Genetic analysis of over one million people identifies 535 new loci associated with blood pressure traits.

## Table of Contents

## Supplementary Table Legends

## Supplementary Table Legends

**Supplementary Table 1a**: Descriptive summary statistics of included participants and total sample size, UK Biobank and UK BiLEVE. Table (i) shows mean (SD) for age, body mass index and % for sex, blood pressure-lowering meds (medication) and hypertension. Table ii) shows mean (SD) for systolic blood pressure, diastolic blood pressure and pulse pressure.

**Supplementary Table 1b**: Summary of blood pressure phenotypes and covariates for all studies contributing to ICBP meta-analysis.

**Supplementary Table 1c**: Summary of replication cohorts: i) Descriptive statistics, ii) Genotyping and Quality control, iii) SNP-level QC and iv) sample level QC.

**Supplementary Table 2a**: Association results for all 130 SBP novel replicated variants from two-stage analysis. SNPs are ordered by *P*-value from the combined meta-analysis.

**Supplementary Table 2b**: Association results for all 91 DBP novel replicated variants from two-stage analysis. SNPs are ordered by *P*-value from the combined meta-analysis.

**Supplementary Table 2c**: Association results for all 104 PP novel replicated variants from two-stage analysis. SNPs are ordered by *P*-value from the combined meta-analysis.

**Supplementary Table 3a**: Association results of 60 novel variants identified through the one-stage design. Results are ordered by *P*-value from the discovery meta-analysis.

**Supplementary Table 3b**: Association results of 103 novel variants identified through the one-stage design. Results are ordered by *P*-value from the discovery meta-analysis.

**Supplementary Table 3c**: Association results of 47 novel variants identified through the one-stage design. Results are ordered by *P*-value from the discovery meta-analysis.

**Supplementary Table 4**: Previously reported Blood Pressure Loci. Results for the 357 replicated SNPs from 274 published blood pressure loci. The reported *P*-values have been derived from the discovery meta-analysis (N=757,601) using inverse variance fixed effects models.

**Supplementary Table 5**: Ninety-two newly confirmed variants that have been previously reported but have not been replicated elsewhere. SNPs are ordered by replicated BP trait,

then by *P*-value from the combined meta-analysis. The *P*-values were calculated using inverse variance fixed effects meta-analysis.

**Supplementary Table 6:** Results for the 163 additional secondary signals from conditional analysis.

**Supplementary Table 7:** Percentage of variance explained for BP traits in Airwave.

**Supplementary Table 8**: Bioinformatics annotation of all SNPs in LD with the 535 novel sentinel SNPS, showing all annotation at SNP-level.

**Supplementary Table 9:** 535 novel loci GTEx eQTL associations per tissue. "X" represents an eQTL observed for at least one variant within the locus (according to all variants in linkage disequilibrium of $r^2 \geq 0.8$ with the sentinel variant). The number of tissues a given loci has eQTL in is indicated on the right and the number of eQTLs per tissue on the bottom.

**Supplementary Table 10a:** DEPICT tissue enrichment results across published (N=357) and all known and novel SNPs with P-value $<1x10^{-12}$ ($N_{novel}$=227). The gene set enrichment analysis algorithm is described in Pers et al[66]. Only results with FDR <0.05 are shown.

**Supplementary Table 10b:** DEPICT gene set enrichment results across known (N=357 and all known and novel SNPs with P-value $<1x10^{-12}$ ($N_{novel}$=227). The gene set enrichment analysis algorithm is described in Pers et al[66]. Results are shown for GO categories, mouse KO phenotypes and protein-protein interaction subnetworks separately. Only results with FDR <0.01 are shown.

**Supplementary Table 11:** Druggability annotation for all BP associated genes.

**Supplementary Table 12:** Associations of all loci with lifestyle and anthropometric traits in UK Biobank with $P < 1x10^{-6}$ using GeneAtlas (N=408,455) (http://geneatlas.roslin.ed.ac.uk/) and Global Biobank Engine (N=337,199) (https://biobankengine.stanford.edu). The associations were performed using linear mixed models in BOLT-LMM (GeneAtlas) or linear models in PLINK2 (Global Biobank Engine).

**Supplementary Table 13:** Novel Loci Associations with Other Traits from PhenoScanner and GWAS Catalog: Novel loci where the sentinel SNP or proxies ($r^2 \geq 0.8$) is associated with other traits at genome-wide significance, from either the PhenoScanner database or the GWAS Catalog. "Trait" shows the name of trait listed from the database; "Category" corresponds to the disease/trait category used for the Figure plot (Fig. 5) illustrating cross-trait associations.

**Supplementary Table 14:** Metabolomics signatures of novel SNPs from (a) lipid fraction profiling of serum $^1$H Nuclear Magnetic Resonance spectroscopy lipidomics data from 2,022 Airwave participants (Bonferroni corrected *P*-value = $4.76x10^{-4}$); (b) Mass Spectroscopy using Metabolon platform from 1,942 Airwave participants (Bonferroni corrected *P*-value = $4.36x10^{-5}$). *P*-values were obtained from linear regression analysis.

**Supplementary Table 15:** Metabolomics signatures of novel SNPs derived from publicly available genome-wide association studies using PhenoScanner.

**Supplementary Table 16:** Associations of the genetic risk score (GRS) with blood pressure traits, hypertension and cardiovascular events in unrelated individuals of European descent in UK Biobank. Tables a) and b) show associations of continuous GRS with SBP/DBP/PP and HTN; tables c), d) and e) compare top vs bottom quintiles of the GRS distribution for SBP/DBP/PP, HTN and CVD/MI/Stroke whereas tables f) g) and h) compare top vs bottom deciles of the GRS distribution for SBP/DBP/PP, HTN and CVD/MI/Stroke.

**Supplementary Table 17:** Associations of the genetic risk score (GRS) with blood pressure (BP) traits and hypertension (HTN) in the Airwave Study. Tables a) and b) show associations of continuous GRS with SBP/DBP/PP and HTN; tables c) and d) compare top vs bottom quintiles of the GRS distribution for SBP/DBP/PP (non-stratified and stratified by age) whereas tables e) and f) compare top vs bottom quintiles of the GRS distribution for HTN (non-stratified and stratified by age).

**Supplementary Table 18**: Full association results for all novel and previously published SNPs in other ancestries using linear mixed modelling in BOLT-LMM (Africans and South Asians) and inverse variance fixed effects models (discovery meta-analysis). Results are shown for (i) Africans; (iii) South Asians (iii) discovery meta-analysis. SNPs are ordered by chromosome and position.

**Supplementary Table 19a:** Associations of the genetic risk score (GRS) with blood pressure traits and hypertension in unrelated Africans in UK Biobank. Tables a) and b) show associations of continuous GRS with SBP/DBP/PP and HTN; tables c) and d) compare top vs bottom quintiles of the GRS distribution for SBP/DBP/PP and HTN whereas tables e) and f) compare top vs bottom deciles of the GRS distribution for SBP/DBP/PP and HTN.

**Supplementary Table 19b:** Associations of the genetic risk score (GRS) with blood pressure traits and hypertension in unrelated South Asians in UK Biobank. Tables a) and b) show associations of continuous GRS with SBP/DBP/PP and HTN; tables c) and d) compare top vs bottom quintiles of the GRS distribution for SBP/DBP/PP and HTN whereas tables e) and f) compare top vs bottom deciles of the GRS distribution for SBP/DBP/PP and HTN.

**Supplementary Table 20a:** Study design summary information for each of the studies contributing to ICBP meta-analysis.

**Supplementary Table 20b:** Summaries of methods used to adjust for population stratification and kinship for all studies contributing to ICBP meta-analysis.

**Supplementary Table 20c:** Summary of genotyping and imputation strategy for all studies contributing to ICBP meta-analysis.

**Supplementary Table 21:** Full association results for all 1,062 lookup variants from GWAS discovery meta-analysis. Results are shown for (i) discovery meta-analysis; (ii) UKB-only; (iii) ICBP-only; (iv) replication meta-analysis and (v) combined meta-analysis for all 3 traits.

Results in (i), (iii), (iv) and (v) have been computed using inverse variance fixed effects models. Results in (ii) have been computed from linear mixed models using BOLT-LMM. SNPs are ordered by chr:pos.

**Supplementary Table 22:** Lookup variants from GWAS discovery meta-analysis that required proxies for replication in MVP and EGCUT data. Discovery association results are shown for both the original sentinel ("Lead") variant and the proxy variant used as a lookup in the replication meta-analysis. Summary effects are from inverse variance fixed effects meta-analysis.

**Supplementary Table 23:** Fifty-nine previously reported loci that have not been replicated. SNPs are ordered by chromosome and position (CP)**.** Results are from inverse variance fixed effects meta-analysis.

**Supplementary Table 24:** List of a) published and b) novel variants and their effects that have been used as weights for the genetic risk score. The reported results have been derived from ICBP meta-analysis (N=299,024) and the combined meta-analysis (N=757,601) using fixed effects inverse variance models.

**Supplementary Table 25.** UK Biobank CVD algorithm for defining cardiovascular disease cases.

## Supplementary Figure Legends

**Supplementary Figure 1**: **GWAS discovery Manhattan plots.** Manhattan plots (a), (b) and (c) for systolic blood pressure (SBP), diastolic blood pressure (DBP) and pulse pressure (PP) respectively. P-value results from the GWAS discovery meta-analysis (N=757,601), were derived using inverse variance fixed effects meta-analysis and they are plotted on a – log10 scale for all SNPs with Minor Allele Frequency (MAF) ≥ 1%. SNPs within the 274 known loci (±500kb; Linkage Disequilibrium $r^2 \geq 0.1$) are highlighted in green.

**Supplementary Figure 2: Effect Sizes of all Blood Pressure associated loci**. Plot (A) shows strong correlation between the published effect size estimates (x-axis) from literature vs the effect sizes from our discovery meta-analysis (y-axis), for known SNPs, colour-coded according to the published primary trait from the first published report. From the 357 validated SNPs listed in Supplementary Table 4 from the 274 published loci, 327 are available within the MAF ≥ 1% HRC-imputed data. For comparison of effect sizes, we only consider 299 such SNPs which have been identified from main-effect genetic association studies within Europeans (i.e. excluding any SNPs from interaction/stratified/multi-phenotype analysis, or from studies of other ancestries). For reliable comparison of effect sizes, we further restrict to the 284 known SNPs which reach genome-wide significance within the discovery meta-analysis for at least one BP trait. The $r^2$ value is presented to

show the correlation between published and observed effect sizes. Plots (B), (C) and (D) are trait-specific plots for SBP, DBP and PP, respectively (SBP: systolic blood pressure; DBP: diastolic blood pressure; PP: pulse pressure). Across all plots, the 284 "known" SNPs (black squared) from plot (A) are compared against the 325 novel sentinel SNPs from the 2-stage analysis (red circles), the 210 novel sentinel SNPs from the 1-stage analysis (green triangles), and the 92 SNPs (blue diamonds) replicated for the first time from Hoffman et al[9]. Each SNP is only plotted in one of the trait-specific plots, according to the published primary trait for the known SNPs, or the primary trait for the novel / replicated SNPs. For all SNPs we show the relationship between Minor Allele Frequency (MAF) on the x-axis and the effect size (mmHg) on the y-axis, where results are taken from the UKB+ICBP discovery meta-analysis. All meta-analysis results were computed using inverse variance fixed effects models. The different symbols and colours distinguish the "known" vs "novel-2stage" vs "novel-1stage" vs "replicated-Hoffman" SNPs, and show that in general, the novel SNPs have smaller effect sizes than known SNPs, and that there is no significant difference (P=0.447) between the effect sizes of the 1-stage (N=757,601) and 2-stage (N=1,006,863) novel SNPs. (UKB: UK Biobank; ICBP: International Consortium of Blood Pressure).

**Supplementary Figure 3:** For all 535 novel loci, we show the blood pressure traits associated with each locus. We present the two-stage loci first, followed by the one-stage loci. The locus names provided in alphabetical order correspond to the nearest annotated gene. SNPs: Single nucleotide polymorphisms; SBP: systolic blood pressure; DBP: diastolic blood pressure; PP: pulse pressure; UKB: UK Biobank; ICBP: International Consortium of Blood Pressure.

**Supplementary Figure 4: Overview of functional annotation and prioritisation of genome-wide associated variants and genes**. SNPs: single nucleotide polymorphisms; LD: Linkage Disequilibrium; eQTL: expression Quantitative Trait Loci; UCSC: University of California Santa Cruz (UCSC) genome browser; IPA: Ingenuity Pathway Analysis (IPA) software (IPA®,QIAGEN Redwood City,www.qiagen.com/ingenuity); DEPICT: Data-driven Expression Prioritized Integration for Complex Traits; GREAT: Genomic Regions Enrichment of Annotations Tool.

**Supplementary figure 5: DEPICT enrichment analysis.** DEPICT software was used to investigate enrichment of a range of biological properties, in each case we compared known sentinel SNPS (N=357) to all known and novel SNPs with P-value $<1x10^{-12}$ (N=227). The gene set enrichment analysis algorithm is described in Pers et al[66]. Enrichment –log p value is reported for both groups, we also present delta –log p value as a measure of novelty introduced by novel associations reported. Enrichment categories are as follows. a) Enrichment of tissues and cell types. b) GO annotation. c) Protein-protein interaction subnetwork annotation. d) Mammalian phenotype annotation.

**Supplementary Fig 6: Enrichments of eQTLs**. 535 novel blood pressure associated SNPs and the SNPs in LD $r^2$>0.8 were annotated for their effect on gene expression using the GTEx portal. The number of eGenes associated with BP SNPs in a given tissue/ cell type was normalised with the total number of eGenes in that tissue and z-score was calculated using the trimmed mean and standard deviation of the normalised scores. Tissues of the same tissue group were coloured the same.

**Supplementary Figure 7**: **FORGE Dnase I sensitive region enrichment in known sentinel SNPs, compared to known and novel sentinel SNP associations for blood pressure**. Sentinel SNPs were investigated for enrichment in ENCODE DNase I regulatory regions using FORGE. The background probability of overlap is determined from the 1000 background set overlap counts and the probability of the observed test result under a binomial distribution is calculated. The *P*-value thresholds of 0.05 and 0.01 are corrected for multiple testing by division by the number of tissue groupings tested, and the corrected threshold is used. Strongest enrichment in known SNPs was seen in vasculature (Human Aortic artery fibroblast (AoAF) and also Human Villous Mesenchymal Fibroblasts (HVMF) found in placenta). Enrichment in all known and novel SNPS was increased across vasculature (AoAF; HMVEC, Human microvascular endothelial cells) and highly vascularised tissues. Tissues in red are significant after correction for false discovery.

**Supplementary Figure 8**: **Ingenuity pathway analysis of BP genes.** For genes mapped to 357 sentinel SNPs at 274 known loci and genes mapped to all 901 loci. Sentinel gene mapping is compared to genes identified by extended LD ($r^2$>0.8). Pathway enrichment is represented as – log p value. A) Canonical pathway enrichment. B) Upstream regulator enrichment. C) Disease and Biofunction enrichment.

**Supplementary Figure 9: Exploring known and novel drug mechanisms in blood pressure.** The figure summarises known and novel target opportunities highlighted by blood pressure genetics. Ingenuity pathway analysis was used to create a network of 6,562 genes showing direct interaction with 145 known blood pressure target genes. This network was compared with all genes that are either directly associated with BP or linked by LD ($r^2$>0.8). Overlap between genetic associated genes and the BP drug interactome demonstrates genetic support for known drug mechanisms. Drugged or druggable genes showing genetic association with BP, but no interaction with the known BP drug interactome, represent potentially new mechanisms in blood pressure drug development and repositioning. Number of known and novel drugged/druggable gene associations are shown in parentheses.

**Supplementary figure 10: Comparison of beta effect sizes between individuals of European (N=757,601), African (N=7,782) and South Asian (N=10,323) ancestry.**

Scatterplots showing the direction of the standardized regression coefficient (beta) of novel (red) and known (grey) BP variants between Europeans and Africans (a,b,c) and South Asians (d,e,f), on the three studied BP phenotypes.

**Supplementary Figure 11**: **Correlation and distribution of minor allele frequencies (MAF) of BP variants in individuals of European (N=757,601), African (N=7,782) and South Asian (N=10,323) ancestry.** Scatterplot showing the correlation and the distribution of MAF of novel (red) and known (grey) BP variants between a) Europeans and Africans and b) Europeans and South Asians. ρ is the Pearson correlation coefficient.

**Supplementary Figure 12: Ethnicity clustering performed using PCA.** PC1 is plotted against PC2 for all N=486,683 UK Biobank participants post-QC, colour-coded according to the five ethnic clusters created from our K-means PCA clustering, from which only "White" Caucasians are selected for analysis of individuals of European ancestry. Plot (A) shows the clustering for all subjects, whereas plot (B) only shows the subsets of individuals selected for race-stratified analysis, after combining information together from both the PCA clustering and the self-reported ethnicity. PCA: Principal Component Analysis; QC: Quality Control; PCs: Principal Components.

**Supplementary Figure 13: Quantile-Quantile plots.** QQ plots of results for (A) systolic blood pressure (SBP), (B) diastolic blood pressure (DBP), (C) pulse pressure (PP) from GWAS discovery (N=757,601). The black curves are based on all SNPs in the corresponding analysis, with Minor Allele Frequency ≥ 1%. The green curves are results after excluding SNPs within the 274 known loci (±500kb; Linkage Disequilibrium $r^2 \geq 0.1$). The *P*-values have been derived from inverse variance fixed effects meta-analysis.

## Supplementary Methods

### 1. UK Biobank data

The UK Biobank cohort includes ~500,000 volunteers aged 40-69 years of age ascertained through NHS registers[1]. Following informed consent participants completed a standardised questionnaire on life course exposures, medical history and treatments and underwent a standardised portfolio of phenotypic tests including two blood pressure (BP) measurements taken seated after two minutes rest using an appropriate cuff and an Omron HEM-7015IT digital blood pressure monitor. A manual sphygmometer was used if the standard automated device could not be employed. Body mass index (BMI) was calculated as weight (kg) divided by height squared ($m^2$) with weight measured using an electronic weighing scale (Tanita BC-418). The participants undergo longitudinal life course linkage to electronic health data including Hospital Episode Statistics and Office for National Statistics cause of death data.

The UK Biobank and UK BiLEVE genotyping arrays were purpose-designed specifically for the UK Biobank project and share 95% marker content. Variants were imputed centrally by UKB using a reference panel that merged the UK10K and 1000 Genomes Phase 3 panel as well as the Haplotype Reference Consortium (HRC) panel[2]. For current analysis only SNPs imputed from the HRC panel were analysed (N=39,235,157), of which ~7.1 million SNPs with minor allele frequency (MAF) >1% and imputation quality INFO > 0.1 are analysed here for GWAS.

### 2. UKB Quality Control

All SNPs had passed central Quality Control (QC) checks, such as departures from Hardy-Weinberg Equilibrium, batch and plate effects, sex effects, array effects and discordance across control replicates. The SNPs that failed QC were set to missing for all individuals in the corresponding batch within the final genetic data files provided. Likewise, the QC performed centrally for each sample tested for heterozygosity and missing rates. Genotypes of 488,377 UKB participants were released after the QC. Full details of the QC of the genetic data performed centrally by UK Biobank are available[2].

Additionally, we excluded 968 individuals listed as QC outliers for heterozygosity or missingness and 378 individuals with sex discordance resulting to 486,683 individuals in both the post-QC genetic data and the overall phenotype data to consider for analysis, using information and data in the sample QC files provided centrally by UKB.

We restricted our data to a subset of European ancestry individuals for analysis. First, we excluded anyone who self-reported as a non-European ancestry keeping only those with self-reported ethnicity categorised as either White, other, mixed, or missing (NA). Then, using principal components (PC) provided by UK Biobank, we performed a 4-means clustering according to each of PC1 and PC2 separately using the *kmeans* algorithm in R statistical software, corresponding to four ethnic groups (White, Black, Asian, Chinese) and created an intersection of these two clusterings, to create five final clusters (White,

Black, Asian, Chinese, Mixed/Other) (**Supplementary Fig. 12**). Finally, we combined the information from the self-reported (White, other, mixed or NA) and the PCA-ancestry data to get an intersection of a total N=460,468 Europeans.

### 3. Phenotypic data

We performed analyses for systolic (SBP) and diastolic (DBP) BP and for pulse pressure. We calculated the mean SBP and DBP values from two automated (N=418,755) or two manual (N=25,888) BP measurements. For individuals with one manual and one automated BP measurement (N=13,521), we used the mean of these two values. For individuals with only one available BP measurement (N=413), we used this single value. These phenotypic sample exclusions were applied to all individuals who had passed the above genetic QC. Following both genetic and phenotypic data QC and by excluding pregnant women (n=372) and those individuals who had withdrawn consent (N=36), the sample size for analysis therefore included N=458,577 and N=458,575 European ancestry individuals for SBP and DBP, respectively.

Analysis of the summary descriptive statistics of the UK Biobank sample shows there were small but significant differences when comparing the UK Biobank vs UK BiLEVE participants, for age and BMI, due to large sample sizes. UK BiLEVE participants were slightly older and heavier compared to the UK Biobank participants. Moreover males and females were equally represented in the UK BiLEVE sample whereas more females (54.7%) were included in UK Biobank data (**Supplementary Table 1a).**

### 4. UKB analysis

For the UKB GWAS we performed linear mixed model (LMM) association testing under an additive genetic model of the three (untransformed) continuous, medication-adjusted BP traits (SBP, DBP, PP) for all measured and imputed genetic variants in dosage format using the BOLT-LMM (v2.3) software[3]. We used genotyped SNPs filtered for MAF > 5%; HWE $P > 1x10^{-6}$; missingness < 0.015, to estimate the parameters of the linear mixed model, for the initial modelling step only. Within the association analysis, we adjusted for the following covariates: sex, age, age$^2$, BMI and a binary indicator variable for UKB vs UK BiLEVE to account for the different genotyping chips and different study ascertainment. The association analysis performed by BOLT-LMM (v2.3) corrects for population structure and cryptic relatedness in very large datasets such as UKB. The genome-wide association analysis of all imputed SNPs was restricted to variants with MAF ≥ 1% and INFO > 0.1. We ran all analyses independently in parallel by two analysts across different sites and results were compared for consistency.

## 5. Genomic inflation and confounding

We applied the univariate LD score regression method (LDSR)[4] to test for genomic inflation that is expected for polygenic traits like BP, with large sample sizes, and especially also from analyses of such dense genetic data with so many SNPs in high-LD[5]. LDSR intercepts were 1.217 (standard error (se) 0.018), 1.219 (se:0.020) and 1.185 (se:0.017) for SBP, DBP and PP respectively, and were used to adjust the UKB GWAS results for genomic inflation, prior to the meta-analysis**.**

## 6. International Consortium for Blood Pressure (ICBP) GWAS

ICBP GWAS is an international consortium to investigate BP genetics[6-8]. We combined previously reported post-quality control (QC) GWAS data from 54 studies (N=150,134)[8], 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). Full study names, cohort information and general study methods are provided in supplementary tables **(Supplementary Table 1b; Supplementary Tables 20a-c)**.

Definition of phenotype data and GWAS analyses of SBP, DBP and PP were as per our previous ICBP protocol for 54 studies[8], extended to the additional 23 studies for which new data were available. Residuals were calculated for each trait using a normal linear regression of the medication-adjusted[9] trait values (mmHg) on sex, age, age$^2$ and BMI, with optional

inclusion of additional covariates to account for population stratification. Methods to account for relatedness within a study were used where appropriate (**Supplementary Table 20b**). Association testing was carried out by linear regression under an additive genetic model.

Prior to meta-analysis of all 77 ICBP GWAS studies, we undertook central QC checks across all studies. This included checks to ensure allele frequency consistency (across studies and with reference populations), checks of effect size and standard error distributions (i.e. to highlight phenotype issues) and generation of quantile-quantile (QQ) plots and genomic inflation factor lambdas to check for over- or under-inflation of test statistics. Genomic control was applied (if lambda>1) at study-level. Variants with imputation quality <0.3 were excluded prior to meta-analysis. EasyQC was used for the QC process[10]. Finally, data were filtered to SNPs with MAF ≥ 1% and N effective sample size > 60%. Meta-analysis was performed using METAL and inverse variance weighted fixed effects models. Between-study heterogeneity was assessed using the Cochran's Q statistic and we performed additional filtering removing heterogeneous variants with Cochran's Q $p<1x10^{-4}$.

### 7. Meta-analyses of discovery datasets

We performed a fixed-effects inverse variance weighted meta-analysis[11] using METAL[12] to obtain summary results from the combined UKB and ICBP GWAS, for up to N=757,601 participants and ~7.1 M SNPs with MAF>1% present in both the HRC-imputed UKB data and ICBP meta-analysis for all three traits. To verify concordance of the MAF between the two datasets in the discovery stage we checked the consistency of the effect allele frequencies between the two datasets.

### 8. Linkage Disequilibrium calculations

Linkage Disequilibrium (LD) was calculated between all published and novel sentinel SNPs within the full genetic dataset using PLINK2 software[13]. In order to do this, all genetic data with INFO > 0.1, were converted from BGEN format to PLINK binary files in best-guess genotype format, using PLINK2 software with default parameters. For any given SNP for which LD calculations were performed, the LD was estimated for all variants within a 5 mb window downstream and upstream of this reference SNP. All variants in LD with the reference SNP reaching an $r^2 \geq 0.1$ threshold were identified. LD calculations were done within HRC imputed data only and proxies were used for 13 of the 357 published SNPs not contained within the HRC imputed UKB data. These proxies were obtained using previous LD data from the interim UKB 150K release which contained additional 1000G/UK10K imputations. They were selected as the SNP with highest $r^2$, contained within HRC (within 500kb). To ensure that all published loci were captured, we also calculated LD for lower frequency published SNPs using the full UKB BGEN dataset even though our analysis was restricted to MAF ≥ 1%.

### 9. Locus definition

SNPs within +/-500kb of sentinel SNPs were extracted from the PLINK best-guess genotype format files using PLINK2 software. SNPs were filtered for those included in the HRC imputation panel and were combined with the SNPs in LD ($r^2 \geq 0.1$) per sentinel SNP, to obtain the list of all unique SNPs within a locus.

### 10. Previously reported variants

We compiled from the peer-reviewed literature all 357 SNPs previously reported to be associated with BP at the time that our analysis was completed, that have been identified and validated as the sentinel SNP from primary analyses from previous BP genetic association studies[7,14-19].[20]. These 357 published SNPs correspond to 274 distinct loci, according to locus definition of: (i) SNPs within ±500kb distance of each other; (ii) SNPs in Linkage Disequilibrium (LD), according to a threshold of $r^2 \geq 0.1$. We then augment this list to all SNPs present within our data, which are contained within these 274 published BP loci, i.e. all SNPs which are located ± 500kb from each of the 357 published SNPs and/or in LD with any of the 357 previously reported SNPs ($r^2 \geq 0.1$). This allows us to exclude all SNPs in published BP loci from our discovery meta-analysis, in order to only consider novel findings. Additionally, we are able to extract results from the discovery meta-analysis for all SNPs in LD with the 357 published SNPs, for each of the three BP traits to confirm that the previously reported BP associations show support for association with at least one of the three BP traits within our data too. The HLA region was considered as known and due to complexity, we excluded the whole region completely (chr 6:25-34 mb).
All SNPs within the loci were mapped to genes (GRCh37.75) when the variant localized within 5kb of the start or end of the gene's transcription (bedtools v2.17). Any genes which were annotated from previously reported-LD variants were listed and referred to as previously reported BP genes **(Supplementary Table 5)**.

### 11. Genomic inflation and polygenicity (QC)

We checked Quantile-Quantile plots (**Supplementary Fig. 13**) of the overall meta-analysis with and without exclusion of all SNPs within published loci. After inspection of our QQ-plots, we applied the LD score regression approach[21] to determine whether any inflation was due to polygenicity or underlying population stratification. We calculated the LDSR intercept, after the exclusion of the SNPs at published loci in the discovery meta-analysis of UKB and ICBP. It was 1.090 (0.017), 1.097 (0.017) and 1.064 (0.0146) for SBP, DBP and PP respectively; hence showing very little inflation in the discovery GWAS after the exclusion of published loci.
We also used the LD score regression to estimate the potential overlap between UKB and UK-based cohorts in ICBP by calculating the inflation of the effect estimates due to confounding from overlapping samples. The overlap is minimal as provided by the estimated calculation of the ratio ($intercept_{(LDSR)}$-1)/Mean $\chi^2$-1). Specifically, the ratios were 0.072,

0.073 and 0.062 for SBP, DBP and PP respectively, suggests that ~93-94 of the inflation for BP traits is due to polygenicity rather than population stratification, cryptic relatedness or technical artefacts.

## 12. Selection of variants for follow-up

After exclusion of all SNPs within the 274 published loci, we considered for follow-up any SNPs remaining with $P < 1x10^{-6}$ from the discovery meta-analysis for any of the three BP traits, and concordant direction of effect between UKB vs ICBP. By ranking the SNPs by significance in order of minimum *P*-value across all BP traits, we performed an iterative algorithm to determine the number of novel signals: with the top-ranked SNP being the sentinel (most significant) SNP of the 1st signal and subsequently removing all SNPs ±500kb within the same locus; then the top-ranked SNP of the remaining list becoming the sentinel SNP of the 2nd signal, etc, until the list is empty. Secondly, we calculated the pairwise LD of all sentinel SNPs, and merged together any SNPs in LD ($r^2 \geq 0.1$) belonging to the same locus. Our final selection of lookup SNPs for follow-up hence contained the sentinel SNP from each association signal, with all selected SNPs being pairwise-independent by LD ($r^2 < 0.1$). To aid replication meta-analyses across different datasets with different genetic coverage, we also selected up to two of the best proxies for each sentinel SNP. The proxies were selected as the two SNPs tagging the same signal in high LD ($r^2 \geq 0.8$) which are also associated with at least one BP trait ($P < 1x10^{-5}$) and at a similar level of significance as the sentinel SNP (<1.5-fold difference in the ratio of the $-\log_{10}$(P-value) results of the sentinel and proxy SNP). Of the 1,062 novel loci containing previously unreported SNPs with MAF ≥ 1%, and $P < 1x10^{-6}$, 1021 lead SNPs were available in both datasets from the replication resources (MVP and EGCUT); for the remaining 41 SNPs we used one of the two best proxies **(Supplementary Table 22)**. Overall, all 1,062 signals considered for lookups were available for replication analysis, either by the sentinel SNP or a good proxy **(Supplementary Table 23)**.

## 13. Replication datasets and meta-analysis for two-stage design

We used two independent external data sets for replication. We considered SNPs with MAF ≥ 1% and performed a reciprocal replication exchange with the Million Veteran's Program (max N = 220,520). We also sought independent replication in the Estonian Genome Center, University of Tartu (EGCUT) Biobank (max N=28,742). This provides a total of N = 249,262 independent samples of European descent available for replication.

The Million Veteran Program (MVP) is a large cohort of fully consented participants who were recruited from the patient populations of approximately 51 Veteran's Administration (VA) medical facilities. Recruitment began in 2011 and is conducted in-person, which is initiated by an invitation letter and completed by answering baseline and lifestyle questionnaires, providing a blood sample, and providing access to medical records, and giving permission for re-contact. Consent to participate is provided after counselling by research staff and mailing of informational materials. All documents and protocols have

been approved by the VA Central Institutional Review Board. Blood samples are collected by phlebotomists and banked at the VA Central Biorepository in Boston, MA. Genotyping was conducted using a customized Affymetrix Axiom Biobank Array chip with additional content added to provide coverage of African and Hispanic haplotypes, as well as markers for common diseases in the VA population. Researchers are provided with de-identified versions of these data, and do not have the ability or authorization to link these details with a participants' identity.

For this analysis, we selected adults (age ≥ 18) and used the earliest median eligible non-Emergency Department outpatient measured SBP in the EHR, and also used the corresponding DBP from this measure. Measures are ineligible if they occur at or after an ICD-9 code from the groups 585, 405, or 428. If pain scores were available, we censored BP measures taken during encounters when a pain score ≥5 was recorded. For measures taken while a patient was on an antihypertensive medication we added 15 mm Hg to SBP and 10 mm Hg to DBP. We adjusted linear regression models analysing SNP associations for age at BP measure, age$^2$, sex, BMI measured within 1 year of BP measure, and 10 principal components of ancestry in analyses. Primary analyses were conducted using SNPTEST by self-reported race/ethnicity for MVP (White non-Hispanic).

Similarly, EGCUT ran a GWAS for unrelated individuals of European descent using the same model in EPACTS for dosage data that were imputed using the EstRef imputation panel[22]. Allele frequencies and strand alignments were tracked for consistency with the two datasets. The two independent datasets were quality controlled to ensure concordant MAF with those of the discovery effort and were synthesized using fixed effect inverse variance weighted meta-analysis.

### 14. Combined meta-analyses for two-stage design

In the combined two-stage meta-analysis we synthesized the results of the discovery meta-analysis (UKB and ICBP GWAS) with the lookup results from the two independent replication studies (replication meta-analysis) using fixed effect inverse variance weighted meta-analysis. For consistency with the SNPs present in the replication data sets, the results for the proxy SNPs were used within the discovery input for the variants that required proxies or alternative SNPs to the sentinel SNPs for the lookups. Genomic control had already been applied in both the UKB and ICBP datasets at study-level. No further GC corrections were applied in METAL for our combined meta-analyses.

### 15. Significance thresholds for two-stage design

All of the following criteria must be satisfied for a signal to be reported as a novel signal of association with BP using our two-stage design:

(i) the sentinel SNP shows significance ($P < 1\times10^{-6}$) in the discovery meta-analysis of UKB and ICBP, with concordant direction of effect between UKB and ICBP;

(ii) the sentinel SNP is genome-wide significant ($P < 5\times10^{-8}$) in the combined meta-analysis of discovery and replication (MVP and EGCUT) (replication, described below);

(iii) the sentinel SNP shows support ($P < 0.01$) in the replication meta-analysis of MVP and EGCUT alone (Supplementary Methods);

(iv) the sentinel SNP has concordant direction of effect between the discovery and the replication meta-analyses;

(v) the sentinel SNP must not be located within any of the 274 previously reported loci described above.

The primary replicated trait was then defined as the replicated BP trait with the most significant association from the combined meta-analysis of discovery and replication (in the case of many SNPs replicating for more than one BP trait).

We note that the standard genome-wide significance threshold ($P < 5 \times 10^{-8}$) is appropriate for our combined meta-analysis. The UK Biobank-GWAS analysis follows up 7 million SNPs with MAF ≥ 1% and coverage in HRC data. In addition, we require replication support of $P < 0.01$ which is more stringent than a range of thresholds calculated according to False Discovery Rate (FDR) which gives FDR thresholds of $0.01 < P < 0.04$ using the approaches proposed by Benjamini and Hochberg [23]and Benjamini and Yekutieli[24] respectively. As a further protection against false positive findings, we require concordance in direction of effect between the discovery and replication meta-analysis results.

**16. <u>Significance thresholds from one-stage design</u>**

Variants that were looked-up but did not replicate according to the two-stage criteria were considered in a one-stage design. All of the following criteria must be satisfied for a signal to be reported as a novel signal of association with BP using our one-stage criteria:

i) the sentinel SNP has $P < 5\times10^{-9}$ in the discovery (UKB+ICBP) meta-analysis

ii) the sentinel SNP shows support ($P < 0.01$) in the UKB GWAS alone

iii) the sentinel SNP shows support ($P < 0.01$) in the ICBP GWAS alone

iv) the sentinel SNP has concordant direction of effect between UKB and ICBP datasets

v) The sentinel SNP must not be located within any of the 274 previously reported loci described above or the recently reported non-replicated loci from Hoffman et al[17]

We selected the one-stage *P*-value threshold to be more stringent than a genome-wide significance *P*-value, in order to ensure robust findings and to minimize false positives. The threshold of $P < 5 \times 10^{-9}$ has been proposed as a more conservative statistical significance threshold, e.g. for whole-genome sequencing-base studies[25]. This is even more conservative than the number of independent statistical tests performed in our data that they were calculated by assessing the correlation between nearby test statistics empirically[26].

Selection of variants from the meta-analysis of UKB and ICBP was performed as described above for the two-stage design.

**17. <u>Conditional analysis</u>**

For conditional analysis, we used two different methodological approaches, each using the Genome-wide Complex Traits Analysis (GCTA) software: (i) genome-wide conditional analysis; and (ii) locus-specific conditional analysis.

*(i) Genome-wide 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. The summary statistics from the GWAS discovery meta-analysis of UKB and ICBP were used as the input summary data. Three sets of analyses were performed, one for each BP trait, using the trait-specific meta-analysis results. The UKB genetic data was used as the reference genotype-level data, in PLINK format, restricted to MAF ≥ 1%. As the UKB genetic data is stored individually by chromosome, the GCTA analyses were performed separately per chromosome. With the combination of these two input data files, analysis was therefore restricted to the ~7 million HRC-imputed SNPs with MAF ≥ 1% in common to both UKB and ICBP data from the GWAS discovery meta-analysis. Within the UKB genetic data, LD was calculated between all pairwise SNPs. By using the UKB data as our reference genotype data, our analysis is equivalent to a full multiple regression analysis between the actual genotype data and the BP phenotypic traits. Within the GCTA analysis, a p-value cut-off of $5x10^{-8}$ was used as the selection threshold, in order to identify secondary signals at the genome-wide significance level. For the collinearity threshold we used the default cut-off 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 all the 22 chromosome output files together, each trait-specific analysis results in a distinct set of jointly independent significant signals. Then, for each BP trait, by excluding all SNPs which are in LD ($r^2 \geq 0.1$) with any of the 357 published SNPs (**Supplementary Table 4**) or any of the sentinel SNPs at the 535 novel loci (**Supplementary Tables 2a-c and Supplementary Tables 3a-c**) or 92 newly replicated loci (**Supplementary Table 5**), all remaining SNPs are additional, independent secondary signals associated with the given BP trait. However, after merging together all genome-wide results across all three BP traits, some signals could be duplicated across traits, so pairwise LD was calculated 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 BP 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 901 BP loci separately.

Within each of the 535 novel loci (**Supplementary Tables 2a-c and Supplementary Tables 3a-c)** and the 92 loci replicated for the first time (**Supplementary Table 5**), we searched for any potential secondary signals, which are independently associated in addition to the sentinel SNP. Conditional analysis was performed on all HRC-imputed SNPs with MAF ≥ 1% within the 1Mb locus region centred ±500kb around the sentinel SNP, conditioning on the sentinel SNP. Analysis was performed for association of the primary validated BP trait of the sentinel SNP.

For known loci (**Supplementary Table 4**) we initially considered all 274 loci containing all 357 published SNPs. However, we excluded the HLA region (chr 6:25-34MB), as the long-range LD within this region could potentially confound the conditional analysis. Furthermore, 28 of the 357 exact SNPs were not present within the UKB HRC-imputed data at MAF ≥ 1%. For 14 of these SNPs we were able to use a good proxy SNP instead ($r^2 \geq 0.8$), but the remaining 14 SNPs were excluded due to lack of signal coverage, which excluded some loci completely if no other published SNP existed within this locus. Of the 329 published SNPs present in the UKB data, only 6 SNPs did not reach Bonferroni significance for any BP trait within the GWAS discovery meta-analysis (although the locus was still confirmed as SNPs in LD with $r^2 \geq 0.1$ within the 1Mb locus region reached nominal significance with $P < 0.01$). These 6 SNPs were therefore excluded, with the entire locus also being excluded from analysis if no other published SNPs were present. Overall this resulted in a total of 259 known loci being analysed. For loci containing only one published SNP, the 1Mb locus region centred ±500kb around the published SNP was used for analysis. For loci containing multiple published 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 published SNP, conditional analysis was performed on all HRC-imputed SNPs with MAF ≥ 1% within the 1Mb region, conditioning on the single published SNP within the locus, testing for association of the primary, most significantly associated BP trait of this published SNP from the GWAS discovery meta-analysis (see **Supplementary Table 4**). For known loci containing more than one published SNP, conditional analysis was performed within the wider locus region, conditioning jointly on all published SNPs within the locus. If any pairs of SNPs at a locus were in high LD ($r^2 \geq 0.9$) beyond the collinearity cut-off, the most significant SNP with the minimum P-value across all

BP traits from the GWAS discovery meta-analysis was selected. If all published SNPs at the locus had the same primary BP trait, GCTA testing was only performed for one trait, but if the primary traits differed across multiple published SNPs within a locus, then GCTA was run for each of the primary associated BP traits. (The post-GCTA filtering removes any duplication of signals across traits afterwards.) Of the 259 known loci analysed, 50 loci conditioned on multiple SNPs, and of these 14 loci were tested for more than one BP trait.

All these 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 UKB 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 BP itself, at genome-wide significance level

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

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

(d) not in LD with any of the 357 published SNPs or any of the sentinel SNPs at the 535 novel loci or 92 newly replicated loci ($r^2 < 0.1$)

All significant independent SNPs meeting the above criteria, from all loci across all chromosomes are 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.

The outputs from the two different approaches are therefore combined together. For robustness, a secondary signal is only claimed if the SNP is obtained from both approaches. By only selecting SNPs from approach (i) if they are also present from the results of approach (ii), we ensure that all secondary signals belong to one of the 901 BP-associated loci confirmed from our primary analysis, either from a validated novel locus, or a published

locus. By only selecting from the many SNPs from approach (ii), those which are in the final GCTA model from approach (i), we reduce the list of SNPs to signals and ensure that these signals are jointly independent. As a final check, we confirmed again, by LD calculation within the UKB genetic data using PLINK, that all the final secondary SNPs are pairwise-LD-independent ($r^2 < 0.1$). In all cases the UKB genetic data used for LD calculation was restricted to individuals of European ancestry only.

The final list of secondary SNPs which we report are therefore:

- significantly associated with BP
- statistically independent in addition to the novel sentinel SNPs and published SNPs from our primary analysis
- independent to the novel sentinel SNPs and published SNPs by LD
- jointly independent with each other statistically
- pairwise-LD-independent from each other
- within 500kb from one of the novel sentinel SNPs or published SNPs, and hence contained within one of the 901 loci from our primary analysis

The curated list of 357 published SNPs (**Supplementary Table 4**) excluded from our discovery GWAS was restricted to published SNPs which had been identified as the sentinel SNP from primary GWAS analyses and validated with independent replication. Other published BP genetics studies have also reported secondary signals from secondary analyses using a variety of different approaches. We calculated pairwise LD between all of our secondary SNPs versus all previously reported. For any of our secondary SNPs overlapping with previous findings ($r^2 \geq 0.1$), the corresponding publication is cited.

### 18. Look ups in non-European ancestries

As a secondary analysis, we look up all known and novel SNPs in non-European ancestry samples of UK Biobank. These analyses are stratified by ancestry, according to two main non-European ancestry ethnicity categories within UKB: Africans and South Asians. As with the identification of European ancestry (described above), participants were selected according to both self-reported ethnicity data and PCA ancestry clustering, e.g. for Africans, those with self-reported ethnicity (Africans, other, mixed or NA) and PCA-ancestry = Africans, giving a total of N=7,782 Africans; similarly for 10,322 South Asians and 2,156 Chinese ancestry. An equivalent GWAS-LMM analysis is performed using BOLT-LMM for this subset of variants within each stratified ancestry. (Note that analysis for 2,156 Chinese is not performed as BOLT-LMM is only recommended for N>5,000)

### 19. Genetic risk scores analyses

We calculated a genetic risk score (GRS) to provide an estimate of the combined effect of the BP raising variants on BP and risk of hypertension, and applied this to the UK Biobank

data. We first created two trait-specific weighted GRSs (i.e. SBP, DBP), for all pairwise-independent, LD-filtered ($r^2 < 0.1$) previously reported variants and 535 novel sentinel variants combined**.** For the previously reported variants, we weighted BP increasing alleles by the trait-specific beta coefficients from the ICBP meta-analysis GWAS that is part of the discovery stage **(Supplementary Table 24)**. For the novel and the 92 previously reported variants that have not been replicated before, beta coefficients of the replication meta-analysis for each BP trait were used as independent, unbiased weights. We then derived a single BP GRS as the average of the GRS for SBP and DBP, and standardize it to have mean zero and standard deviation of one. We assessed the association of the continuous GRS variable on BP by simple linear regression, and we used logistic regression to examine the association of the GRS with risk of hypertension, with and without adjustment for sex. We then applied linear and logistic regression to compare BP levels and risk of hypertension, respectively, for individuals in the top vs bottom quintiles of the GRS distribution. Similar analyses were performed for the top vs bottom deciles of the GRS distribution. We restricted our analysis to unrelated individuals of European ancestry from UKB excluding 1$^{st}$ and 2$^{nd}$ degree relatives. Specifically, we used the centrally provided kinship data from UKB to remove one of each pair of related individuals, related of 1st, and 2$^{nd}$ degree relationship. We used kinship index of 0.08838835 to exclude individuals (n=36,182) who were related to at least one individual in the final database for analysis. This resulted to N=392,092 unrelated individuals.

We also assessed the association of the GRS with cardiovascular disease in unrelated participants in UKB data, based on self-reported medical history, and linkage to hospitalization and mortality data. We used logistic regression with binary outcome variables for composite incident cardiovascular disease, incident myocardial infarction and incident stroke (using the algorithmic UKB definitions) and GRS as explanatory variable (with and without sex adjustment) **(Supplementary Table 25)**.

As a secondary analysis, and to see whether BP-associated SNPs identified from GWAS predominantly in Europeans area also associated with BP in populations of non-European ancestry, we also performed GRS analyses in non-European ancestry samples of UK Biobank. As with the identification of European ancestry (see section '2. UKB Quality Control'), participants were selected according to both self-reported ethnicity data and PCA ancestry clustering, e.g. for Africans, those with self-reported ethnicity (Africans other, mixed or NA) and PCA-ancestry = Africans, giving a total of N=7,782 participants of African descent; similarly, for 10,322 South Asians. As the sample size for participants of Chinese ancestry was much lower, we focused analyses only on Africans and South Asians. For the GRS analyses using simple linear regression, the samples also had to be restricted to a set of unrelated subjects, by using the centrally provided kinship data from UKB, as described above. This resulted in final sample sizes for analysis of 6,264 unrelated Africans and 7,881 unrelated South Asians.

## 20. Airwave study data

The Airwave Health Monitoring Study (Airwave)[27] was used as an independent cohort for the GRS analyses, the analysis of metabolomics data and calculation of the percentage of the variance explained. The Airwave analyses included 14,004 participants with high quality HRC imputed genetic data. Systolic and diastolic blood pressures were measured as three consecutive readings using a digital blood pressure monitor (Omron HEM 705-CP digital BP monitor). Mean SBP and DBP adjusted for medication (as previously defined) were calculated from available readings and were used as dependent variables in the analyses. We calculated a GRS to provide an estimate of the combined effect of the BP raising variants on BP and risk of hypertension, and applied this to the Airwave study data[27] to assess effect in an independent cohort, thus avoiding any over-fitting or bias by "winner's curse".
To calculate the percent of variance in BP explained by genetic variants in an independent dataset, we generated the residuals from a regression of each trait against age, age$^2$, sex and body mass index in Airwave. We then fit a second linear model for the trait residuals with all the variants in the GRS plus the top 10 principal components, and estimated the percentage variance of the dependent (BP) variable explained by the GRS. We considered three different levels of the GRS: (i) all pairwise-independent, LD-filtered ($r^2 < 0.1$) published SNPs within the known loci; (ii) all known SNPs and sentinel SNPs at novel loci; (iii) all independent signals at all 901 known and novel loci including the 163 secondary SNPs.
From the Airwave plasma $^1$H NMR metabolomics we use $^1$H NMR lipidomics data on plasma from a subset of 2,022 participants. For each replicated BP-associated SNP we ran association tests with the lipidomics data using linear regression analyses, adjusted for age and sex. We computed significance thresholds using a Bonferroni correct *P*-value ($4.7 \times 10^{-4}$). We also examined associations between each replicated SNP and a subset of 1,941 participants in Airwave with data from Metabolon platform.

## 21. Cardiovascular outcomes data in UK Biobank

To classify cardiovascular disease (CVD) outcomes we used self-reported baseline information on CVD prevalence available in UKB, and linkage to Hospital Episodes Statistics (HES) and mortality data **(Supplementary Table 25)**. HES provides detailed information for participants admitted to hospital and includes coded data on diagnoses and operations. Coronary artery disease and stroke were classified using International Classification of Disease (ICD) 9 and 10 codes and operation codes. All events occurring before assessment were categorized as pre-existing disease and were excluded from the CVD analysis (n=22,829). Final sample used for CVD analysis, consisted of 392,092 European ancestry individuals. The large UK Biobank cohort with sufficient numbers of cardiovascular events enables the assessment of cardiovascular risk within the same data set, noting that results are still independent, as the variants within the GRS are selected for their association with BP, not for cardiovascular outcomes.

## 22. Functional analyses

We used an integrative bioinformatics approach to collate functional annotation at both the variant and gene level for each SNP within the reported blood pressure loci. SNPs in LD $r^2 \geq 0.8$ with the blood pressure-associated SNPs are considered after extraction using PLINK. At the variant level we use Variant Effect Predictor (VEP) to obtain comprehensive characterization of variants, including consequence (e.g. downstream or non-coding transcript exon), information on nearest genomic features and, where applicable, amino acid substitution functional impact, based on SIFT and PolyPhen. The biomaRt R package is used to further annotate the nearest genes.

GTEx database: We evaluate all SNPs in LD ($r^2 \geq 0.8$) with our novel sentinel SNPs for evidence of mediation of expression quantitative trait loci (eQTL) in all 44 tissues using the Genotype-Tissue Expression (GTEx) database (www.gtexportal.org), to highlight specific tissue types which show eQTLs for a larger than expected proportion of novel loci. We further seek to identify novel loci with the strongest evidence of eQTL associations in arterial tissue, in particular. A locus is annotated with a given eGene only if the most significant eQTL SNP for the given eGene is in high LD ($r^2 \geq 0.8$) with the lead SNP, suggesting that the eQTL signal co-localises with the sentinel SNP.

Fantom5: We annotated nearest genes, eGenes and Hi-C interactors with HUVEC, HVSMC and HAEC expression from the Fantom5 project (fantom.gsc.riken.jp/5). Genes that had higher then median expression levels in the given cell types were indicated as expressed.

DeepSEA: To identify SNPs in the novel loci that have a non-coding functional effect (influence binding of transcription factors or RNA polymerase, or influence DNase hypersensitivity sites or histone modifications), we used DeepSEA, a deep learning algorithm, that learnt the biding and modification patterns of ~900 cell/factor combinations[28]. A change > 0.1 in the binding score predicted by DeepSEA for the reference and alternative alleles respectively has been shown to have high true positive rate ~80-95% and low false positive rate ~5-10% therefore we used this cut-off to find alleles with non-coding functional effect.

Hi-C analysis: We identified potential target genes of regulatory SNPs using long-range chromatin interaction (Hi-C) data from HUVECs, aorta, adrenal glands, neural progenitor and mesenchymal stem cell, which are tissues and cell types that are considered relevant for regulating blood pressure. Hi-C data are corrected for genomic biases and distance using the Hi-C Pro and Fit-Hi-C pipelines according to Schmitt et al. (40kb resolution – correction applied to interactions with 50kb-5Mb span). We found the most significant promoter interactions for all potential regulatory SNPs (RegulomeDB score$\leq$5)[29] in LD ($r^2 \geq 0.8$) with our novel sentinel SNPs and published SNPs, and choose the interactors with the SNPs of highest regulatory potential to annotate the loci. We then performed overall enrichment testing across all loci.

DEPICT: Firstly, we used DEPICT[30] (Data-driven Expression Prioritized Integration for Complex Traits) to identify highly expressed tissues and cells within the blood pressure loci. DEPICT uses a large number of microarrays (~78k) to identify cells and tissues where the

genes are highly expressed and uses precomputed GWAS phenotypes to adjust for co-founding sources. Secondly, we used DEPICT to test for enrichment in gene sets associated with biological annotations (manually curated and molecular pathways, phenotype data from mouse KO studies). Using the co-expression data DEPICT calculates a probability for each gene to belong to a given gene set and uses this weight the enrichment of the genes present in the tested loci. DEPICT provides a *P*-value of enrichment and false discovery rates adjusted *P*-values for each tissue/cells or gene set tested based on the algorithm described in Pers et al[30]. We report significant enrichments with a false discovery rate <0.01. The variants tested were the 357 published blood pressure associated SNPs at the time of analysis and a set including all BP (for novel: combined *P*-value $<1 \times 10^{-12}$) variants. The *P*-value threshold was selected in order to include as many as possible variants in the analysis
FORGE: Furthermore, to investigate cell type specific enrichment within DNase I sites, we used FORGE, which tests for enrichment of SNPs within DNase I sites in 123 cell types from the Epigenomics Roadmap Project and ENCODE[31]. The results of analyses from only the published sentinel SNPs were compared to the results from analyses of SNPs at all loci, including also the novel loci discovered in our study in order to evaluate the overall tissue specific enrichment of blood pressure associated variants.

Functional Analyses: Genes

IPA: At the gene level, we used Ingenuity Pathway Analysis (IPA) software (IPA®,QIAGEN Redwood City,www.qiagen.com/ingenuity) to review genes with prior links to blood pressure, based on annotation with the "Disorder of Blood Pressure", "Endothelial Development" and "Vascular Disease" Medline Subject Heading (MESH) terms. We used the Mouse Genome Informatics (MGI) tool (http://www.informatics.jax.org/batch) to identify BP and cardiovascular relevant mouse knockout phenotypes for all genes linked to all BP variants. We also used IPA to identify genes which interact with known targets of anti-hypertensive drugs. Genes were also evaluated for evidence of small molecule druggability or known drugs based on queries of the Drug Gene Interaction database (dgidb.genome.wustl.edu).

### 23. Cross-trait lookups

PhenoScanner and GWAS catalog: We query SNPs against PhenoScanner[32] and GWAS catalog[33] to investigate trait pleiotropy, extracting all association results with genome-wide significance at $P < 5 \times 10^{-8}$, for all SNPs in high LD ($r^2 \geq 0.8$) with all 535 sentinel novel SNPs, to highlight the novel loci with strongest evidence of association with other traits.
DisGeNET: We further evaluated pleiotropic effects using DisGeNET, a resource that integrates data from expert curated repositories, GWAS catalogues, animal models and the literature[34,35]. At the SNP level, overlaps with DisGeNET terms were computed, with roughly the same number of markers in the published and novel BP loci. Thus, given the expected saturation of the overlaps, a more than double increase in it from published-only, to all BP loci, indicates that pleiotropy is more frequent in the novel BP loci. At the gene level,

overrepresentation enrichment analysis (ORA) with WebGestalt[36] (on the nearest genes to all BP loci) was carried out.

Global Biobank Engine and GeneATLAS: We tested sentinel SNPs at all (n=901) BP loci for association with lifestyle related data including food, water and alcohol intake, anthropomorphic traits and urinary sodium, potassium and creatinine excretion using the recently developed Stanford Global Biobank Engine and the Gene ATLAS. Both are search engines for GWAS findings for multiple phenotypes in UK Biobank. We used a Bonferroni corrected significance threshold of $P < 1x10^{-6}$ to deem significance.

**Study Funding**

The AGES-Reykjavik study was supported by contract from the National Institute on Aging (N01-AG-1-2100 and HHSN27120120022C), the Icelandic Heart Association (Hjartavernd)and the Icelandic Parliament (Althingi)

ASCOT**:** This work was funded by the National Institutes for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Barts Biomedical Research Unit and the NIHR Biomedical Research Centre at Imperial College, the International Centre for Circulatory Health Charity and the Medical Research Council through G952010. We thank all ASCOT trial participants, physicians, nurses, and practices in the participating countries for their important contribution to the study.

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Funding support for the Genetic Epidemiology of Causal Variants Across the Life Course (CALiCo) program was provided through the NHGRI PAGE program (U01 HG007416). Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The authors thank the staff and participants of the ARIC study for their important contributions.

The ASPS research reported in this article was funded by the Austrian Science Fond (FWF) grant number P20545-P05 and P13180. The Medical University of Graz supports the databank of the ASPS. The authors thank the staff and the participants of the ASPS for their valuable contributions. The authors thank Birgit Reinhart for her long-term administrative commitment and Ing Johann Semmler for the technical assistance

The B58C acknowledge use of phenotype and genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. Genotyping for the B58C-WTCCC subset was funded by the Wellcome Trust grant 076113/B/04/Z. The B58C-T1DGC genotyping utilized resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. B58C-T1DGC GWAS data were deposited by the Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research (CIMR), University of Cambridge, which is funded by Juvenile Diabetes Research Foundation International, the Wellcome Trust and the National Institute for Health Research Cambridge Biomedical Research Centre;

the CIMR is in receipt of a Wellcome Trust Strategic Award (079895). The B58C-GABRIEL genotyping was supported by a contract from the European Commission Framework Programme 6 (018996) and grants from the French Ministry of Research.
BRIGHT**:** This work was funded by the Medical Research Council of Great Britain (grant number: G9521010D). 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 themes contributing to the translational research portfolio for the NIHR Barts Cardiovascular Biomedical Research Unit.
This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HL130114, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, and R01HL120393 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 the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
The Cilento study was supported by the Italian Ministry of Education Universities and Research (Interomics Flagship Project, PON03PE_00060_7), FP6 (Vasoplus-037254), the Assessorato Ricerca Regione Campania, the Fondazione con il SUD (2011-PDR-13), and the Istituto Banco di Napoli - Fondazione to MC. The Cilento study address special thanks to the populations of Cilento for their participation in the study.
The CoLaus study was and is supported by research grants from GlaxoSmithKline (GSK), the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 3200B0-105993, 3200B0- 118308, 33CSCO-122661, and 33CS30-139468. The CoLaus study thank all participants, involved physicians and study nurses to the CoLaus cohort.
The CROATIA-Vis, CROATIA-Korcula and CROATIA-Split studies in the Croatian islands of Vis and Korcula and mainland city of Split were supported by grants from the Medical Research Council (UK); the Ministry of Science, Education, and Sport of the Republic of Croatia (grant number 108-1080315-0302); the European Union framework program 6 European Special Populations Research Network project (contract LSHG-CT-2006-018947) and the Research Centre of Excellence in Personalized Health Care. The CROATIA studies would like to acknowledge the invaluable contributions of the recruitment teams (including those from the Institute of Anthropological Research in Zagreb) in Vis, Korcula and Split, the administrative teams in Croatia and Edinburgh, and the people of Vis, Korcula and Split. SNP genotyping of the CROATIA-Vis samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, WGH, Edinburgh, Scotland. SNP genotyping

for CROATIA-Korcula was performed by Helmholtz ZentrumMünchen, GmbH, Neuherberg, Germany. The SNP genotyping for the CROATIA-Split cohort was performed by AROS Applied Biotechnology, Aarhus, Denmark .
EPIC-CVD has been supported by the European Union Framework 7 (HEALTH-F2-2012-279233), the European Research Council (268834), the UK Medical Research Council (G0800270 and MR/L003120/1), the British Heart Foundation (SP/09/002 and RG/08/014 and RG13/13/30194) and the UK National Institute of Health Research. We thank all EPIC participants and staff for their contribution to the study. We thank staff from the EPIC-CVD and EPIC-InterAct Coordinating Centres for carrying out sample preparation and data-handling work, particularly Sarah Spackman (EPIC-CVD Data Manager), and Cambridge Genomic Services for genotyping.

EPIC-CVD consortium:
Kim Overvad[1,2], Anne Tjønneland[3], Francoise Clavel-Chapelon[4], Rudolf Kaaks[5], Heiner Boeing[6], Antonia Trichopoulou[7,8], Pietro Ferrari[9], Domenico Palli[10], Vittorio Krogh[11], Salvatore Panico[12], Rosario Tumino[13], Giuseppe Matullo[14,15], Jolanda Boer[16], Yvonne van. der Schouw[17], Elisabete Weiderpass[18,19,20,21], J. Ramon Quiros[22], María-José Sánchez[23,24], Carmen Navarro[25], Conchi Moreno-Iribas[26], Larraitz Arriola[27], Olle Melander[28], Patrik Wennberg[29], Nicholas J. Wareham[30], Timothy J. Key[31], Elio Riboli[32], Adam S. Butterworth[33,34], Joanna M M Howson[33], John Danesh[33,34,35]
1. Department of Public Health, Section for Epidemiology, Aarhus University, Aarhus, Denmark
2. Department of Cardiology, Aalborg University Hospital, Aalborg, Denmark
3. Diet, Genes and Environment, Danish Cancer Society Research Center, Copenhagen, Denmark
4. INSERM, Centre for Research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones, and Women's Health Team, Institut Gustave Roussy, Villejuif, France
5. Division of Cancer Genetic Epidemiology, German Cancer Research Centre (DKFZ), im Neuenheimer Feld 581, 69120 Heidelberg, Germany
6. Department of Epidemiology, German Institute of Human Nutrition (DIfE), Potsdam-Rehbrücke, Germany
7. WHO Collaborating Center for Nutrition and Health, Unit of Nutritional Epidemiology and Nutrition in Public Health, Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, Athens, Greece
8. Hellenic Health Foundation, Athens, Greece
9. International Agency for Research on Cancer, Lyon, France
10. Molecular and Nutritional Epidemiology Unit, Centro per lo Studio e la Prevenzione Oncologica-Scientific Institute of Tuscany, Florence, Italy
11. Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
12. Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy

13. Cancer Registry and Histopathology Unit, Civic- M.P.Arezzo Hospital, ASP Ragusa, Italy
14. Human Genetics Foundation, Turin, Italy
15. Department of Medical Sciences, University of Turin, Italy
16. Centre for Nutrition, Prevention and Health Services, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
17. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands
18. Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, The Arctic University of Norway, Tromsø, Norway
19. Department of Research, Cancer Registry of Norway, Institute of Population-Based Cancer Research, Oslo, Norway
20. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
21. Genetic Epidemiology Group, Folkhälsan Research Center, Helsinki, Finland
22. Public Health Directorate, Asturias, Spain
23. Escuela Andaluza de Salud Pública, Instituto de Investigación Biosanitaria ibs.GRANADA. Hospitales Universitarios de Granada/Universidad de Granada, Granada, Spain
24. CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
25. Epidemiology Department, Murcia Health Authority, Murcia, Spain
26. Public Health Institute of Navarra, Pamplona, Spain
27. Public Health Division of Gipuzkoa, Instituto Bio-Donostia, Basque Government, CIBERESP, Spain
28. Department of Clinical Sciences, Hypertension & Cardiovascular Disease, Clinical Research Centre, Malmö University Hospital, Malmö, Sweden
29. Department of Public Health and Clinical Medicine, Family Medicine, Umeå University, Umeå, Sweden
30. Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, UK
31. Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, UK
32. School of Public Health, Imperial College London, UK
33. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health & Primary Care, University of Cambridge, UK
34. The National Institute for Health Research Blood and Transplant Unit (NIHR BTRU) in Donor Health and Genomics at the University of Cambridge, UK
35. Wellcome Trust Sanger Institute, Genome Campus, Hinxton, UK

EPIC-Norfolk, Fenland-OMICS, Fenland-GWAS, InterAct-GWAS were supported by the 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).

The Erasmus Rucphen Family (ERF) was funded from: FP6 STRP 018947 [LSHG-CT-2006-01947]); European Network of Genomic and Genetic Epidemiology (ENGAGE) from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4- 2007-201413; the JPND PERADES project (grant number 733051021), BBMRI–NL (184.021.007 and 184.033.111) and Netherlands Heart Foundation CVON2012-03.
For the Rotterdam study there is additional funding: EU Horizon 2020 (grant agreements No 667375 and No 645740, the Netherlands Organization of Scientific Research NWO Investments (175.010.2005.011, 911-03-012, 050-060-810, 014-93-015-RIDE2). The ERF study is grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection.
The FINRISK PREDICT CVD study has been funded by the Academy of Finland (grant numbers 139635, 129494, 118065, 129322, 250207), the Orion-Farmos Research Foundation, the Finnish Foundation for Cardiovascular Research, and the Sigrid Jusélius Foundation. The FINRISK PREDICT CVD study is grateful for the THL DNA laboratory for its skillful work to produce the DNA samples used in this study. We thank the Sanger Institute genotyping facilities for genotyping the samples.
Support for FUSION was provided by NIH grants R01-DK062370 (to M.B.) and intramural project number ZIA-HG000024 (to F.S.C.). Genome-wide genotyping was conducted by the Johns Hopkins University Genetic Resources Core Facility SNP Center at the Center for Inherited Disease Research (CIDR), with support from CIDR NIH contract no. N01-HG-65403.
Generation Scotland Scottish Family Health Study (GS:SFHS) received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the UK's Medical Research Council. Ethics approval for the study was given by the NHS Tayside committee on research ethics (reference 05/S1401/89). Generation Scotland Scottish Family Health Study is grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses. Ethics approval for the study was given by the NHS Tayside committee on research ethics (reference 05/S1401/89).
The GRAPHIC Study was funded by the British Heart Foundation (BHF/RG/2000004). CPN and NJS are supported by the British Heart Foundation and NJS is a NIHR Senior Investigator. This work falls under the portfolio of research supported by the NIHR Leicester Biomedical Research Centre.
The Health 2000 Study was funded by the National Institute for Health and Welfare (THL), the Finnish Centre for Pensions (ETK), the Social Insurance Institution of Finland (KELA), the Local Government Pensions Institution (KEVA) and other organizations listed on the website of the survey (http://www.terveys2000.fi). The Health 2000 Study are grateful for the THL

DNA laboratory for its skillful work to produce the DNA samples used in this study. We thank the Sanger Institute genotyping facilities for genotyping the GenMets subcohort.
The Health ABC Study was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106 and, in part, by the NIA Intramural Research Program. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (http://biowulf.nih.gov).
The HTO study was funded by the Wellcome Trust, Medical Research Council and British Heart Foundation. The HTO study thank all the families who participated in the study
INGI-CARL and INGI-FVG studies were supported by the Italian Ministry of Health RF2010 to Paolo Gasparini, RC2008 to Paolo Gasparini
The INGI-Val Borbera population is a collection of 1,664 genotyped samples collected in the Val Borbera Valley, a geographically isolated valley located within the Appennine Mountains in Northwest Italy1. The valley is inhabited by about 3,000 descendants from the original population, living in 7 villages along the valley and in the mountains. Participants were healthy people 18-102 years of age that had at least one grandfather living in the valley. The study plan and the informed consent form were reviewed and approved by the institutional review boards of San Raffaele Hospital in Milan. The INGI-Val Borbera thank the inhabitants of the VB that made this study possible, the local administrations, the MD of the San Raffaele Hospital and Prof Clara camaschea for clinical data collection. The INGI-Val Borbera also thank Fiammetta Viganò for technical help, Corrado Masciullo and Massimiliano Cocca for building and mantaining the analysis platform. Traglia, M. et al. Heritability and demographic analyses in the large isolated population of Val Borbera suggest advantages in mapping complex traits genes. PLoS One 4, e7554 (2009) Colonna V, et al. Small effective population size and genetic homogeneity in the Val Borbera isolate. Eur J Hum Genet. 2):89-94. 2013. The research was supported by funds from Compagnia di San Paolo, Torino, Italy; Fondazione Cariplo, Italy and Ministry of Health, Ricerca Finalizzata 2008 and CCM 2010, PRIN 2009 and Telethon, Italy to DT. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
The KORA study was initiated and financed by the Helmholtz Zentrum München—German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The KORA-Study Group consists of A. Peters (speaker), J.Heinrich, R.Holle, R.Leidl, C.Meisinger, K.Strauch and their co-workers, who are responsible for the design and conduct of the KORA studies. We gratefully acknowledge the contribution of all members of field staff conducting the KORA F4

study.For this publication, biosamples from the KORA Biobank as part of the Joint Biobank Munich have been used.
Phenotype collection in the Lothian Birth Cohort 1921 was supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC), The Royal Society, and The Chief Scientist Office of the Scottish Government. Genotyping was funded by the BBSRC. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC and Medical Research Council (MRC) is gratefully acknowledged. The Lothian Birth Cohort 1921 thank the Lothian Birth Cohort 1921 participants and team members who contributed to these studies.
Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Genotyping was funded by the BBSRC. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC and Medical Research Council (MRC) is gratefully acknowledged. The Lothian Birth Cohort 1936 thank the Lothian Birth Cohort 1936 participants and team members who contributed to these studies.
The LifeLines Cohort Study, and generation and management of GWAS genotype data for the LifeLines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. The authors wish to acknowledge the services of the Lifelines Cohort Study, the contributing research centers delivering data to Lifelines, and all the study participants.
We thank U. Bultmann [(1)], J.M. Geleijnse [(2)], P. van der Harst [(3)], S. Mulder [(4)], J.G.M. Rosmalen [(5)], E.F.C. van Rossum [(6)], H.A. Smit [(7)], M. Swertz [(8)], E.A.L.M. Verhagen [(9)]
[(1)] Department of Social Medicine, University Medical Center Groningen, University of Groningen, The Netherlands
[(2)] Department of Human Nutrition, Wageningen University, The Netherlands
[(3)] Department of Cardiology, University Medical Center Groningen, University of Groningen, The Netherlands
[(4)] Lifelines Cohort Study, The Netherlands
[(5)] Interdisciplinary Center of Psychopathology of Emotion Regulation (ICPE), Department of Psychiatry, University Medical Center Groningen, University of Groningen, The Netherlands
[(6)] Department of Endocrinology, Erasmus Medical Center, Rotterdam, The Netherlands
[(7)] Department of Public Health, University Medical Center Utrecht, The Netherlands

[8] Department of Genetics, University Medical Center Groningen, University of Groningen, The Netherlands
[9] Department of Public and Occupational Health, VU Medical Center, Amsterdam, The Netherlands

The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966, G0700931), the Wellcome Trust (084723/Z/08/Z, 090532 & 098381) the NIHR (RP-PG-0407-10371), the NIHR Official Development Assistance (ODA, award 16/136/68), the European Union FP7 (EpiMigrant, 279143) and H2020 programs (iHealth-T2D, 643774). We acknowledge support of the MRC-PHE Centre for Environment and Health, and the NIHR Health Protection Research Unit on Health Impact of Environmental Hazards. The work was carried out in part at the NIHR/Wellcome Trust Imperial Clinical Research Facility. The views expressed are those of the author(s) and not necessarily those of the Imperial College Healthcare NHS Trust, the NHS, the NIHR or the Department of Health. We thank the participants and research staff who made the study possible. JC is supported by the Singapore Ministry of Health's National Medical Research Council under its Singapore Translational Research Investigator (STaR) Award (NMRC/STaR/0028/2017). We thank the participants and research staff who made the study possible.
MDC was supported by the Knut and Alice Wallenberg Foundation
MESA research was supported by the Multi-Ethnic Study of Atherosclerosis (MESA) contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420. Funding for MESA Share genotyping was provided by NHLBI Contract N02-HL-6-4278. This publication was partially developed under a STAR research assistance agreement, No. RD831697 (MESA Air), awarded by the U.S Environmental Protection Agency. It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors and the EPA does not endorse any products or commercial services mentioned in this publication. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TRO1881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The authors thank the participants of the MESA study, the Coordinating Center, MESA investigators, and study staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.The MICROS study was supported by the Ministry of Health and Department of Innovation, Research and University of the Autonomous Province of Bolzano, the South Tyrolean Sparkasse Foundation, and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). For the MICROS study, we thank the primary care

practitioners Raffaela Stocker, Stefan Waldner, Toni Pizzecco, Josef Plangger, Ugo Marcadent, and the personnel of the Hospital of Silandro (Department of Laboratory
Funding for NESDA was obtained from the Netherlands Organization for Scientific Research (Geestkracht program grant 10-000-1002); the Center for Medical Systems Biology (CSMB, NWO Genomics), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL), VU University's Institutes for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam, University Medical Center Groningen, Leiden University Medical Center, National Institutes of Health (NIH, R01D0042157-01A, MH081802, Grand Opportunity grants 1RC2 MH089951 and 1RC2 MH089995). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health.Computing was supported by BiG Grid, the Dutch e-Science Grid, which is financially supported by NWO
The Northern Swedish Population Health Study (NSPHS) was funded by the Swedish Medical Research Council (Project Number K2007-66X-20270-01-3, 2011-5252, 2012-2884 and 2011-2354), the Foundation for Strategic Research (SSF). NSPHS as part of EUROSPAN (European Special Populations Research Network) was also supported by the European Commission FP6 STRP grant number 01947 (LSHG-CT-2006-01947). This work has also been supported by the Swedish Society for Medical Research (SSMF), and the Swedish Medical Research Council (#2015-03327). The Northern Swedish Population Health Study are grateful for the contribution of district nurse Svea Hennix for data collection and Inger Jonasson for logistics and coordination of the health survey. We also thank all the participants from the community for their interest and willingness to contribute to this study.
NTR funding was obtained from the Netherlands Organization for Scientific Research (NWO) and The Netherlands Organisation for Health Research and Development (ZonMW) grants 904-61-090, 985-10-002, 904-61-193,480-04-004, 400-05-717, Addiction-31160008, Middelgroot-911-09-032, Spinozapremie 56-464-14192, Biobanking and Biomolecular Resources Research Infrastructure (BBMRI –NL, 184.021.007), NWO-Groot 480-15-001/674 ; the Netherlands Heart Foundation grants 86.083 and 88.042 and 90.313; the VU Institute for Health and Care Research (EMGO+ ); the European Community's Seventh Framework Program (FP7/2007-2013), ENGAGE (HEALTH-F4-2007-201413); the European Research Council (ERC Advanced, 230374), the Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the Avera Institute, Sioux Falls, South Dakota (USA) and the National Institutes of Health (NIH, R01D0042157-01A, MH081802, R01 DK092127-04; Grand Opportunity grant 1RC2 MH089951). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health. Computing was supported by BiG Grid, the Dutch e-Science Grid, which is financially supported by NWO.
ORCADES was supported by the Chief Scientist Office of the Scottish Government, the Royal Society, the MRC Human Genetics Unit, Arthritis Research UK and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). DNA extractions were performed at the Wellcome Trust Clinical Research Facility in Edinburgh.

We would like to acknowledge the invaluable contributions of Lorraine Anderson and the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney. PIVUS was supported by Knut and Alice Wallenberg Foundation (Wallenberg Academy Fellow), European Research Council (ERC Starting Grant), Swedish Diabetes Foundation (grant no. 2013-024), Swedish Research Council (grant no. 2012-1397), and Swedish Heart-Lung Foundation (20120197). The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project b2011036. Genetic data analysis was funded by the Wellcome Trust under awards WT098017 and WT090532. The PIVUS project thank the SNP&SEQ Technology Platform in Uppsala (www.genotyping.se) for excellent genotyping.
PROCARDIS was supported by the European Community Sixth Framework Program (LSHM-CT- 2007-037273), AstraZeneca, the British Heart Foundation, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the Swedish Heart-Lung Foundation, the Torsten and Ragnar Söderberg Foundation, the Strategic Cardiovascular Program of Karolinska Institutet and Stockholm County Council, the Foundation for Strategic Research and the Stockholm County Council (560283).
The PROSPER study was supported by an investigator initiated grant obtained from Bristol-Myers Squibb. Prof. Dr. J. W. Jukema is an Established Clinical Investigator of the Netherlands Heart Foundation (grant 2001 D 032). Support for genotyping was provided by the seventh framework program of the European commission (grant 223004) and by the Netherlands Genomics Initiative (Netherlands Consortium for Healthy Aging grant 050-060-810).
The generation and management of GWAS genotype data for the Rotterdam Study (RS I, RS II, RS III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The Rotterdam Study thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Karol Estrada, PhD, Yurii Aulchenko, PhD, and Carolina Medina-Gomez, MSc, for the creation and analysis of imputed data. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg- West Pomerania. The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH.

The Mount Sinai IPM Biobank Program is supported by The Andrea and Charles Bronfman Philanthropies.

TRAILS (TRacking Adolescents' Individual Lives Survey) is a Dutch collaborative project involving various departments of the University Medical Center and University of Groningen, the University of Utrecht, the Radboud Medical Center Nijmegen, and the Parnassia Bavo group. TRAILS has been financially supported by grants from the Netherlands Organization for Scientific Research NWO (Medical Research Council program grant GB-MW 940-38-011; ZonMW Brainpower grant 100-001-004; ZonMw Risk Behavior and Dependence grant 60-60600-97-118; ZonMw Culture and Health grant 261-98-710; Social Sciences Council medium-sized investment grants GB-MaGW 480-01-006 and GB-MaGW 480-07-001; Social Sciences Council project grants GB-MaGW 452-04-314 and GB-MaGW 452-06-004; NWO large-sized investment grant 175.010.2003.005; NWO Longitudinal Survey and Panel Funding 481-08-013); the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-NL (CP 32), the participating universities, and Accare Center for Child and Adolescent Psychiatry. We are grateful to all adolescents, their parents and teachers who participated in this research and to everyone who worked on this project and made it possible. Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation.

ULSAM was supported by Knut and Alice Wallenberg Foundation (Wallenberg Academy Fellow), European Research Council (ERC Starting Grant), Swedish Diabetes Foundation (grant no. 2013-024), Swedish Research Council (grant no. 2012-1397), and Swedish Heart-Lung Foundation (20120197). The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project b2011036. Genotyping was funded by the Wellcome Trust under award WT064890. Analysis of genetic data was funded by the Wellcome Trust under awards WT098017 and WT090532. The ULSAM project thank the SNP&SEQ Technology Platform in Uppsala (www.genotyping.se) for excellent genotyping.

WGHS is supported by the National Heart, Lung, and Blood Institute (HL043851, HL080467, HL099355) and the National Cancer Institute (CA047988 and UM1CA182913) with

collaborative scientific support and funding for genotyping provided by Amgen. Genetic analysis in the JUPITER trial was funded by a grant from AstraZeneca (DIC and PMR, Co-PIs). The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; and Diabetes Research Foundation of Finnish Diabetes Association. The expert technical assistance in the statistical analyses by Irina Lisinen is gratefully acknowledged.
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.
The Airwave Health Monitoring Study is funded by the UK Home Office, (Grant number 780-TETRA) with additional support from the National Institute for Health Research Imperial College Health Care NHS Trust and Imperial College Biomedical Research Centre. P.E. is supported by the UK Dementia Research Institute which receives its funding from UK DRI Ltd funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK; and is an Associate Director of the Health Data Research UK (HDR-UK) London Centre which receives its funding from a consortium led by the UK Medical Research Council. We thank all participants in the Airwave Health Monitoring Study. This work used computing resources provided by the MRC- funded UK MEDical Bioinformatics partnership programme (UK MED-BIO) (MR/L01632X/1).
UKHLS**:** These data are from Understanding Society: The UK Household Longitudinal Study, which is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The data were collected by NatCen and the genome wide scan data were analysed by the Wellcome Trust Sanger Institute. The Understanding Society DAC have an application system for genetics data and all use of the data should be approved by them. The application form is at: https://www.understandingsociety.ac.uk/about/health/data. UKHLS was founded by Grant 1 - ES/M008592/1. We thank Michaela Benzeval[1], Jonathan Burton[1], Nicholas Buck[1], Annette Jäckle[1], Meena Kumari[1], Heather Laurie[1], Peter Lynn[1], Stephen Pudney[1], Birgitta Rabe[1], Dieter Wolke[1]
[(1)] Institute for Social and Economic Research

[2] University of Warwick"

EGCUT study was funded by Estonian Research Council Grant IUT20-60 and PUT1660, EU H2020 grant 692145, and European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012) GENTRANSMED.

Million Veteran Program: Consortium Acknowledgement

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)

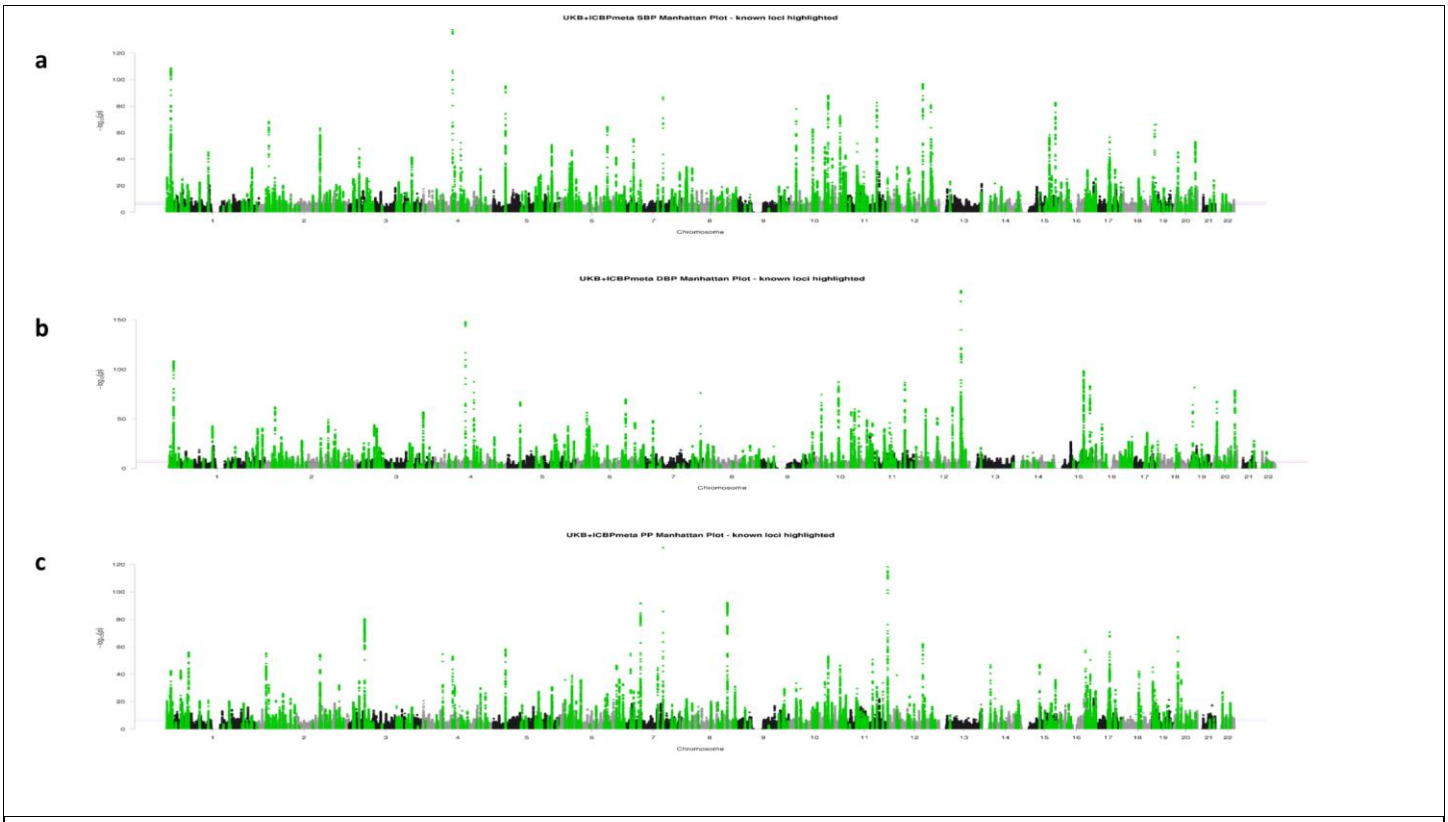
MVP: This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by VA awards I01BX003340 (P.S.T.), I01BX003362 (P.W.F.W), I01BX003360 (A.M.H.), and NHLBI-R21HL121429 (T.L.E., D.R.V.E). This publication does not represent the views of the Department of Veterans Affairs or the United States Government.

## References

1. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
2. Bycroft, C.F., C; Petkova, D; Band, G; Elliot, LT; Sharp, K; Motyer, A; Vukcevic, D; Delaneau, O; O'Conell, J; Cortes, A; Welsh, S; McVean, G; Leslie, S; Donelly, P; Marchini, J. Genome-wide geentic data on 500,000 UK Biobank Participants. *bioRxiv* **166298**(2017).
3. Loh, P.R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-90 (2015).
4. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).

5. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186-190 (2017).
6. International Consortium for Blood Pressure Genome-Wide Association, S. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
7. Wain, L.V. *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* **43**, 1005-11 (2011).
8. Wain, L.V. *et al.* Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension* (2017).
9. Tobin, M.D., Sheehan, N.A., Scurrah, K.J. & Burton, P.R. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat Med* **24**, 2911-35 (2005).
10. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-212 (2014).
11. Evangelou, E. & Ioannidis, J.P. Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet* **14**, 379-89 (2013).
12. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
13. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
14. 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).
15. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet* **41**, 666-76 (2009).
16. Warren, H.R. *et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet* **49**, 403-415 (2017).
17. Hoffmann, T.J. *et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet* **49**, 54-64 (2017).
18. Wain, L.V. *et al.* Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension* (2017).
19. Newton-Cheh, C. *et al.* Association of common variants in NPPA and NPPB with circulating natriuretic peptides and blood pressure. *Nat Genet* **41**, 348-53 (2009).
20. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat Genet* **48**, 1162-70 (2016).
21. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
22. Mitt, M. *et al.* Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet* **25**, 869-876 (2017).
23. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
24. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. 1165-1188 (2001).
25. Pulit, S.L., de With, S.A. & de Bakker, P.I. Resetting the bar: Statistical significance in whole-genome sequencing-based association studies of global populations. *Genet Epidemiol* **41**, 145-151 (2017).
26. Zheng, H.F. *et al.* Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature* **526**, 112-7 (2015).

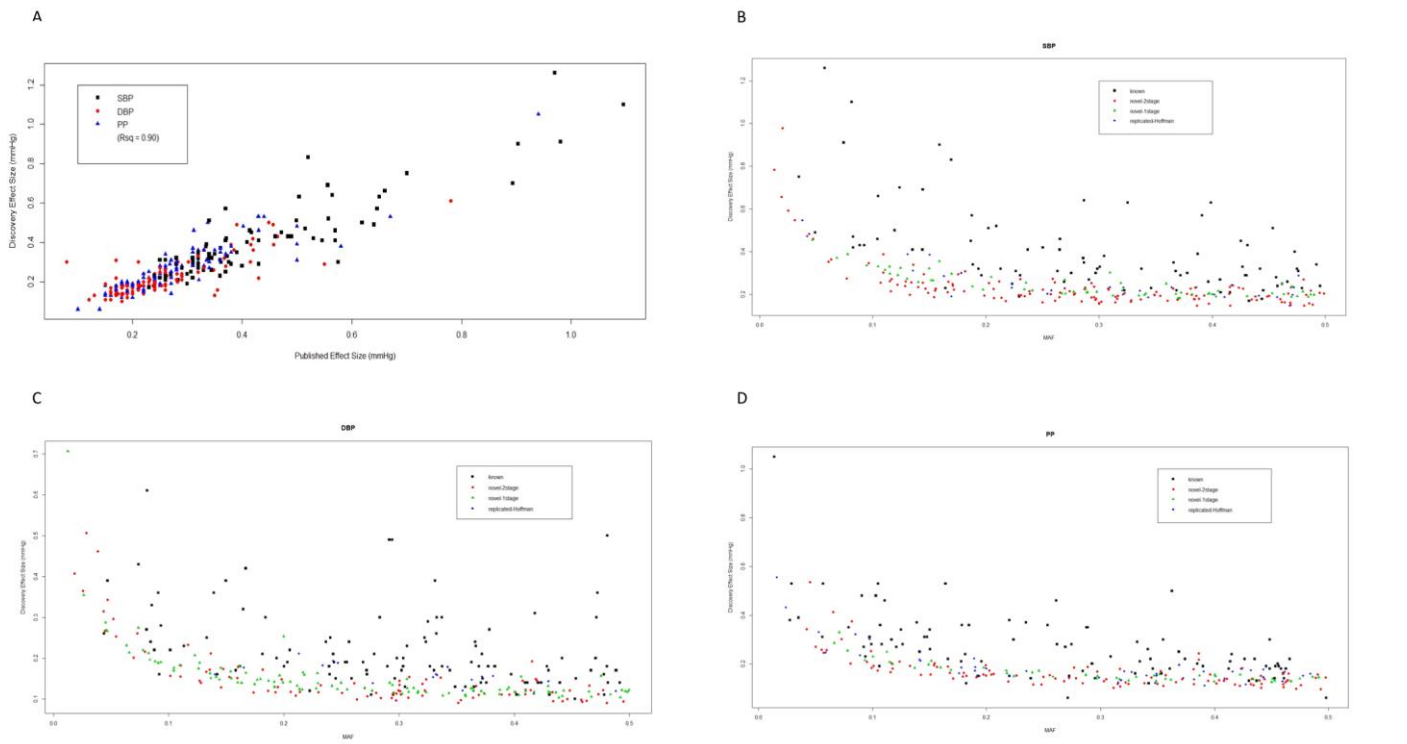
27. Elliott, P. *et al.* The Airwave Health Monitoring Study of police officers and staff in Great Britain: rationale, design and methods. *Environ Res* **134**, 280-5 (2014).
28. Zhou, J. & Troyanskaya, O.G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods* **12**, 931-4 (2015).
29. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-7 (2012).
30. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
31. Dunham, I.K., E.; Iotchkova, V.; Morganella, S.; Birney, E. FORGE: A tool to discover cell specific enrichments of GWAS associated SNPs in regulatory regions. *F1000Research* **4**(2015).
32. Staley, J.R. *et al.* PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* **32**, 3207-3209 (2016).
33. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896-D901 (2017).
34. Pinero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833-D839 (2017).
35. Pinero, J. *et al.* DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford)* **2015**, bav028 (2015).
36. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res* (2017).




**Supplementary Figure 1**

**GWAS discovery Manhattan plots**.

Manhattan plots (a), (b) and (c) for systolic blood pressure (SBP), diastolic blood pressure (DBP) and pulse pressure (PP) respectively. P-value results from the GWAS discovery meta-analysis (N=757,601), were derived using inverse variance fixed effects meta-analysis and they are plotted on a – log10 scale for all SNPs with Minor Allele Frequency (MAF) ≥ 1%. SNPs within the 274 known loci (±500kb; Linkage Disequilibrium $r^2 \geq 0.1$) are highlighted in green.




**Supplementary Figure 2**

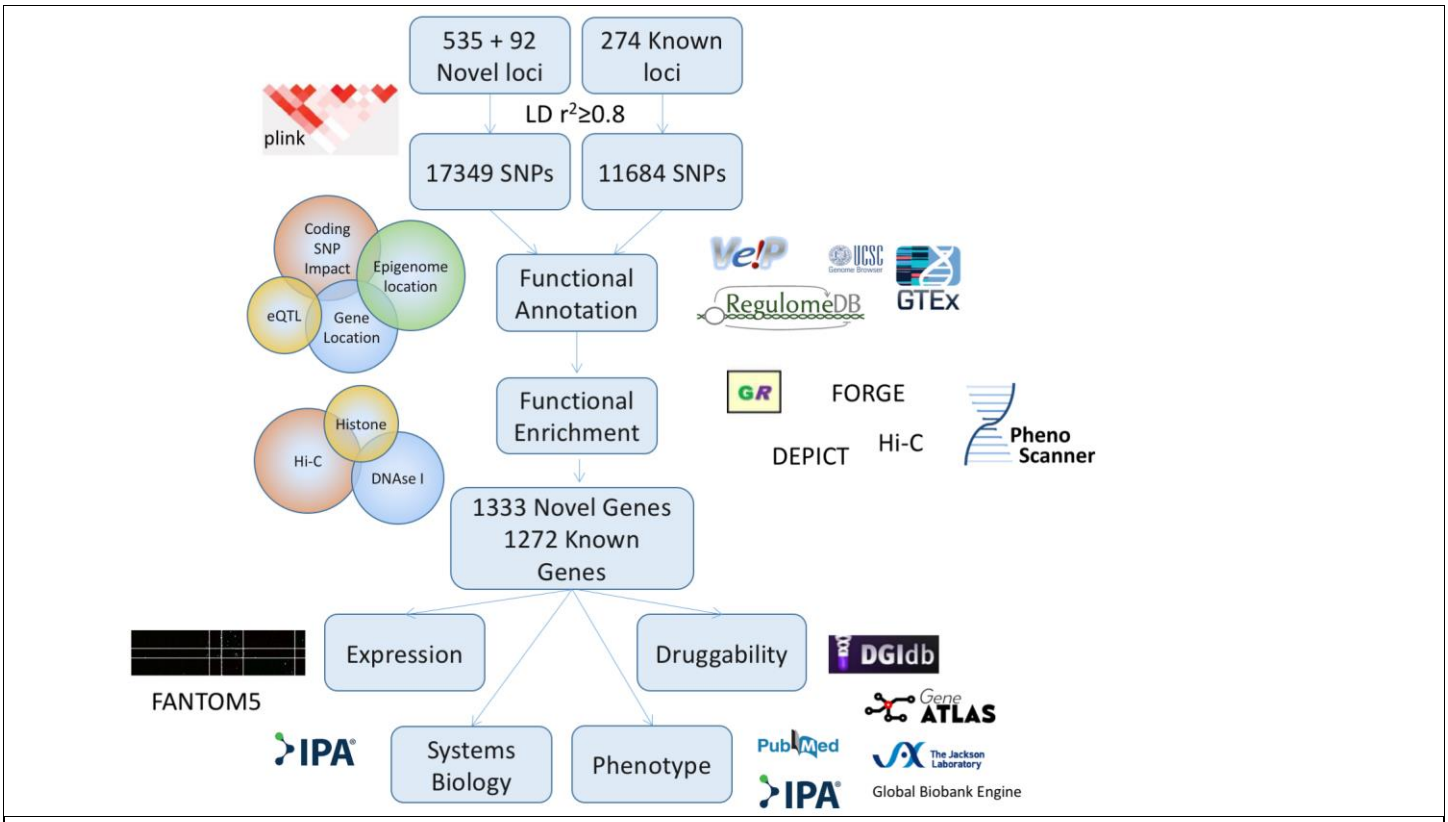
**Effect Sizes of all Blood Pressure associated loci.**

Plot (A) shows strong correlation between the published effect size estimates (x-axis) from literature vs the effect sizes from our discovery meta-analysis (y-axis), for known SNPs, colour-coded according to the published primary trait from the first published report. From the 357 validated SNPs listed in Supplementary Table 4 from the 274 published loci, 327 are available within the MAF ≥ 1% HRC-imputed data. For comparison of effect sizes, we only consider 299 such SNPs which have been identified from main-effect genetic association studies within Europeans (i.e. excluding any SNPs from interaction/stratified/multi-phenotype analysis, or from studies of other ancestries). For reliable comparison of effect sizes, we further restrict to the 284 known SNPs which reach genome-wide significance within the discovery meta-analysis for at least one BP trait. The $r^2$ value is presented to show the correlation between published and observed effect sizes. Plots (B), (C) and (D) are trait-specific plots for SBP, DBP and PP, respectively (SBP: systolic blood pressure; DBP: diastolic blood pressure; PP: pulse pressure). Across all plots, the 284 "known" SNPs (black squared) from plot (A) are compared against the 325 novel sentinel SNPs from the 2-stage analysis (red circles), the 210 novel sentinel SNPs from the 1-stage analysis (green triangles), and the 92 SNPs (blue diamonds) replicated for the first time from Hoffman et al[9]. Each SNP is only plotted in one of the trait-specific plots, according to the published primary trait for the known SNPs, or the primary trait for the novel / replicated SNPs. For all SNPs we show the relationship between Minor Allele Frequency (MAF) on the x-axis and the effect size (mmHg) on the y-axis, where results are taken from the UKB+ICBP discovery meta-analysis. All meta-analysis results were computed using inverse variance fixed effects models. The different symbols and colours distinguish the "known" vs "novel-2stage" vs "novel-1stage" vs "replicated-Hoffman" SNPs, and show that in general, the novel SNPs have smaller effect sizes than known SNPs, and that there is no significant difference (P=0.447) between the effect sizes of the 1-stage (N=757,601) and 2-stage (N=1,006,863) novel SNPs. (UKB: UK Biobank; ICBP: International Consortium of Blood Pressure).


**Supplementary Figure 3**

**Venn diagram of novel loci results.**

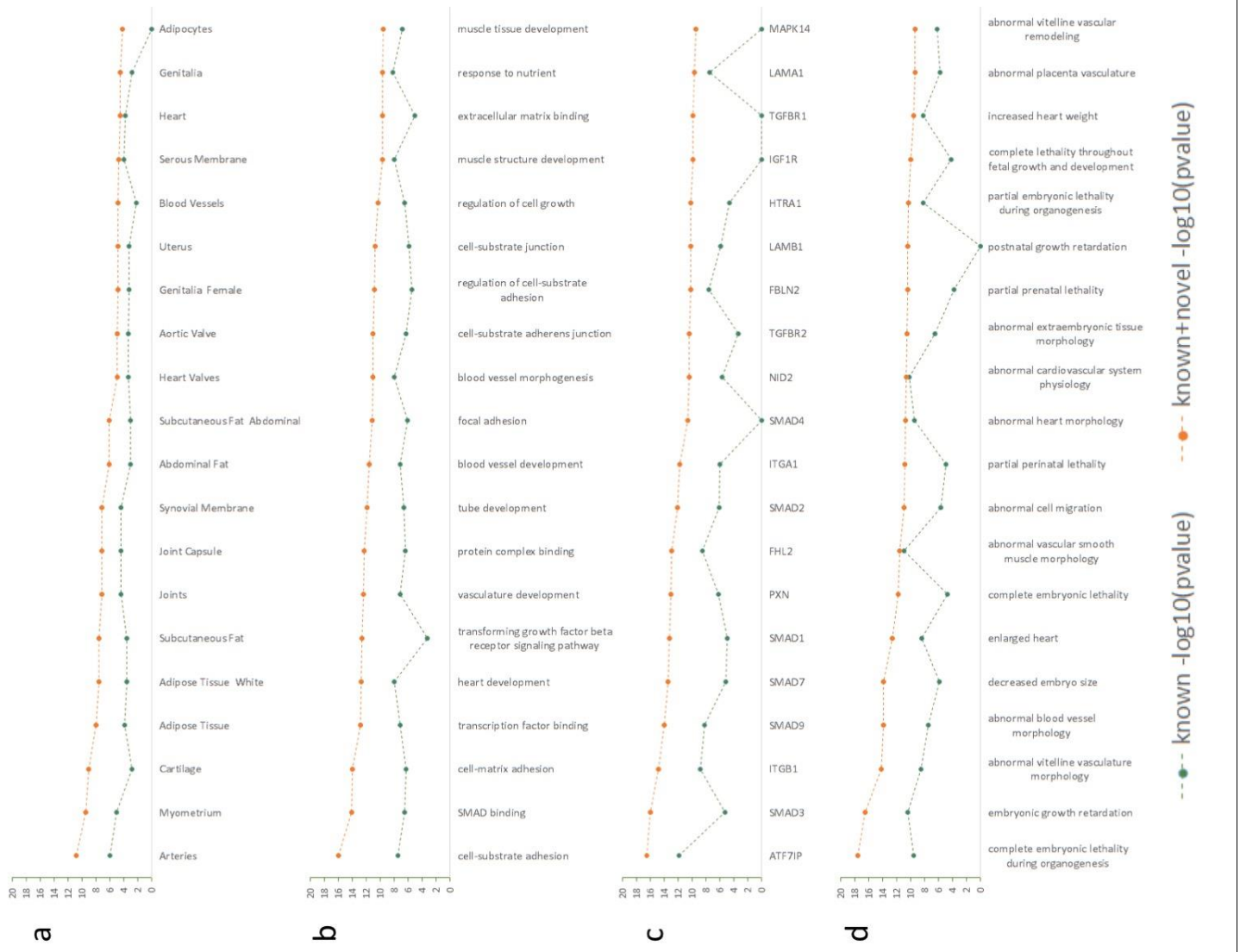
For all 535 novel loci, we show the blood pressure traits associated with each locus. We present the two-stage loci first, followed by the one-stage loci. The locus names provided in alphabetical order correspond to the nearest annotated gene. SNPs: Single nucleotide polymorphisms; SBP: systolic blood pressure; DBP: diastolic blood pressure; PP: pulse pressure; UKB: UK Biobank; ICBP: International Consortium of Blood Pressure.




**Supplementary Figure 4**

**Overview of functional annotation and prioritisation of genome-wide associated variants and genes.**

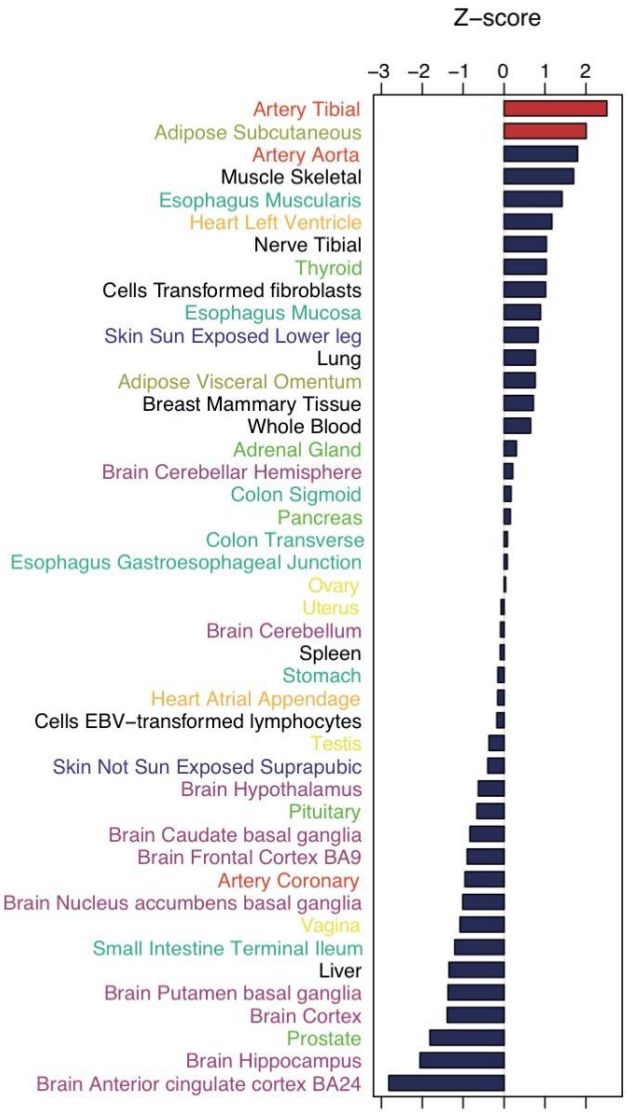
SNPs: single nucleotide polymorphisms; LD: Linkage Disequilibrium; eQTL: expression Quantitative Trait Loci; UCSC: University of California Santa Cruz (UCSC) genome browser; IPA: Ingenuity Pathway Analysis (IPA) software (IPA®,QIAGEN Redwood City,www.qiagen.com/ingenuity); DEPICT: Data-driven Expression Prioritized Integration for Complex Traits; GREAT: Genomic Regions Enrichment of Annotations Tool.




**Supplementary Figure 3**

**DEPICT enrichment analysis.**

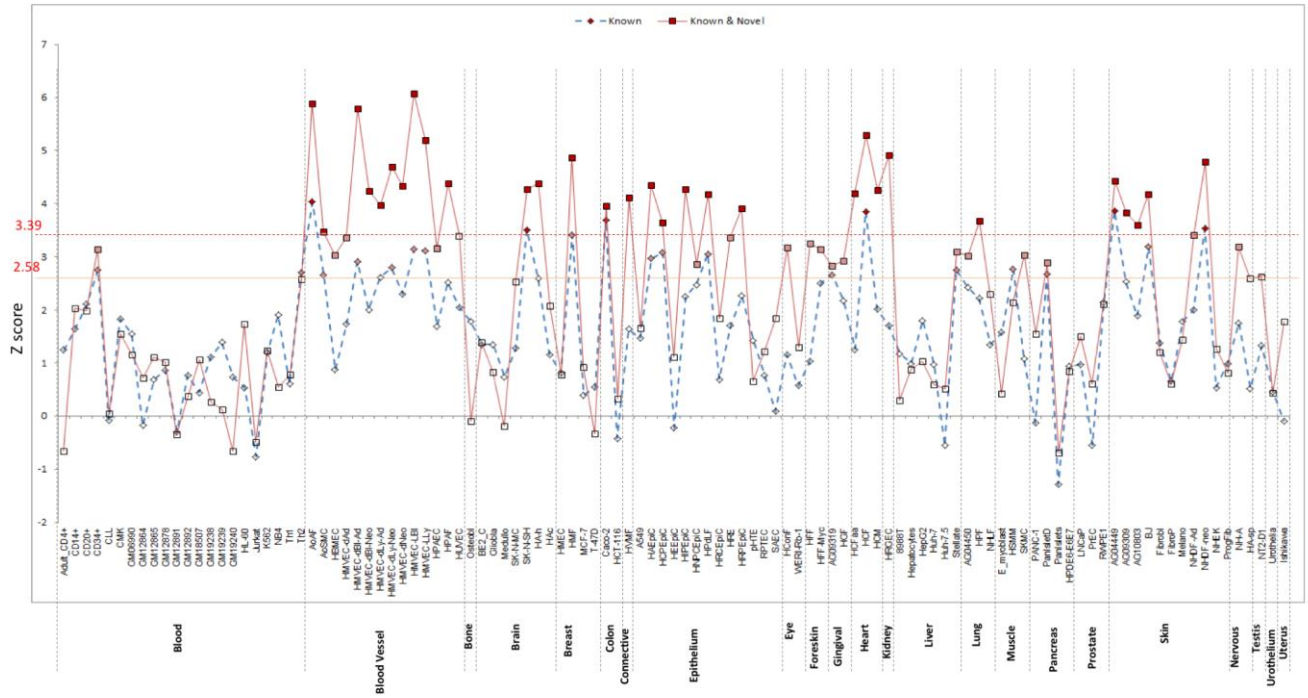
DEPICT software was used to investigate enrichment of a range of biological properties, in each case we compared known sentinel SNPS (N=357) to all known and novel SNPs with P-value $<1x10^{-12}$ (N=227). The gene set enrichment analysis algorithm is described in Pers et al[66]. Enrichment –log p value is reported for both groups, we also present delta –log p value as a measure of novelty introduced by novel associations reported. Enrichment categories are as follows. a) Enrichment of tissues and cell types. b) GO annotation. c) Protein-protein interaction subnetwork annotation. d) Mammalian phenotype annotation.




**Supplementary Figure 6**

**Enrichments of eQTLs.**

535 novel blood pressure associated SNPs and the SNPs in LD $r^2$>0.8 were annotated for their effect on gene expression using the GTEx portal. The number of eGenes associated with BP SNPs in a given tissue/ cell type was normalised with the total number of eGenes in that tissue and z-score was calculated using the trimmed mean and standard deviation of the normalised scores. Tissues of the same tissue group were coloured the same.




**Supplementary Figure 7**

**FORGE Dnase I sensitive region enrichment in known sentinel SNPs, compared to known and novel sentinel SNP associations for blood pressure.**

Sentinel SNPs were investigated for enrichment in ENCODE DNase I regulatory regions using FORGE. The background probability of overlap is determined from the 1000 background set overlap counts and the probability of the observed test result under a binomial distribution is calculated. The *P*-value thresholds of 0.05 and 0.01 are corrected for multiple testing by division by the number of tissue groupings tested, and the corrected threshold is used. Strongest enrichment in known SNPs was seen in vasculature (Human Aortic artery fibroblast (AoAF) and also Human Villous Mesenchymal Fibroblasts (HVMF) found in placenta). Enrichment in all known and novel SNPS was increased across vasculature (AoAF; HMVEC, Human microvascular endothelial cells) and highly vascularised tissues. Tissues in red are significant after correction for false discovery.

a)

| Canonical Pathway | known sentinel genes | known LD genes | All sentinel genes | All LD genes |
|---|---|---|---|---|
| Gap Junction Signaling | 9.1 | 6.8 | 6.5 | 4.7 |
| Cellular Effects of Sildenafil (Viagra) | 8.8 | 7.0 | 7.2 | 3.9 |
| Nitric Oxide Signaling in the CV System | 6.6 | 5.5 | 7.6 | 4.9 |
| G-Protein Coupled Receptor Signaling | 7.0 | 4.4 | 7.6 | 4.2 |
| Protein Kinase A Signaling | 5.4 | 4.5 | 7.2 | 5.4 |
| P2Y Purigenic Receptor Signaling Pathway | 6.8 | 5.4 | 5.8 | 4.2 |
| Relaxin Signaling | 6.1 | 5.3 | 6.7 | 3.7 |
| Dopamine-DARPP32 Feedback in cAMP Sign | 7.6 | 5.7 | 4.5 | 3.3 |
| PPARα/RXRα Activation | 5.4 | 5.1 | 5.6 | 4.7 |
| Role of NFAT in Cardiac Hypertrophy | 5.1 | 3.6 | 7.4 | 4.6 |
| Cardiac β-adrenergic Signaling | 5.8 | 4.6 | 6.2 | 3.6 |
| Sperm Motility | 7.0 | 4.9 | 4.8 | 3.2 |
| Axonal Guidance Signaling | 5.7 | 3.2 | 7.5 | 3.4 |
| CREB Signaling in Neurons | 6.0 | 4.4 | 5.9 | 3.3 |
| Molecular Mechanisms of Cancer | 4.5 | 4.4 | 5.9 | 4.4 |
| Chronic Myeloid Leukemia Signaling | 3.4 | 3.1 | 6.7 | 5.4 |
| PEDF Signaling | 5.7 | 4.4 | 4.8 | 3.3 |
| Corticotropin Releasing Hormone Signaling | 5.7 | 4.1 | 4.9 | 2.2 |
| Synaptic Long Term Potentiation | 5.4 | 3.8 | 4.5 | 3.0 |
| HER-2 Signaling in Breast Cancer | 3.8 | 3.6 | 4.8 | 4.3 |

b)

| Upstream regulators | known sentinel genes | known LD genes | All sentinel genes | All LD genes |
|---|---|---|---|---|
| AGT | 9.13 | 11 | 13 | 11.8 |
| ESR1 | 5.82 | 6.68 | 7.44 | 7 |
| NPR1 | 7.56 | 7.01 | 5.61 | 4.66 |
| nifedipine | 7.41 | 5.85 | 7.2 | 3.55 |
| losartan potassium | 4.15 | 4.34 | 6.54 | 5.38 |
| candesartan | 5.81 | 5.04 | 4.55 | 4.71 |
| KMT2A | 2.31 | 4.46 | 7.17 | 5.73 |
| AGTR1 | 3.88 | 3.55 | 6.05 | 5.41 |
| TGFB1 | 0 | 2.88 | 7.34 | 8.42 |
| SP1 | 5.01 | 4.64 | 5.26 | 3.72 |
| COMMD3-BMI1 | 1.81 | 8.73 | 1.81 | 5.99 |
| darusentan | 4.69 | 3.9 | 5.08 | 4.08 |
| PIAS1 | 3.05 | 3.89 | 4.65 | 6.14 |
| NR3C2 | 5.02 | 3.01 | 6.54 | 2.86 |
| progesterone | 4.52 | 3.67 | 5.51 | 2.9 |
| MAPK7 | 2.45 | 3 | 5.08 | 5.78 |
| PCGF2 | 5.24 | 4.75 | 3.58 | 2.75 |
| VCAN | 4.32 | 4.19 | 4.09 | 3.55 |
| HOXA10 | 1.56 | 5.25 | 3.55 | 5.38 |
| REN | 4.6 | 4.39 | 3.38 | 2.78 |

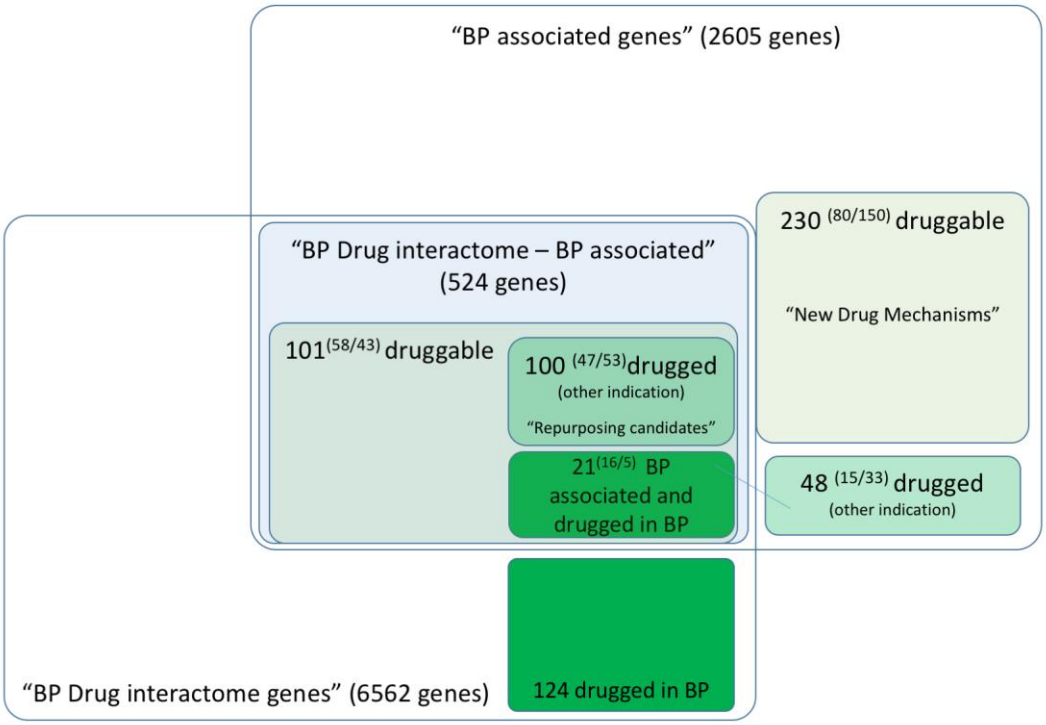
c)

| Diseases and Bio Functions | known sentinel genes | known LD genes | All sentinel genes | All LD genes |
|---|---|---|---|---|
| morphology of cardiovascular system | 14.73 | 13.83 | 20.85 | 17.01 |
| morphology of heart | 15.19 | 12.9 | 20.14 | 14.7 |
| morphology of body cavity | 15.87 | 13.17 | 17.94 | 11.99 |
| organismal death | 11.8 | 10.28 | 20.58 | 13.26 |
| morbidity or mortality | 11.88 | 10.16 | 20.18 | 12.7 |
| morphology of heart ventricle | 12.1 | 11.61 | 16.39 | 12.42 |
| skin lesion | 6.696 | 4.877 | 23 | 16.96 |
| morphology of muscle | 11.71 | 11.66 | 14.11 | 12.66 |
| development of vasculature | 10.35 | 9.228 | 19.05 | 10.96 |
| abnormal morphology of cardiovascular | 10.22 | 9.081 | 14.07 | 10.49 |
| development of body trunk | 8.922 | 8.063 | 16.08 | 10.52 |
| angiogenesis | 8.331 | 6.831 | 17.92 | 10.14 |
| primary hypertension | 11.22 | 9.003 | 12.44 | 10.15 |
| familial cardiovascular disease | 8.063 | 5.735 | 16.96 | 10.31 |
| hypertrophy of heart | 7.684 | 7.93 | 14.02 | 11.43 |
| abnormal morphology of body cavity | 11.55 | 9.621 | 11.94 | 7.423 |
| liver lesion | 4.977 | 4.891 | 16.44 | 14.09 |
| vasculogenesis | 6.576 | 6.361 | 15.89 | 10.8 |
| morphology of cardiac muscle | 9.097 | 8.926 | 11.18 | 10.03 |
| morphology of cardiomyocytes | 8.775 | 8.932 | 11.07 | 9.968 |

**Supplementary Figure 8**

**Ingenuity pathway analysis of BP genes.**

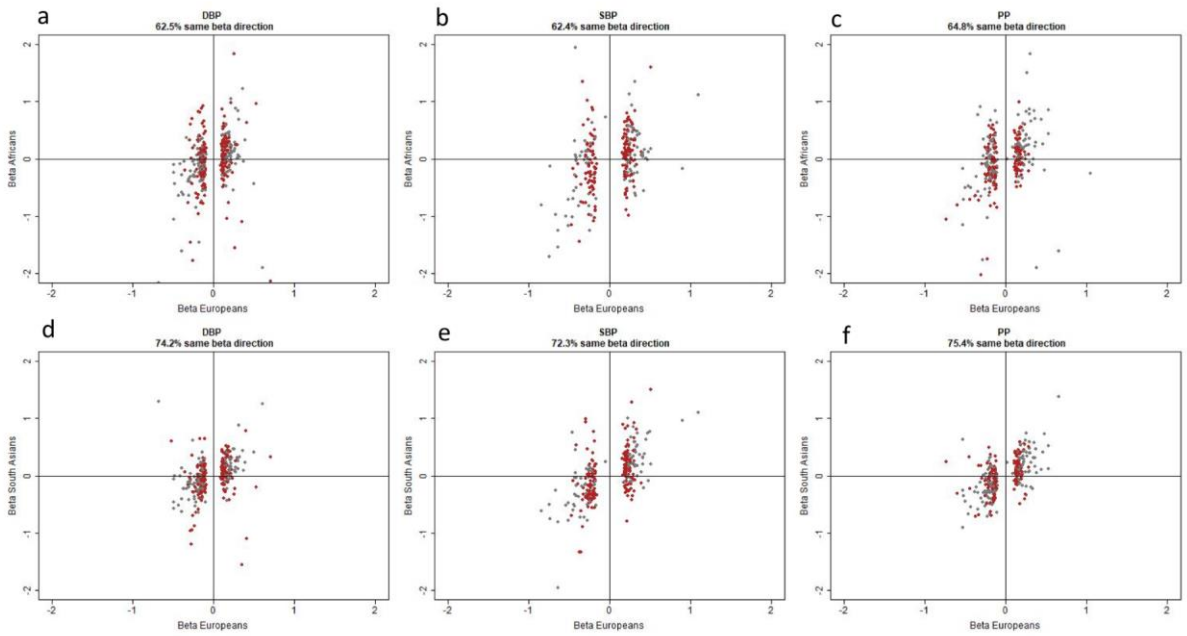
For genes mapped to 357 sentinel SNPs at 274 known loci and genes mapped to all 901 loci. Sentinel gene mapping is compared to genes identified by extended LD ($r^2$>0.8). Pathway enrichment is represented as – log p value. A) Canonical pathway enrichment. B) Upstream regulator enrichment. C) Disease and Biofunction enrichment.




**Supplementary Figure 9**

**Exploring known and novel drug mechanisms in blood pressure.**

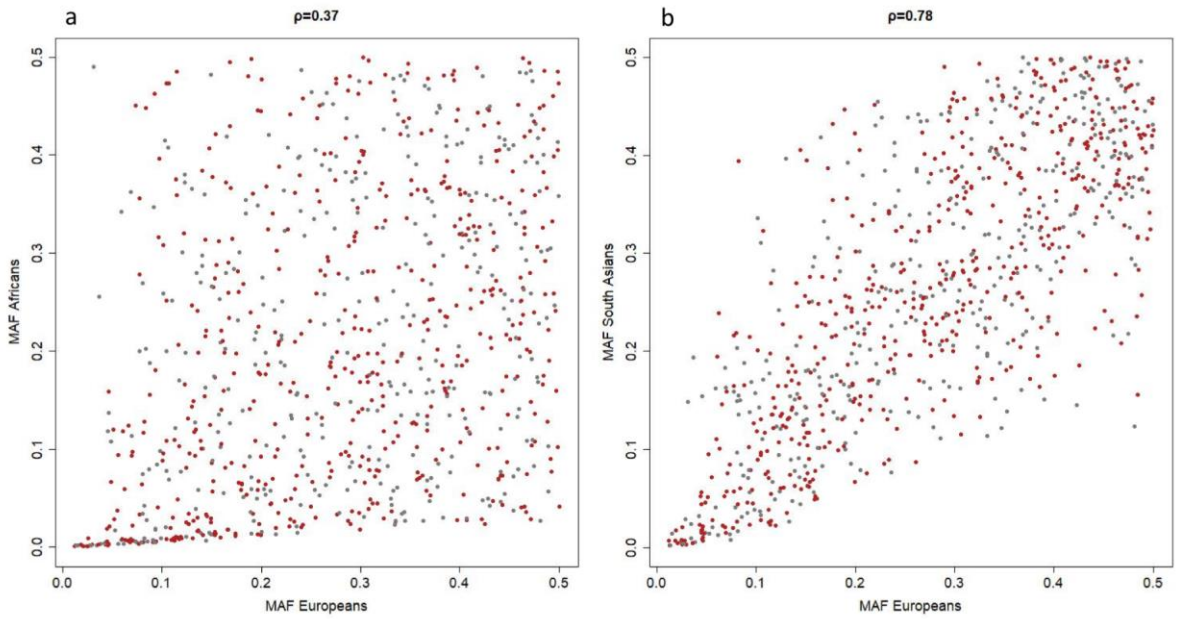
The figure summarises known and novel target opportunities highlighted by blood pressure genetics. Ingenuity pathway analysis was used to create a network of 6,562 genes showing direct interaction with 145 known blood pressure target genes. This network was compared with all genes that are either directly associated with BP or linked by LD ($r^2$>0.8). Overlap between genetic associated genes and the BP drug interactome demonstrates genetic support for known drug mechanisms. Drugged or druggable genes showing genetic association with BP, but no interaction with the known BP drug interactome, represent potentially new mechanisms in blood pressure drug development and repositioning. Number of known and novel drugged/druggable gene associations are shown in parentheses.




**Supplementary Figure 10**

**Comparison of beta effect sizes between individuals of European (N=757,601), African (N=7,782) and South Asian (N=10,323) ancestry.**

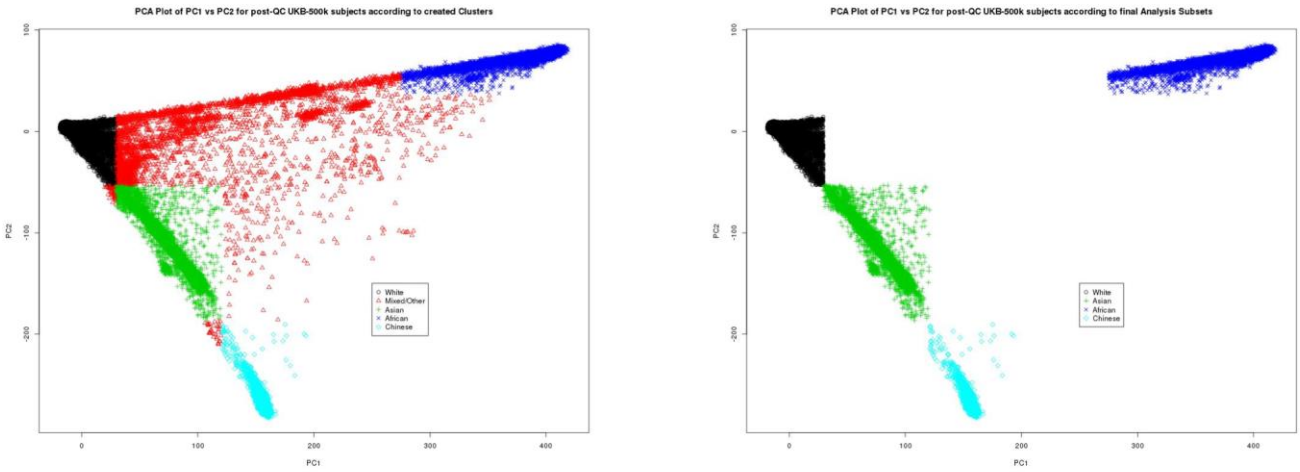
Scatterplots showing the direction of the standardized regression coefficient (beta) of novel (red) and known (grey) BP variants between Europeans and Africans (a,b,c) and South Asians (d,e,f), on the three studied BP phenotypes.




**Supplementary Figure 11**

**Correlation and distribution of minor allele frequencies (MAF) of BP variants in individuals of European (N=757,601), African (N=7,782) and South Asian (N=10,323) ancestry**.

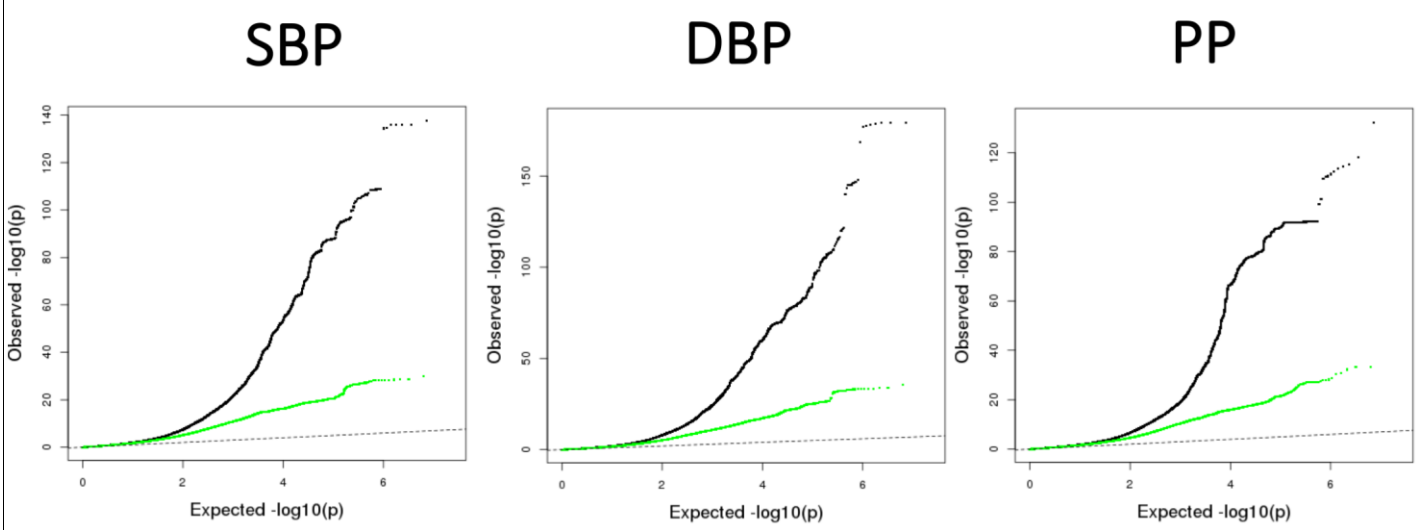
Scatterplot showing the correlation and the distribution of MAF of novel (red) and known (grey) BP variants between a) Europeans and Africans and b) Europeans and South Asians. ρ is the Pearson correlation coefficient.




**Supplementary Figure 12**

**Ethnicity clustering performed using PCA.**

PC1 is plotted against PC2 for all N=486,683 UK Biobank participants post-QC, colour-coded according to the five ethnic clusters created from our K-means PCA clustering, from which only “White” Caucasians are selected for analysis of individuals of European ancestry. Plot (A) shows the clustering for all subjects, whereas plot (B) only shows the subsets of individuals selected for race-stratified analysis, after combining information together from both the PCA clustering and the self-reported ethnicity. PCA: Principal Component Analysis; QC: Quality Control; PCs: Principal Components.




**Supplementary Figure 13**

**Quantile-Quantile plots**.

QQ plots of results for (A) systolic blood pressure (SBP), (B) diastolic blood pressure (DBP), (C) pulse pressure (PP) from GWAS discovery (N=757,601). The black curves are based on all SNPs in the corresponding analysis, with Minor Allele Frequency ≥ 1%. The green curves are results after excluding SNPs within the 274 known loci (±500kb; Linkage Disequilibrium $r^2 \geq 0.1$). The *P*-values have been derived from inverse variance fixed effects meta-analysis.