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

Understanding Society Scientific Group; Giri, Ayush; Hellwege, Jacklyn N.; Keaton, Jacob M.; Park, Jihwan; Qiu, Chengxiang; Warren, Helen R.; Torstenson, Eric S.; Kovesdy, Csaba P.; Sun, Yan V.; Wilson, Otis D.; Robinson-Cohen, Cassianne; Roumie, Christiane L.; Chung, Cecilia P.; Birdwell, Kelly A.; Damrauer, Scott M.; DuVall, Scott L.; Klarin, Derek; Cho, Kelly; Wang, Yu; Evangelou, Evangelos; Cabrera, Claudia P.; Wain, Louise V.; Shrestha, Rojesh; Mautz, Brian S.; Akwo, Elvis A.; Sargurupremraj, Muralidharan; Debette, Stéphanie; Boehnkne, Michael; Scott, Laura J.; Luan, Jian'an; Zhao, Jing-Hua; Willems, Sara M.; Thériault, Sébastien; Shah, Nabi; Oldmeadow, Christopher; Almgren, Peter; Li-Gao, Ruifang; Verweij, Niek; Boutin, Thibaud S.; Mangino, Massimo; Ntalla, Ioanna; Feofanova, Elena; Surendran, Praveen; Cook, James P.; Karthikeyan, Savita; Lahrouchi, Najim; Liu, Chunyu; Sepúlveda, Nuno; Palmer, Colin Neil Alexander; Wilson, Peter W. F.; Hung, Adriana M.; Edwards, Todd L.

Published in:
Nature Genetics

DOI:
10.1038/s41588-018-0303-9

Publication date:
2019

Document Version
Peer reviewed version

Link to publication in Discovery Research Portal
Supplementary Table Legends

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 (PP). 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; PP (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 and build 37 position; Effect allele - allele corresponding to measured effect on the outcome; Other allele - allele not corresponding to measured effect on the outcome; EAF - effect allele frequency in the discovery meta-analysis; Effect - measured effect in the discovery meta-analysis; SE - standard error of the measured effect in the discovery meta-analysis; P-value - association p-value for the measured effect in the 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.

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 and build 37 position; Effect allele - allele corresponding to measured effect on the outcome; Other allele - allele not corresponding to measured effect on the outcome; EAF - effect allele frequency in the discovery meta-analysis; Effect - measured effect in the discovery meta-analysis; SE - standard error of the measured effect in the discovery meta-analysis; P-value - association p-value for the measured effect in the 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.

Supplementary Table 2c. Association results for all sentinel PP SNPs in previously reported loci. SNPs are ordered by chromosome and position. rsID - dbSNP accession number; CHR:BP - chromosome and build 37 position; Effect allele - allele corresponding to measured effect on the outcome; Other allele - allele not corresponding to measured effect on the outcome; EAF - effect allele frequency in the discovery meta-analysis; Effect - measured effect in the discovery meta-analysis; SE - standard error of the measured effect in the discovery meta-analysis; P-value - association p-value for the measured effect in the 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.

**Supplementary Table 3a. Association results for all 128 SBP novel replicated variants from two-stage analysis.** SNPs are ordered by tier of replication, then by chromosome and position. rsID - dbSNP accession number; CHR:BP - chromosome and 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 on the outcome; Other allele - allele not corresponding to measured effect on the outcome; EAF<sub>comb</sub> - effect allele frequency in the combined discovery and replication meta-analysis; Effect<sub>comb</sub> - measured effect in the combined discovery and replication meta-analysis; SE<sub>comb</sub> - standard error of the measured effect in the combined discovery and replication meta-analysis; P-value<sub>comb</sub> - association p-value for the measured effect in the combined discovery and replication meta-analysis; Neff<sub>comb</sub> - effective number of subjects in the combined discovery and replication meta-analysis; P<sub>het</sub> - value for Cochran’s Q test of statistical heterogeneity in the combined discovery and replication meta-analysis; I<sup>2</sup> - percentage of total variation across studies that is due to statistical heterogeneity rather than chance; Tier - evidentiary tier for association 1) Genome-wide significance in the discovery stage, and Bonferroni-corrected significance in replication and consistent trait-specific direction of effect across stages 2) Genome-wide significance in the discovery stage, and P-value ≤ 0.05 in the replication stage and consistent trait-specific direction of effect across stages or 3) Variants had P-value less than 1x10^-6 and > 5x10^-8 in the discovery stage, and had P-value < 0.05 in the replication stage and had consistent trait-specific direction of effect across stages and was genome-wide significant after final analysis; EAF<sub>disc</sub> - effect allele frequency in the discovery meta-analysis; Effect<sub>disc</sub> - measured effect in the discovery meta-analysis; SE<sub>disc</sub> - standard error of the measured effect in the discovery meta-analysis; P-value<sub>disc</sub> - association p-value for the measured effect in the discovery meta-analysis; Effect<sub>rep</sub> - measured effect in the replication meta-analysis; SE<sub>rep</sub> - standard error of the measured effect in the replication meta-analysis; P-value<sub>rep</sub> - association p-value for the measured effect in the replication meta-analysis; N - number of subjects in the ethnicity-specific discovery meta-analysis.

**Supplementary Table 3b. Association results for DBP novel replicated sentinel variants from two-stage analysis.** SNPs are ordered by tier of replication, then by chromosome and position. rsID - dbSNP accession number; CHR:BP - chromosome and 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 on the outcome; Other allele - allele not corresponding to measured effect on the outcome; EAF<sub>comb</sub> - effect allele frequency in the combined discovery and replication meta-analysis; Effect<sub>comb</sub> - measured effect in the combined discovery and replication meta-analysis; SE<sub>comb</sub> - standard error of the measured effect in the combined discovery and replication meta-analysis; P-value<sub>comb</sub> - association p-value for the measured effect in the combined discovery and replication meta-analysis; Neff<sub>comb</sub> - effective number of subjects in the combined discovery and replication meta-analysis; P<sub>het</sub> - value for Cochran’s Q test of statistical heterogeneity in the combined discovery and replication meta-analysis; I<sup>2</sup> - percentage of total variation across studies that is due to statistical heterogeneity.
rather than chance; Tier - evidentiary tier for association 1) Genome-wide significance in the discovery stage, and Bonferroni-corrected significance in replication and consistent trait-specific direction of effect across stages 2) Genome-wide significance in the discovery stage, and P-value ≤ 0.05 in the replication stage and consistent trait-specific direction of effect across stages or 3) Variants had P-value less than 1x10^{-6} and > 5x10^{-8} in the discovery stage, and had P-value < 0.05 in the replication stage and had consistent trait-specific direction of effect across stages and was genome-wide significant after final analysis; EAF_{disc} - effect allele frequency in the discovery meta-analysis; Effect_{disc} - measured effect in the discovery meta-analysis; SE_{disc} - standard error of the measured effect in the discovery meta-analysis; P-value_{disc} - association p-value for the measured effect in the discovery meta-analysis; EAF_{rep} - measured effect in the replication meta-analysis; Effect_{rep} - measured effect in the replication meta-analysis; SE_{rep} - standard error of the measured effect in the replication meta-analysis; P-value_{rep} - association p-value for the measured effect in the replication meta-analysis; N - number of subjects in the ethnicity-specific discovery meta-analysis.

Supplementary Table 3c. Association results for all 126 PP novel replicated sentinel variants from two-stage analysis. SNPs are ordered by tier of replication, then by chromosome and position. rsID - dbSNP accession number; CHR:BP - chromosome and 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 on the outcome; Other allele - allele not corresponding to measured effect on the outcome; EAF_{comb} - effect allele frequency in the combined discovery and replication meta-analysis; Effect_{comb} - measured effect in the combined discovery and replication meta-analysis; SE_{comb} - standard error of the measured effect in the combined discovery and replication meta-analysis; P-value_{comb} - association p-value for the measured effect in the combined discovery and replication meta-analysis; N_{eff,comb} - effective number of subjects in the combined discovery and replication meta-analysis; P_{het} - value for Cochran’s Q test of statistical heterogeneity in the combined discovery and replication meta-analysis; I^2 - percentage of total variation across studies that is due to statistical heterogeneity rather than chance; Tier - evidentiary tier for association 1) Genome-wide significance in the discovery stage, and Bonferroni-corrected significance in replication and consistent trait-specific direction of effect across stages 2) Genome-wide significance in the discovery stage, and P-value ≤ 0.05 in the replication stage and consistent trait-specific direction of effect across stages or 3) Variants had P-value less than 1x10^{-6} and > 5x10^{-8} in the discovery stage, and had P-value < 0.05 in the replication stage and had consistent trait-specific direction of effect across stages and was genome-wide significant after final analysis; EAF_{disc} - effect allele frequency in the discovery meta-analysis; Effect_{disc} - measured effect in the discovery meta-analysis; SE_{disc} - standard error of the measured effect in the discovery meta-analysis; P-value_{disc} - association p-value for the measured effect in the discovery meta-analysis; Effect_{rep} - measured effect in the replication meta-analysis; SE_{rep} - standard error of the measured effect in the replication meta-analysis; P-value_{rep} - association p-value for the measured effect in the replication meta-analysis; N - number of subjects in the ethnicity-specific discovery meta-analysis.

Supplementary Table 4. Conditional association results for all jointly conditional SNPs. SNPs are ordered by chromosome and position. rsID - dbSNP accession number; CHR:BP - chromosome and build 37 position; Nearest Gene - most proximal gene within 250kb of sentinel SNP; Distance - distance in base pairs from
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 of main text with and without conditioning on select SNPs in MVP Whites Discovery Analysis. 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. SNPs are ordered by chromosome and position.

rsID - dbSNP accession number; CHR:BP - chromosome and build 37 position; Nearest Gene - most proximal gene within 250kb of sentinel SNP; Effect allele - allele corresponding to measured effect on the outcome; Other allele - allele not corresponding to measured effect on the outcome; Trait - Trait used in conditional analysis was chosen if the trait was significant in both rare and common variant analysis. If more than one trait was significant then the trait with the most significant pair (common and rare) was chosen; EAF_{disc} - effect allele frequency in the combined discovery and replication meta-analysis; Effect_{disc} - measured effect in the discovery meta-analysis; P-value_{disc} - association p-value for the measured effect in the discovery meta-analysis; Lead SNP - SNP with most significant association p-value for the measured effect in the discovery meta-analysis; Novel/Known - indicator of whether locus was previously reported or novel in our analyses; Nearest Gene - 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-value_{cond} - association p-value for the measured effect of SNP in the rsID column in the genome-wide joint conditional analysis.
Supplementary Table 7. Rare variant results from meta-analysis with UKB 500K. Table is sorted by tier of replication, then by chromosome and position. rsID - dbSNP accession number; CHR:BP - chromosome and 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 on the outcome; Other allele - allele not corresponding to measured effect on the outcome; EAF - effect allele frequency in the combined discovery and replication meta-analysis; Neff - effective number of subjects in the 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; Tier of Replication - evidentiary tier for association 1) Genome-wide significance in the discovery stage, and Bonferroni-corrected significance in replication and consistent trait-specific direction of effect across stages 2) Genome-wide significance in the discovery stage, and P-value ≤ 0.05 in the replication stage and consistent trait-specific direction of effect across stages or 3) Variants had P-value less than 1x10^-6 and > 5x10^-8 in the discovery stage, and had P-value < 0.05 in the replication stage and had consistent trait-specific direction of effect across stages and was genome-wide significant after final analysis; 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 the 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-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.
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 \( W' \times G \times W \) (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.

Supplementary Table 8c. S-PrediXcan results across 45 tissues with PP. 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 \( W' \times G \times W \) (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.

Supplementary Table 9a. Expression from single-cell RNA sequencing of murine kidney cell types of mouse homologs of colocalized genes associated with SBP. Gene - mouse homolog of significant gene identified in kidney tissue (see ST7a). 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.

Supplementary Table 9b. Expression from single-cell RNA sequencing of murine kidney cell types of mouse homologs of colocalized genes associated with DBP. Gene - mouse homolog of significant gene identified in kidney tissue (see ST7b). 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.

Supplementary Table 9c. Expression from single-cell RNA sequencing of murine kidney cell types of mouse homologs of colocalized genes associated with PP. Gene - mouse homolog of significant gene identified in kidney tissue (see ST7c). 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.

**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 scRNA-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 BP SNP - most strongly associated SNP previously reported within 250kb of S-PrediXcan window for predicted expressed gene; Previously Reported BP Gene - nearest gene to previously reported BP 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 BP SNP - most strongly associated SNP previously reported within 250kb of S-PrediXcan window for predicted expressed gene; Previously Reported BP Gene - nearest gene to previously reported BP SNP. Primary Indication – primary indication for prescription of the drug as annotated from BIDD TTD.
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 BP SNP - most strongly associated SNP previously reported within 250kb of S-PrediXcan window for predicted expressed gene; Previously Reported BP Gene - nearest gene to previously reported BP SNP; ADE - adverse drug event as annotated from SIDER.

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 BP-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 group; 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_Covars - 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.

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; \( \chi^2 \) - 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<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<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<0.05) DEPICT tissue enrichment results across all significant (P<5x10^-8) known and novel PP 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<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<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<0.05) DEPICT gene set enrichment results across all significant (P<5x10^-8) known and novel PP 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 Figure 1. Comparison of effect sizes for known and novel sentinel SNPs identified with SBP, DBP, and PP across whites, blacks, and Hispanics. Sentinel SNPs from final meta-analysis for each BP trait (left to right) were compared for consistency between UKB whites and MVP whites (row 1), MVP blacks and MVP whites (row 2), MVP Hispanics and MVP whites (row 3), and MVP Hispanics and MVP blacks (row 4). Blue dots denote sentinel SNPs from known loci and red dots denote from novel loci.

Supplementary Figure 2. Juxtaposed mirror plots for S-PrediXcan (–log10p) and GWAS (log10p) for BP-trait. Log-10p-values for associations between genetically predicted gene expression (GPGE) analyses with BP-traits in 44 tissues are juxtaposed with –log10p-values from GWAS analyses for SBP (a); DBP (b) and PP (c). All GWAS plots represent discovery + replication samples included. GPGE analysis with S-PrediXcan was also performed the full discovery + replication summary statistics.

Supplementary Figure 3. Comparison of effect sizes for significant PheWAS results identified with SBP, DBP, and PP across whites, blacks, and Hispanics. Genetic risk scores (GRS) weighted for SBP, DBP and PP were regressed onto the clinical phenome in whites, blacks and Hispanics separately. Effect estimates for phenotypes that were significant in whites, blacks or Hispanics were compared across three ethnicities. Comparison of effect estimates are presented in the following order: blacks and whites (row 1), Hispanics and whites (row 2) and Hispanics and Blacks (row 3) for SBP, DBP and PP (left to right). R^2 denotes correlation between effect estimates calculated from a linear regression model.

Supplementary Figure 4. Venn Diagram of associations from PheWAS for BP-trait specific GRS. Shows overlap of associations between SBP, DBP and PP w-GRS. PheWAS analysis was conducted in self-reported/administratively assigned white MVP participants only. W-GRS were constructed using statistically significant SNPs using weights from the UK Biobank dataset.

Figure 5. Subcellular layout of top network from IPA analysis of significant SBP genes in aorta. 20 out of 45 molecules are represented by genes significant in S-PrediXcan analyses as indicated by node coloring. Arrows indicate direction of relationship while solid lines indicate direct interaction (e.g. phosphorylation) and broken lines indicate indirect relationships (e.g. activation). Interactions without direction (e.g. protein-protein) do not have an arrow. Nodes outlined in purple indicate overlay of cardiovascular disease (enrichment P=7.16x10^{-6}) and cardiovascular system and development (enrichment P=7.73x10^{-5}) pathways.

Figure 6. Subcellular layout of top network from IPA analysis of significant DBP genes in aorta. 11 out of 27 molecules are represented by genes significant in S-PrediXcan analyses as indicated by node coloring. Arrows indicate direction of relationship while solid lines indicate direct interaction (e.g. phosphorylation) and broken lines indicate indirect relationships (e.g. activation). Interactions without direction (e.g. protein-protein) do not have an arrow. Nodes
outlined in purple indicate overlay of hematopoiesis (enrichment $P=6.57 \times 10^{-7}$) and hematological system and development (enrichment $P=6.57 \times 10^{-7}$) pathways.

**Figure 7. Subcellular layout of top network from IPA analysis of significant PP genes in aorta.** 18 out of 36 molecules are represented by genes significant in S-PrediXcan analyses as indicated by node coloring. Arrows indicate direction of relationship while solid lines indicate direct interaction (e.g. phosphorylation) and broken lines indicate indirect relationships (e.g. activation). Interactions without direction (e.g. protein-protein) do not have an arrow. Nodes outlined in purple indicate overlay of cardiovascular disease (enrichment $P=9.53 \times 10^{-4}$) and cardiovascular system and development (enrichment $P=9.53 \times 10^{-4}$) pathways.

**Supplementary Figure 8. Quantile-quantile (QQ) plots for discovery meta-analysis GWAS of BP-traits.** Shows QQ plots for SBP (a), DBP (b) and PP (c). Genomic inflation statistic Lambda is presented for discovery meta-analysis for each BP-trait.
Supplementary Data

Replication in International Consortium for Blood Pressure (ICBP)

ICBP GWAS is an international consortium to investigate BP genetics\textsuperscript{1-3}. We combined previously reported post-quality control (QC) GWAS data from 54 studies (N=150,134)\textsuperscript{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), EGCGUT (n=6395), EGCGUT2 (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), MICROG (n=1148), MIGEN (n=1214), NESDA (n=2336), NSPSH (n=1005), NTR (n=1490), PHASE (n=4535), PIVUS (n=945), PROCARDIS (n=1652), SHIP (n=4068), ULSAM (n=1114), WGS (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\textsuperscript{1}, Jonathan Burton\textsuperscript{1}, Nicholas Buck\textsuperscript{1}, Annette Jäckle\textsuperscript{1}, Meena Kumari\textsuperscript{1}, Heather Laurie\textsuperscript{1}, Peter Lynn\textsuperscript{1}, Stephen Pudney\textsuperscript{1}, Birgitta Rabe\textsuperscript{1}, Dieter Wolke\textsuperscript{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\textsuperscript{2} that was established for studies of exome content and BP. We combined previously reported post-quality control (QC) exome array summary BP association results\textsuperscript{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), INCPE (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).
Relationship between t-statistic, chi-square and $R^2$ statistic

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

Rearranging the equation in terms of $r^2$

\[
r^2 = \frac{t^2}{(n-2) + t^2}
\]  
(2)

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

\[
r^2 \approx \frac{z^2}{(n-2) + z^2}
\]  
(3)

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

\[
r^2 \approx \frac{\chi^2}{(n-2) + \chi^2}
\]  
(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 $n \to \infty$

\[
r^2 \approx \frac{\chi^2}{n}
\]  
(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.

References


