CHR:BP - chromosome, build 37 position; Nearest Gene - most proximal gene within 250kb of sentinel SNP; Distance - distance in base pairs from sentinel SNP to nearest gene; Location - location of sentinel SNP relative to nearest gene; Effect allele - allele corresponding to measured effect; Trait(best) - blood pressure trait for which association p-value for measured effect in discovery meta-analysis was most significant; EAF_{disc} - effect allele frequency in combined discovery and replication meta-analysis; $Effect_{disc}$ - measured effect in discovery meta-analysis; SE_{disc} - standard error of measured effect in discovery meta-analysis; P-

$value_{disc}$ - association p-value for measured effect in discovery meta-analysis; Lead SNP(s) - SNP(s) with most significant association p-value for measured effect in discovery meta-analysis on which the SNP in the rsID column was conditioned; Novel/Known - indicator of whether locus was previously reported or novel in our analyses; Nearest Gene(s) - most proximal gene within 500kb of Lead SNP(s); R^2 - linkage disequilibrium correlation between SNP in rsID column and Lead SNP(s); $Effect_{cond}$ - measured effect of SNP in the rsID column in the genome-wide joint conditional analysis; SE_{cond} - standard error of the measured effect of SNP in the rsID column in the genome-wide joint conditional analysis; $P\text{-}value_{cond}$ - association p-value for the measured effect of SNP in the rsID column in the genome-wide joint conditional analysis. Maximum effective n=455,994 biologically independent samples. Two-sided Wald test performed to obtain z-scores and p-values. Genome-wide significance threshold ($P<5x10^{-8}$) used to correct for multiple testing.

Supplementary Table 5. Summary statistics for regression of effect estimates between MVP race/ethnic groups at known and novel SNPs. Comparison - description of ethnicities compared; N SNPs - number of SNPs available for comparison between ethnicities; All - slope of the best-fit line or coefficient of determination from regression of effect estimates between ethnicities for SNPs from previously reported loci and loci novel in our analyses; Novel - slope of the best-fit line or coefficient of determination from regression of effect estimates between ethnicities for SNPs from loci novel in our analyses; Known - slope of the best-fit line or coefficient of determination from regression of effect estimates between ethnicities for SNPs from previously reported loci.

Supplementary Table 6. Conditional analysis of missense variants identified from meta-analysis. Table presents results from analyses for coding variants presented in Table 2 with and without conditioning on select SNPs in MVP Whites Discovery Analysis. SNPs are ordered by chromosome and position. rsID - dbSNP accession number; CHR:BP - chromosome, build 37 position; Nearest Gene - most proximal gene within 250kb of sentinel SNP; Effect allele - allele corresponding to measured effect; Other allele - allele not corresponding to measured effect; Trait- Trait used in conditional analysis; EAF_{disc} - effect allele frequency in combined discovery and replication meta-analysis; $Effect_{disc}$ - measured effect in discovery meta-analysis; SE_{disc} - standard error of measured effect in discovery meta-analysis; $P\text{-}value_{disc}$ - association p-value for measured effect in discovery meta-analysis; Lead SNP - SNP with most significant association p-value for measured effect in discovery meta-analysis on which the SNP in the rsID column was conditioned; Novel/Known - indicator of whether locus was previously reported or novel in our analyses; R^2 - linkage disequilibrium correlation between SNP in rsID column and Lead SNP; $Effect_{cond}$ - measured effect of SNP in the rsID column in conditional analysis; SE_{cond} - standard error of the measured effect of SNP in the rsID column in conditional analysis; $P\text{-}value_{cond}$ - association p-value for the measured effect of SNP in the rsID column in conditional analysis. U = Unknown, relationship not found in the LDLINK tool. Maximum effective n=864,699 biologically independent samples. Two-sided Wald test performed to obtain z-scores and p-values. Genome-wide significance threshold ($P<5x10^{-8}$) used to correct for multiple testing.

Supplementary Table 7. Rare variant results from meta-analysis with UKB 500K. Table is sorted by replication tier, chromosome, and position. rsID - dbSNP accession number; CHR:BP - chromosome, build 37 position; Nearest Gene - most proximal gene within 250kb of sentinel SNP; Distance - distance in base pairs from sentinel SNP to nearest gene; Location - location of sentinel SNP relative to nearest gene; Effect allele - allele corresponding to measured effect; Other allele - allele not corresponding to measured effect; EAF - effect allele frequency in combined discovery and replication meta-analysis; N_{eff} - effective number of subjects in combined discovery and replication meta-analysis; Info - weighted average imputation info scores across all available discovery and replication datasets; Best Trait - blood pressure trait for which the association p-value for the measured effect in the discovery+replication meta-analysis was most significant; Replication tier - evidentiary tier for association; Effect - measured effect in the discovery+replication meta-analysis; SE - standard error of the measured effect in the discovery+replication meta-analysis; P-value - association p-value for the measured effect in the discovery+replication meta-analysis; N – number of subjects in ethnicity-specific discovery meta-analysis; EAF_{disc} - effect allele frequency in the ethnicity-specific discovery meta-analysis; $Effect_{disc}$ - measured effect in the ethnicity-specific discovery meta-analysis; SE_{disc} - standard error of the measured effect in the ethnicity-specific discovery meta-analysis; $P\text{-value}_{disc}$ - association p-value for the measured effect in the ethnicity-specific discovery meta-analysis.

Supplementary Table 8a. S-PrediXcan results across 45 tissues with SBP. Table is sorted by p-value. Tissue - transcriptome tissue source from GTEx v6 release or Ko et al (kidney); Gene - gene name from the transcriptome model mapped to ensemble genes, generally extracted from Genquant; Z-score - S-PrediXcan's association result for the gene; Effect - S-PrediXcan 's association effect size for the gene; P-value - P-value of the aforementioned statistic; var_g - variance of the gene expression, calculated as W' * G * W (where W is the vector of SNP weights in a gene's model, W' is its transpose, and G is the covariance matrix); pred_perf_r2: R2 of tissue model's correlation to gene's measured transcriptome (prediction performance); pred_perf_pval: pval of tissue model's correlation to gene's measured transcriptome (prediction performance); pred_perf_qval: qval of tissue model's correlation to gene's measured transcriptome (prediction performance); N_SNPs_used: number of SNPs from GWAS that got used in S-PrediXcan analysis; N_SNPs_inModel: number of SNPs in the model. Two-sided Wald test performed to obtain z-scores and p-values. Bonferroni significance threshold ($P<2.5\text{x}10^{-7}$) was used to account for multiple testing.

Supplementary Table 8b. S-PrediXcan results across 45 tissues with DBP. Table is sorted by p-value. Tissue - transcriptome tissue source from GTEx v6 release or Ko et al (kidney); Gene - gene name from the transcriptome model mapped to ensemble genes, generally extracted from Genquant; Z-score - S-PrediXcan's association result for the gene; Effect - S-PrediXcan 's association effect size for the gene; P-value - P-value of the aforementioned statistic; var_g - variance of the gene expression, calculated as W' * G * W (where W is the vector of SNP weights in a gene's model, W' is its transpose, and G is the covariance matrix); pred_perf_r2: R2 of tissue model's correlation to gene's measured transcriptome (prediction performance); pred_perf_pval: pval of tissue model's correlation to gene's measured transcriptome (prediction performance); pred_perf_qval: qval of tissue model's correlation to gene's measured

transcriptome (prediction performance); N_SNPs_used: number of SNPs from GWAS that got used in S-PrediXcan analysis; N_SNPs_inModel: number of SNPs in the model. Two-sided Wald test performed to obtain z-scores and p-values. Bonferroni significance threshold ($P<2.5x10^{-7}$) was used to account for multiple testing.

Supplementary Table 8c. S-PrediXcan results across 45 tissues with pulse pressure. Table is sorted by p-value. Tissue - transcriptome tissue source from GTEx v6 release or Ko et al (kidney); Gene - gene name from the transcriptome model mapped to ensemble genes, generally extracted from Genquant; Z-score - S-PrediXcan's association result for the gene; Effect - S-PrediXcan 's association effect size for the gene; P-value - P-value of the aforementioned statistic; var_g - variance of the gene expression, calculated as W' * G * W (where W is the vector of SNP weights in a gene's model, W' is its transpose, and G is the covariance matrix); pred_perf_r2: R2 of tissue model's correlation to gene's measured transcriptome (prediction performance); pred_perf_pval: pval of tissue model's correlation to gene's measured transcriptome (prediction performance); pred_perf_qval: qval of tissue model's correlation to gene's measured transcriptome (prediction performance); N_SNPs_used: number of SNPs from GWAS that got used in S-PrediXcan analysis; N_SNPs_inModel: number of SNPs in the model. Two-sided Wald test performed to obtain z-scores and p-values. Bonferroni significance threshold ($P<2.5x10^{-7}$) was used to account for multiple testing.

Supplementary Table 9a. Expression from single-cell RNA sequencing of murine kidney cell types of mouse homologs of S-PrediXcan genes associated with SBP in kidney tissue. Gene - mouse homolog of significant gene identified in kidney tissue (see ST8a). Endo - endothelial; Podo - podocyte; PT - proximal tubule; LOH - Loop of Henle; DCT - distal convoluted tubule; CD-PC - collecting duct principal cell; CD-IC - collecting duct intercalated cell; Fib - fibroblast; Macro - macrophage; Neutro - neutrophil; B lymph - B lymphocyte; T lymph - T lymphocyte; NK - natural killer cell. n=7 biologically independent samples. Two-sided Wald test performed to obtain z-scores with no correction for multiple testing.

Supplementary Table 9b. Expression from single-cell RNA sequencing of murine kidney cell types of mouse homologs of S-PrediXcan genes associated with DBP in kidney tissue. Gene - mouse homolog of significant gene identified in kidney tissue (see ST8b). Endo - endothelial; Podo - podocyte; PT - proximal tubule; LOH - Loop of Henle; DCT - distal convoluted tubule; CD-PC - collecting duct principal cell; CD-IC - collecting duct intercalated cell; Fib - fibroblast; Macro - macrophage; Neutro - neutrophil; B lymph - B lymphocyte; T lymph - T lymphocyte; NK - natural killer cell. n=7 biologically independent samples. Two-sided Wald test performed to obtain z-scores with no correction for multiple testing.

Supplementary Table 9c. Expression from single-cell RNA sequencing of murine kidney cell types of mouse homologs of S-PrediXcan genes associated with pulse pressure in kidney tissue. Gene - mouse homolog of significant gene identified in kidney tissue (see ST8c). Endo - endothelial; Podo - podocyte; PT - proximal tubule; LOH - Loop of Henle; DCT - distal convoluted tubule; CD-PC - collecting duct principal cell; CD-IC - collecting duct intercalated cell; Fib - fibroblast; Macro - macrophage; Neutro - neutrophil; B lymph - B lymphocyte; T

lymph - T lymphocyte; NK - natural killer cell. n=7 biologically independent samples. Two-sided Wald test performed to obtain z-scores with no correction for multiple testing.

Supplementary Table 10. Expression of genes from mouse scRNA-seq in human kidney from the Human Protein Atlas. Table is sorted by gene name. Gene - human homolog of mouse scRNS-seq gene available in the Human Protein Atlas; Glomeruli expression - detected gene expression in human glomerular tissue categorized as high, medium, or low; Tubule expression - detected gene expression in human renal tubule tissue categorized as high, medium, or low.

Supplementary Table 11. Known targets for anti-hypertension drugs significant by S-PrediXcan and a summary of the most significant S-PrediXcan result across tissues and blood pressure traits. Table is sorted by p-value. Previously reported blood pressure genes and SNPs listed in last two columns. Gene - gene target of known antihypertensive drug; Known Hypertension Drug – medications with a primary indication for hypertension as identified using the SIDER Side Effect Resource and the DEB2 database; Gene-drug relationship - mechanism by which the drug acts with regard to the targeted gene; Sources- source of information for gene-drug relationship as annotated from DGIdb; Effect - effect size for association of the most significant tissue's predicted gene expression and blood pressure trait from S-PrediXcan; P-value for association of the most significant tissue's predicted gene expression and blood pressure trait from S-PrediXcan; Tissue - the most significant tissue for this gene-trait pair from S-PrediXcan; Blood Pressure Trait - the trait which resulted in the most significant S-PrediXcan gene-tissue pair; Previously Reported Blood Pressure SNP - most strongly associated SNP previously reported within 250kb of S-PrediXcan window for predicted expressed gene; Previously Reported Blood Pressure Gene - nearest gene to previously reported Blood Pressure SNP.

Supplementary Table 12. Significant S-PrediXcan genes that have positive effect sizes in any tissue and are targeted by a non-hypertension drug, with the name of the drug and the primary indication for treatment. Table is sorted by p-value. Previously reported blood pressure genes and SNPs listed in last two columns. Gene - gene target of known antihypertensive drug; Drug - any medications without a primary indication for hypertension as identified using the SIDER Side Effect Resource and the DEB2 database; Gene-drug relationship - mechanism by which the drug acts with regard to the targeted gene; Sources- source of information for gene-drug relationship as annotated from DGIdb; Effect - effect size for association of the most significant tissue's predicted gene expression and blood pressure trait from S-PrediXcan; P-value for association of the most significant tissue's predicted gene expression and blood pressure trait from S-PrediXcan; Tissue - the most significant tissue for this gene-trait pair from S-PrediXcan; Blood Pressure Trait - the trait which resulted in the most significant S-PrediXcan gene-tissue pair; Previously Reported Blood Pressure SNP - most strongly associated SNP previously reported within 250kb of S-PrediXcan window for predicted expressed gene; Previously Reported Blood Pressure Gene - nearest gene to previously reported Blood Pressure SNP. Primary Indication – primary indication for prescription of the drug as annotated from BIDD TTD. Two-sided Wald test performed to obtain z-scores and p-values. Bonferroni significance threshold ($P<2.5\text{x}10^{-7}$) was used to account for multiple testing.

Supplementary Table 13. Genes that are significant by S-PrediXcan, are targeted by a drug, and that have an ADE involving hypertension or hypotension. Table is sorted by p-value. Gene - gene target of drug with an ADE involving hypertension or hypotension; Drug - medications with an ADE involving hypertension or hypotension as identified using the SIDER Side Effect Resource; Gene-drug relationship - mechanism by which the drug acts with regard to the targeted gene; Sources- source of information for gene-drug relationship as annotated from DGIdb; Effect - effect size for association of the most significant tissue's predicted gene expression and blood pressure trait from S-PrediXcan; P-value for association of the most significant tissue's predicted gene expression and blood pressure trait from S-PrediXcan; Tissue - the most significant tissue for this gene-trait pair from S-PrediXcan; Blood Pressure Trait - the trait which resulted in the most significant S-PrediXcan gene-tissue pair; Previously Reported Blood Pressure SNP - most strongly associated SNP previously reported within 250kb of S-PrediXcan window for predicted expressed gene; Previously Reported Blood Pressure Gene - nearest gene to previously reported Blood Pressure SNP; ADE - adverse drug event as annotated from SIDER. Two-sided Wald test performed to obtain z-scores and p-values. Bonferroni significance threshold ($P<2.5x10^{-7}$) was used to account for multiple testing.

Supplementary Table 14. Gene-drug relationships for all significant S-PrediXcan genes. Table is sorted by locus. Gene - gene target of drug significant by S-PrediXcan; Drug - medication targeting gene as annotated from DGIdb; Gene-drug relationship - mechanism by which the drug acts with regard to the targeted gene; Sources- source of information for gene-drug relationship as annotated from DGIdb; Novel or Known Blood Pressure Locus - indicator of whether gene was previously reported as a blood pressure locus or was novel in our analyses.

Supplementary Table 15. Significant phenome-wide associations of blood pressure-trait specific genetic risk score (GRS) in unrelated individuals in MVP by race/ethnicity. Table is sorted by best p-value in whites. PheCode - PheWAS code, a hierarchical grouping of International Classification of Disease, 9th edition (ICD9) codes applied to EMR data, which loosely follow the 3-digit (category) and section groupings defined with the ICD9 code system itself, and have been revised based on statistical co-occurrence, code frequency, and human review; Description - full name of PheCode grouping; Phenotype Group - physiological system to which the PheCode is assigned; N_{Total} - total number of individuals not excluded in analysis of PheCode; N_{Cases} - number of individuals with one or more diagnosis codes corresponding to the PheCode; $N_{controls}$ - number of individuals lacking diagnosis codes or exclusion criteria corresponding to the PheCode; Effect - measured effect size of association between the weighted GRS and PheCode; SE - standard error of the measured effect; P-value - p-value for association of the weighted GRS and the PheCode. Results provided for whites (maximum n=188,008 biologically independent samples), blacks (maximum n=52,530 biologically independent samples) and Hispanics (maximum n=16,735 biologically independent samples) separately. Two-sided Wald test performed to obtain z-scores and resulting p-values. Bonferroni significance threshold ($P<2.75x10^{-5}$; 0.05/1,813) used to account for multiple testing.

Supplementary Table 16. Chi-square test for tissue-specific enrichment by trait. Tissue - transcriptome tissue source from GTEx v6 release or Ko et al (kidney); Significant Genes -

number of significant results for specified tissue-trait combination from ST7a-c; χ^2 - chi-squared statistic comparing the proportion of significant genes in a given tissue to the proportion of significant genes in all other tissues; P-value - p-value for chi-squared statistic with one degree of freedom.

Supplementary Table 17a. Significant (FDR *Q*<0.05) DEPICT tissue enrichment results across all significant (P<5x10^{-8}) known and novel SBP GWAS loci. Table is sorted by p-value. MeSH Tree Numbers - Medical Subject Heading (MeSH) tissue and cell type annotations for which genes are highly expressed; MeSH first level term - name of the most specific term in the MeSH Tree Number; MeSH second level term - name of the least specific term in the MeSH Tree Number; P - nominal p-value for enrichment; FDR < 5% - yes/no indicator of whether the false discovery rate q-value was less than 5%.

Supplementary Table 17b. Significant (FDR *Q*<0.05) DEPICT tissue enrichment results across all significant (*P*<5x10^{-8}) known and novel DBP GWAS loci. Table is sorted by p-value. MeSH Tree Numbers - Medical Subject Heading (MeSH) tissue and cell type annotations for which genes are highly expressed; MeSH first level term - name of the most specific term in the MeSH Tree Number; MeSH second level term - name of the least specific term in the MeSH Tree Number; P - nominal p-value for enrichment; FDR < 5% - yes/no indicator of whether the false discovery rate q-value was less than 5%.

Supplementary Table 17c. Significant (FDR *Q*<0.05) DEPICT tissue enrichment results across all significant (P<5x10^{-8}) known and novel pulse pressure GWAS loci. Table is sorted by p-value. MeSH Tree Numbers - Medical Subject Heading (MeSH) tissue and cell type annotations for which genes are highly expressed; MeSH first level term - name of the most specific term in the MeSH Tree Number; MeSH second level term - name of the least specific term in the MeSH Tree Number; P - nominal p-value for enrichment; FDR < 5% - yes/no indicator of whether the false discovery rate q-value was less than 5%.

Supplementary Table 18a. Significant (FDR *Q*<0.05) DEPICT gene set enrichment results across all significant (P<5x10^{-8}) known and novel SBP GWAS loci. Table is sorted by p-value. Original gene set ID - name of DEPICT gene set tested for enrichment; Original gene set description - description of DEPICT gene set tested for enrichment; P - nominal p-value for enrichment; FDR < 5% - yes/no indicator of whether the false discovery rate q-value was less than 5%.

Supplementary Table 18b. Significant (FDR *Q*<0.05) DEPICT gene set enrichment results across all significant (P<5x10^{-8}) known and novel DBP GWAS loci. Table is sorted by p-value. Original gene set ID - name of DEPICT gene set tested for enrichment; Original gene set description - description of DEPICT gene set tested for enrichment; P - nominal p-value for enrichment; FDR < 5% - yes/no indicator of whether the false discovery rate q-value was less than 5%.

Supplementary Table 18c. Significant (FDR *Q*<0.05) DEPICT gene set enrichment results across all significant (P<5x10^{-8}) known and novel pulse pressure GWAS loci. Table is sorted by p-value. Original gene set ID - name of DEPICT gene set tested for enrichment; Original gene

set description - description of DEPICT gene set tested for enrichment; P - nominal p-value for enrichment; FDR < 5% - yes/no indicator of whether the false discovery rate q-value was less than 5%.

Million Veteran Program (MVP):

MVP Executive Committee

- Co-Chair: J. Michael Gaziano, M.D., M.P.H.
- Co-Chair: Rachel Ramoni, D.M.D., Sc.D.
- Jim Breeling, M.D. (ex-officio)
- Kyong-Mi Chang, M.D.
- Grant Huang, Ph.D.
- Sumitra Muralidhar, Ph.D.
- Christopher J. O'Donnell, M.D., M.P.H.
- Philip S. Tsao, Ph.D.

MVP Program Office

- Sumitra Muralidhar, Ph.D.
- Jennifer Moser, Ph.D.

MVP Recruitment/Enrollment

- Recruitment/Enrollment Director/Deputy Director, Boston – Stacey B. Whitbourne, Ph.D.; Jessica V. Brewer, M.P.H.
- MVP Coordinating Centers
    - Clinical Epidemiology Research Center (CERC), West Haven – John Concato, M.D., M.P.H.
    - Cooperative Studies Program Clinical Research Pharmacy Coordinating Center, Albuquerque - Stuart Warren, J.D., Pharm D.; Dean P. Argyres, M.S.
    - Genomics Coordinating Center, Palo Alto – Philip S. Tsao, Ph.D.
    - Massachusetts Veterans Epidemiology Research Information Center (MAVERIC), Boston - J. Michael Gaziano, M.D., M.P.H.
    - MVP Information Center, Canandaigua – Brady Stephens, M.S.
- Core Biorepository, Boston – Mary T. Brophy M.D., M.P.H.; Donald E. Humphries, Ph.D.
- MVP Informatics, Boston – Nhan Do, M.D.; Shahpoor Shayan
- Data Operations/Analytics, Boston – Xuan-Mai T. Nguyen, Ph.D.

MVP Science

- Genomics - Christopher J. O'Donnell, M.D., M.P.H.; Saiju Pyarajan Ph.D.; Philip S. Tsao, Ph.D.
- Phenomics - Kelly Cho, M.P.H, Ph.D.
- Data and Computational Sciences – Saiju Pyarajan, Ph.D.
- Statistical Genetics – Elizabeth Hauser, Ph.D.; Yan Sun, Ph.D.; Hongyu Zhao, Ph.D.

MVP Local Site Investigators

- Atlanta VA Medical Center (Peter Wilson)
- Bay Pines VA Healthcare System (Rachel McArdle)
- Birmingham VA Medical Center (Louis Dellitalia)
- Cincinnati VA Medical Center (John Harley)
- Clement J. Zablocki VA Medical Center (Jeffrey Whittle)
- Durham VA Medical Center (Jean Beckham)
- Edith Nourse Rogers Memorial Veterans Hospital (John Wells)
- Edward Hines, Jr. VA Medical Center (Salvador Gutierrez)
- Fayetteville VA Medical Center (Gretchen Gibson)
- VA Health Care Upstate New York (Laurence Kaminsky)
- New Mexico VA Health Care System (Gerardo Villareal)
- VA Boston Healthcare System (Scott Kinlay)
- VA Western New York Healthcare System (Junzhe Xu)
- Ralph H. Johnson VA Medical Center (Mark Hamner)
- Wm. Jennings Bryan Dorn VA Medical Center (Kathlyn Sue Haddock)
- VA North Texas Health Care System (Sujata Bhushan)
- Hampton VA Medical Center (Pran Iruvanti)
- Hunter Holmes McGuire VA Medical Center (Michael Godschalk)
- Iowa City VA Health Care System (Zuhair Ballas)
- Jack C. Montgomery VA Medical Center (Malcolm Buford)
- James A. Haley Veterans' Hospital (Stephen Mastorides)
- Louisville VA Medical Center (Jon Klein)
- Manchester VA Medical Center (Nora Ratcliffe)
- Miami VA Health Care System (Hermes Florez)
- Michael E. DeBakey VA Medical Center (Alan Swann)
- Minneapolis VA Health Care System (Maureen Murdoch)
- N. FL/S. GA Veterans Health System (Peruvemba Sriram)
- Northport VA Medical Center (Shing Shing Yeh)
- Overton Brooks VA Medical Center (Ronald Washburn)
- Philadelphia VA Medical Center (Darshana Jhala)
- Phoenix VA Health Care System (Samuel Aguayo)
- Portland VA Medical Center (David Cohen)
- Providence VA Medical Center (Satish Sharma)

- Richard Roudebush VA Medical Center (John Callaghan)
- Salem VA Medical Center (Kris Ann Oursler)
- San Francisco VA Health Care System (Mary Whooley)
- South Texas Veterans Health Care System (Sunil Ahuja)
- Southeast Louisiana Veterans Health Care System (Amparo Gutierrez)
- Southern Arizona VA Health Care System (Ronald Schifman)
- Sioux Falls VA Health Care System (Jennifer Greco)
- St. Louis VA Health Care System (Michael Rauchman)
- Syracuse VA Medical Center (Richard Servatius)
- VA Eastern Kansas Health Care System (Mary Oehlert)
- VA Greater Los Angeles Health Care System (Agnes Wallbom)
- VA Loma Linda Healthcare System (Ronald Fernando)
- VA Long Beach Healthcare System (Timothy Morgan)
- VA Maine Healthcare System (Todd Stapley)
- VA New York Harbor Healthcare System (Scott Sherman)
- VA Pacific Islands Health Care System (Gwenevere Anderson)
- VA Palo Alto Health Care System (Philip Tsao)
- VA Pittsburgh Health Care System (Elif Sonel)
- VA Puget Sound Health Care System (Edward Boyko)
- VA Salt Lake City Health Care System (Laurence Meyer)
- VA San Diego Healthcare System (Samir Gupta)
- VA Southern Nevada Healthcare System (Joseph Fayad)
- VA Tennessee Valley Healthcare System (Adriana Hung)
- Washington DC VA Medical Center (Jack Lichy)
- W.G. (Bill) Hefner VA Medical Center (Robin Hurley)
- White River Junction VA Medical Center (Brooks Robey)
- William S. Middleton Memorial Veterans Hospital (Robert Striker)

This publication does not represent the views of the Department of Veterans Affairs or the United States Government.

Replication in International Consortium for Blood Pressure (ICBP)

ICBP GWAS is an international consortium to investigate blood pressure genetics[1–3]. We combined previously reported post-quality control (QC) GWAS data from 54 studies (N=150,134)[2], with newly available GWAS data from a further 23 independent studies (N=148,890) using a fixed effects inverse variance weighted meta-analysis. All study participants were of European descent and were imputed to either the 1000 Genomes Project Phase 1 integrated release version 3 [March 2012] all ancestry reference panel or the Haplotype Reference Consortium (HRC) panel. The final enlarged ICBP GWAS dataset included 77 studies comprising data from 299,024 individuals from the following cohorts: The initial ICBP GWAS included: AGES (n=3215), ARIC (n=9402), ASPS (n=828), B58C (n=6458), BHS (n=4492),

CHS (n=3254), Cilento study (n=999), COLAUS (n=5404), COROGENE-CTRL (n=1878), CROATIA-Vis (n=945), CROATIA-Split (n=494), CROATIA-Korcula (n=867), EGCUT (n=6395), EGCUT2 (n=1844), EPIC (n=2100), ERF (n=2617), Fenland (n=1357), FHS (n=8096), FINRISK-ctrl (n=861), FINRISK CASE (n=839), FUSION (n=1045), GRAPHIC (n=1010), H2000-CTRL (n=1078), HealthABC (n=1661), HTO (n=1000), INGI-CARL (n=456), INGI-FVG (n=746), INGI-VB (n=1775), IPM (n=300), KORAS3 (n=1590), KORAS4 (n=3748), LBC1921 (n=376), LBC1936 (n=800), LOLIPOP-EW610 (n=927), MESA (n=2678), MICROS (n=1148), MIGEN (n=1214), NESDA (n=2336), NSPHS (n=1005), NTR (n=1490), PHASE (n=4535), PIVUS (n=945), PROCARDIS (n=1652), SHIP (n=4068), ULSAM (n=1114), WGHS (n=23049), YFS (n=1987), ORCADES (n=1908), RS1 (n=5645), RS2 (n=2152), RS3 (n=3018), TRAILS (n=1262), TRAILS-CC (n=282) and TWINGENE (n=9789). The enhanced dataset includes ASCOT-SC (n=2462), ASCOT-UK (n=3803), BRIGHT (n=1791), Dijon 3C (n=4061), EPIC-CVD (n=8375), GAPP (n=1685), HCS (n=2112), GS:SFHS (n=19429), Lifelines (n=13292), JUPITER (n=8719), PREVEND (n=3619), TWINSUK (n=4973), Fenland-GWAS (n=1358), InterAct-GWAS (n=6675) OMICS-EPIC (n=17850) OMICS-Fenland (n=8526) UKHLS (n=7462) GoDARTS-Illumina and GoDarts-Affymetrix (n=7413), NEO (n=5731), MDC (n=5271), SardiNIA (n=6021), METSIM (n=8262).

Understanding Society Scientific Group (USSG) group members contributed to this data :

Michaela Benzeval[1], Jonathan Burton1, Nicholas Buck[1], Annette Jäckle1, Meena Kumari[1], Heather Laurie[1], Peter Lynn[1], Stephen Pudney[1], Birgitta Rabe[1], Dieter Wolke[2]

1) Institute for Social and Economic Research; 2) University of Warwick

Replication in the Blood Pressure-International Consortium for Exomechip

The BP-ICE is a working group in ICBP[2] that was established for studies of exome content and blood pressure. We combined previously reported post-quality control (QC) exome array summary blood pressure association results[4,5] and newly available exome array data and GWAS from up to 87 studies, using a fixed effects meta-analysis. Results were provided for meta-analyses of up to 361,375 participants of European descent and up to 420,704 participants from ALL ancestry analyses. Summary statistics from the ALL ancestry analyses were included in this analysis. The studies involved were:

1) Published exome chip EUR datasets: ASCOT (n=5,703), 1958BC (n=5,864), BRIGHT (n=1,230), CROATIA-Korcula (n=814), DIABNORD (n=912), EGCUT (n=1,785), FINRISK97/02 (n=5,152), GS:SFHS (n=9,832), GLACIER (n=922), GRAPHIC (n=1,887), HELIC-MANOLIS (n=944), HUNT (n=4,735), INCIPE (n=1,995), LBC1921 (n=359), LBC1936 (n=783), LIFELINES (n=1,948), MDC (n=8,268), NFBC1966 (n=1,353), OXBB (n=4,440), PIVUS/ULSAM (n=1,998), TWINS UK (n=689), UHP (n=2,306), ADDITION (n=2,307), SDR/ANDIS (n=2,636), DPS (n=416), DR's EXTRA Study (n=740), FIN-D2D 2007 (n=2,580), FINRISK 2007 (n=1,088), FUSION (n=4,237), Health 2006-2008 (n=3,674), INTER99 (n=5,986), METSIM (8,411), PPP-Botania (n=4,766), SDC (n=498), Veijle (n=1,996),

CCHS (n=8,070), CGPS (n=11,784), CIHDS (n=1,436), EPIC-CVD (n=15,676), MORGAM (n=5,757), PROSPER (n=1,275), WOSCOPS (n=1,337), AGES (5,526), ARIC (N=10,864), BioVu (N=19,885),CARDIA (n=2,175), CHS (n=4,113) ERF (n=1,153), FamHS (n=3,722), FHS (n=7,495), GAPP (n=1,947), HRS (n=9,621), MESA (n=2,505), IPM (n=1,337), RS (n=2,875), SHIP (n=7,161), WGHS (n=21,964) and WHI (n=21,841).

2) Newly available or updated EUR datasets: Airwave (n=13,102), ALSPAC (n=6,529), GoDarts (n=4,824), HELIC-POMAK (n=565), NEO (n=6,117), NFBC1986 (n=3,639), UKHLS (n=7,462), Fenland-CoreExome (n=1,040), InterAct-CoreExome (n=10,915), EPIC-Norfolk(n=17,850), Fenland-Omics (n=8,526), Fenland-GWAS (n=1,358) and EPIC-InterAct-GWAS (n=6,675).

The ALL ancestry sample for the final enlarged BP-ICE exome dataset comprised data from 74 studies of EUR ancestry (described above), and data from individuals of African ancestry: the Gambia (n=605); African American ancestry: ARIC (n=3,354), BioVu (n=2,018), CARDIA (n=1,975), CHS (n=789), JHS (n=2,300), HRS (n=2,026), MESA (n=1,658), IPM (n=2,835), WHI (n=3,515); South Asian ancestry: BRAVE (n=5,250), PROMIS (n=25,012), LOLIPOP (n=2,641) East Asians: MESA (n=770) and Hispanic Ancestry: MESA (n=1,440) and IPM (n=3,141).

FULL ACKNOWLEDGEMENTS AND FUNDING STATEMENTS

GoDARTS

Acknowledgements and Funding

The Wellcome Trust United Kingdom Type 2 Diabetes Case Control Collection (GoDARTS) was funded by The Wellcome Trust (072960/Z/03/Z, 084726/Z/08/Z, 084727/Z/08/Z, 085475/Z/08/Z, 085475/B/08/Z) and as part of the EU IMI-SUMMIT program.

The Women's Genome Health Study (WGHS)

Acknowledgements and Funding

WGHS is supported by the National Heart, Lung, and Blood Institute (HL043851 and HL080467) and the National Cancer Institute (CA047988 and UM1CA182913) with collaborative scientific support and funding for genotyping provided by Amgen. JUPITER: The JUPITER trial was funded by AstraZeneca, who also provided collaborative support for genotyping.

Malmö Diet and Cancer Study

Acknowledgements and Funding

Malmö Diet and Cancer Study received funding by the Knut and Alice Wallenberg Foundation.

Netherlands Epidemiology in Obesity (NEO)

Acknowledgements and Funding

The authors of the NEO study thank all individuals who participated in the Netherlands Epidemiology in Obesity study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group, Pat van Beelen, Petra Noordijk and Ingeborg de Jonge for the coordination, lab and data management of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area Vascular and Regenerative Medicine. Dennis Mook-Kanamori is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023).

CROATIA_Vis ,CROATIA_Korcula and CROATIA_Split

Acknowledgements and Funding

The CROATIA_Vis ,CROATIA_Korcula and CROATIA_Split studies were funded by grants from the Medical Research Council (UK), European Commission Framework 6 project

EUROSPAN (Contract No. LSHG-CT-2006-018947) and Republic of Croatia Ministry of Science, Education and Sports research grants. (108-1080315-0302). We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools, Institute for Anthropological Research in Zagreb and Croatian Institute for Public Health.

Generation Scotland

Acknowledgements and Funding

Generation Scotland received core funding from the Chief Scientist Office of the Scottish Government Health Directorate CZD/16/6 and the Scottish Funding Council HR03006. Genotyping of the GS:SFHS samples was carried out by staff at the Genetics Core Laboratory at the Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Wellcome Trust 104036/Z/14/Z (STRADL, Stratifying Resilience and Depression Longitudinally) and the UK's Medical Research Council. We thank all families and practitioners involved in the recruitment process as well as the entirety of Generation Scotland team; interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses.

TwinsUKstudy

Acknowledgements and Funding

The TwinsUK study was funded by the Wellcome Trust, Medical Research Council, and European Union. The study also receives support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR.

KWLPS (Gambia) cohort

Acknowledgements and funding.

The KWLPS (Gambia) cohort is supported through funding was received from the UK Medical Research Council (MRC) and the UK Department for International Development (DFID), under the MRC/DFID Concordat agreement (MC-A760-5QX00, U105960371 and U123261351). We thank all residents of the villages of Kiang West, The Gambia, for their willingness to participate in our studies. Thanks also go to field, laboratory, clinical, data, and administrative staff at MRC Keneba, and in particular Mohammed Ngum as well as past and present members of the Keneba Biobank team, who facilitated the collection and processing of data and samples in The Gamiba that form the basis of these analyses. Thanks are further due to Josyf C Mychaleckyj and Uma Nayak (University of Virginia, USA), Matt Silver and Modou Jobe (MRC Unit The Gambia),

Vickie S. Braithwaite (MRC Human Nutrition Research, Cambridge, UK) and Kerra Pearce (UCL Genomics) for their assistance with genotyping and/or data analyses.

Avon Longitudinal Study of Parents and Children Study (ALSPAC)

Acknowledgements and funding

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is a programme lead in the MRC Integrative Epidemiology Unit (MC_UU_12013/3) and works within the University of Bristol NIHR Biomedical Research Centre (BRC). TGR is a UKRI Innovation Research Fellow (MR/S003886/1).

Hellenic Isolated Cohorts - Minoan Isolates Study

Acknowledgements and funding

Hellenic Isolated Cohorts - Minoan Isolates Study was funded by the Wellcome Trust (098051) and the European Research Council (ERC-2011-StG 280559-SEPI). The MANOLIS cohort is named in honour of Manolis Giannakakis, 1978-2010. We thank the residents of the Mylopotamos villages for taking part. The HELIC study has been supported by many individuals who have contributed to sample collection (including A. Athanasiadis, O. Balafouti, C. Batzaki, G. Daskalakis, E. Emmanouil, C. Giannakaki, M. Giannakopoulou, A. Kaparou, V. Kariakli, S. Koinaki, D. Kokori, M. Konidari, H. Koundouraki, D. Koutoukidis, V. Mamakou, E. Mamalaki, E. Mpamiaki, M. Tsoukana, D. Tzakou, K. Vosdogianni, N. Xenaki, E. Zengini), data entry (T. Antonos, D. Papagrigoriou, B. Spiliopoulou), sample logistics (S. Edkins, E. Gray), genotyping (R. Andrews, H. Blackburn, D. Simpkin, S. Whitehead), research administration (A. Kolb-Kokocinski, S. Smee, D. Walker) and informatics (M. Pollard, J. Randall).

Cardiovascular Health Study (CHS)

Acknowledgements and funding

CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, and R01HL130114 with additional contribution from the National Institute of Neurological Disorders and Stroke

(NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by CTSI grant UL1TR000124 and DK063491. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Individual Acknowledgements and funding

Todd L. Edwards as supported by NIH/NHLBI grant HL121429. This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by award I01BX003360 to Adriana Hung (AH). AH is also supported by VA grant 1I01CX000982.This work was supported using resources and facilities of the VA Informatics and Computing Infrastructure (VINCI), VA HSR RES 13-457. Philip S. Tsao, and Scott M. Damrauer were funded by the *Genetics of Cardiometabolic Diseases in the VA Population* Veterans Affairs Office of Research and Development (BX-003362-01) grant. Peter Wilson was funded by the Veterans Affairs Merit Award I01-01BX003340 (Wilson)

Ayush Giri is supported by the Building Interdisciplinary Research Careers in Women's Health career development program's 2K12HD043483-17 (PI: Katherine E Hartmann). Digna R. Velez Edwards received support from NIH/NHLBI (HL121429). Jacklyn N. Hellwege and Brian S. Mautz were supported by the Vanderbilt Molecular and Genetic Epidemiology of Cancer (MAGEC) training program, funded by T32CA160056 (PI: X.-O. Shu).
Jacob M Keaton is supported by the Vanderbilt Genomic Medicine Training Program. The Vanderbilt Genomic Medicine Training Program is supported by an institutional training grant (T32HG008341) from the National Human Genome Research Institute of the National Institute of Health.

Yan V. Sun was supported by Veterans Affairs Merit Award I01-BX003340 (to PWFW) and NIH grant NR013520.

Cassianne Robinson-Cohen was supported by NIH/NIDDK K01DK109019.

Cecilia P. Chung was supported by K23AR064768 (NIAMS), and the Rheumatology Research Foundation.

Louise Wain holds a GSK / British Lung Foundation Chair in Respiratory Research. This Article presents independent research funded partially by the UK National Institute for Health Research (NIHR). The views expressed are our own and not necessarily those of the NHS, the NIHR, or the UK Department of Health.

Stéphanie Debette was supported for this work by grants from the European Research Council (ERC), the EU Joint Programme - Neurodegenerative Disease Research (JPND), from the European Union's Horizon 2020 research and innovation programme under grant agreements No 643417 & No 640643, and by the Agence Nationale de la Recherche (ANR).

Eleftheria Zeggini was supported by the Wellcome Trust (WT098051).

Patricia Munroe was supported by the Medical Research Council of Great Britain (grant number G9521010D); and by the British Heart Foundation (grant number PG/02/128). Martin Farrall acknowledges support from the Wellcome Trust core award (090532/Z/09/Z) and the Oxford British Heart foundation Centre for Research Excellence (RE/13/1/30181). A.F.D. was supported by the British Heart Foundation (grant numbers RG/07/005/23633, SP/08/005/25115); and by the European Union Ingenious HyperCare Consortium: Integrated Genomics, Clinical Research, and Care in Hypertension (grant number LSHM-C7-2006-037093). The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. This work forms part of the research portfolio for the National Institute for Health Research Barts Biomedical Research Centre.

Peter Sever is the recipient of NIHR Senior Investigator Award. Peter Sever and Neil R Poulter were supported by Pfizer, New York, NY, USA, for the ASCOT study and the collection of the ASCOT DNA repository; by Servier Research Group, Paris, France; and by Leo Laboratories, Copenhagen, Denmark. We thank all ASCOT trial participants, physicians, nurses, and practices in the participating countries for their important contribution to the study. In particular, we thank Clare Muckian and David Toomey for their help in DNA extraction, storage, and handling. This work forms part of the research programme of the NIHR Cardiovascular Biomedical Research Unit at Barts.

Joanna M.M. Howson, Praveen Surendran, and Savita Karthikeyan was supported by UK Medical Research Council (G0800270), British Heart Foundation (SP/09/002), UK National Institute for Health Research Cambridge Biomedical Research Centre, European Research Council (268834), European Commission Framework Programme 7 (HEALTH-F2-2012-279233). We thank all EPIC participants and staff for their contribution to the study, the laboratory teams at the Medical Research Council Epidemiology Unit for sample management and Cambridge Genomic Services for genotyping, Sarah Spackman for data management, and the team at the EPIC-CVD Coordinating Centre for study coordination and administration.

Nicholas Wareham and Claudia Langenberg were supported by Medical Research Council UK (G1000143; MC_UU_12015/1; MC_PC_13048; MC_U106179471), Cancer Research UK (C864/A14136), EU FP6 Programme (LSHM_CT_2006_037197).

Sébastien Thériault was supported by Canadian Institutes of Health Research; Laval University (Quebec City, Canada).

John Attia and Christopher Oldmeadow would like to acknowledge the Vincent Fairfax Family Fund, the University of Newcastle Strategic Initiative Fund and all the participants who volunteered their time.

Pim van der Harst was supported by Marie Sklodowska-Curie GF (call: H2020-MSCA-IF-2014, Project ID: 661395).

Niek Verweij was supported by Marie Sklodowska-Curie GF (call: H2020-MSCA-IF-2014, Project ID: 661395) and NWO VENI (016.186.125).

David Schlessinger was supported by Intramural Research Program of the National Institute on Aging, NIH.

Marjo-Riitta Jarvelin and Karl-Heinz Herzig were supported by NFBC1966 and NFBC1986: NFBC1966 and 1966 received financial support from the Academy of Finland (project grants 104781, 120315, 129269, 1114194, 24300796, Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), NIHM (MH063706, Smalley and Jarvelin), Juselius Foundation, NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), ENGAGE project and grant agreement HEALTH-F4-2007-201413, EU FP7 EurHEALTHAgeing -277849, the Medical Research Council, UK (G0500539, G0600705, G1002319, PrevMetSyn/SALVE) and the MRC, Centenary Early Career Award. The program is currently being funded by the H2020 DynaHEALTH action (grant agreement 633595) and academy of Finland EGEA-project (285547). The DNA extractions, sample quality controls, biobank up-keeping and aliquotting was performed in the National Public Health Institute, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki. We thank the late Professor Paula Rantakallio (launch of NFBCs), and Ms Outi Tornwall and Ms Minttu Jussila (DNA biobanking). The authors would like to acknowledge the contribution of the late Academian of Science Leena Peltonen.

Paul Elliott acknowledges support from the Medical Research Council (MRC) and Public Health England (PHE) Centre for Environment and Health (MR/L01341X/1), and additional support from the NIHR Biomedical Research Centre at Imperial College Healthcare NHS Trust and Imperial College London, and the NIHR Health Protection Research Unit in Health Impact of Environmental Hazards (HPRU-2012
10141). This work used computing resources of the UK MEDical BIOinformatics partnership- (UK MED-BIO) supported by the Medical Research Council (MR/L01632X/1). P.E. is a UK Dementia Research Institute (DRI) professor, UK DRI at Imperial College London, funded by the MRC, Alzheimer's Society and Alzheimer's Research UK and an Associate Professor of the London Health Data Research UK Centre.

Cecilia Lindgren received funding from Wellcome Trust (086596/Z/08/Z) and is supported by Li Ka Shing Foundation, WT-SSI/John Fell funds, the NIHR Biomedical Research Centre, Oxford, Widenlife and NIH (5P50HD028138-27)

Christopher Newton-Cheh was Supported by R01HL113933 and R01HL124262.

Daniel Chasman received funding for genotyping of the exome chip and collaborative scientific support from Amgen.

Najim Lahrouchi received support from The Dutch Heart Foundation CVON-PREDICT project (CVON2012-10).

Wei-Qi Wei received funding from NIH R01 (R01 HL133786)

Helen R. Warren was funded by the National Institute for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Biomedical Research Unit at Barts and The London School of Medicine and Dentistry.

Claudia P. Cabrera was funded by the National Institute for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Biomedical Research Unit at Barts and The London School of Medicine and Dentistry.

Mark J. Caulfield was funded by the National Institute for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Biomedical Research Unit at Barts and The London School of Medicine and Dentistry. M.J.C. is a National Institute for Health Research (NIHR) senior investigator.

Yaomin Xu and Yu Wang were supported by the Biostatistics Development Award of the Department of Biostatistics of Vanderbilt University Medical Center.

Katalin Susztak was funded by R01DK076077, R01DK087635 and DP3DK108220.

Michael Boehnke and Laura J Scott were funded by DK062370.

One of the datasets used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding, the 1S10RR025141-01 instrumentation award, and by the CTSA grant UL1TR000445 from NCATS/NIH. Additional funding provided by the NIH through grants P50GM115305 and U19HL065962. The authors wish to acknowledge the expert technical support of the VANTAGE and VANGARD core facilities, supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA068485) and Vanderbilt Vision Center (P30 EY08126).

Relationship between t-statistic, chi-square and R^2 statistic[6]

$$t = r \times \sqrt{\frac{n-2}{1-r^2}} \qquad (1)$$

Rearranging the equation in terms of r^2

$$r^2 = \frac{t^2}{(n-2)+t^2} \qquad (2)$$

When n is large enough (n > 20) t-distribution approximates the z distribution

$$r^2 \approx \frac{z^2}{(n-2)+z^2} \qquad (3)$$

The square of a z distribution is the Chi-square distribution

$$r^2 \approx \frac{\chi^2}{(n-2)+\chi^2} \qquad (4)$$

Equation 1 describes the relationship between a student's t statistic, correlation coefficient r, and r^2. Equation 4 describes the transformed equation that describes r^2 in terms of the chi-square.

As n → ∞

$$r^2 \approx \frac{\chi^2}{n} \qquad (5)$$

When the sample size is sufficiently large enough, the variance explained by each SNP can then be well approximated by equation 5.

$$R^2 \approx \sum_{i=1}^{m} \frac{{\chi_i}^2}{n_i}$$

Where m = total number of independent SNPs in the study, R^2 is the total variance explained by independent SNPs, n_i and ${\chi_i}^2$ represent the number individuals in the analysis and the square of the Wald z-statistic for the given SNP, respectively.

Conditional Analysis

For conditional analysis of common variants we used two parallel approaches implemented in the Genome-wide Complex Traits Analysis (GCTA) software: (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[7,8]. The summary statistics from the GWAS discovery meta-analysis of MVP and UKP were used as the input summary data (separate by trait), and the imputed, hard-called BioVU EA 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 $5x10^{-8}$ 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 three blood pressure traits to exclude signals which were duplicated across traits (identified via pairwise LD for the list of all unique SNPs). For any sets of SNPs in LD ($r^2 \geq 0.1$), we selected the most significant SNP with the minimum p-value across all blood pressure traits from the GCTA joint model. Hence all final SNPs are pairwise-LD-independent.

(ii) Locus-specific conditional analysis

Here we considered each of the 505 blood pressure loci separately.

Within each of the Tier 1 or 2 replicated novel loci (N = 131 [69 SBP, 2 DBP, 68 pulse pressure]; **Supplementary Tables 2a-c)**, 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.

For known loci (**Supplementary Table 4**) we considered all 304 loci detected by our discovery analysis as well as 265 previously published SNPs included in the construction of our GRS. Overall this resulted in a total of 344 known loci being analyzed. 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 $\geq$ 1% within the 1Mb region, conditioning on the single published SNP within the locus, testing for association of all three blood pressure traits. For known loci containing more than one identified SNP, conditional analysis was performed within the wider locus region, conditioning 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 blood pressure traits from the GWAS discovery meta-analysis was selected. Of the 344 known loci analyzed, 197 loci conditioned on multiple SNPs.

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 four criteria:

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

(b) $Pc < 5x10^{-8}$ from the conditional analysis, so that the SNP is also significantly associated with blood pressure 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

(d) not in LD with any of the reported SNPs or any of the sentinel SNPs at the 201 novel loci ($r^2 < 0.1$)

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 blood pressure locus (either novel or known). For robustness, a secondary signal was only claimed if the SNP is validated from both approaches.

Rare-variant conditional analyses

We performed conditional analysis for coding rare-variants (Table 2) detected in the final meta-analysis. Since all of the rare variants from Table 2 were not well represented in the BioVU dataset (N ~ 19,500) used to approximate the LD matrix for GCTA, conditional analysis of rare-variants were performed using the largest available primary discovery dataset (MVP whites). The purpose of these analyses were two-fold: 1) to determine whether signals from these rare-variants were independent of GWAS-significant common variants for that locus, and 2) to determine whether rare variants were independent of each other for instances where more than one rare variant was noted for a given locus. Conditional analyses were performed using SNPTEST by regressing variants while adjusting for common or rare variants in the logistic regression model, in addition to adjustment of all other covariates. Effect estimates from models before and after conditioning, and R^2 between variants were inspected to determine independence of signals. If more than one trait was significant then the trait with the most significant pair (common and rare) was chosen. If only one rare variant was present per gene, analyses were conditioned upon the most significant (index) common variant within the locus, defined as +- 500KB of a signal. If more than one rare variant was found to be statistically significant, then the variant was adjusted for the index common variant in that locus and also mutually adjusted for other rare variants. Conditional analysis was not performed for SNP rs61760904, the significant rare variant in the RRAS gene as no common variants reaching GWAS significance level was found for this locus.

Enrichment and Pathway Analyses

We investigated whether one or more of the 45 tissues evaluated with S-PrediXcan were enriched for statistically significant genes (S-PrediXcan tissue-wide Bonferroni P-value < 2.5×10^{-7}). Enrichment analyses were formally constructed using a one-degree freedom chi-square test by comparing the proportion of significant genes in a given tissue to the proportion of significant genes detected across all other tissues.

Enrichment analyses in DEPICT 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.

Z-statistics from significant genes from the top enriched S-PrediXcan tissue (aorta) for each trait were then evaluated with the Ingenuity Pathway Analysis (IPA) software (IPA®,QIAGEN Redwood City) to report relationships between genes in the top networks for each trait (Supplementary Figure 5, 6 and 7). IPA software is based on algorithms testing functional connectivity of genes through direct and indirect relationships utilizing information from the proprietary IPA database. Output includes a p-value for enrichment of disease and functional pathways.

References

1. 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,** 103–109 (2011).
2. Wain, L. V. *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat. Genet.* **43,** 1005–1011 (2011).
3. 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. *Hypertens. Dallas Tex 1979* (2017). doi:10.1161/HYPERTENSIONAHA.117.09438
4. 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).
5. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat. Genet.* **48,** 1162–1170 (2016).
6. Kutner, M., Nachtsheim, C., Neter, J. & Li, W. Inferences on Regression and Correlation Analysis. in *Applied Linear Statistical Methods* 89 (Mc Graw Hill).
7. 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).
8. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44,** 369–375, S1-3 (2012).