Supplementary Information

87 rare variants associated with blood pressure regulation in meta-analysis of ~1.3 million individuals

Power calculations

Power estimation was performed in R (https://genome.sph.umich.edu/wiki/Power_Calculations:_Quantitative_Traits) and the additive variance formula came from Falconer D.S.[1]. With our Pan-Ancestry meta-analysis of up to 1,318,884 individuals, we have 80% power to detect association with a variant with H2 of 0.003%, which corresponds to a variant with MAF of 0.01 and effect size of 0.039, or a variant with MAF=0.05 and effect size of 0.018 (Supplementary Figure 2). The effect sizes in our analyses are in terms of standard deviation (SD) units.

EAWAS Study design

We curated a list of 362 BP-associated loci that were known at the time of the analyses and conservatively defined known loci using both distance (±500kb) and LD such that variants outside of the known loci had $r^2 < 0.1$ (in 1000 Genomes EUR) with the previously reported variants (Methods; Supplementary Table 4). Single variant association summaries for 382 SNVs with $P<5x10^{-8}$ (derived from two-sided tests) outside of these regions (Stage 1) was requested from MVP, deCODE and GENOA. Results obtained from MVP, deCODE and GENOA was meta-analysed. Meta-analyses of Stage 1 and the results from meta-analyses of MVP, deCODE and GENOA was performed and any variant with *P*-value$<5x10^{-8}$ and consistent direction of effects with no evidence for heterogeneity were considered new.

Three hundred and forty-four SNVs (200 genomic regions; eight rare SNVs, 25 low-frequency SNVs; Methods) of the 382 BP-associated SNVs (91%) were associated with one or more BP traits at $P<5x10^{-8}$ in the combined EUR (Stage 2) meta-analyses involving up to ~1.165 million individuals (Table 1, Supplementary Table 5, Figure 2). An additional seven SNVs from seven genomic regions were only genome-wide significant in the PA (Stage 2) meta-analyses of ~1.3 million individuals (Supplementary Table 5), bringing the total number of BP-associated SNVs in Stage 2 to 355. Of the novel EUR BP-associated SNVs, 41 (30 loci; three rare SNVs, four low-frequency SNVs) were associated with an additional BP trait in the PA meta-analyses in addition to the EUR associated trait. All the associations had consistent directions of effect across Stage 1 and also across Stage 2 and no evidence of heterogeneity ($P>0.0001$; Supplementary Table 5).

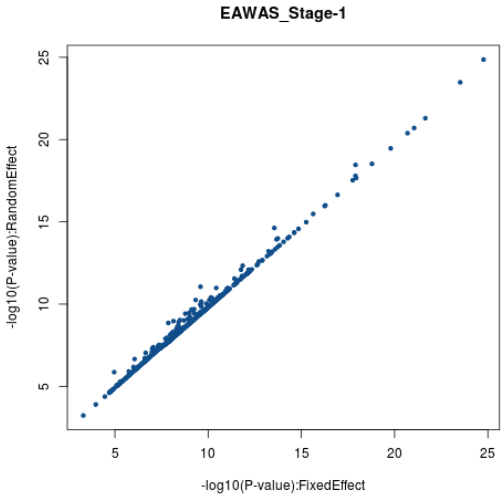
Quality Control of novel BP-associated variants from EAWAS and RV-GWAS

We adopted a single discovery-stage meta-analysis study design for both the EAWAS and RV-GWAS primarily for reasons of statistical power. The data request studies were not statistically powered on their own to detect the effects of the subset of SNVs we selected for data request from

MVP/deCODE/GENOA (EAWAS) or MVP (RV-GWAS) since these studies involved only around half the samples of the discovery. For a replication study, a sample size similar to, or larger than that used for the discovery, is required to have sufficient statistical power. In the absence of a well powered replication dataset, we have taken exhaustive measures to ensure the robustness of our findings.

We ensured that novel BP-associated variants that we claim were not driven by a single study. All reported variants had data from ≥ 19 studies in the Stage 1 EAWAS and 2 studies in the RV-GWAS, reducing the likelihood of a false association. In addition, all the novel BP-associated variants we report had consistent directions of effect in the Stage 1 studies and the data request studies (MVP+deCODE+GENOA for EAWAS, MVP alone for RV-GWAS). We verify the assumption of the fixed effects meta-analysis model, we ensured there was no evidence of heterogeneity across the effect estimates from contributing studies. In addition, we performed random effects meta-analysis (Han and Eskin's AJHG 2011 Random Effects Model) of novel BP-associated variants to minimise false discoveries due to study heterogeneity. The below plot (Supplementary Figure A) compares the -$\log_{10}$(*P*-values) from the fixed effect and random effects meta-analyses for all the variants in the EAWAS for which data were requested in the look up studies (see Supplementary Table 5a). There is strong concordance, suggesting that a fixed effects meta-analysis model is appropriate.

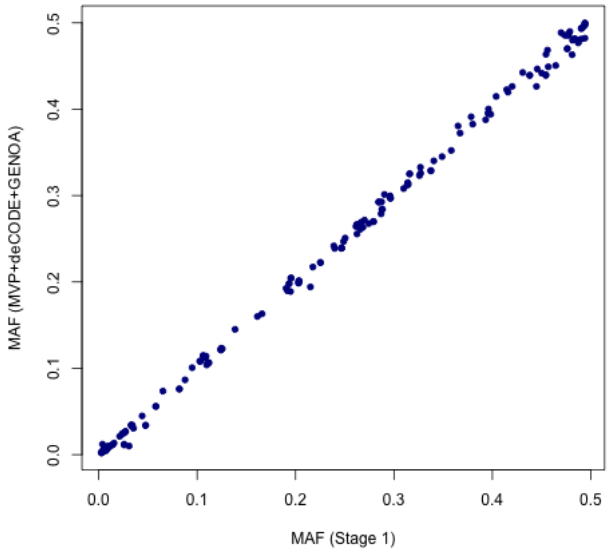
Supplementary Figure 4a: Comparison of the *P*-values for association of the novel BP SNVs from the random effects meta-analyses and the fixed effects meta analyses as provided in Supplementary Table 5. -$\log_{10}$(*P*-values) are plotted.



To ensure that the frequency of variants are not a result of inaccurate clustering/genotype calling, we confirmed that the allele frequencies were in the expected range by comparing the allele frequencies between Stage 1 and the data request studies (MVP+deCODE+GENOA for the EAWAS and MVP alone for RV-GWAS). In addition, we compared the allele frequencies to those in the reference

datasets (gnomAD, UCSC, and 1000 Genomes). Allele frequencies were plotted to check for consistency and those not consistent were removed *e.g.* rs7775698. The plot below shows the comparison of MAFs of novel variants in EAWAS between Stage 1 and MVP+deCODE+GENOA.

Supplementary Figure 4b: Minor Allele Frequencies (MAF) of the novel BP-associations from the from Stage 1 of the EAWAS and the data request studies (MVP+deCODE+GENOA) restricted to EUR.



Where variants were only available in a small number of studies, we checked the cluster plots of the studies involved and such variants as rs201702041, rs200510006, rs142360750 and rs143226982 that were poorly clustered in the PROMIS study were removed.

Within UK Biobank we performed our own QC for the genotyped variants rather than using the QC'd data as provided by UK Biobank, as we were specifically interested in the rare variants and knew that these were most vulnerable to clustering errors. Also described in detail within the section: "UK Biobank specific analyses" in this document. For the RV-GWAS and the FINEMAP analyses of UK Biobank we were able to perform additional checks for some of the variants. We compared the minor allele frequencies of the variants genotyped by arrays or imputed with those genotyped using whole exome sequencing. For the three novel BP-associated variants we identified in UK Biobank (rather than the EAWAS), the MAF was consistent between the imputed and WES data, suggesting the genotyping was robust.

Variants 1: Chromosome: 1; Position: 198,222,215
rsID: rs55833332
MAF in WES (both versions of calling/QC): 0.00747
MAF for the imputed variant in UKBB: 0.00816
MAF of variant in gnomAD v2.1.1 (for reference): 0.006475 (exomes), 0.008991 (genomes) and 0.009749 (European non-Finnish)

Variant 2: Chromosome: 20; Position: 61,050,522
rsID: rs200383755
MAF in WES (both versions of calling/QC): 0.00680
MAF for the imputed variant in UKBB: 0.00601
MAF of variant in gnomAD v2.1.1 (for reference): 0.003412 (exomes), 0.003479 (genomes) and 0.005443 (European non-Finnish)

Variant 3 (was imputed): Chromosome:14; Position: 100,143,685
rsID: rs149250178
MAF in WES (both versions of calling/QC): 0.00020
MAF for the imputed variant in UKBB: 0.00036
MAF of variant in gnomAD v2.1.1 (for reference): no variant (exomes), 0.003479 (genomes) and 0.001104 (European non-Finnish)

We compared the minor allele frequency (MAF) calculated using genotyped genotypes and imputed genotypes of the rare variants both genotyped and imputed in UKBB. We looked at this distribution as a function of the INFO score and identified that the MAF of the imputed variants with INFO>0.3 had an almost perfect correlation (ρ>0.9998) with the MAF of genotyped variants. Based on this comparison we only analysed rare variants with an INFO>0.3 in UKBB. We checked imputation quality for any BP-associated variant that was claimed and imputed. All variants we claim had imputation info score >0.8 in all Stage 1 studies.

Associations of previously reported variants in the Stage 1 EAWAS and UKBB

Of the 362 BP-associated loci reported prior to our analyses (*i.e.* pre-2018; Methods; Supplementary Table 4), 291 (80%) had one or more genome-wide significant associations in our UKBB GWAS that were in LD with the previously reported variant and 124 were genome-wide significant in the EAWAS. We confirmed 332 known loci at $P \leq 5 \times 10^{-5}$ and 344 (95%) were nominally significant ($P \leq 0.05$).

Comparison of conditional analyses in the EAWAS and UKBB GWAS

For eight of the known regions in Table 2 the common BP-associated SNVs were not available on Exome array, but independently associated rare/low-frequency variants had been identified. We therefore verified that these associations were valid using the dense genomic coverage in UKBB. At *NOX4*, *ZFAT*, *GEM*, *MYO1C* and *LTBP4* the same variants (or proxies r^2>0.9) were identified with FINEMAP in UKBB (Table 3) as with GCTA for the EAWAS (Table 2). At *GEM* and *NOX4* two rare BP-associated SNVs were identified in both genes in addition to the previously reported common variant associations (Table 3; Supplementary Table 10). At *FBXL19*, a rare missense variant was independent of the common variant signal in the EAWAS, (Table 2, Supplementary Table 10) while in the FINEMAP analyses in UKBB, an intron variant in *STX4* was in LD (r^2=0.88) with the *FBXL19* missense variant. (A second rare SNV, rs2234710, upstream of *BCL7C*, was independent of the *STX4* and common variant associations at this locus, in UKBB.) At *FOXS1*, a rare missense variant was identified as the top association in the EAWAS, while in the FINEMAP UKBB analyses an intronic

variant in *MYLK2*, which is in LD (r^2=1 in 1000 genomes EUR) with the *FOXS1* variant was identified, and although the *FOXS1* SNV is a more attractive candidate causal variant as it is missense, *MYLK2* is an attractive candidate gene as it is targeted by the drug Fostamatinib, which is used for the treatment of chronic immune thrombocytopenia and hypertension is reported as a side effect of Fostamatinib. Therefore it is likely that the rare/low-frequency associations at these loci are valid and independent of the established common variant associations.

Annotation of BP-associated variants

Across all 589 BP loci considered, 45% of the independent BP-associated rare variants were coding, while amongst the common variants, 20% were coding, in part reflecting the exome-centric design of the EAWAS. Twenty-one rare and 43 low-frequency variants were within regulatory elements including enhancers, promoters, CTCF binding sites, transcription factor binding sites and open chromatin regions highlighting genetic control of BP levels through gene expression.

Gene-based association tests sensitivity analyses

Amongst the genes that map to our newly identified BP-associated loci, ten from the EAWAS (*SCMH1*, *FILIP1L*, *CEP97*, *G6PC2*, *PHC3*, *HAUS6*, *PLCB3*, *TBX5*, *SOS2*, *NEK9*) and four from the RV-GWAS (*NEK7*, *PHC3*, *TBX5*, *GATA5*) were associated with BP ($P<2.5\text{x}10^{-6}$). Analyses conditional on the top SNV in the gene showed that the associations were attributable to a single rare variant identified in the single variant analyses and not likely to be due to multiple rare SNVs (Supplementary Table 11).

We tested the genes that mapped to the 362 previously reported BP loci. In the EAWAS, 21 genes within known loci, were associated with BP ($P<2.5\text{x}10^{-6}$; Supplementary Table 11) and ten genes (two not in the EAWAS list, *ZNF646* and *COL17A1*) were associated in the RV-GWAS ($P<2.5\text{x}10^{-6}$; Supplementary Table 11). Analyses conditional on the top SNV in the gene, showed that six of these gene associations were due to multiple rare SNV associations (*GEM*, *NPR1*, *DBH*, *COL21A1*, *NOX4* and *AGT*: SKAT conditional $P<1\text{x}10^{-4}$; Supplementary Table 11). To test whether the associations were due to LD with known common BP-associated variants, we also performed SKAT tests conditional on the known common variants in the individual loci. Five of the genes, *NPR1*, *DBH*, *COL21A1*, *NOX4*, *GEM*, were associated with BP independently of both the common variant associations and the top SNV in the gene ($P\leq1\text{x}10^{-5}$; Supplementary Table 11) confirming the findings in the single variant conditional analyses (Supplementary Table 10).

To assess sensitivity to the MAF threshold, we repeated the gene-based tests using a MAF<0.05 threshold. No genes with multiple rare/low-frequency SNVs were identified outside of known or novel regions (conditional SKAT $P>0.0001$; Supplementary Table 11). Of the 27 genes that were associated in the novel loci ($P<2.5\text{x}10^{-6}$), the association at *PLCB3* with DBP was due to multiple DBP-associated SNVs ($P=2.63\text{x}10^{-6}$; Supplementary Table 11) consistent with the conditional single variant analyses that identified one rare and one low-frequency variant associated in this gene (Supplementary Table

10). Of the 67 genes associated in known regions, nine (*NPR1*, *DBH*, *COL21A1*, *NOX4*, *CEP120*, *LARP4*, *PLCE1*, *NOS3* and *TBC1D32*) were due to multiple SNVs, and the associations with *NPR1*, *COL21A1*, and *CEP120* were not due to common variant associations (conditional SKAT $P<1x10^{-5}$; Supplementary Table 11, 12). In total, seven genes, one in a novel region (*PLCB3*) and six in known regions (*NPR1*, *DBH*, *COL21A1*, *NOX4*, *GEM* and *CEP120*) were implicated in BP regulation with multiple SNVs associated in the genes that were not due to LD with established common SNV-BP associations.

Rare variant gene-set enrichment analyses

Lists of genes representing various pathways and biological processes were constructed from the following sources: GO (download from http://geneontology.org/ on December 9, 2018, using the files go-basic.obo and goa_human.gaf), GTEX (download from https://gtexportal.org on December 9, 2018, using the file GTEx_Analysis_2016-01-15_v7_RNASeQCv1.1.8_gene_median_tpm.gct.gz), KEGG (downloaded from ftp.pathways.jp on December 9, 2018 using the files hsa.list and map_title.tab), MGI (downloaded from http://www.informatics.jax.org downloads/reports on December 9, 2018, using the files MPheno_OBO.ontology.obo, HMD_HumanPhenotype.rpt and MGI_PhenoGenoMP.rpt) and Orphanet (downloaded from http://www.orphadata.org/data/ORDO/ on December 9, 2018, using the files ordo.owl). For GTEx, a gene set for a tissue was defined as the set of all genes with highest expression in that tissue. In the cases of the ontologies (GO, MGI, Orpha) gene sets were constructed by first collecting the genes annotated to each specific node and then rolling these annotations up to each parent node recursively to the top of the ontology. For the MGI data the mouse to human orthology mappings provided in the source files were used. All gene references were mapped to entrez IDs using Homo_sapiens.gene_info file obtained from ftp://ftp.ncbi.nih.gov/gene/DATA/GENE_INFO/Mammalia. Genes not listed as "protein-coding" genes in entrez genes were omitted, as were genes with no chromosomal mappings in the hg38 reference genome assembly. Gene sets with only a single gene were eliminated from further consideration.

We tested whether genes near rare BP-associated SNVs were enriched in gene sets from Gene Ontology (GO), KEGG, Mouse Genome Informatics (MGI) and Orphanet (Methods; Supplementary Table 25). These (rare variant) genes from both known and novel loci were enriched in BP-related pathways (Bonferroni adjusted $P<0.05$, Methods; Supplementary Table 23) including "regulation of blood vessel size" (GO) and "renin secretion" (KEGG). Genes implicated by rare SNVs at known loci were enriched in "tissue remodeling" (GO) and "artery aorta" (GO). Genes implicated by rare SNVs at new BP-loci were enriched in rare circulatory system diseases (that include hypertension and rare renal diseases) in Orphanet.

Information on new BP genes

Below is provided some information on some interesting genes harbouring or neighbouring new BP-associated rare/low-frequency variants.

ZFHX3

The low frequency missense variant rs62051555 (p.Gln2014His), located in exon eight of the transcription factor, zinc finger homeobox 3 (*ZFHX3*), is associated with increased levels of SBP and

PP. Interestingly, *ZFHX3* plays a role in the left-right patterning of cardiac atria during development, with changed expression of genes important for sidedness [2]. Mice with cardiac-restricted knockdown of *ZFHX3* have cardiomyopathy, impaired left ventricular function, atrial enlargement, altered atrial electrophysiology properties (increased conduction velocity) [2] and abnormalities in calcium homeostasis [3,2]. They also have severely dilated and fibrosed atria with a large mass consistent with thrombus and a significantly shorter life span compared to control animals [2]. The above abnormalities can increase susceptibility to atrial fibrillation (AF) [2]. *ZFHX3* has been reported multiple times to be associated with AF [4,5,6,7,8,9], a major risk factor for cardioembolic stroke[10] [11,12]. The association between AF and an increased risk for cardiovascular morbidity and mortality cannot be explained by thromboembolism alone, and patients with AF have increased beat-to-beat BP variability, which may adversely affect vascular structure and function [13], which can potentially influence BP.

LAMA5

Two low-frequency missense variants, rs11699758 (p.Val1757Ile) and rs13039398 (p.Arg1667Trp), residing in *LAMA5*, are associated with decreased SBP and PP. *LAMA5* encodes an extracellular matrix laminin α5 chain. Laminins are a group of α/β/γ glycoprotein heterotrimers, which constitute the main noncollagenous component of basement membranes [14]. Laminin α5 plays an important role in embryogenesis, and *Lama5-/-* mice embryos do not survive until birth [14]. Particularly, laminin heterotrimers containing laminin α5 chain are involved in glomerulogenesis, and are essential for the formation of the glomerular basement membrane, so that *Lama5-/-* embryos have failed vascularization of glomeruli in kidneys and even present with kidney agenesis [15].

Moreover, endothelial cell basement membrane laminin α5 is required for a normal shear response by resistance arteries [16]. The loss of laminin α5 from endothelial basement membranes in Tek-Cre::*Lama5-/-* mice results in an almost complete elimination of dilation in response to increased shear stress, which correlates with decreased endothelial cell cortical stiffness, decreased size of integrin beta1-positive/vinculin-positive focal adhesions and decreased junctional association of actin–myosin II [16]. *In vitro* experiments suggest that arterial endothelial cells directly bind to laminin α5/β1/γ1 via β1 integrins and that this binding increases VE-cadherin stabilization at cell-cell junctions, required for an adequate shear response [16].

LAMA5 is also a target of a therapy under investigation for treatment of stroke (Supplementary Table 24).

HSPA4

The missense variant of *HSPA4* (rs61755724, p.Ala159Thr) is associated with increase in DBP. Heat shock protein HSPA4 is a member of the HSP110 family and acts as a nucleotide exchange factor of HSP70 chaperones [17]. Upregulated expression of Hspa4 is observed in murine hearts exposed to pressure overload and in failing human hearts [17]. Furthermore, *Hspa4-/-* mice developed cardiac concentric hypertrophy and fibrosis with elevated expression levels of hypertrophic markers and an accumulation of polyubiquitinated proteins in neonatal hearts, suggesting that *Hspa4-/-* plays a role in protein quality control [17].

MCL1

The missense variant rs11580946 (p.Ala227Val), belonging to apoptosis regulator MCL1, is associated with decreased levels of SBP and PP. MCL1 participates in survival of haematopoietic stem

cells [18], progenitor cells, effector lymphocytes and cardiomyocytes [19]. Given its role in cell survival, MCL1 is a drug target for cancer-related phenotypes, with the small molecule inhibitor (antagonist) currently in 1 phase II trials and also for emergency treatment of acute angle-closure glaucoma and other conditions in which rapid reduction in intraocular pressure and vitreous volume is indicated (Supplementary Table 24). Cardiac-specific ablation of *Mcl-1* in mice results in a rapidly fatal dilated cardiomyopathy, preceded by loss of myofibrils and cardiac contractility, abnormal mitochondria ultrastructure, defective mitochondrial respiration, and impaired autophagy [20].

TBX5

The newly identified rare variant rs77357563 (p.Asp111Tyr; predicted deleterious by SIFT) in *TBX5*, is adjacent to the known *TBX3* region[21-23] and highlights *TBX5* as an additional candidate gene. *TBX5* is essential for normal cardiac development. Mutations in *TBX5* are known to cause various congenital heart diseases[24] and arrhythmias including Holt Oram syndrome and are associated with atrial fibrillation[25].

TGFB2

We observed rare variants in both intergenic and intronic regions, one rare intergenic variant rs12135454 is located near *TGFB2*. Prior work has indicated the TGFβ pathway as important in the genetics of BP traits[26]. Mutations in *TGFB2* cause Loeys-Dietz syndrome 4, a condition which includes aortic aneurysm, bicuspid aortic valve and arterial tortuosity.[27]

Mendelian Randomisation to assess the effect of metabolites on BP

We tested for pleiotropic effects of the IVs used for the 3-methylglutarylcarnitine(2) using two models. Firstly, we included any of the 14 metabolites in the analyses that shared at least one IV with 3-methylglutarylcarnitine(2) in a multi-variable MR model (three metabolites in total). Secondly, we included glycine in a multi-variable MR model with 3-methylglutarylcarnitine(2) as these two metabolites shared several IVs but glycine was not in our list of 14 metabolites analysed and we have recently shown that glycine is causal for BP[28]. 3-methylglutarylcarnitine(2) was consistently and significantly associated with DBP ($P < 0.05$) in the multi-variable MR models. Notably, we found that 3-methylglutarylcarnitine(2) was independently associated with DBP adjusting for the effect of glycine. Sensitivity analysis from multi-variable MR-Egger showed little evidence that the Egger intercept was deviated from zero for both models ($P_{\text{intercept}} > 0.01$).

We found genetically determined 3-methylglutarycarnitine (2) was predictive of DBP in both univariable and multivariable MR analyses (Supplementary Table 18). 3-methylglutarylcarnitine belongs to the class of organic compounds known as acyl carnitines involved in long-chain fatty acid metabolism in mitochondria and in leucine metabolism. It is a diagnostic metabolite of 3-hydroxy-3-methylglutaryl-coenzyme A lyase deficiency, an inborn error of metabolism in which the body cannot process leucine or generate ketones[29], with dilated cardiomyopathy as a complication[30]. Leucine has been shown to increase hypothalamic mTORC1 leading to an increase in BP[31]. A prospective clinical study also found that 3-methylglutarycarnitine was significantly lower in maternal first-trimester serum of fetal congenital heart defects (CHDs) than healthy controls[32].

Kidney expression data

Datasets, expression and SNP genotyping

The *cis*-eQTL meta-analysis was carried out using data from two projects: TRANScriptome of renaL humAn TissuE (TRANSLATE) Study (N=186) and The Cancer Genome Atlas (TCGA) study (N=99). The same quality control filters, data processing and analyses methods were applied to both datasets. Gene expression was quantified in terms of transcripts per million (TPM) using Kallisto[33]. Outlier samples were removed based on a statistic described in Wright *et al.* [34] or based on pairwise correlation between samples, where samples with median correlation <0.8 were excluded as per 't Hoen *et al.* [35]. Only genes on autosomal chromosomes were selected for the analysis. Gene expression threshold was set at TPM>0.1 in at least 20% of samples within each study/sequencing batch and read counts ≥ 6. A gene was also removed if its interquartile range was zero. Only genes that passed all of the above RNA-seq quality control filters in both studies were used in the analysis.

Gene-level TPM values were normalised as follows. First, $\log_2$ of TPM values were normalised across samples using robust quantile normalisation. Second, the normalised gene expression values were transformed using rank-based inverse normal transformation. Third, to account for hidden variation in RNA-seq data due to technical processing (such as batch effects or sample processing in pre-sequencing stage), we used probabilistic estimation of expression residuals (PEER) method[36] and estimated 30 hidden factors for TRANSLATE Study and 15 for TCGA. The numbers of hidden factors were chosen based on sample sizes of each dataset as recommended in GTEx eQTL analyses[37,38].

In TRANSLATE Study, genotyping was done using Infinium HumanCoreExome-24 BeadChip arrays and the allele calls were made using Genome Studio. In TCGA, genotyping was done using Affymetrix Genome-Wide Human SNP Array 6.0 and the allele calls were made using Birdseed. The following quality control filters were applied to genotype data. Samples were excluded if their genotyping rate was <95%, their heterozygosity rate was outside ± 3 standard deviations from the mean, they had cryptic relatedness with other individuals, were of non-white European genetic ancestry or had discordant sex information (inconsistency between declared and genotyped sex). Genetic variants were excluded if their genotyping rate was <95%, they mapped to Y chromosome or mitochondrial DNA, they had ambiguous chromosomal location, they violated Hardy-Weinberg equilibrium (HWE) (P<0.001) or if their minor allele frequency (MAF) was <5%.

Genotype imputation was conducted using *minimac3*[39] with Haplotype Reference Consortium data as the reference panel. The imputation was performed on Michigan Imputation Server[39]. Post-imputation, we excluded duplicate variants, non-SNPs, variants with low imputation coefficient (R^2<0.4), low frequency variants (MAF<5%) and variants that violated HWE ($P<10^{-6}$).

cis-eQTL meta-analysis

The association between gene expression and genotype was conducted using multiple linear regression with normalised gene expression as the dependent variable and genotype dosage, sex, top three genotype-derived principal components and the estimated hidden factors (30 for TRANSLATE Study and 15 for TCGA) as independent variables. The estimated coefficients from both studies were combined using inverse variance method. Only SNPs within 1Mb from the closest bound of a gene were considered. The correction for multiple testing for analysis of each gene with its in-*cis* SNPs was conducted using the permutation test, where the distribution of the smallest meta-combined P-value was determined using 2,000 permutations. At each permutation, the genotype sample labels were permuted but kept coupled with the sample labels of the top three genotype principal components for TRANSLATE Study data and TCGA data, separately. For each gene, the associations between its expression and its in-*cis* SNPs were re-estimated and the smallest meta-combined *P*-value recorded. Finally, for each gene the SNP with the smallest meta-combined P-value was identified and adjusted using the corresponding empirical distribution of the smallest meta-combined P-values for that gene. False discovery rate was determined using q-values from the *qvalue* R package. The permutation corrected P-values were used for calculating the false discovery rate (FDR) with a cut-off of 5%.

A threshold for nominal meta-combined *P*-values for SNPs that did not have the smallest meta-combined *P*-values was calculated as follows. First, a global permutation-adjusted P-value, p_t, was chosen to be the permutation-adjusted P-value for the gene with FDR closets to 5%. Then for each gene, a threshold for meta-combined nominal P-values was chosen to be the probability of observing a value less than or equal to p_t using the gene's empirical distribution of the smallest meta-combined *P*-values.

In total, 16,333 genes with at least one in-*cis* SNP and 4,862,143 SNPs with at least one in-*cis* gene were used in the analysis, resulting in 60,984,484 models. After correction for multiple testing, 4,431 genes passed FDR 5% cut-off. There were 425,096 statistically significant gene-SNP pairs that passed nominal P-value cut-offs: 317,425 unique SNPs associated with 4,431 genes.

The BP SNVs (N= 358 at 214 loci, see Supplementary Table 4b) were considered or proxies (r^2>0.8) if the sentinel SNV was not available. For reporting we only considered genes passing the 5% FDR cut-off and significant *cis*-eQTL signal(s) at $P < 5 \times 10^{-8}$. We reviewed the results for the most strongly associated *cis*-eQTL for the corresponding transcript. If the BP SNV and the eQTL were the same or in high LD (r^2>0.7), the BP SNV was reported as an eQTL. The results are summarized in Supplementary Table 21.

Colocalisation of BP associations and eQTL

Colocalisation analyses using the common variant results identified 32 unique BP-associated loci where the new BP-associated variant colocalised with the eQTL for 54 unique genes in GTEx tissues highlighting potential candidate genes. Many of the novel BP variants in genes including those in *PHACTR1, TIE1, CTSK, LTBP1, CRIM1, TIPARP* that colocalised with gene expression in GTEx in specific cardiovascular tissues, are also associated with CVD related phenotypes[40-53]. *TIE1* is involved in angiopoietin function in vascular remodelling and inflammation[54]. In the mouse, mutations in *Tie1* cause many cardiovascular phenotypes including small heart development, abnormal vascular endothelial cell morphology, abnormal endocardium morphology and abnormal heart atrium morphology[44,55]. Together these observations make *TIE1* a plausible candidate gene. *Crim1* KST264/KST264 mice implicate Crim1 in the regulation of vascular endothelial growth factor-A activity during glomerular vascular development[52]. *Tiparp* negative mice have kidney defects, including defects in smooth muscle cell number and location[56].

Tissue and cell enrichment analyses using DEPICT

We used DEPICT (Data-driven Expression Prioritized Integration for Complex Traits) as a complementary enrichment analysis to (1) identify tissues and cells in which genes at novel and previously reported BP loci are highly expressed and 2) to test for enrichment in gene sets associated with biological annotations, which included molecular pathways and phenotype data from mouse knockout studies. Two analyses were performed one involved all BP variants reported previously for BP traits (that were genome-wide significant in our dataset; Supplementary Table 9, 10) and a second set including all previously reported BP variants and variants at new loci, i.e. newly validated genome-wide significant SNVs (including the rare variants identified in the RV-GWAS) and any independent variants at these loci (Supplementary Tables 5, 6, 8). We report significant enrichments with a false discovery rate of 1%. We found the most significant enrichments were observed for the urogenital system (P=1.25x10^{-16}), cardiovascular system (P=2.01x10^{-13}) and endocrine system (P=1.78x10^{-11}) (Supplementary Table 23).

Enrichment of BP-associated SNVs in DNase I-hypersensitive sites

To investigate cell-type-specific enrichment within DNase I-hypersensitive sites we used FORGE, which tests for enrichment of SNVs within DNase I-hypersensitive sites in 299 cell types from the Epigenomics Roadmap Project and 125 cell lines from ENCODE[57]. All common and rare **non-coding** novel and conditionally independent validated variants from EAWAS, and SNVs from the RV-GWAS

(all $P<5.0x10^{-8}$) were included (Supplementary Tables 5, 6, 8). BP-trait specific analyses were not performed. We supplemented this listing to include all novel rare, low frequency and common variants from FINEMAP (variants not in LD ($r^2>0.6$) with a previously reported BP SNV (851 variants; Supplementary Table 10). In total 1,055 variants were included in the input from which 37 that were not in 1000 genomes Phase I and 64 that were in LD (r2>0.8) with the data were excluded leaving 954 for analysis. Enrichment was calculated by taking the Bonferroni corrected *P*-values from a binomial test comparing overlap of the supplied SNPs with 100 background SNP sets.

Significant results (Bonferroni corrected *P*-value<0.01) were observed across 15 tissues (Supplementary Table 23) in the ENCODE dataset. The strongest enrichments were in blood vessels, heart, skin, connective tissue, lung and epithelium (Z-score >6). These enriched tissues are similar to those reported for common BP associated SNVs[26]. Testing for enrichment in the Epigenomics Roadmap project indicated striking enrichment of BP SNVs in fetal kidney and fetal lung tissues (renal pelvis, renal cortex, renal kidney and lung, Z score=300) and significant enrichment across a further 12 tissues (new Supplementary Table 23).

Phenome-wide associations of the new common SNV BP loci

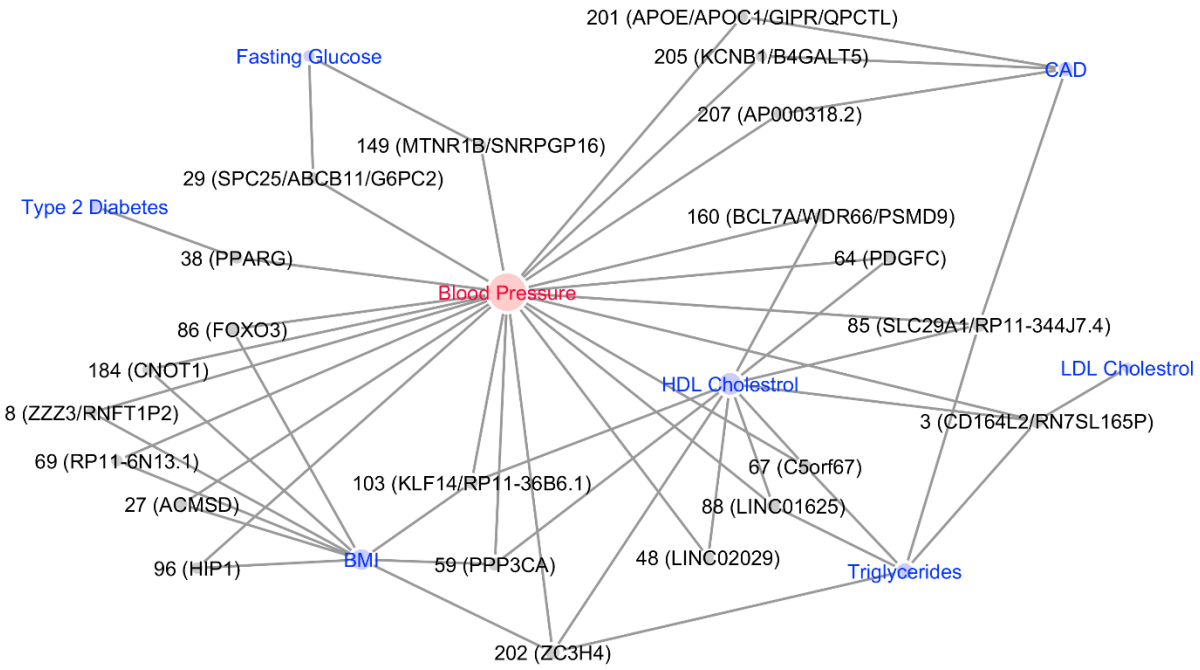
Two BP-associated loci were in high LD ($r^2>0.8$) with alcohol consumption variants. Variants at four new BP loci were in high LD with red blood cell trait associated SNVs, in particular haemoglobin, and one of these was also shared with iron traits (Figure 3). One locus was in LD with platelet traits and one with a plasminogen related trait. The new BP variants were also in high LD with variants associated with eye diseases for which hypertension is a risk factor: two with age-related macular degeneration and two with exfoliation glaucoma. The BP associated variant in *CASC16* was shared with Parkinson's disease. Telomere length has also been linked to aging and a variant at the *MYNN* locus was in LD with a telomere length associated variant.

Colocalization of BP-associated SNVs with cardiometabolic traits in the EAWAS

High blood pressure is one of several risk factors that act in concert increase risk for cardiovascular disease (CVD). To explore the genetic relations between blood pressure and other CVD risk factors (obesity, elevated blood total cholesterol, low density lipoprotein cholesterol [LDL], and triglyceride levels, high density lipoprotein [HDL] cholesterol levels, and diabetes), we conducted colocalization analyses using our blood pressure genetic results in conjunction with summary GWAS of other risk factors (body mass index[58], LDL cholesterol[59], triglycerides[59], HDL cholesterol[59], fasting glucose[60], type 2 diabetes[61] and coronary artery disease (CAD)[62]) using the COLOC package[63] in R to determine whether the same causal variant at each locus was associated with both blood pressure and CVD risk factor (Methods). At a posterior probability of both traits colocalising (H4) >90% (Supplementary Table 15), we found that blood pressure (DBP, SBP, PP) shared associated SNVs with CAD on chromosome 6 (SLC29A1/RP11-344J7.4 locus) , chromosome 19 (APOE/APOC1/GIPR/QPCTL), chromosome 20 (KCNB1/B4GALT5), chromosome 21(AP000318.2); with lipids (HDL cholesterol , LDL cholesterol and triglycerides) on chromosome 1(CD164L2), chromosome 3 (LINC02029), chromosome 4 (PPP3CA and PDGFC), chromosome 5 (C5orf67), chromosome 6 (SLC29A1 and

LINC01625), chromosome 7 (KLF14), chromosome 12 (BCL7A), chromosome 19 (ZC3H4); with BMI on chromosome 1(ZZZ3), chromosome 2 (ACMSD),chromosome 4(PPP3CA), chromosome 5 (RP11-6N13.1), chromosome 6 (FOXO3), chromosome 7(HIP1 and KLF14), chromosome 16(CNOT1), chromosome 19 (ZC3H4); with fasting glucose on chromosome 2 (SPC25/ABCB11/G6PC2), chromosome 11 (MTNR1B/SNRPGP16); and with type 2 diabetes on chromosome 3 (PPARG).

Supplementary Figure 5. Co-localisation of the newly identified BP-associated loci with cardiometabolic traits using the UKBB GWAS data The locus number is provided for the novel locus with the nearest gene(s) in parentheses



Mendelian Randomisation (MR) analyses of CVDs

We applied Mendelian randomisation (MR) to estimate the effects of blood pressure on CVD traits in a two-sample MR framework. The MR approach was based on the following assumption: (i) the genetic variants used as instrumental variables (IVs) are associated with blood pressure. (ii) the genetic variants are not associated with any confounders of the exposure-outcome relationship. (iii) the genetic variants are associated with the outcome only through change in BP *i.e.* a lack of pleiotropy.

The inverse-variance weighted (IVW) method with a multiplicative random-effect model[64], MR-Egger and MR-PRESSO were used. We also performed several sensitivity analyses to assess the robustness of our results to potential violations of the Mendelian Randomisation assumptions given these analyses have different assumptions for validity. To assess instrument strength, we computed the F statistic[65] for the association of genetic variants with SBP, DBP and PP, respectively. MR-Egger regression generates valid estimates even if not all the genetic instruments are valid, as long as the InSIDE (Instrument Strength Independent of Direct Effect) assumption holds[66] and also test if there is unbalanced pleiotropy. MR-PRESSO permits removal of outlier IVs[67]. To minimise pleiotropy, we removed SNVs associated with cardiovascular traits, including cholesterol level (LDL/HDL/triglycerides), smoking, Type 2 diabetes (T2D) and Atrial Fibrillation (AF) (Supplementary Table 16c). Although these methods may have different statistical power, the rationale is that if these methods give a similar conclusion regarding the association of BP and CVD, then we are more confident in inferring that the positive results are unlikely driven by violation of the MR assumptions[68].

We performed a genetic analysis of BP plus BP trait specific analyses of SBP, DBP, PP (online methods) using both previously published and newly identified BP SNVs. We considered any stroke, any ischemic stroke, large artery stroke, cardioembolic stroke, small vessel stroke and coronary artery disease (CAD) (online methods). As expected, blood pressure was positively associated with increased stroke (any stroke) risk (odds ratio (95% confidence interval) = 1.42 (1.36 - 1.49) per increase of one standard deviation in inverse-normal transformed of generic blood pressure (BPgeneric), $P = 5.70 \times 10^{-50}$; 1.71 (1.61 - 1.82) per increase of one standard deviation of inverse-normal transformed of SBP, $P = 1.35 \times 10^{-67}$; 1.53 (1.44 - 1.64) per increase of one standard deviation in inverse-normal transformed of DBP, $P = 2.34 \times 10^{-37}$; 1.39 (1.31 - 1.47) per increase of one standard deviation of inverse-normal transformed of PP, $P = 3.62 \times 10^{-28}$). MR-EGGER and MR-PRESSO gave similar results (Supplementary Table 16) and no significant pleiotropy was detected (P>0.01 for the MR-EGGER intercept; Supplementary Table 16). The positive association with stroke subtypes were statistically significant (P<0.00069; Figure 4, Supplementary Table 16), with the largest effect size of blood pressure on large artery stroke while smallest effect was with cardioembolic stroke. SBP was the primary association - with the largest effect size, with any of the CVD traits investigated (Figure 4, Supplementary Table 16), suggesting that SBP is the most sensitive BP measure, consistent with clinical practice.

In MR-Egger, we tested if the intercept estimate deviated from zero for the inference of genetic pleiotropy, i.e. where certain genetic variants affect the outcome through a different biological pathway from BP. In practice, there was little evidence that the MR-Egger intercept deviated from zero for any BP traits and any CVD traits, e.g. SBP and large artery stroke (intercept = 0.0026, SE = 0.0025, P = 0.31).

With MR-PRESSO, we used the outlier test embedded in the R package 'MR-PRESSO' to remove outlier due to pleiotropy and estimated the causal effects by IVW method before and after outlier removal. The causal effects (OR) after outlier-corrected were similar to the 'raw' estimates (Supplementary Table 16: with MR-PRESSO results), indicating that there was little evidence for genetic pleiotropy.

To quantify the strength of the selected instrumental variants for each "exposure (BP) – outcome (CVD)" pairs, we computed F-statistics (Supplementary Table 16). The F-statistics for the 964 SNVs for the "BP generic – Any Stroke" ranged from 11 to 767 with a median of 44, well above the threshold of $F > 10$ typically recommended for MR analysis [69].

When performing a multi-variable MR analyses including both SBP and DBP in the model for the inference of their effects on stroke, we found that the effect of SBP is still significant after adjusting for DBP, but not vice versa. Interestingly, we found that the effect of SBP on large artery stroke (P=7.21 × 10^{-23}; OR(95%CI)=2.62 (2.16, 3.17)per increase of one standard deviation of inverse-normal transformed of SBP) after adjusting for DBP is larger than the univariate MR estimation (P=1.30 × 10^{-33}; 2.19 (1.93, 2.48)), while the effect of DBP becomes negatively associated with stroke risk (P=6.28 × 10^{-2}; 0.832 (0.686, 1.01)) adjusting for SBP (although this did not pass our P-value threshold for significance). This is consistent with the findings from the univariable MR analysis of PP on stroke risk, which showed that PP has the largest effect on large artery stroke.

We also performed sensitivity analysis using multivariable MR-Egger to correct for pleiotropy[70]. Similar to the univariable MR-Egger results, there was little evidence that the multi-variable MR-Egger intercept deviated from zero for any BP traits and any CVD outcomes ($P_{\text{intercept}} > 0.01$).

Variance explained by BP-associated SNVs

We used 5,390 individuals from the Danish cohort within EPIC-CVD[71] to calculate variance explained as these participants were not used as part of the discovery set, genotyped using the Illumina Human CoreExome BeadChip array. SBP and DBP were measured twice at baseline and the average was used. Using a genetic risk score to represent all the known and new BP associations, we fitted a linear regression of each transformed BP trait against age, age^2, sex, BMI, top 10 genetic principle components, and CVD event (defined as any first CVD event) as a factor to obtained the variance explained by covariates ($R^2_{covariants}$). We then fit a second linear model for the transformed BP trait with all covariates plus a GRS to obtain the variance explained by all variables (R^2_{all}). Thus, the variance explained by GRS of BP genetic variants was:

$R^2_{GRS} = R^2_{all} - R^2_{covariants}$

We considered five different levels of GRS for each BP trait: (i) all independent common variants (MAF >= 0.01); (ii) all independent rare variants (MAF < 0.01); (iii) all independent SNVs within known loci; (iv) all independent SNVs within novel loci; (v) all independent SNVs.

We found the percentage of variance in BP explained by the BP-associated SNVs were consistent with previous reports.

Supplementary Table 26:

Percentage of variance explained for BP traits in the EPIC-CVD Danish cohort.

BP trait	Number of SNPs for constructing the GRS				
	ALL	COMM	RARE	KNOWN	NOVEL
SBP	778	734	44	507	271
DBP	742	708	34	494	248
PP	802	760	42	569	233

BP trait	% variance explained by GRS				
	ALL	COMM	RARE	KNOWN	NOVEL
SBP	4.54	4.55	0.17	4.54	0.62
DBP	3.541	3.421	0.183	3.311	0.601
PP	5.39	5.4	0.05	5.09	0.59

ALL = GRS of all associated variants for any BP trait
COMM = GRS of all common and low-frequency variants (MAF >= 0.01)
RARE = GRS of all rare variants (MAF < 0.01)
KNOWN = GRS of all known variants
NOVEL = GRS of novel variants identified in current study

UK Biobank specific analyses

The UK Biobank (UKBB) is a large prospective study of 502,642 participants aged 40–69 years when recruited between 2006–2010 at 22 assessment centres across the United Kingdom[72,73]. The study has collected and continues to collect a large amount of phenotypic measurements including systolic and diastolic blood pressure (BP).

Processing, quality control and analyses of the data provided by UK Biobank, were performed at two sites independently and were confirmed to be concordant at each step of the process.

Blood pressure measurement

BP was measured twice in a seated position after two minutes rest with a one minute rest before the second measurement [UK Biobank. UKB : Resource 100225 - Blood-pressure measurement procedures using ACE - Version 1.0. Available at: http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=100225. Accessed October 2, 2017]. An appropriate cuff and an Omron 705IT digital BP monitor, was used to measure BP in the majority of participants (UK Biobank data fields: SBP: f.4080.0.0 and f.4080.0.1; DBP: f.4079.0.0 and f.4079.0.1). If the largest cuff size was too small for the participant, or the electronic BP monitor failed, a sphygmomanometer with an inflatable cuff was used in conjunction with a stethoscope to perform a manual measurement (UK Biobank data fields: SBP: f.93.0.0 and f.93.0.1; DBP: f.93.0.0 and f.93.0.1). Of the 502,642 UKBB participants, 488,366 had both BP measurements and genotype data available, we therefore restricted phenotype quality control (QC) to these individuals. At baseline there were 446,611 participants with two automated BP measurements; 14,133 participants with one automated and one manual measurement and 26,615 with both manual measurements. The 1,007 samples with only one blood pressure measurement at baseline were excluded. Comparison of the BP distributions obtained using automated and manual approaches were concordant and reassured us both approaches were accurate. Individuals missing SBP or DBP at baseline assessment were removed (n=1,834). The mean of both measurements at baseline for a given participant was calculated to create an overall measure for SBP, DBP and PP. Phenotype QC was performed in R version v3.3.

Blood pressure measurement quality control Participants were excluded from analysis if

1. the difference between the first and second blood pressure measurement > 99.9th percentile (n=857);
2. covariates were missing: Age (n=0), gender (n=0), BMI (n=3105) using respectively UK Biobank data fields: f.21003.0.0, f.31.0.0 and f.21001.0.0;
3. they were pregnant at time of blood pressure measurement (n=131) UK Biobank data field: f.3140.0.0;
4. BMI >99.9th or <0.01 percentile (n=970).

In total 483,515 participants remained following quality control.

Adjustment of BP measurement for treatment effect For all UKBB participants that were on anti-hypertensive medication at time of blood pressure measurement (n=48,800) we added 15mmHg to the mean observed SBP, 10mmHg to the mean observed DBP and 5mmHg to the mean observed PP.

Definition of hypertension UKBB participants were defined as having hypertension when at least one of the following criteria was met:

1. Mean observed SBP ≥ 140 mmHg
2. Mean observed DBP ≥ 90 mmHg
3. History of hypertension: which was defined using the "non-cancer illnesses and associated first diagnosis timestamp" collected through the verbal interview (UK Biobank data field: f.20002.0.0) at baseline assessment for each UKBB participant. That is, where the following codes: "1065 hypertension", "1072 essential hypertension" are present in data field f.20002.0.0. No ICD codes were used to define hypertension.

4. Use of anti-hypertensive medication: at a baseline survey, we used responses to the "Medication for cholesterol, blood pressure or diabetes" question for males and responses to the "Medication for cholesterol, blood pressure, diabetes, or take exogenous hormones" question for females, both collected through the touchscreen questionnaire and providing information on regular medication use (UK Biobank data fields: f.6177.0.0 and f.6153.0.0, respectively). If a participant selected "2 Blood pressure medication" we defined this participant as having a current status of taking anti-hypertensive medication (27,931 females, 22,630 males).

255,794 individuals were defined as hypertensive and 227,721 were non-hypertensive.

Genotype quality control (Supplementary Figure)

We used both the Affymetrix UK Biobank/BiLEVE array genotypes and the Human Reference Consortium imputed genotypes [ref Bartlett et al.]. Genotype QC was performed using PLINK1.9 and R v3.3

Defining a European set of UK Biobank participants Approximately 22,000 UKBB participants had a self-reported ethnic background outside of Europe[73]. Deviation from Hardy Weinberg Equilibrium (HWE) is often an indicator of a poorly genotyped variant. However, due to the ethnic diversity of the UKBB cohort, deviations from HWE could also be due to violation of the assumptions of HWE *e.g.* large differences in allele frequency in an ethnically mixed cohort. We therefore sought to define a genetically European group of UKBB participants using principal component analyses (PCA) with FlashPCA2[74]. High-quality autosomal variants were selected for PCA based on an overall call rate ≥99%; minor allele frequency (MAF) >=0.05 and HWE $P \geq 10^{-5}$. Regions of the genome known to exhibit long-range linkage disequilibrium (LD) were removed (chr6:25–33.5 Mb, chr8:8–12 Mb, chr17:40.4–42.4 Mb) to ensure the PCs were picking up ancestry and not LD. These variants were then LD pruned so no pair of variants within a 100 variant window had R^2>0.2. A final round of LD pruning was performed in a 1000 variant window.

Having generated 50 PCs, we adopted the method of Astle et al.[75], to identify ancestral outliers to be remove. In brief, a 'genetic distance',

$$d(i) = \sqrt{\sum_{m=1}^{15} E_m (P_{im} - C_m)^2}$$

, between individual *i* and a hypothetical median ''white British'' participant was calculated, where E_m represents the eigenvalue corresponding to PC, m (*i.e.* the genetic variance explained by PC_m), P_{im} represents the score of individual i on PC_m, C_m represents the median score on PC_m of participants with self-reported White ancestry (defined as "British", "Irish", "White" or "Any other White background").

We used a threshold of genetic distance > 0.2 to identify non-Europeans, which resulted in the exclusion of 23,511 non-European participants.

Batch level variant and sample QC Genotype QC was performed with the above defined European subset of participants, separately for each of the 106 UKBB genotyping batches. The following thresholds were applied to remove variants: call rate ≤ mean (call rate) - [3 x SD (call rate)]; HWE *P*-value < $1x10^{-12}$ (MAF<0.01) or HWE *P*-value < $1x10^{-6}$ (MAF≥0.01). Variants that failed either call rate or HWE within a batch were excluded from the corresponding batch prior to batch-level sample QC. Within batch, samples with call rate < mean (call rate) - [3 x SD (call rate)] or Heterozygosity > (mean +/- 3SD) were removed (n=11,944).

Variant and sample QC across all batches Variants that failed QC in >48 batches (UKBB array) or > 3 batches (UK BiLEVE array) were excluded (n=23,221 SNVs). We excluded samples who's genetic sex and phenotypically defined sex (as provided by the UKBB) were discordant (n=136 samples).

After variant and sample QC across all batches we performed a second PCA with FlashPCA2[74] using the same approach to select variants for PCA as described above. A genetic distance measure of 0.175 calculated using 8PCs (as described above) was used to remove a further 3,015 individuals of non-European ancestry.

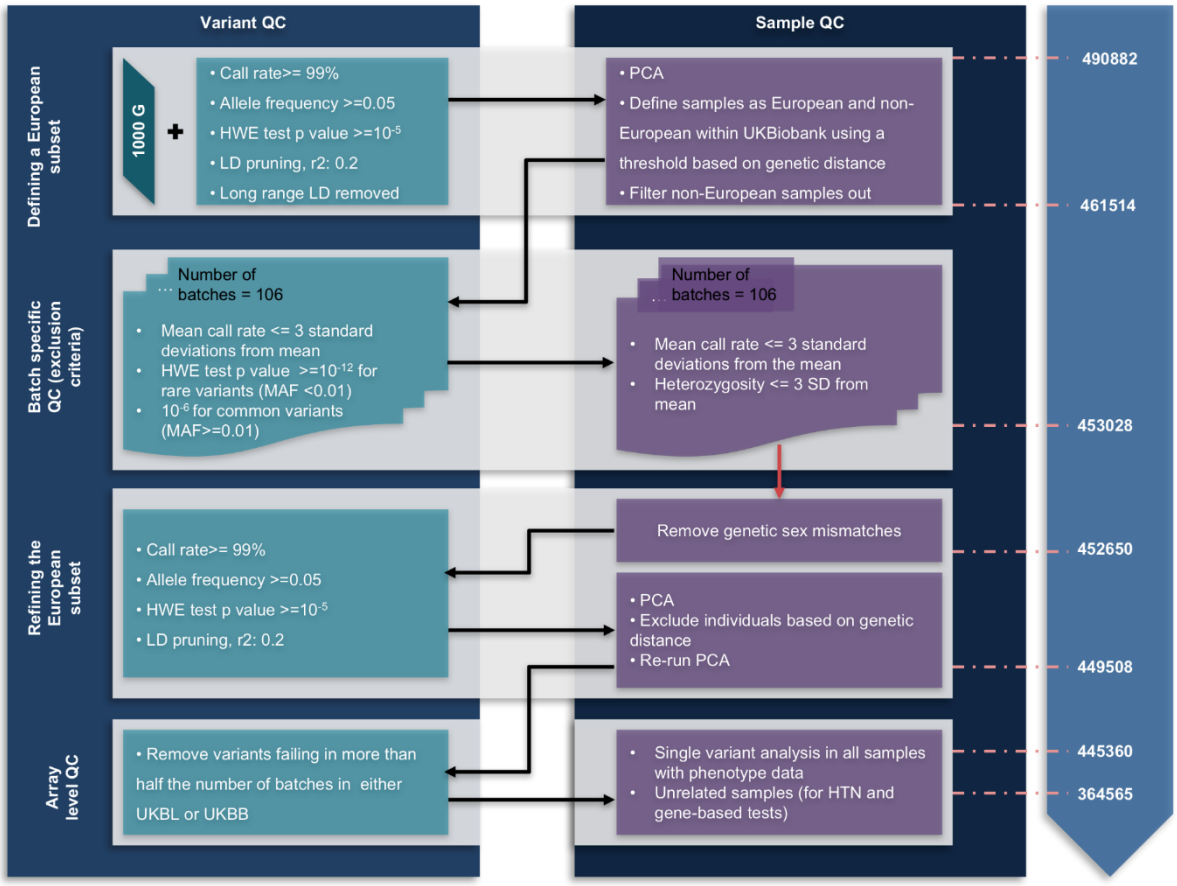
Definition of an unrelated set of UK Biobank participants For analyses of hypertension, we chose not to use a mixed effects model due to limitations with calculating a full kinship matrix. Therefore, using the fully QC'd data, we defined a subset of unrelated UKBB participants using the kinship information provided by UKBB that lists the kinship coefficient of pairs of individuals up to 3rd degree relatives. We calculated sample call rate to guide which participant within a pair of relatives to remove. All pairs that shared individual(s) were aggregated into families. From each of these families the sample with the highest call rate was retained. If individuals within the family had the same call rate we chose the one that occurred first in the file.

Imputation The pre-imputation variant QC, phasing and imputation performed on the combined UKBB and UK BiLEVE data has been described in detail elsewhere[73]. The genetic data were imputed using the Haplotype Reference Consortium (HRC) panel. Additional variants were available in the interim release of imputed using 1000G/UK10K data in 150,000 UKBB participants but were not part of the HRC imputation panel. We extracted 30,315 variants that were readily available in the first release UKBB imputation dataset and were genotyped on the exome array but not either of the Affymetrix arrays used by UKBB. After QC of these variants and using an information score threshold >0.3, 157,666 variants were available for analysis in ~150,000 participants from the interim release.
Variants for which both genotype and imputation data were available, we used the imputed variant if the genotyping call rate was <0.98 and the variant was imputed with an information score >0.7. We used the genotyped data for all variants that did not satisfy these criteria. All variants that passed QC and were available in either the genotyped or imputed data alone were also analysed.

In total, 39,312,035 imputed variants with info>0.3 of which 31,835,351 were low frequency or rare were analysed in GWAS of UKBB. A further 784055 genotyped variants were analysed of which 405,033 were rare or low-frequency.

Final dataset used for exome content analyses Following QC, 156,481 variants from the UK-Biobank full release and 18,947 variants from the interim release were analyzed in 364,510 European participants with SBP, DBP and PP measurements. Following QC and transformation, 157,666 Exome array variants (62,032 genotyped and 95,634 imputed) were tested for association with HTN in up to 364,565 unrelated European participants.

Supplementary Figure 6. Flowchart summarizing quality control procedures applied to genetic data in UKBB



Analyses of SBP DBP and PP

Each of the continuous traits (SBP, DBP and PP) were regressed on baseline age, baseline age squared, gender, BMI and genotyping array using the lm function in R. The residuals from these regression models were rank transformed and inverse normalised and the resulting transformed SBP, DBP and PP residuals were analysed using linear mixed models implemented in BOLT-LMM (Version: v2.3). The set of QCd variants used for the second PCA were also used for BOLT-LMM model building.

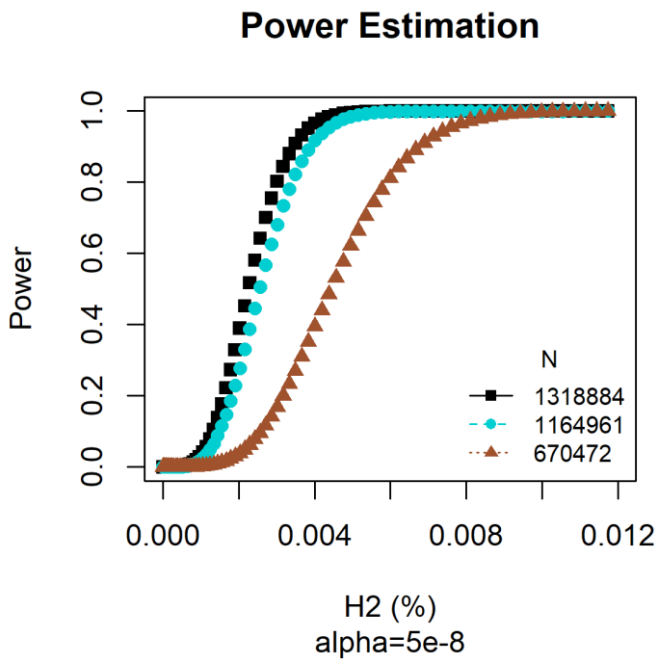
In total, 784,045 directly genotyped and 39,312,035 imputed variants (175,430 were Exome array variants of which 59,824 variants were genotyped and 115,606 variants were imputed) were analysed for association with SBP, DBP and PP in up to 445,415 individuals of European ancestry from UKBB.

Analyses of hypertension

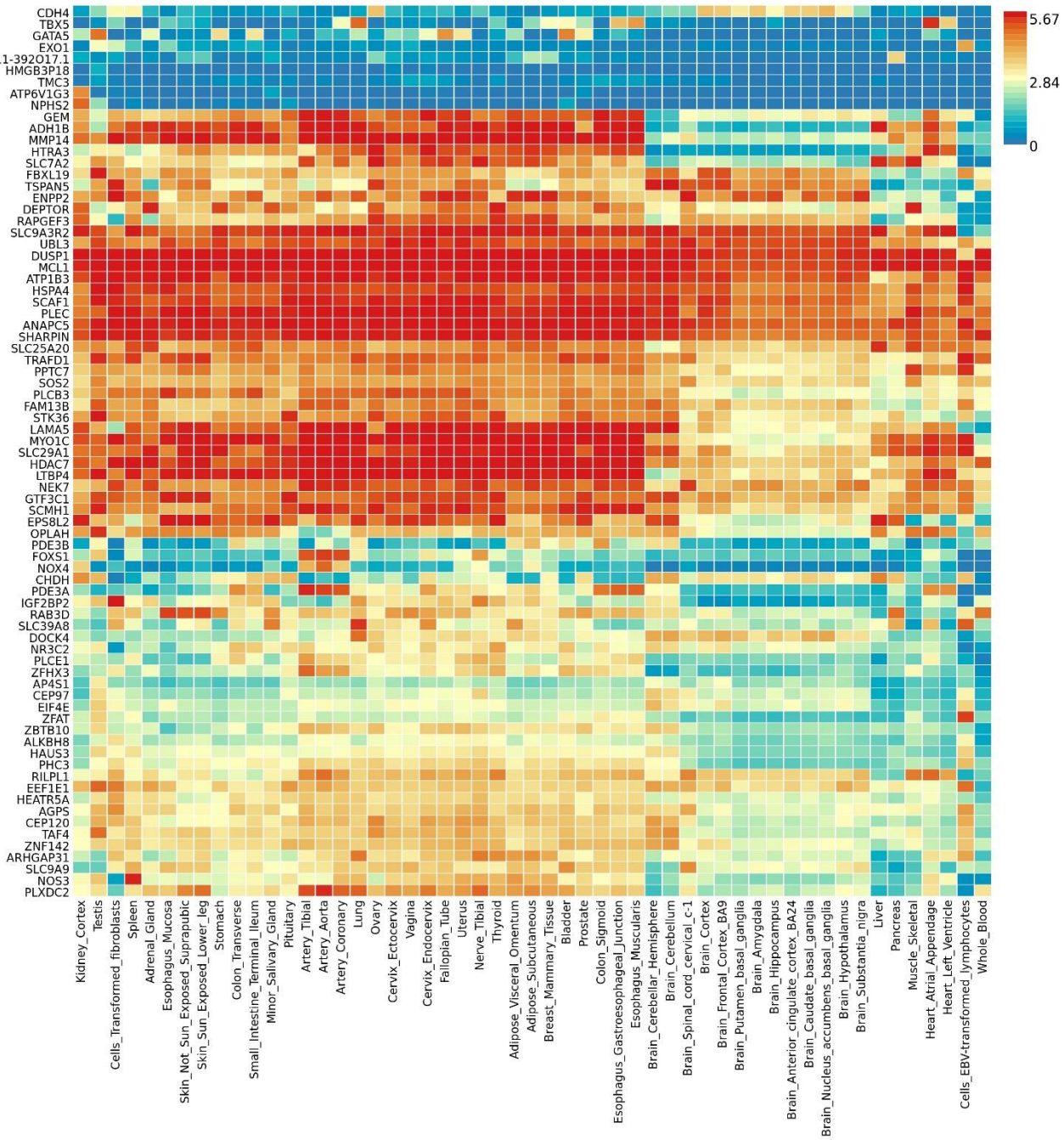
Genetic analysis of exome array variants was performed for hypertension as a binary outcome in 364,510 unrelated individuals (192,235 hypertensive cases and 172,275 controls) of European ancestry using SNPTEST (Version: v2.5.4-beta3). Analyses were adjusted for baseline age, baseline age squared, gender, BMI, genotyping array and the first eight ancestry principal components (PCs).

Supplementary Figures

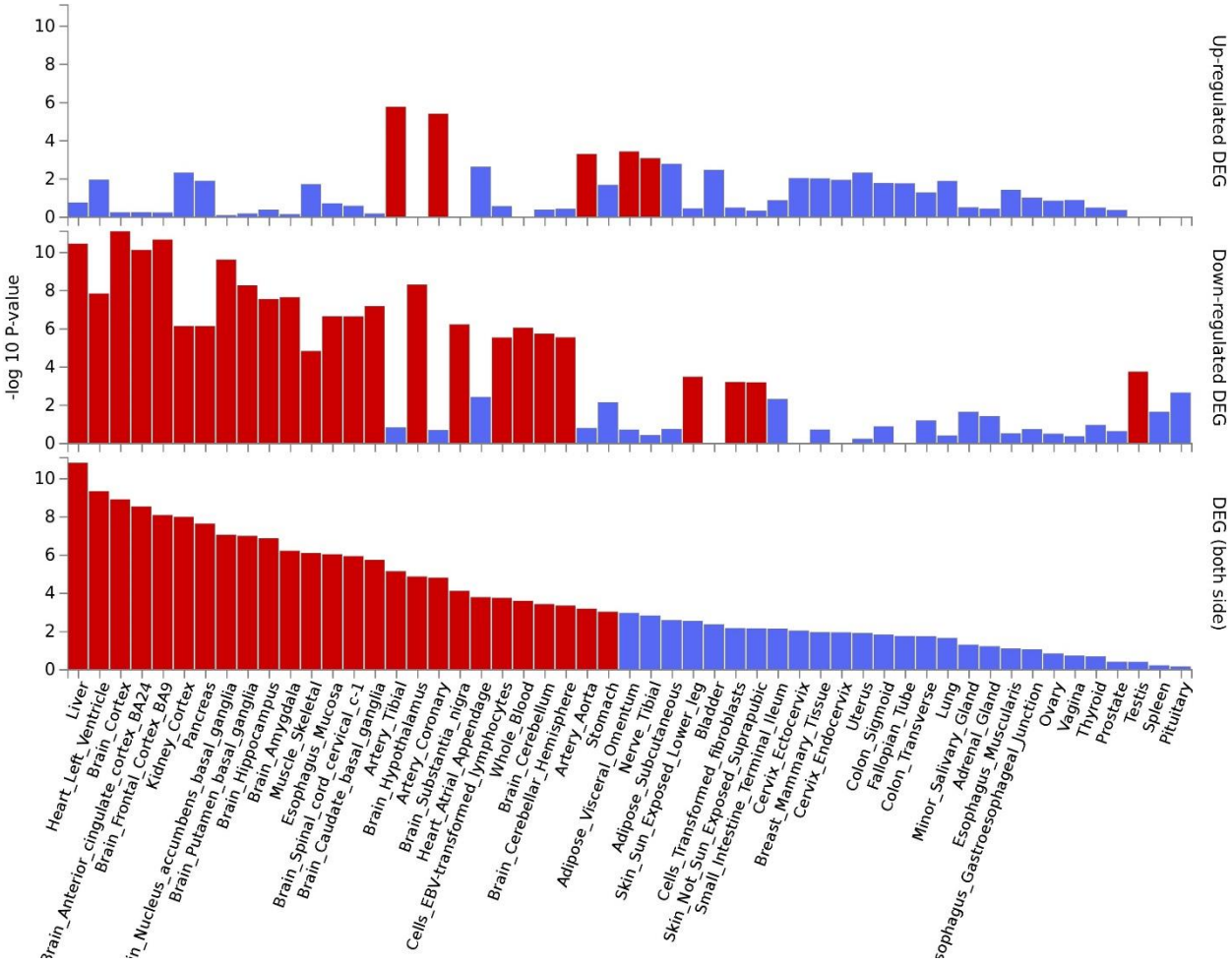
Supplementary Figure 2. Power estimation for Stage 2 meta-analyses (EAWAS: PA – 1,318,884 participants; EA – 1,164,961 participants; RV-GWAS: 670,472 participants).



Supplementary Figure 3 (a) Expression of genes implicated by the rare SNVs in GTEx v7 tissues. We used FUMA GWAS.

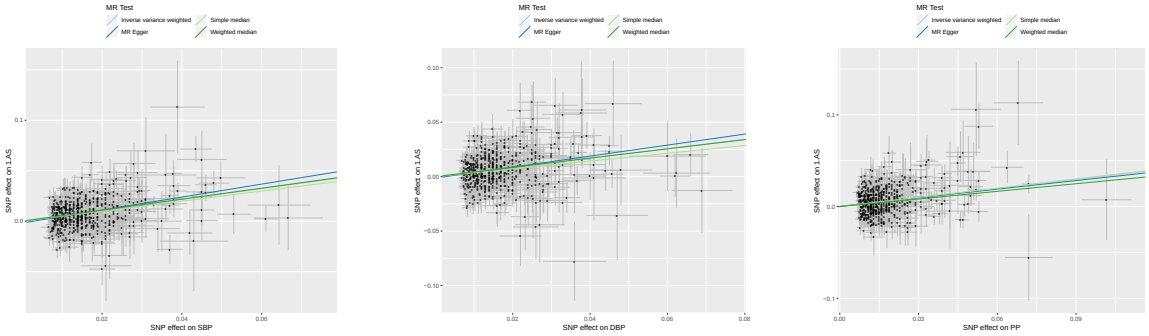


Supplementary Figure 3 (b) Tissue enrichment of rare variant gene expression levels in GTEx v7. We used FUMA GWAS to perform these analyses.

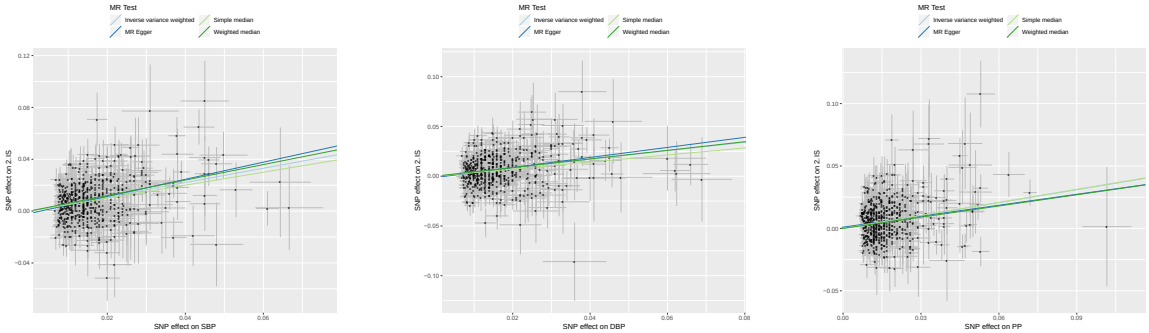


Supplementary Figure 4. Mendelian randomization analysis for blood pressure level and risk of cardiovascular diseases.

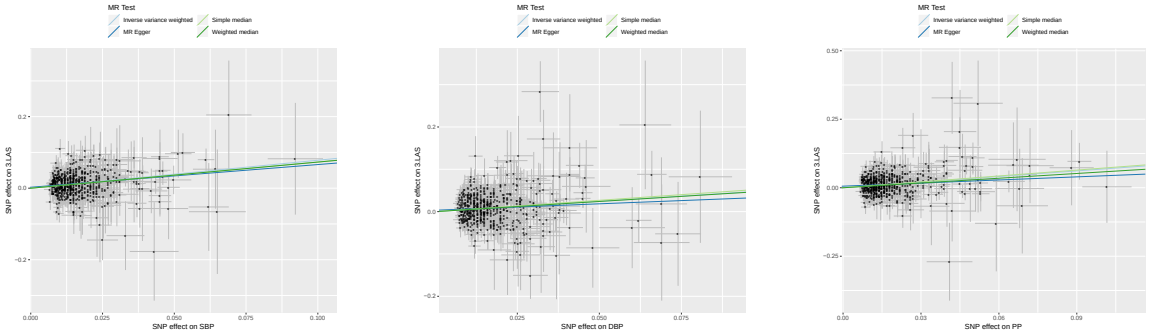
a. SBP -> All Stroke | **DBP -> All Stroke** | **PP -> All Stroke**



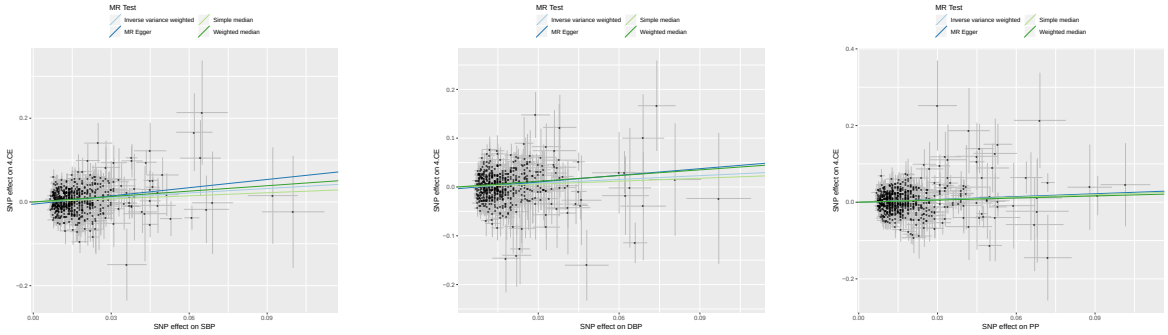
b. SBP -> Ischemic Stroke | **DBP -> Ischemic Stroke** | **PP -> Ischemic Stroke**



c. SBP -> Large Artery Stroke | **DBP -> Large Artery Stroke** | **PP -> Large Artery Stroke**



d. SBP -> Cardioembolic Stroke | **DBP -> Cardioembolic Stroke** | **PP -> Cardioembolic Stroke**



Supplementary Figure 4. Mendelian randomization analysis for blood pressure level and risk of cardiovascular diseases.

Associations between genetically determined blood pressure traits (SBP, DBP and PP) and risk of All Stroke (a), Ischemic Stroke (b), Large Artery Stroke (c), Cardioembolic Stroke (d), Small Vessel Stroke (e) and Coronary Artery Disease (f) based on four MR methods: IVW, MR-Egger, Simple median and Weighted median.

References

1. Falconer, D.S. *Introduction to Quantitative Genetics*, (Longman Green/John Wiley & Sons, UK/New York, 1989).
2. Hanley, A. *et al.* 56 Role of ZFHX3 in atrial fibrillation. *Heart* **104**, A42-A43 (2018).
3. Kao, Y.H. *et al.* ZFHX3 knockdown increases arrhythmogenesis and dysregulates calcium homeostasis in HL-1 atrial myocytes. *Int J Cardiol* **210**, 85-92 (2016).
4. Ellinor, P.T. *et al.* Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat Genet* **44**, 670-5 (2012).
5. Benjamin, E.J. *et al.* Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry. *Nat Genet* **41**, 879-81 (2009).
6. Christophersen, I.E. *et al.* Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat Genet* **49**, 946-952 (2017).
7. Low, S.K. *et al.* Identification of six new genetic loci associated with atrial fibrillation in the Japanese population. *Nat Genet* **49**, 953-958 (2017).
8. Lee, J.Y. *et al.* Korean atrial fibrillation network genome-wide association study for early-onset atrial fibrillation identifies novel susceptibility loci. *Eur Heart J* **38**, 2586-2594 (2017).
9. Gudbjartsson, D.F. *et al.* A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet* **41**, 876-8 (2009).
10. Kamel, H. & Healey, J.S. Cardioembolic Stroke. *Circ Res* **120**, 514-526 (2017).
11. Hauer, A.J. *et al.* A replication study of genetic risk loci for ischemic stroke in a Dutch population: a case-control study. *Sci Rep* **7**, 12175 (2017).
12. Traylor, M. *et al.* Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol* **11**, 951-62 (2012).
13. Olbers, J. *et al.* High beat-to-beat blood pressure variability in atrial fibrillation compared to sinus rhythm. *Blood Pressure* **27**, 249-255 (2018).
14. Miner, J.H., Cunningham, J. & Sanes, J.R. Roles for laminin in embryogenesis: exencephaly, syndactyly, and placentopathy in mice lacking the laminin alpha5 chain. *J Cell Biol* **143**, 1713-23 (1998).
15. Miner, J.H. & Li, C. Defective glomerulogenesis in the absence of laminin alpha5 demonstrates a developmental role for the kidney glomerular basement membrane. *Dev Biol* **217**, 278-89 (2000).
16. Di Russo, J. *et al.* Endothelial basement membrane laminin 511 is essential for shear stress response. *EMBO J* **36**, 1464 (2017).
17. Mohamed, B.A. *et al.* Targeted disruption of Hspa4 gene leads to cardiac hypertrophy and fibrosis. *J Mol Cell Cardiol* **53**, 459-68 (2012).
18. Opferman, J.T. *et al.* Obligate role of anti-apoptotic MCL-1 in the survival of hematopoietic stem cells. *Science* **307**, 1101-4 (2005).
19. Czabotar, P.E., Lessene, G., Strasser, A. & Adams, J.M. Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat Rev Mol Cell Biol* **15**, 49-63 (2014).
20. Wang, X. *et al.* Deletion of MCL-1 causes lethal cardiac failure and mitochondrial dysfunction. *Genes Dev* **27**, 1351-64 (2013).
21. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat Genet* **41**, 677-87 (2009).
22. 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).

23. Kato, N. *et al.* Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet* **43**, 531-8 (2011).
24. Kokkinopoulos, I. *et al.* Single-Cell Expression Profiling Reveals a Dynamic State of Cardiac Precursor Cells in the Early Mouse Embryo. *PLoS One* **10**, e0140831 (2015).
25. Zhang, R. *et al.* Common Variants in the TBX5 Gene Associated with Atrial Fibrillation in a Chinese Han Population. *PLoS One* **11**, e0160467 (2016).
26. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* **50**, 1412-1425 (2018).
27. Lindsay, M.E. *et al.* Loss-of-function mutations in TGFB2 cause a syndromic presentation of thoracic aortic aneurysm. *Nat Genet* **44**, 922-7 (2012).
28. Wittemans, L.B.L. *et al.* Assessing the causal association of glycine with risk of cardio-metabolic diseases. *Nat Commun* **10**, 1060 (2019).
29. Roe, C.R., Millington, D.S. & Maltby, D.A. Identification of 3-methylglutarylcarnitine. A new diagnostic metabolite of 3-hydroxy-3-methylglutaryl-coenzyme A lyase deficiency. *J Clin Invest* **77**, 1391-4 (1986).
30. Leung, A.A., Chan, A.K., Ezekowitz, J.A. & Leung, A.K. A Case of Dilated Cardiomyopathy Associated with 3-Hydroxy-3-Methylglutaryl-Coenzyme A (HMG CoA) Lyase Deficiency. *Case Rep Med* **2009**, 183125 (2009).
31. Harlan, S.M., Guo, D.F., Morgan, D.A., Fernandes-Santos, C. & Rahmouni, K. Hypothalamic mTORC1 signaling controls sympathetic nerve activity and arterial pressure and mediates leptin effects. *Cell Metab* **17**, 599-606 (2013).
32. Bahado-Singh, R.O. *et al.* Metabolomic prediction of fetal congenital heart defect in the first trimester. *Am J Obstet Gynecol* **211**, 240 e1-240 e14 (2014).
33. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-7 (2016).
34. Wright, F.A. *et al.* Heritability and genomics of gene expression in peripheral blood. *Nat Genet* **46**, 430-7 (2014).
35. t Hoen, P.A. *et al.* Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nat Biotechnol* **31**, 1015-22 (2013).
36. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* **7**, 500-7 (2012).
37. Consortium, G.T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
38. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
39. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
40. Paquette, M., Dufour, R. & Baass, A. PHACTR1 genotype predicts coronary artery disease in patients with familial hypercholesterolemia. *J Clin Lipidol* **12**, 966-971 (2018).
41. Rodriguez-Perez, J.M. *et al.* Possible role of intronic polymorphisms in the PHACTR1 gene on the development of cardiovascular disease. *Med Hypotheses* **97**, 64-70 (2016).
42. Gupta, R.M. *et al.* A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. *Cell* **170**, 522-533 e15 (2017).
43. Woo, K.V. & Baldwin, H.S. Role of Tie1 in shear stress and atherosclerosis. *Trends Cardiovasc Med* **21**, 118-23 (2011).
44. Puri, M.C., Partanen, J., Rossant, J. & Bernstein, A. Interaction of the TEK and TIE receptor tyrosine kinases during cardiovascular development. *Development* **126**, 4569-80 (1999).
45. Woo, K.V. *et al.* Tie1 attenuation reduces murine atherosclerosis in a dose-dependent and shear stress-specific manner. *J Clin Invest* **121**, 1624-35 (2011).

46. Zhao, G. *et al.* Increased Circulating Cathepsin K in Patients with Chronic Heart Failure. *PLoS One* **10**, e0136093 (2015).
47. Vassalle, C. & Iervasi, G. Cathepsin K--a classical bone biomarker in cardiovascular disease: the heart is not alone anymore. *Atherosclerosis* **228**, 36-7 (2013).
48. Todorovic, V. *et al.* Long form of latent TGF-beta binding protein 1 (Ltbp1L) is essential for cardiac outflow tract septation and remodeling. *Development* **134**, 3723-32 (2007).
49. Todorovic, V. *et al.* Long form of latent TGF-beta binding protein 1 (Ltbp1L) regulates cardiac valve development. *Dev Dyn* **240**, 176-87 (2011).
50. Horiguchi, M., Todorovic, V., Hadjiolova, K., Weiskirchen, R. & Rifkin, D.B. Abrogation of both short and long forms of latent transforming growth factor-beta binding protein-1 causes defective cardiovascular development and is perinatally lethal. *Matrix Biol* **43**, 61-70 (2015).
51. Iyer, S. *et al.* Crim1 has cell-autonomous and paracrine roles during embryonic heart development. *Sci Rep* **6**, 19832 (2016).
52. Wilkinson, L. *et al.* Crim1KST264/KST264 mice implicate Crim1 in the regulation of vascular endothelial growth factor-A activity during glomerular vascular development. *J Am Soc Nephrol* **18**, 1697-708 (2007).
53. Humblet, O., Birnbaum, L., Rimm, E., Mittleman, M.A. & Hauser, R. Dioxins and cardiovascular disease mortality. *Environ Health Perspect* **116**, 1443-8 (2008).
54. La Porta, S. *et al.* Endothelial Tie1-mediated angiogenesis and vascular abnormalization promote tumor progression and metastasis. *J Clin Invest* **128**, 834-845 (2018).
55. Patan, S. TIE1 and TIE2 receptor tyrosine kinases inversely regulate embryonic angiogenesis by the mechanism of intussusceptive microvascular growth. *Microvasc Res* **56**, 1-21 (1998).
56. Schmahl, J., Raymond, C.S. & Soriano, P. PDGF signaling specificity is mediated through multiple immediate early genes. *Nat Genet* **39**, 52-60 (2007).
57. Breeze, C.E. *et al.* eFORGE: A Tool for Identifying Cell Type-Specific Signal in Epigenomic Data. *Cell Rep* **17**, 2137-2150 (2016).
58. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641-3649 (2018).
59. Willer, C.J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-1283 (2013).
60. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-16 (2010).
61. Scott, R.A. *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* **66**, 2888-2902 (2017).
62. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121-1130 (2015).
63. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
64. Burgess, S. *et al.* Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol* **30**, 543-52 (2015).
65. Pierce, B.L., Ahsan, H. & Vanderweele, T.J. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* **40**, 740-52 (2011).
66. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512-25 (2015).
67. Verbanck, M., Chen, C.Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* **50**, 693-698 (2018).

68. Lawlor, D.A., Tilling, K. & Davey Smith, G. Triangulation in aetiological epidemiology. *Int J Epidemiol* **45**, 1866-1886 (2016).
69. Palmer, T.M. *et al.* Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research* **21**, 223-242 (2012).
70. Rees, J.M.B., Wood, A.M. & Burgess, S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat Med* **36**, 4705-4718 (2017).
71. Danesh, J. *et al.* EPIC-Heart: the cardiovascular component of a prospective study of nutritional, lifestyle and biological factors in 520,000 middle-aged participants from 10 European countries. *Eur J Epidemiol* **22**, 129-41 (2007).
72. 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).
73. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
74. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33**, 2776-2778 (2017).
75. Astle, W.J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429 e19 (2016).

ACKNOWLEDGEMENTS

UK Biobank

This research was conducted using the UK Biobank Resource under Application Numbers 20480 and 15293.

CHD Exome+ Consortium

This work was supported by core funding from: the UK Medical Research Council (G0800270; MR/L003120/1), the British Heart Foundation (SP/09/002; RG/13/13/30194; RG/18/13/33946) and the National Institute for Health Research [Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust]*. Work was also funded by the European Research Council (268834), the European Commission Framework Programme 7 (HEALTH-F2-2012-279233), Pfizer, Novartis and Merck.

MORGAM

This work has been sustained by the MORGAM Project's recent funding: European Union FP 7 projects ENGAGE (HEALTH-F4-2007-201413), CHANCES (HEALTH-F3-2010-242244) and BiomarCaRE (278913). This has supported central coordination, workshops and part of the activities of the The MORGAM Data Centre, at THL in Helsinki, Finland. MORGAM Participating Centres are funded by regional and national governments, research councils, charities, and other local sources.

BRAVE

The BRAVE study genetic epidemiology working group is a collaboration between the Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, UK, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh and the National Institute of Cardiovascular Diseases, Dhaka, Bangladesh.

CCHS, CGPS, CIHDS

We thank participants and staff of the Copenhagen City Heart Study, the Copenhagen General Population Study and the Copenhagen Ischemic Heart Disease Study for their important contributions.

EPIC-CVD

This work was supported by core funding from: the UK Medical Research Council (G0800270; MR/L003120/1), the British Heart Foundation (SP/09/002; RG/13/13/30194; RG/18/13/33946) and the National Institute for Health Research [Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust] [*]. Work was also funded by the European Research Council (268834), the European Commission Framework Programme 7 (HEALTH-F2-2012-279233), Pfizer, Novartis and Merck.

*The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

We thank all EPIC participants and staff for their contribution to the study, the laboratory teams at the Medical Research Council Epidemiology Unit for sample management and Cambridge Genomic Services for genotyping, Sarah Spackman for data management, and the team at the EPIC-CVD Coordinating Centre for study coordination and administration.

WOSCOPS/PROSPER (CHD Exome+ Consortium)

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° HEALTH-F2-2009-223004

PROMIS (CHD Exome+ Consortium)

We are thankful to all the study participants in Pakistan. Recruitment in PROMIS was funded through grants available to investigators at the Center for Non-Communicable Diseases, Pakistan (Danish Saleheen and Philippe Frossard) and investigators at the University of Cambridge, UK (Danish Saleheen and John Danesh). Field-work, genotyping, and standard clinical chemistry assays in PROMIS were principally supported by grants awarded to the University of Cambridge from the British Heart Foundation, UK Medical Research Council, Wellcome Trust, EU Framework 6-funded Bloodomics Integrated Project, Pfizer, Novartis, and Merck. We would like to acknowledge the contributions made by the following individuals who were involved in the field work and other administrative aspects of the study: Mohammad Zeeshan Ozair, Usman Ahmed, Abdul Hakeem, Hamza Khalid, Kamran Shahid, Fahad Shuja, Ali Kazmi, Mustafa Qadir Hameed, Naeem Khan, Sadiq Khan, Ayaz Ali, Madad Ali, Saeed Ahmed, Muhammad Waqar Khan, Muhammad Razaq Khan, Abdul Ghafoor, Mir Alam, Riazuddin, Muhammad Irshad Javed, Abdul Ghaffar, Tanveer Baig Mirza, Muhammad Shahid, Jabir Furqan, Muhammad Iqbal Abbasi, Tanveer Abbas, Rana Zulfiqar, Muhammad Wajid, Irfan Ali, Muhammad Ikhlaq, Danish Sheikh and Muhammad Imran.

EPIC-InterAct

Funding for the InterAct project was provided by the EU FP6 programme (grant number LSHM_CT_2006_037197). We thank all EPIC participants and staff for their contribution to the study. We thank the lab team at the MRC Epidemiology Unit for sample management and Nicola Kerrison for data management.

Fenland

The Fenland Study is funded by the Wellcome Trust and the Medical Research Council (MC_U106179471). We are grateful to all the volunteers for their time and help, and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland Study Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Data and Laboratory teams. We further acknowledge support from the Medical research council (MC_UU_12015/1).

EPIC Norfolk

The EPIC-Norfolk study (https://doi.org/10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1 and MC-UU_12015/1) and Cancer Research UK (C864/A14136). The genetics work in the EPIC-Norfolk study was funded by the Medical Research Council (MC_PC_13048). Metabolite measurements in the EPIC-Norfolk study were supported by

the MRC Cambridge Initiative in Metabolic Science (MR/L00002/1) and the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement no. 115372. We are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research.

GoT2D Consortium

GoT2D Funding for the GoT2D and T2D-GENES studies was provided by grants NIH U01s DK085526, DK085501, DK085524, DK085545, and DK085584 (Multiethnic Study of Type 2 Diabetes Genes) and DK088389 (Low-Pass Sequencing and High-Density SNP Genotyping for Type 2 Diabetes).

GoT2D Genotyping of the METSIM and DPS studies, and part of the FUSION study, was conducted at the Genetic Resources Core Facility (GRCF) at the Johns Hopkins Institute of Genetic Medicine.

GoT2D The Broad Genomics Platform for genotyping of the FIN-D2D 2007, FINRISK 2007, DR'sEXTRA, and FUSION studies.

ADDITION

The Danish Diabetes Academy is funded by the Novo Nordisk Foundation. The ADDITION-DK study was supported by the National Health Service in the counties of Copenhagen, Aarhus, Ringkoebing, Ribe, and South Jutland; the Danish Council for Strategic Research; the Danish Research Foundation for General Practice; Novo Nordisk Foundation; the Danish Center for Evaluation and Health Technology Assessment; the Diabetes Fund of the National Board of Health; the Danish Medical Research Council; and the Aarhus University Research Foundation. ADDITION-DK has been given unrestricted grants from Novo Nordisk A/S, Novo Nordisk Scandinavia AB, Novo Nordisk UK, ASTRA Denmark, Pfizer Denmark, GlaxoSmithKline Pharma Denmark, Servier Denmark A/S, and HemoCue Denmark A/S. The ADDITION-PRO study was funded by an unrestricted grant from the European Foundation for the Study of Diabetes/Pfizer for Research into Cardiovascular Disease Risk Reduction in Patients with Diabetes (74550801), by the Danish Council for Strategic Research and by research and equipment funds from Steno Diabetes Center.

The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (www.metabol.ku.dk).

DPS

The DPS has been financially supported by grants from the Academy of Finland (117844 and 40758, 211497, and 118590 (MU); The EVO funding of the Kuopio University Hospital from Ministry of Health and Social Affairs (5254), Finnish Funding Agency for Technology and Innovation (40058/07), Nordic Centre of Excellence on ‘Systems biology in controlled dietary interventions and cohort studies, SYSDIET (070014), The Finnish Diabetes Research Foundation, Yrjö Jahnsson Foundation (56358), Sigrid Juselius Foundation and TEKES grants 70103/06 and 40058/07.

"DR's EXTRA Study"

The DR's EXTRA Study was supported by grants to Rainer Rauramaa by the Ministry of Education and Culture of Finland (627;2004-2011), Academy of Finland (102318; 123885), Kuopio University Hospital, Finnish Diabetes Association, Finnish Heart Association, Päivikki and Sakari Sohlberg Foundation and by grants from European Commission FP6 Integrated Project (EXGENESIS); LSHM-CT-2004-005272, City of Kuopio and Social Insurance Institution of Finland (4/26/ 2010).

"FIN-D2D 2007"

The FIN-D2D 2007 study was supported by funds from the hospital districts of Pirkanmaa; Southern Ostrobothnia; North Ostrobothnia; Central Finland and Northern Savo; the Finnish National Public Health Institute; the Finnish Diabetes Association; the Ministry of Social Affairs and Health in Finland; Finland's Slottery Machine Association; the Academy of Finland [grant number 129293] and Commission of the European Communities, Directorate C-Public Health [grant agreement no. 2004310].

FUSION

The FUSION study was supported by DK093757, DK072193, DK062370, and 1Z01 HG000024.

Health 2006/2008

Health 2006: The Health2006 was financially supported by grants from the Velux Foundation; The Danish Medical Research Council, Danish Agency for Science, Technology and Innovation; The Aase and Ejner Danielsens Foundation; ALK-Abello A/S, Hørsholm, Denmark, and Research Centre for Prevention and Health, the Capital Region of Denmark. Health 2008: This work was supported by the Timber Merchant Vilhelm Bang's Foundation, the Danish Heart Foundation (Grant number 07-10-R61-A1754-B838-22392F), and the Health Insurance Foundation (Helsefonden) (Grant number 2012B233).

Health 2006/2008 The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (www.metabol.ku.dk).

Inter99

The Inter99 was initiated by Torben Jørgensen (PI), Knut Borch-Johnsen (co-PI), Hans Ibsen and Troels F. Thomsen. The steering committee comprises the former two and Charlotta Pisinger. The study was financially supported by research grants from the Danish Research Council, the Danish Centre for Health Technology Assessment, Novo Nordisk Inc., Research Foundation of Copenhagen County, Ministry of Internal Affairs and Health, the Danish Heart Foundation, the Danish Pharmaceutical Association, the Augustinus Foundation, the Ib Henriksen Foundation, the Becket Foundation, and the Danish Diabetes Association.

Inter99 The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (www.metabol.ku.dk).

METSIM

The METSIM study was supported by the Academy of Finland (contract 124243), the Finnish Heart Foundation, the Finnish Diabetes Foundation, Tekes (contract 1510/31/06), and the Commission of the European Community (HEALTH-F2-2007 201681), and the US National Institutes of Health grants DK093757, DK072193, DK062370, and 1Z01 HG000024.

SDC

The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (www.metabol.ku.dk).

Vejle (Cases and controls)

The Vejle Diabetes Biobank was supported by The Danish Research Council for Independent Research.

CHARGE

FamHS

This study was supported in part by the NHLBI grant R01HL117078

AGES

The Age, Gene/Environment Susceptibility Reykjavik Study has been funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). We would like to thank the participants in the study for their contribution.

SHIP

We thank all SHIP and SHIP-TREND participants and staff members as well as the genotyping staff involved in the generation of the SNP data.

BioMe

The Mount Sinai BioMe Biobank is supported by The Andrea and Charles Bronfman Philanthropie

GENOA

Support for the Genetic Epidemiology Network of Arteriopathy (GENOA) was provided by the National Heart, Lung and Blood Institute (HL054464, HL054457, HL054481, HL087660, HL086694, HL119443). Genotyping was performed at the University of Texas Health Sciences Center (Eric Boerwinkle, Megan Grove-Gaona) and the Center for Inherited Disease Research (CIDR). We would like to thank the families that participated in the GENOA study.

HRS

is supported by the National Institute on Aging (NIA U01AG009740). The genotyping was funded separately by the National Institute on Aging (RC2 AG036495, RC4 AG039029), and the analysis was funded in part by R03 AG046389. Our genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and

final preparation of the data were performed by the Genetics Coordinating Center at the University of Washington.

ARIC

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). Funding support for "Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium" was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). The authors thank the staff and participants of the ARIC study for their important contributions.

FHS

The FHS is supported by NHLBI/NIH contract #N1-HC-25195, NIH NIDDK R01 DK078616 and K24 DK080140, and from the Boston University School of Medicine

GAPP

The GAPP study is supported by the Swiss National Science Foundation (PP00P3_133681), the Liechtenstein Government, the Commission for Technology and Innovation, the Swiss Heart Foundation, the University of Basel, the University Hospital Basel and the Hanela Foundation.

WGHS

The WGHS is supported by the National Heart, Lung, and Blood Institute (HL043851 and HL080467) and the National Cancer Institute (CA047988 and UM1CA182913) with funding for genotyping provided by Amgen.

WHI

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The WHI study was funded in part by R21 HL123677 (to NF). The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Long%20List.pdf

Cardiovascular Health Study (CHS)

This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, and U01HL130114 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 UL1TR001881, 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 Jackson Heart Study (JHS)

The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD). The authors also wish to thank the staffs and participants of the JHS.

MESA

MESA and the MESA SHARe projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by 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, UL1-TR-001420. MESA Family is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support is provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071258, R01HL071259, by the National Center for Research Resources and Grant UL1RR033176. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, 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.

Exome BP

Airwave

The Airwave Health Monitoring Study was funded by the UK Home Office (780- TETRA, 2003-2018) and is currently funded by the MRC and ESRC (MR/R023484/1) with additional support from the NIHR Imperial College Biomedical Research Centre in collaboration with Imperial College NHS Healthcare Trust. We thank all participants in the Airwave Health Monitoring Study. Paul Elliott acknowledges support from the NIHR Biomedical Research Centre at Imperial College Healthcare NHS Trust and Imperial College London, the NIHR Health Protection Research Unit in Health Impact of Environmental Hazards (HPRU-2012-10141), and the Medical Research Council (MRC) and Public Health England (PHE) Centre for Environment and Health (MR/L01341X/1). P.E. is a

UK Dementia Research Institute (DRI) professor, UK DRI at Imperial College London, funded by the MRC, Alzheimer's Society and Alzheimer's Research UK. PE is associate director of Health Data Research UK-London, funded by a consortium led by MRC. David Mosen-Ansorena is supported by the Medical Research Council [grant number MR/L01632X.1]. He Gao was funded by the NIHR Imperial College Health Care NHS Trust and Imperial College London Biomedical Research Centre

ALSPAC

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.

This study was supported by the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. Authors were supported by the UK Medical Research Council (MRC Integrative Epidemiology Unit, MC_UU_00011/1, MC_UU_00011/4, MC_UU_00011/5). NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is a programme lead in the MRC Integrative Epidemiology Unit (MC_UU_12013/3) and works within the University of Bristol NIHR Biomedical Research Centre (BRC). T.G.R. is a UKRI Innovation Research Fellow (MR/S003886/1) and supported by the Elizabeth Blackwell Institute Proximity to Discovery award (EBI 424).

ASCOT

This work was supported by Pfizer, New York, NY, USA, for the ASCOT study and the collection of the ASCOT DNA repository; by Servier Research Group, Paris, France; and by Leo Laboratories, Copenhagen, Denmark. We thank all ASCOT trial participants, physicians, nurses, and practices in the participating countries for their important contribution to the study. In particular we thank Clare Muckian and David Toomey for their help in DNA extraction, storage, and handling. We also acknowledge support from the NIHR Barts Biomedical Research Centre and Queen Mary University of London, UK.

1958BC

We are grateful for using the British 1958 Birth Cohort DNA collection. Sample collection funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. Genotyping was funded by the Wellcome Trust.

BRIGHT study

This work was supported by the Medical Research Council of Great Britain (grant number G9521010D); and by the British Heart Foundation (grant number PG/02/128). A.F.D. was supported

by the British Heart Foundation (grant numbers RG/07/005/23633, SP/08/005/25115); and by the European Union Ingenious HyperCare Consortium: Integrated Genomics, Clinical Research, and Care in Hypertension (grant number LSHM-C7-2006-037093). The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. We would also like to thank the Barts Genome Centre staff for their assistance with this project. We also acknowledge support from the NIHR Barts Biomedical Research Centre and Queen Mary University of London, UK.

CROATIA-Korcula

The CROATIA-Korcula study was 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); and the European Union framework program 6 European Special Populations Research Network project (contract LSHG-CT-2006-018947). We would like to acknowledge the invaluable contributions of the recruitment team in Korcula, the administrative teams in Croatia and Edinburgh, and the people of Korcula. Exome genotyping for CROATIA-Korcula was performed by by the Genetics Core Laboratory at the Clinical Research Facility, WGH, University of Edinburgh, Scotland

DIABNORD

We are grateful to the study participants who dedicated their time and samples to these studies. We also thank the VHS, the Swedish Diabetes Registry and Umeå Medical Biobank staff for biomedical data and DNA extraction. We also thank M Sterner, G Gremsperger and P Storm for their expert technical assistance with genotyping and genotype data preparation. The current study was funded by Novo Nordisk, the Swedish Research Council, Påhlssons Foundation, the Swedish Heart Lung Foundation, and the Skåne Regional Health Authority (all to PWF).

EGCUT

This study was supported by EU H2020 grants 692145, 676550, 654248, Estonian Research Council Grant IUT20-60, NIASC and EIT – Health and EU through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012 GENTRANSMED).

FINRISK97/02

Dr. Salomaa was supported by the Finnish Foundation for Cardiovascular Research

Kiang West Longitudinal Population Study (KWLPS) Gambia The KWLPS cohort is supported through funding from the UK Medical Research Council (MRC) and the UK Department for International Development (DFID), under the MRC/DFID Concordat agreement (MC-A760-5QX00, U105960371 and U123261351). We thank all residents of the villages of Kiang West, The Gambia, for their willingness to participate in our studies. Thanks also go to field, laboratory, clinical, data, and administrative staff at MRC Keneba, and in particular members of the Keneba Biobank team, who faciliated the collection and processing of data and samples in The Gambia that form the basis of these analyses. Thanks are further due to Josyf C Mychaleckyj and Uma Nayak (University of Virginia, USA), Matt Silver and Modou Jobe (MRC Unit The Gambia) for their advice on data analyses and finally Kerra Pearce (UCL Genomics) for coordinating the genotyping.

GS:SFHS

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 Clinical Research Facility, WGH, University of 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 are 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.

GLACIER controls

We are indebted to the study participants who dedicated their time and samples to these studies. We J Hutiainen and Å Ågren (Umeå Medical Biobank) for data organization and K Enquist and T Johansson (Västerbottens County Council) for technical assistance with DNA extraction. We also thank M Sterner, G Gremsperger and P Storm for their expert technical assistance with genotyping and genotype data preparation. The current study was funded by Novo Nordisk, the Swedish Research Council, Påhlssons Foundation, the Swedish Heart Lung Foundation, and the Skåne Regional Health Authority (all to PWF).

GoDARTS

We acknowledge the support of the Health Informatics Centre, University of Dundee for managing and supplying the anonymised data and NHS Tayside, the original data owner. We are grateful to all the participants who took part in the Go-DARTS study, to the general practitioners, to the Scottish School of Primary Care for their help in recruiting the participants, and to the whole team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses.

GRAPHIC

N.J.S. has funding from the BHF, the Transatlantic Networks of Excellence Award (12CVD02) from the Leducq Foundation and EU-FP7/2007-2013 grant HEALTH-F2-2013-601456 and is an NIHR Senior Investigator.

HELIC-MANOLIS/ HELIC-POMAK

This work was funded by the Wellcome Trust (098051) and the European Research Council (ERC-2011-StG 280559-SEPI). The MANOLIS cohort is named in honour of Manolis Giannakakis, 1978-2010. We thank the residents of the Mylopotamos villages, and of the Pomak villages, for taking part. The HELIC study has been supported by many individuals who have contributed to sample collection (including A. Athanasiadis, O. Balafouti, C. Batzaki, G. Daskalakis, E. Emmanouil, C. Giannakaki, M. Giannakopoulou, A. Kaparou, V. Kariakli, S. Koinaki, D. Kokori, M. Konidari, H. Koundouraki, D. Koutoukidis, V. Mamakou, E. Mamalaki, E. Mpamiaki, M. Tsoukana, D. Tzakou, K. Vosdogianni, N. Xenaki, E. Zengini), data entry (T. Antonos, D. Papagrigoriou, B. Spiliopoulou),

sample logistics (S. Edkins, E. Gray), genotyping (R. Andrews, H. Blackburn, D. Simpkin, S. Whitehead), research administration (A. Kolb-Kokocinski, S. Smee, D. Walker) and informatics (M. Pollard, J. Randall).

INCIPE

The study was co-sponsored by Fondazione Cassa di Risparmio di Verona, Azienda Ospedaliera di Verona, and University of Veronas

LBC1921

We thank the cohort participants and team members who contributed to these studies. 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. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Genotyping was supported by Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, and the Royal Society of Edinburgh. 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.

LBC1936

We thank the cohort participants and team members who contributed to these studies. 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. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Genotyping was supported by Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, and the Royal Society of Edinburgh. 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.

LIFELINES

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. N. Verweij is supported by the Netherlands Heart Foundation (grant NHS2010B280).

LOLIPOP

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).

NEO

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

NFBC1966/ NFBC1986

Eero Kajantie: Academy of Finland (grants 127437, 129306, 130326, 134791,263924, 315690); the Finnish Foundation for Pediatric Research; the Juho Vainio Foundation; the Novo Nordisk Foundation; the Signe and Ane Gyllenberg Foundation; the Sigrid Jusélius Foundation;and the Yrjö Jahnsson Foundation. Academy of Finland (project grants 104781, 120315, 129269, 1114194, 24300796, Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), NIHM (MH063706, Smalley and Jarvelin), Juselius Foundation, NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), ENGAGE project and grant agreement HEALTH-F4-2007-201413, EU FP7 EurHEALTHAgeing -277849, the Medical Research Council, UK (G0500539, G0600705, G1002319, G0802782, PrevMetSyn/SALVE) and the MRC, Centenary Early Career Award. The program is currently being funded by the H2020 DynaHEALTH action (grant agreement 633595), EU H2020-HCO-2004 iHEALTH Action (grant agreement 643774), EU H2020-PHC-2014 ALEC Action (grant agreement No. 633212), EU H2020-SC1-2016-2017 LifeCycle Action (grant agreement No 733206), EU H2020-MSCA-ITN-2016 CAPICE Action (grant agreement 721567), Academy of Finland EGEA-project (285547) and MRC Grant nro

MR/M013138/1. We thank the late Professor Paula Rantakallio (launch of NFBCs), and Ms Outi Tornwall and Ms Minttu Jussila (DNA biobanking).

OxBB

The Oxford Biobank is supported by the Oxford Biomedical Research Centre and part of the National NIHR Bioresource. M.I.M. is a Wellcome Trust Senior Investigator (WT098381); and a National Institute of Health Research Senior Investigator.

TWINSUK

TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

UHP

UHP (LRGP) infrastructure is financed through various (semi-) governmental funding, genotyping by BBMRI. We thank participating inhabitants of "Leidsche Rijn" for sharing their data.

ULSAM/ PIVUS

ULSAM and PIVUS: This work was funded by the Wellcome Trust (098051) and the European Research Council (ERC-2011-StG 280559-SEPI). PIVUS and ULSAM are supported by the Swedish Research Council, Swedish Heart-Lung Foundation, Swedish Diabetes Foundation and Uppsala University. The investigators express their deepest gratitude to the study participants. Genotyping and analysis was funded by the Wellcome Trust under awards WT064890, WT090532 and WT098017.

UKHLS

The UK Household Longitudinal Study was funded by grants from the Economic & Social Research Council (ES/H029745/1) and the Wellcome Trust (WT098051). UKHLS 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 survey was conducted by NatCen and the genome-wide scan data were analysed and deposited by the Wellcome Trust Sanger Institute. Information on how to access the data can be found on the Understanding Society website https://www.understandingsociety.ac.uk/.

Understanding Society Scientific Group:

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(2)

(1) Institute for Social and Economic Research

(2) University of Warwick

Million Veteran Program (MVP)

MVP: MVP-VA grant BX003360 to AMH & VA CX000982 to AMH, HL121429 to TLE and DRVE from NIH/NHLBI

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)

EPIC-CVD (extended information)

CHD case ascertainment and validation, genotyping, and clinical chemistry assays in EPIC-CVD were principally supported by grants awarded to the University of Cambridge from the EU Framework Programme 7 (HEALTH-F2-2012-279233), the UK Medical Research Council (G0800270 and MR/L003120/1) and British Heart Foundation (SP/09/002 and RG/13/13/30194), and the European Research Council (268834). We thank all EPIC participants and staff for their contribution to the study, the laboratory teams at the Medical Research Council Epidemiology Unit for sample management and Cambridge Genomic Services for genotyping, Sarah Spackman for data management, and the team at the EPIC-CVD Coordinating Centre for study coordination and administration.

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 T. van. der Schouw[149,150], 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 University, 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-InterAct

Funding for the InterAct project was provided by the EU FP6 programme (grant number LSHM_CT_2006_037197). We thank all EPIC participants and staff for their contribution to the study. We thank the lab team at the MRC Epidemiology Unit for sample management and Nicola Kerrison for data management.

EPIC-InterAct (extended information)

Claudia Langenberg(1), Robert A Scott(1), Stephen J Sharp(1), Nita G Forouhi(1), Nicola D Kerrison(1), Matt Sims(1), Debora ME Lucarelli(1), InÍs Barroso(2,3), Panos Deloukas(2), Mark I McCarthy(4,5,6), Larraitz Arriola(7,8,9), Beverley Balkau(10,11), Aurelio Barricarte(12,13,9), Heiner Boeing(14), Paul W Franks(15,16), Carlos Gonzalez(17), Sara Grioni(18), Rudolf Kaaks(19), Timothy J Key(20), Carmen Navarro(21,9,22), Peter M Nilsson(15), Kim Overvad(23,24), Domenico Palli(25), Salvatore Panico(26), J. RamÛn QuirÛs(27), Olov Rolandsson(16), Carlotta

Sacerdote(28,29), MarÌa-JosÈ S·nchez(30,9,31), Nadia Slimani(32), Anne Tjonneland(33), Rosario Tumino(34), Daphne L van der A(35), Yvonne T van der Schouw(36), Elio Riboli(37), Nicholas J Wareham(1)

(1) MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom, (2) The Wellcome Trust Sanger Institute, Cambridge, United Kingdom, (3) University of Cambridge Metabolic Research Laboratories, Cambridge, UK, (4) Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM), University of Oxford, UK, (5) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK, (6) Oxford NIHR Biomedical Research Centre, Oxford, UK, (7) Public Health Division of Gipuzkoa, San Sebastian, Spain, (8) Instituto BIO-Donostia, Basque Government, San Sebastian, Spain, (9) CIBER EpidemiologÌa y Salud P ̈blica (CIBERESP), Spain, (10) Inserm, CESP, U1018, Villejuif, France, (11) Univ Paris-Sud, UMRS 1018, Villejuif, France, (12) Navarre Public Health Institute (ISPN), Pamplona, Spain, (13) Navarra Institute for Health Research (IdiSNA) Pamplona, Spain, (14) German Institute of Human Nutrition Potsdam-Rehbruecke, Germany, (15) Lund University, Malmˆ, Sweden, (16) UmeÂ University, UmeÂ, Sweden, (17) Catalan Institute of Oncology (ICO), Barcelona, Spain, (18) Epidemiology and Prevention Unit, Milan, Italy, (19) German Cancer Research Centre (DKFZ), Heidelberg, Germany, (20) University of Oxford, United Kingdom, (21) Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain, (22) Unit of Preventive Medicine and Public Health, School of Medicine, University of Murcia, Spain, (23) Department of Public Health, Section for Epidemiology, Aarhus University, Aarhus, Denmark, (24) Aalborg University Hospital, Aalborg, Denmark, (25) Cancer Research and Prevention Institute (ISPO), Florence, Italy, (26) Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy, (27) Public Health Directorate, Asturias, Spain, (28) Unit of Cancer Epidemiology, Citta' della Salute e della Scienza Hospital-University of Turin and Center for Cancer Prevention (CPO), Torino, Italy, (29) Human Genetics Foundation (HuGeF), Torino, Italy, (30) Andalusian School of Public Health, Granada, Spain, (31) Instituto de InvestigaciÛn Biosanitaria de Granada (Granada.ibs), Granada (Spain), (32) International Agency for Research on Cancer, Lyon, France, (33) Danish Cancer Society Research Center, Copenhagen, Denmark, (34) ASP Ragusa, Italy, (35) National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands, (36) University Medical Center Utrecht, Utrecht, the Netherlands, (37) School of Public Health, Imperial College London, UK

INTERVAL (Metabolite measurement using Metabolon HD4 platform)

Participants in the INTERVAL randomised controlled trial were recruited with the active collaboration of NHS Blood and Transplant England (www.nhsbt.nhs.uk), which has supported field work and other elements of the trial. DNA extraction and genotyping was co-funded by the National Institute for Health Research (NIHR), the NIHR BioResource (http://bioresource.nihr.ac.uk/) and the NIHR [Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust] [*].The academic coordinating centre for INTERVAL was supported by core funding from: NIHR Blood and Transplant Research Unit in Donor Health and Genomics (NIHR BTRU-2014-10024), UK Medical Research Council (MR/L003120/1), British Heart Foundation (SP/09/002; RG/13/13/30194; RG/18/13/33946) and the NIHR [Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust] [*]. A complete list of the investigators and contributors to the INTERVAL trial is provided in reference [**]. The academic

coordinating centre would like to thank blood donor centre staff and blood donors for participating in the INTERVAL trial.

This work was supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome.

*The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

**Di Angelantonio E, Thompson SG, Kaptoge SK, Moore C, Walker M, Armitage J, Ouwehand WH, Roberts DJ, Danesh J, INTERVAL Trial Group. Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors. Lancet. 2017 Nov 25;390(10110):2360-2371.

EPIC Norfolk (Metabolite measurement using Metabolon HD4 platform)

The EPIC-Norfolk study (https://doi.org/10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1 and MC-UU_12015/1) and Cancer Research UK (C864/A14136). The genetics work in the EPIC-Norfolk study was funded by the Medical Research Council (MC_PC_13048). Metabolite measurements in the EPIC-Norfolk study were supported by the MRC Cambridge Initiative in Metabolic Science (MR/L00002/1) and the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement no. 115372. We are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research.