Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits

Evangelou, Evangelos
2018-10

Million Vet Program 2018, 'Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits' Nature Genetics, vol. 50, no. 10, pp. 1412-+. https://doi.org/10.1038/s41588-018-0205-x

http://hdl.handle.net/10138/251517
https://doi.org/10.1038/s41588-018-0205-x

Downloaded from Helda, University of Helsinki institutional repository.
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits

High blood pressure is a highly heritable and modifiable risk factor for cardiovascular disease. We report the largest genetic association study of blood pressure traits (systolic, diastolic and pulse pressure) to date in over 1 million people of European ancestry. We identify 535 novel blood pressure loci that not only offer new biological insights into blood pressure regulation but also highlight shared genetic architecture between blood pressure and lifestyle exposures. Our findings identify new biological pathways for blood pressure regulation with potential for improved cardiovascular disease prevention in the future.

H igh blood pressure is a leading heritable risk factor for stroke and coronary artery disease, responsible for an estimated 7.8 million deaths and 148 million disability life years lost worldwide in 2015 alone. Blood pressure is determined by complex interactions between life-course exposures and genetic background. Previous genetic association studies have identified and validated variants at 274 loci with modest effects on population blood pressure, explaining in aggregate ~3% of the trait variance. Here we report genome-wide discovery analyses of blood pressure traits—systolic blood pressure (SBP), diastolic blood pressure (DBP) and pulse pressure (PP)—in people of European ancestry drawn from UK Biobank (UKB) and the International Consortium of Blood Pressure Genome Wide Association Studies (ICBP).

We adopted a combination of a one- and two-stage study design to test common and low-frequency single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) ≥1% associated with blood pressure traits. In all, we studied over 1 million people of European descent, including replication data from the US Million Veteran Program (MVP, n = 220,520) and the Estonian Genome Centre, University of Tartu (EGCUT, n = 28,742) Biobank.

UKB is a prospective cohort study of ~500,000 richly phenotyped individuals, including blood pressure measurements, with genotyping by customized array and imputation from the Haploype Reference Consortium (HRC) panel, yielding ~7 million SNPs (imputation quality score (INFO) ≥0.1 and MAF ≥1%). We performed genome-wide association studies (GWAS) of blood pressure traits (n = 458,577 Europeans) under an additive genetic model (Supplementary Table 1a). Following linkage disequilibrium (LD) score regression, genomic control was applied to the UKB data before meta-analysis (Methods).

In addition, we performed GWAS analyses for blood pressure traits in newly extended ICBP GWAS data comprising 77 independent studies of up to 299,024 Europeans genotyped with various arrays and imputed to either the 1000 Genomes Reference Panel or the HRC platforms (Supplementary Table 1b). After quality control, we applied genomic control at the individual study level and obtained summary effect sizes for ~7 million SNPs with INFO ≥0.3 and heterogeneity Cochran’s Q statistic filtered at P ≥ 1 × 10^-4 (Methods). We then combined the UKB and ICBP GWAS results using inverse-variance-weighted fixed-effects meta-analysis (Methods), giving a total discovery sample of up to 757,601 individuals.

In our two-stage design, we attempted replication (in MVP and EGCUT; Supplementary Table 1c) of 1,062 SNPs at P < 1 × 10^-6 from discovery with concordant effect direction between UKB and ICBP, using the sentinel SNP (that is, the SNP with smallest P-value at the locus) after excluding the HLA region (chr. 6: 25–34 MB) and all SNPs in LD (r^2 ≥ 0.1) or ±500 kb from any previously validated blood pressure–associated SNPs at the 274 published loci. Our replication criteria were genome-wide significance (P < 5 × 10^-8) in the combined meta-analysis, P < 0.01 in the replication data, and concordant direction of effect between discovery and replication.

We also undertook a one-stage design to reduce type II error from the two-stage analysis. We used P < 5 × 10^-6 as threshold from the discovery meta-analysis—that is, an order of magnitude more stringent than genome-wide significance—and required an internal replication P < 0.01 in each of the UKB and ICBP GWAS analyses, with concordant direction of effect, to minimize false positive findings.

We carried out conditional analyses using GCTA, a tool for genome-wide complex trait analysis. We then explored putative functions of blood pressure–associated signals using a range of in silico resources and evaluated co-occurrence of blood pressure–associated loci with lifestyle exposures and other complex traits and diseases. Finally, we developed a genetic risk score (GRS) and assessed impact of blood pressure–associated variants on blood pressure, risk of hypertension and other cardiovascular diseases and in other ethnicities.

**Results**

We present a total of 535 novel loci (Fig. 2 and Supplementary Fig. 1): 325 claimed from the two-stage design (Supplementary Tables 2a–c) and an additional 210 claimed from our one-stage design with internal replication (Supplementary Tables 3a–c). Our two-stage design uniquely identified 121 variants, while 204 also met the one-stage criteria (Fig. 3a); many loci would not have been detected by either the one- or two-stage designs alone (Fig. 3a). For SBP, the distributions of effect sizes are similar for the one-stage (median = 0.219 mm Hg per allele; inter-quartile range (IQR) = 0.202–0.278) and two-stage loci (median = 0.224; IQR = 0.195–0.267) (P = 0.447) (Supplementary Fig. 2). Of the 210 loci found only in the one-stage analysis, 186 were also genomewide significant (P < 5 × 10^-8) in the combined meta-analysis, with all variants except 1 having concordant direction of effect between discovery and replication (Supplementary Tables 3a–c); of the remaining 24 SNPs, 10 still had concordant direction of effect.

We find support in our data for all 274 previously published blood pressure loci (Supplementary Figs. 1 and 2 and Supplementary Table 4); >95% of the previously reported SNPs covered within our data are genome-wide significant. Only 6 available SNPs did not reach Bonferroni significance, likely because they were originally

---

A full list of authors and affiliations appear at the end of the paper.
identified in non-European ancestries (for example, rs6749447, rs10474346, rs11564022) or from a gene–age interaction analysis (rs16833934). In addition, we confirmed a further 92 previously reported loci (Supplementary Table 5)\textsuperscript{9}; together with 274 previously reported loci confirmed and 535 novel loci identified here, there are 901 blood pressure–associated loci in total.

Novel genetic loci for blood pressure. Of the 535 independent novel loci, 363 SNPs were associated with one blood pressure trait, 160 with two traits and 12 with all three (Fig. 3b and Supplementary Fig. 3). Using genome-wide complex trait conditional analysis, we further identified 163, genome-wide significant, independent secondary signals with MAF \( \geq 0.1 \)% associated with blood pressure (Supplementary Table 6), of which 19 SNPs were in LD \( r^2 \geq 0.1 \) with previously reported secondary signals. This gives a total of 144 new secondary signals; hence, we now report over 1,000 independent blood pressure signals.

The estimated SNP-wide heritability \( (h^2) \) of blood pressure traits in our data was 0.213, 0.212 and 0.194 for SBP, DBP and PP, respectively, with a gain in percentage of blood pressure variance explained. For example, for SBP, percentage variance explained increased from 2.8% for the 274 previously published loci to 5.7% for SNPs identified at all 901 loci (Supplementary Table 7).

Functional analyses. Our functional analysis approach is summarized in Supplementary Fig. 4. First, for each of the 901 loci, we annotated all SNPs (based on LD \( r^2 \geq 0.8 \)) to the nearest gene within 5 kb of a SNP, identifying 1,333 genes for novel loci and 1,272 genes for known loci. Then we investigated these loci for tissue enrichment, DNase hypersensitivity site enrichment and pathway. At 66 of the 535 novel loci, we identified 97 non-synonymous SNPs, including 8 predicted to be damaging (Supplementary Table 8).

We used chromatin interaction Hi-C data from endothelial cells (HUVEC)\textsuperscript{13}, neural progenitor cells (NPC), mesenchymal stem cells (HVMSC) and tissue from the aorta (HAEC) and adrenal gland\textsuperscript{34} to identify distal associated genes. There were 498 novel loci that contained a potential regulatory SNP, and in 484 of these we identified long-range interactions in at least one of the tissues or cell...
types. We found several potential long-range target genes that did not overlap with the sentinel SNPs in the LD block. For example, the TGFBR2 gene forms a 1.2-Mb regulatory loop with SNPs in the SLC30A10 locus, and the TGFBR1 promoter forms a 100-kb loop with the COL15A1 locus (Supplementary Table 8).

Our expression quantitative trait locus (eQTL) analysis identified 60 novel loci with eQTLs in arterial tissue and 20 in adrenal (Supplementary Table 9), substantially increasing those identified in our previously published GWAS on ~140k UKB individuals. An example is SNP rs31120122, which defines an aortic eQTL affecting expression of the MED8 gene within the SRTZ2 locus. In combination with Hi-C interaction data in HVMSC, this supports a role for MED8 in blood pressure regulation, possibly mediated through expression of smooth muscle cell differentiation. Hi-C interactions provide supportive evidence for involvement of a further 36 arterial eGenes (genes whose expression is affected by the eQTLs) that are distal to their eQTLs (for example, PPXY1, ERAP2, FLRT2, ACVR2A, POU4F1).

Using DeepSEA, we found 198 SNPs in 121 novel loci with predicted effects on transcription factor binding or on chromatin marks in tissues relevant for blood pressure biology, such as vascular tissue, smooth muscle and the kidney (Supplementary Table 8).

We used our genome-wide data at a false discovery rate (FDR) < 1% to robustly assess tissue enrichment of blood pressure loci using DEPICT and identified enrichment across 50 tissues and cells (Supplementary Fig. 5a and Supplementary Table 10a). Enrichment was greater for the cardiovascular system, especially blood vessels (P = 1.5 × 10⁻⁴) and the heart (P = 2.7 × 10⁻⁵). Enrichment was high in adrenal tissue (P = 3.7 × 10⁻⁴), and, for the first time to our knowledge, we observed high enrichment in adipose tissues (P = 9.8 × 10⁻⁶) corroborated by eQTL enrichment analysis (P < 0.05) (Supplementary Fig. 6 and Supplementary Table 9). Evaluation of enriched mouse knockout phenotype terms also pointed to the importance of vascular morphology (P = 6 × 10⁻⁴) and development (P = 2.1 × 10⁻⁴) in blood pressure. With addition of our novel blood pressure loci, we identified new findings from both the gene ontology and protein–protein interaction subnetwork enrichments, which highlight the transforming growth factor-β (TGFβ) (P = 2.3 × 10⁻⁴) and related SMAD pathways (P = 7 × 10⁻⁴) (Supplementary Fig. 5b–d and Supplementary Table 10b).

We used FORGE to investigate the regulatory regions for cell type specificity from DNase I hypersensitivity sites. This showed strongest enrichment (P < 0.001) in the vasculature and highly vascularized tissues, as reported in previous blood pressure genetic studies (Supplementary Fig. 7).

**Potential therapeutic targets.** Ingenuity pathway analysis and upstream regulator assessment showed enrichment of canonical pathways implicated in cardiovascular disease, including pathways targeted by antihypertensive drugs (for example, nitric oxide signaling), and also suggested some potential new targets, such as relaxin signaling. Notably, upstream regulator analysis identified several blood pressure therapeutic targets, such as angiotensinogen, calcium channels, prostaglandin, natriuretic peptide receptor, angiotensin converting enzyme, angiotensin receptors and endothelin receptors (Supplementary Fig. 8).

---

**Fig. 2 | Manhattan plot showing the minimum P-value for the association across all blood pressure traits in the discovery stage excluding known and previously reported variants.** Discovery genome-wide association meta-analysis in 757,601 individuals excluding variants in 274 known loci. Plot gives the minimum P-value, computed using inverse-variance fixed-effects meta-analysis across SBP, DBP and PP. The y axis shows the -log₁₀ P values and the x axis shows their chromosomal positions. Horizontal red and blue lines represent the thresholds of P = 5 × 10⁻⁸ for genome-wide significance and P = 1 × 10⁻⁴ for selecting SNPs for replication, respectively. SNPs in blue are in LD (r² > 0.8) with the 325 novel variants independently replicated from the two-stage design, whereas SNPs in red are in LD (r² > 0.8) with 210 SNPs identified through the one-stage design with internal replication. Any loci in black or gray that exceed the significance thresholds were significant in the discovery meta-analysis but did not meet the criteria of replication in the one- or two-stage designs.

**Fig. 3 | Venn diagrams of novel locus results.** a, Comparison of one-stage and two-stage design analysis criteria. For all 535 novel loci, we compare the results according to the association criteria used for the one-stage and the two-stage design. 210 loci exclusively met the one-stage analysis criteria (P < 5 × 10⁻⁶ in the discovery meta-analysis (n = 757,601), P < 0.01 in UKB (n = 458,577), P < 0.01 in ICBP (n = 299,024) and concordant direction of effect between UKB and ICBP). The P-values for the discovery and the ICBP meta-analyses were calculated using inverse-variance fixed-effects meta-analysis. The P-values in UKB were derived from linear mixed modeling using the software package BOLT-LMM. Of the 325 novel replicated loci from the two-stage analysis (genome-wide significance in the combined meta-analysis, P < 0.01 in the replication meta-analysis and concordant direction of effect), 204 loci would also have met the one-stage criteria, whereas 121 were identified only by the two-stage analysis.

b, Overlap of associations across blood pressure traits. For all 535 novel loci, we show the number of loci associated with each blood pressure trait. We present the two-stage loci first, followed by the one-stage loci.
We developed a cumulative tally of functional evidence at each variant to assist in variant or gene prioritization at each locus and present a summary of the vascularly expressed genes contained within the 535 novel loci, including a review of their potential drug-gability (Supplementary Fig. 9). The overlap between blood pressure–associated genes and those associated with antihypertensive drug targets further demonstrates new genetic support for known drug mechanisms. For example, we report five novel blood pressure associations with targets of five antihypertensive drug classes (Supplementary Table 11), including the PKD2L1, SLC39A8, CACNA1C, CACNB4 and CA7 loci, targeted by potassium-sparing diuretics (amiloride), loop diuretics (bumetanide and furosemide), dihydropyridine, calcium channel blockers, non-dihydropyridines diuretics (amiloride), loop diuretics (bumetanide and furosemide), and thiazide-like diuretics (chlortalidone), respectively. Notably, in all but the last case, functional variants in these genes are the best candidates in each locus.

Concordance of blood pressure variants and lifestyle exposures. We examined association of sentinel SNPs at the 901 blood pressure loci with blood pressure–associated lifestyle traits in UKB using either the Stanford Global Biobank Engine (n = 327,302) or Gene Atlas (n = 408,455). With corrected \( P < 1 \times 10^{-8} \), we found genetic associations of blood pressure variants with daily fruit intake, urinary sodium and creatinine concentration, body mass index (BMI), weight, waist circumference, and intakes of water, caffeine and tea (\( P = 1.0 \times 10^{-7} \) to \( P = 1.3 \times 10^{-46} \)). Specifically, SNP rs13107325 in SLC39A8 is a novel locus for frequency of drinking alcohol (\( P = 3.5 \times 10^{-15} \)) and time spent watching television (\( P = 2.3 \times 10^{-11} \)), as well as being associated with BMI (\( P = 1.6 \times 10^{-15} \)), weight (\( P = 8.8 \times 10^{-16} \)) and waist circumference (\( P = 4.7 \times 10^{-11} \)) (Supplementary Table 12). We used unsupervised hierarchical clustering for the 36 blood pressure loci that showed at least one association at \( P < 1 \times 10^{-8} \) with the lifestyle-related traits in UKB (Fig. 4). The heat map summarizes the locus-specific associations across traits and highlights heterogeneous effects with anthropometric traits across the loci examined. For example, it shows clusters of associations between blood pressure–raising alleles and either increased or decreased adult height and weight. We note that some observed cross-trait associations are in opposite directions to those expected epidemiologically.

Association lookups with other traits and diseases. We further evaluated cross-trait and disease associations using GWAS Catalog 26, Phenomapper 27 and DisGeNET 28,29. The GWAS Catalog and Phenomapper search of published GWAS showed that 77 of our 535 novel loci (using sentinel SNPs or proxies with \( r^2 \geq 0.8 \)) are also significantly associated with other traits and diseases (Fig. 5 and Supplementary Table 13). We identified APOE as a highly cross-related blood pressure locus showing associations with lipid levels, cardiovascular-related outcomes and Alzheimer’s disease, highlighting a common link between cardiovascular risk and cognitive decline (Fig. 5). Other loci overlap with anthropometric traits, including BMI, birth weight and height (Fig. 5), and with DisGeNET terms related to lipid measurements, cardiovascular outcomes and obesity (Fig. 6).

We did lookups of our sentinel SNPs in \(^{1}H\) NMR lipidomics data on plasma (n = 2,022) and data from the Metabolon platform (n = 1,941) in the Airwave Study 30, and used Phenomapper to test SNPs against published significant (\( P < 5 \times 10^{-8} \)) genome- vs. metabolome-wide associations in plasma and urine (Methods). Ten blood pressure SNPs showed association with lipid particle metabolites and a further 31 SNPs (8 also on Phenomapper) showed association with metabolites on the Metabolon platform, highlighting lipid pathways, amino acids (glycine, serine and glutamine),

**Fig. 4 | Association of blood pressure loci with lifestyle traits.** Plot shows unsupervised hierarchical clustering of blood pressure loci based on associations with lifestyle-related factors. For the sentinel SNP at each blood pressure locus (x axis), we calculated the \(-\log(P) \times \text{sign}(\beta)\) (aligned to blood pressure–raising allele) as retrieved from the Gene Atlas catalog. The \( P \)-values in Gene Atlas were calculated applying linear mixed models. Blood pressure loci and traits were clustered according to the Euclidean distance among \(-\log(P) \times \text{sign}(\beta)\). Red squares indicate direct associations with the trait of interest and blue squares inverse associations. Only SNPs with at least one association at \( P < 1 \times 10^{-8} \) with at least one of the traits examined are annotated in the heat-map. All 901 loci are considered, both known and novel; novel loci are in bold.
Articles

We also found that the GRS was associated with increased, sex-adjusted risk of incident stroke, myocardial infarction and all cardiovascular outcomes, comparing top and bottom deciles of the GRS distribution (Supplementary Table 17). Sensitivity analyses in the independent Airwave cohort gave similar results (Supplementary Table 17).

In addition, we observed three-fold sex-adjusted higher risk of hypertension (OR 3.34; 95% CI 3.24 to 3.45; \( P = 1.7 \times 10^{-100} \)) between the top and bottom deciles of the GRS in UKB (Fig. 7a). Sensitivity analyses in the independent Airwave cohort gave similar results (Supplementary Table 17). We also found that the GRS was associated with increased, sex-adjusted risk of incident stroke, myocardial infarction and all incident cardiovascular outcomes, comparing top and bottom deciles of the GRS distribution, with odds ratios of 1.47 (95% CI 1.35 to 1.59, \( P = 1.1 \times 10^{-20} \), 1.50 (95% CI 1.28 to 1.76, \( P = 8.0 \times 10^{-7} \)) and 1.52 (95% CI 1.26 to 1.82, \( P = 7.7 \times 10^{-4} \), respectively (Fig. 7b and Supplementary Table 16).

Extending analyses to other ancestries. We examined associations with blood pressure of both individual SNPs and the GRS among unrelated individuals of African and South Asian descent in UKB for the 901 known and novel loci. Compared to Europeans, 62.4%, 62.5% and 64.8% of the variants among Africans (\( n = 6,970 \)) and South Asians (\( n = 8,827 \)). Blood pressure variants in combination were associated with 6.1 mm Hg (95% CI 4.5 to 7.7; \( P = 4.9 \times 10^{-14} \)) and 7.4 mm Hg (95% CI 6.0 to 8.7; \( P = 1.7 \times 10^{-100} \)) higher, sex-adjusted mean SBP among Africans and South Asians, respectively, comparing top and bottom quintiles of the GRS distribution (Supplementary Table 19a,b).

Discussion

Our study of over 1 million people offers an important step forward in understanding the genetic architecture of blood pressure. We identified over 1,000 independent signals at 901 loci for blood pressure traits, and the 535 novel loci more than triples the number of blood pressure loci and doubles the percentage variance explained, illustrating the benefits of large-scale biobanks. By explaining 27% of the estimated heritability for blood pressure, we make major inroads into the missing heritability influencing blood pressure in the population\(^1\). The novel loci open the vista of entirely new biology and highlight gene regions in systems not previously implicated in blood pressure regulation. This is particularly timely as global prevalence of people with SBP over 110–115 mm Hg, above which cardiovascular risk increases in a continuous graded manner, now exceeds 3.5 billion, of whom over 1 billion are within the treatment range\(^2,3\).

Our functional analysis highlights the role of the vasculature and associated pathways in the genetics underpinning blood pressure traits. We show a role for several loci in the TGFβ pathway, including SMAD family genes and the TGFβ gene locus itself. This pathway affects sodium handling in the kidney and ventricular remodeling, while plasma levels of TGFβ have recently been correlated with hypertension (Fig. 8)\(^4,5\). The activin A receptor type 1C (ACVR1C) gene mediates the effects of the TGFβ family of signaling molecules. A blood pressure locus contains the bone morphogenetic protein 2 (BMP2) gene in the TGFβ pathway, which prevents growth suppression in pulmonary arterial smooth muscle cells and is associated with pulmonary hypertension\(^6\). Another blood pressure locus includes the Kruppel-like family 14 (KLF14) gene of transcription.
factors, which is induced by low levels of TGF\(\beta\) receptor II gene expression and which has also been associated with type 2 diabetes, hypercholesterolemia and atherosclerosis.\textsuperscript{37}

Our analysis shows enrichment of blood pressure gene expression in the adrenal tissue. Autonomous aldosterone production by the adrenal glands is thought to be responsible for 5–10% of all hypertension, rising to ~20% amongst people with resistant hypertension.\textsuperscript{38} Some of our novel loci are linked functionally to aldosterone secretion.\textsuperscript{39,40} For example, the CTNNB1 locus encodes \(\beta\)-catenin, the central molecule in the canonical Wnt signaling system, required for normal adrenocortical development.\textsuperscript{41,42} Somatic adrenal mutations of this gene that prevent serine/threonine phosphorylation lead to hypertension through generation of aldosterone-producing adenomas.\textsuperscript{43,44}

Our novel loci also include genes involved in vascular remodeling, such as vascular endothelial growth factor A (VEGFA), the product of which induces proliferation, migration of vascular endothelial cells and stimulates angiogenesis. Disruption of this gene in mice resulted in abnormal embryonic blood vessel formation, while allelic variants of this gene have been associated with microvascular complications of diabetes, atherosclerosis and the antihypertensive response to enalapril.\textsuperscript{45} We previously reported a

---

**Fig. 6 | Association of blood pressure loci with other traits and diseases.**\(a, b\), Plots showing overlap between variants associated to traits (\(a\)) and diseases (\(b\)) in the manually curated version of the DisGeNET database, and all variants in LD \(r^2 > 0.8\) with the known (red bars) SNPs from the 274 published loci, and all (turquoise bars) blood pressure variants from all 901 loci. Numbers on top of the bars denote the number of SNPs included in DisGeNET for the specific trait or disease. Traits or diseases with an overlap of at least five variants in LD with all markers are shown. The y axis shows the percentage of variants associated with the diseases that is covered by the overlap. For the sake of clarity, the DisGeNET terms for blood pressure and hypertension are not displayed, whereas the following diseases have been combined: coronary artery disease (CAD), coronary heart disease (CHD) and myocardial infarction (MI); prostate and breast carcinoma; Crohn’s and inflammatory bowel diseases (IBD).

---

**Fig. 7 | Relationship of deciles of the genetic risk score (GRS) based on all 901 loci with blood pressure, risk of hypertension and cardiovascular disease in UKB.**\(a, b\), Plots show sex-adjusted mean SBP and odds ratios of hypertension (\(a\); \(n = 364,520\)) and odds ratios of incident cardiovascular disease (CVD), myocardial infarction (MI) and stroke (\(b\); \(n = 392,092\)), comparing each of the upper nine GRS deciles with the lowest decile; dotted lines represent the upper 95% confidence intervals.
fibroblast growth factor (FGF5) gene locus in association with blood pressure. Here, we additionally identify a new blood pressure locus encoding FGF9, which is linked to enhanced angiogenesis and vascular smooth muscle cell differentiation by regulating VEGFA expression.

Several of our novel loci contain lipid-related genes, consistent with the observed strong associations among multiple cardio-metabolic traits. For example, the apolipoprotein E gene (APOE) encodes the major apoprotein of the chylomicron. Recently, APOE serum levels have been correlated with SBP in population-based studies and in murine knockout models; disruption of this gene led to atherosclerosis and hypertension. A second novel blood pressure locus contains the low-density lipoprotein receptor-related protein 4 (LRP4) gene, which may be a target for APOE and is strongly expressed in the heart in mice and humans. In addition, we identified the adrenomedullin (ADM) gene as a blood pressure locus. Adrenomedullin is known to exert differential effects on blood pressure in the brain (vasopressor) and the vasculature (vasodilator). In addition, a locus containing Rho guanine nucleotide exchange factor 25 (ARHGEF25) gene generates a factor that interacts with Rho GTPases involved in contraction of vascular smooth muscle and regulation of responses to angiotensin II.

We evaluated the 901 blood pressure loci for extant or potentially druggable targets. Loci encoding MARK3, PDGFC, TRHR, ADORA1, GABRA2, VEGFA and PDE3A are within systems with existing drugs not currently linked to a known antihypertensive mechanism; they may offer repurposing opportunities, for example, detection of SLC5A1 as the strongest repurposing candidate in a new blood pressure locus targeted by the type 2 diabetes drug canagliflozin. This is important as between 8–12% of patients with hypertension exhibit resistance or intolerance to current therapies and repositioning of a therapy with a known safety profile may reduce development costs.

This study strengthens our previously reported GRS analysis indicating that all blood pressure elevating alleles combined could increase...
SBP by 10 mm Hg or more across quintiles or deciles of the population distribution, substantially increasing risk of cardiovascular events. We previously suggested that genotyping blood pressure elevating variants in the young could lead to targeted lifestyle intervention in early life that might attenuate the blood pressure rise at older ages.

We identified several blood pressure–associated loci that are also associated with lifestyle traits, suggesting shared genetic architecture between blood pressure and lifestyle exposures. We adjusted our blood pressure GWAS analyses for BMI to control for possible confounding effects, though we acknowledge the potential for collider bias. Nonetheless, our findings of possible genetic overlap between loci associated with blood pressure and lifestyle exposures could support renewed focus on altering specific lifestyle measures known to affect blood pressure.

Despite smaller sample sizes, we observed high concordance with direction of effects on blood pressure traits of blood pressure variants in Africans (>62%) and South Asians (>72%). The GRS analyses show that, in combination, blood pressure variants identified in European analyses are associated with blood pressure in non-European ancestries, though effect sizes were 30–40% smaller.

Our use of a two- and one-stage GWAS design illustrates the value of this approach to minimize the effects of stochastic variation and heterogeneity. The one-stage approach included signals that had independent and concordant support (P < 0.01) from both UKB and ICBP, reducing the impact of winners’ curse on our findings. Indeed, all but two of the 210 SNPs discovered in the one-stage analysis reach \( P < 5 \times 10^{-8} \) in either UKB or ICBP. To further minimize the risk of reporting false positive loci within our one-stage design, we set a stringent overall discovery meta-analysis \( P \)-value threshold of \( P < 5 \times 10^{-8} \), an order of magnitude smaller than a genome-wide significance \( P \)-value, in line with thresholds recommended for whole genome sequencing. We found high concordance in direction of effects between discovery data in the one-stage approach and the replication resources, with similar distributions of effect sizes for the two approaches. We note that 24 of the one-stage SNPs that reached \( P < 5 \times 10^{-8} \) in discovery failed to reach genome-wide significance \( (P < 5 \times 10^{-8}) \) in the combined meta-analysis of discovery and replication resources, and hence may still require further validation in future, larger studies.

The new discoveries reported here more than triple the number of loci for blood pressure to a total of 901 and represent a substantial advance in understanding the genetic architecture of blood pressure. The identification of many novel genes across the genome could partly support an omnigenic model for complex traits, where genome-wide association of multiple interconnected pathways is observed. However, our strong tissue enrichment shows particular relevance to the biology of blood pressure and cardiovascular disease, suggesting trait-specificity, which could argue against an omnigenic model. Our confirmation of the impact of these variants on blood pressure level and cardiovascular events, coupled with identification of shared risk variants for blood pressure and adverse lifestyle, could contribute to an early life precision medicine strategy for cardiovascular disease prevention.

References


NIH. A.M.H. was supported by VA Award #101BX00360. C.J.O. was supported by VA Boston Healthcare, Section of Cardiology and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School. The MRC/BHF Cardiovascular Epidemiology Unit is supported by the UK Medical Research Council (MR/L003120/1), British Heart Foundation (RG/13/13/30194) and UK National Institute for Health Research Cambridge Biomedical Research Centre. J. Danesh is a British Heart Foundation Professor and NHIR Senior Investigator. L.V.H. holds a GlaxoSmithKline/British Lung Foundation Chair in Respiratory Research. P.E. acknowledges support from the NHIR Biomedical Research Centre at Imperial College Healthcare NHS Trust and Imperial College London, the NHIR 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 at Imperial College London, funded by the MRC, Alzheimer’s Society and Alzheimer’s Research UK. He is also associate director of Health Data Research—UK London, funded by a consortium led by the Medical Research Council. M.J.C. was funded by the National Institute for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Biomedical Research Center at Barts and The London School of Medicine and Dentistry. M.J.C. is a National Institute for Health Research (NIHR) senior investigator, and this work is funded by the MRC eMedLab award to M.J.C. and M.R.B. and by the NIHR Biomedical Research Centre at Barts.

This research has been conducted using the UK Biobank Resource under application numbers 236 and 10035. This research was supported by the British Heart Foundation (grant SP13/230111). Large-scale comprehensive genotyping of UK Biobank for cardiometabolic traits and diseases: UK CardioMetabolic Consortium (UKCMC). Computing: This work was enabled using the computing resources of (i) the UK Medical Informatics aggregation, integration, visualisation and analysis of large, complex data (UK MED Bio) which is supported by the Medical Research Council (grant number MR/L01632X/1), and (ii) the MRC eMedLab Medical Informatics Infrastructure, supported by the Medical Research Council (grant number MR/L016311/1). The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the US Department of Health and Human Services. C.P.K. is an employee of the US Department of Veterans Affairs. Opinions expressed in this paper are those of the authors and do not necessarily represent the opinion of the Department of Veterans Affairs or the United States Government.

Author contributions


Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0205-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to P.E. or M.J.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1Department of Epidemiology and Biostatistics, Imperial College London, London, UK. 2Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece. 3William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. 4National Institute for Health Research, Barts Cardiovascular Biomedical Research Center, Queen Mary University of London, London, UK. 5MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. 6Department of Mathematics, Imperial College London, London, UK. 7Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA. 8Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center; Tennessee Valley Health Systems VA, Nashville, TN, USA. 9Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA, USA. 10Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA, USA. 11Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA, USA. 12Division of Aging, Department of Medicine, Brigham and Women's Hospital; Department of Medicine, Harvard Medical School, Boston, MA, USA. 13Atlanta VAMC and Emory Clinical Cardiovascular Research Institute, Atlanta, GA, USA. 14VA Palo Alto Health Care System, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, USA. 15Nephrology Section, Memphis VA Medical Center and University of Tennessee Health Science Center, Memphis, TN, USA. 16Estonian Genome Center, University of Tartu, Tartu, Estonia. 17Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. 18Department of Clinical Sciences, Malmö, Lund University, Malmö, Sweden. 19MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, Scotland, UK. 20Department of Neurology, Bordeaux University Hospital, Bordeaux, France. 21Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, CHU Bordeaux, Bordeaux, France. 22Laboratory of Genetics and Genomics, NIA/NIH, Baltimore, MD, USA. 23Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA. 24Hunter Medical Research Institute and Faculty of Health, University of Newcastle, New Lambton Heights, New South Wales, Australia. 25Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. 26Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. 27MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 28MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK. 29Department of Twin Research and Genetic Epidemiology, Kings College London, London, UK. 30NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London, UK. 31Wellcome Trust Sanger Institute, Hinxton, UK. 32Division of Molecular and Clinical Medicine, School of Medicine, University of Dundee, Dundee, UK. 33Department of Pharmacy, COMSATS Institute of Information Technology, Abbottabad, Pakistan. 34Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada. 35Institut universitaire de cardiologie et de pneumologie de Québec-Université Laval, Quebec City, Quebec, Canada. 36Cardiovascular Research Center and Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA. 37University of Groningen, University Medical Center Groningen, Department of Cardiology, Groningen, the Netherlands. 38University of Lille, Inserm, Centre Hosp. Univ. Lille, Institut Pasteur de Lille, UMR1167 – RID-AGE – Risk factors and molecular determinants of aging-related diseases, Epidemiology and Public Health Department, Lille, France. 39University of Dundee, Ninewells Hospital & Medical School, Dundee, UK. 40Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. 41Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. 42Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. 43Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands. 44Impperial Clinical Trials Unit, London, UK. 45School of Medicine, University College Dublin, Dublin, Ireland. 46Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland. 47International Centre for Circulatory Health, Imperial College London, London, UK. 48Center for Statistical Genetics, Department of Biostatistics, SPH II, Washington Heights, Ann Arbor, MI, USA. 49Genetic Epidemiology Unit, Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands. 50Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. 51Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK. 52Sema4, a Mount Sinai venture, Stamford, CT, USA. 53Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Milan, Italy. 54Department of Health Sciences, University of Leicester, Leicester, UK. 55Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle,
Methods

UK Biobank (UKB) data. We performed a genome-wide association study (GWAS) analysis in 458,577 UKB participants1 (Supplementary Note 1). These consist of 408,951 individuals from UKB genotyped at 823,927 variants with a custom Affymetrix UK Biobank Array chip and 49,626 individuals genotyped at 807,411 variants with a custom Affymetrix UK BiLEVE Axiom Array chip from the UK BiLEVE study2, which is a subset of UKB. SNPs were imputed centrally by UKB using a reference panel that merged the UK10k and 1000 Genomes Phase 3 panel as well as the Haplotype Reference Consortium (HRC) panel3. For the current analysis, only SNPs imputed from the HRC panel were considered.

UKB phenotypic data. Following quality control (QC) (Supplementary Note 1), we restricted our data to a subset of post-QC individuals of European ancestry combining information from self-reported and genetic data (Supplementary Note 1), resulting in a maximum of n = 458,577 individuals (Fig. 1 and Supplementary Fig. 12).

Three blood pressure traits were analyzed: systolic blood pressure (SBP), diastolic blood pressure (DBP) and pulse pressure (PP; the difference between SBP and DBP). We calculated the mean SBP and DBP values from two automated (n = 418,755) or two manual (n = 25,886) blood pressure measurements. For individuals with one manual and one automated blood pressure measurement (n = 13,521), we used the mean of these two values. For individuals with only one available blood pressure measurement (n = 413), we used this single value. After calculating blood pressure values, we adjusted for medication use by adding 15 and 10 mm Hg to SBP and DBP, respectively, for individuals reported to be taking blood pressure–lowering medication (n = 94,289). Descriptive summary statistics are shown in Supplementary Table 1a.

UKB analysis models. For the UKB GWAS, we performed linear mixed model (LMM) association testing under an additive genetic model of the three (untransformed) continuous, medication-adjusted blood pressure traits (SBP, DBP, PP) for all measured and imputed genetic variants in dosage format using the BOLT-LMM (v2.3) software6. We also calculated the estimated SNP-wide heritability (h²) in our data. Within the association analysis, we adjust for the following covariates: sex, age, age², BMI and a binary indicator variable for UKB vs. UK BiLEVE to account for the different genotyping chips. The analysis of all HRC-following covariates: sex, age, age², BMI and a binary indicator variable for UKB vs. ICBD.

Following quality control (QC) (Supplementary Note 1), we excluded all published blood pressure variants (see below) in the discovery GWAS that required proxies for replication are shown in Supplementary Table 23. The combined meta-analysis of both the discovery data (n = 757,601) and replication meta-analysis (max n = 249,262) provided a maximum sample size of n = 1,006,863.

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

1. The sentinel SNP shows significance (P < 1 × 10⁻⁶) in the discovery meta-analysis of UKB and ICBP, with concordant direction of effect between UKB and ICBP.
2. The sentinel SNP is genome-wide significant (P < 5 × 10⁻⁸) in the combined meta-analysis of discovery and replication (MVP and EGCUT) (republication, described below).
3. The sentinel SNP shows support (P < 0.01) in the replication meta-analysis of MVP and EGCUT alone (Supplementary Note 1).
4. The sentinel SNP has concordant direction of effect between the discovery and the replication meta-analyses.
5. The sentinel SNP must not be located within any of the 274 previously reported loci described above.

The primary replicated trait was then defined as the blood pressure trait with the most significant association from the combined meta-analysis of discovery and replication (in the case where a SNP was replicated for more than one blood pressure trait).

Identification of novel signals using two-stage and one-stage study designs. To identify novel signals of association with blood pressure, two complementary study designs (which we term here “two-stage design” and “one-stage design”) were implemented in order to maximize the available data and minimize reporting of false positive associations.

Overview of two-stage design. All of the following criteria had to be satisfied for a signal to be reported as a novel signal of association with blood pressure using our two-stage design:

1. The sentinel SNP shows significance (P < 1 × 10⁻⁶) in the discovery meta-analysis of UKB and ICBP, with concordant direction of effect between UKB and ICBP.
2. The sentinel SNP is genome-wide significant (P < 5 × 10⁻⁸) in the combined meta-analysis of discovery and replication (MVP and EGCUT) (republication, described below).
3. The sentinel SNP shows support (P < 0.01) in the replication meta-analysis of MVP and EGCUT alone (Supplementary Note 1).
4. The sentinel SNP has concordant direction of effect between the discovery and the replication meta-analyses.
5. The sentinel SNP must not be located within any of the 274 previously reported loci described above.

The primary replicated trait was then defined as the blood pressure trait with the most significant association from the combined meta-analysis of discovery and replication (in the case where a SNP was replicated for more than one blood pressure trait).

Replication analysis. We considered SNPs with MAF ≥1% for independent replication in MVP (max n = 220,520) and in EGCUT Biobank (n = 28,742) (Supplementary Note 1). This provides a total of n = 249,262 independent samples for individuals of European descent available for replication. Additional information on the analyses of the two replication datasets is provided in the Supplementary Note 1 and Supplementary Table 1c. The two datasets were then combined using fixed-effects inverse-variance-weighted meta-analysis, and summary results for all traits were obtained for the replication meta-data analysis.

Combined meta-analysis of discovery and replication meta-analyses. The meta-analyses were performed within METAL software6 using fixed-effects inverse-variance-weighted meta-analysis (Supplementary Note 1). The variants from the discovery GWAS that required proxies for replication are shown in Supplementary Table 22. The combined meta-analysis of both the discovery data (n = 757,601) and replication meta-analysis (max n = 249,262) provided a maximum sample size of n = 1,006,863.

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

1. The sentinel SNP has P < 5 × 10⁻⁶ in the discovery (UKB + ICBP) meta-analysis.
2. The sentinel SNP shows support (P < 0.01) in the UKB GWAS alone.
3. The sentinel SNP shows support (P < 0.01) in the ICBP GWAS alone.
4. The sentinel SNP has concordant direction of effect between UKB and ICBP datasets.
5. The sentinel SNP must not be located within any of the 274 previously reported loci described above (Supplementary Table 4) or the recently reported non-replicated loci from Hoffman et al.9 (Supplementary Table 23).

We selected the one-stage P-value threshold to be an order of magnitude more stringent than a genome-wide significance P-value so as to ensure robust results.
and to minimize false positive findings. The threshold of $P < 5 \times 10^{-8}$ has been proposed as a more conservative statistical significance threshold, for example, for whole-genome sequencing-based studies.

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

### Conditional analysis

We performed conditional analyses using the GWAS discovery meta-analysis data in order to identify any independent secondary signals in addition to the sentinel SNPs at the 901 loci. We used two different methodological approaches, each using the genome-wide complex traits analysis (GCTA) software:

1. Full genome-wide conditional analysis with joint multivariate analysis and stepwise model selection across all three blood pressure traits, and
2. Locus-specific conditional analysis for the primary blood pressure trait, conditioning on the sentinel SNPs within each locus (Supplementary Note 1).

For robustness, secondary signals are only reported if obtained from both genome-wide approaches. All secondary signals with MAF $\geq 1\%$ were selected at genome-wide significance level and confirmed to be pairwise-LD-independent ($r^2 < 0.1$), as well as not being in LD with any of the published or sentinel SNPs at any of the 901 blood pressure–associated loci ($r^2 < 0.1$). In all cases, the UKB data was used as the reference genetic data for LD calculation, restricted to individuals of European ancestry only.

### Variant-level functional analyses

We used an integrative bioinformatics approach to collate functional annotation at both the variant level (for each sentinel SNP within a given blood pressure–raising variant) and the gene level (using SNPs in LD $r^2 > 0.8$ with the sentinel SNPs). At the variant level, we use Variant Effect Predictor (VEP) to obtain comprehensive characterization of variants, including consequence (for example, downstream or noncoding transcript exon), information on nearest genomic features and, where applicable, amino acid substitution functional impact, based on SIFT and PolyPhen. The biomart R package is used to further annotate the nearest annotated genes.

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

We annotated nearest genes, eGenes (genes whose expression is affected by eQTLs) and Hi-C interactors with HUVEC, HVMSC and HAEC expression from the Fantom5 project. Genes that had higher than median expression levels in the given cell types were indicated as expressed.

To identify SNPs in the novel loci that have a noncoding functional effect (influence binding of transcription factors or RNA polymerase, or influence DNase hypersensitivity sites or histone modifications), we used DeepSEA, a deep learning algorithm, which learned the binding and modification patterns of ~900 cell-factor combinations. A change of $> 0.1$ in the binding score predicted by DeepSEA for the reference and alternative alleles was used as cut-off to find alleles with noncoding functional effect (Supplementary Note 1).

We annotated nearest genes of regulatory SNPs using long-range chromatin interaction (Hi-C) data from HUVECs, aorta, adrenal glands, neural progenitors and mesenchymal stem cells, which are tissues and cell types that are considered relevant for regulating blood pressure. We find the most significant promoter interactions for all potential regulatory SNPs (RegulomeDB score $\geq 5$) in LD ($r^2 > 0.8$) with our novel sentinel SNPs and published SNPs, and choose the interactors with the SNPs of highest regulatory potential to annotate the loci. We then performed overall enrichment testing across all loci. First, we used DEPICT2 (Data-driven Expression Prioritized Integration for Complex Traits) to identify tissues and cells that are highly expressed at genes within the blood pressure loci (Supplementary Note 1). Second, we used DEPICT to test for enrichment in other tissues associated with cardiovascular disease (Supplementary Note 1). We report significant enrichments with a false discovery rate < 0.01. The variants tested were (i) the 357 published blood pressure–associated SNPs at the time of analysis and (ii) a set including all (published and novel) variants with novel SNPs filtered by highest significance, $P < 5 \times 10^{-8}$. Furthermore, to investigate cell type specific enrichment within DNase 1 sites, we used FORGE, which tests for enrichment of SNPs within DNase 1 sites in 123 cell types from the Epigenomics Roadmap Project and ENCODE (Supplementary Note 1). Two analyses were compared (i) using published SNPs only and (ii) using sentinel SNPs at all 901 loci, in order to evaluate the overall tissue specific enrichment of blood pressure–associated variants.

### Genetic risk scores and percentage of variance explained

We calculated a weighted genetic risk score (GRS) (Supplementary Table S24) to provide an estimate of the combined effect of the blood pressure–raising variants on blood pressure and risk of hypertension and applied this to the UKB data (Supplementary Note 1). Our analysis included 423,713 unrelated individuals of European ancestry, of whom 392,092 individuals were free of cardiovascular events at baseline.

We assessed the association of the continuous GRS variable on blood pressure and with the risk of hypertension, with and without adjustment for sex. We then compared blood pressure levels and risk of hypertension, respectively, for individuals in the top vs. bottom quintiles of the GRS distribution. Similar analyses were performed for the top vs. bottom deciles of the GRS distribution. All analyses were restricted to the 392,092 unrelated individuals of European ancestry from UKB. As a sensitivity analysis to assess for evidence of bias in the UKB results, we also carried out similar analyses in Airwave, an independent cohort of $n = 14,004$ unrelated participants of European descent (Supplementary Note 1).

We calculated the association of the GRS with cardiovascular disease in unrelated participants in UKB data on the basis of self-reported medical history and linkage to hospitalization and mortality data (Supplementary Table 25). We use logistic regression with binary outcome variables for composite incident cardiovascular events (Supplementary Note 1), incident myocardial infarction and incident stroke (using the algorithmic UKB definitions) and GRS as explanatory variable (with and without sex adjustment).

We also assessed the association of this GRS with blood pressure in unrelated Africans ($n = 6,970$) and South Asians ($n = 8,827$) from the UKB to see whether blood pressure–associated variants identified from GWAS predominantly in Europeans are also associated with blood pressure in populations of non-European ancestry.

We calculated the percentage of variance in blood pressure explained by genetic variants using the independent Airwave cohort ($n = 14,004$) (Supplementary Note 1). We considered three different levels of the GRS: (i) all pairwise-independent, LD-filtered ($r^2 < 0.1$) published SNPs within the known loci; (ii) all known SNPs and sentinel SNPs at novel loci; and (iii) all independent signals at all 901 known and novel loci including the 163 secondary SNPs.

### Ethics statement

The UKB study has approval from the North West Multi-Centre Research Ethics Committee. Any participants from UKB who withdrew consent have been removed from our analysis. Each cohort within the ICBP meta-analysis, as well as our independent replication cohorts of MVP and EGCUT, had ethical approval locally. More information on the participating cohorts is available in the Supplementary Note 2.

### Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

The genetic and phenotype UKB data are available upon application to the UK Biobank (https://www.ukbiobank.ac.uk). ICBP summary data can be accessed through request to the ICBP steering committee. Contact the corresponding authors to apply for access to the data. The UKB+ICBP summary GWAS discovery data can be accessed by request to the corresponding authors and will be available via LDHub (http://ldsc.broadinstitute.org/lphub). All replication data generated during this study are included in the published article. For example, association
results of look-up variants from our replication analyses and the subsequent combined meta-analyses are contained within the Supplementary Tables.

**References**


Experimental design

1. Sample size
   Describe how sample size was determined.
   Of the total ~500,000 subjects from UK Biobank, we analysed 458,577 subjects which passed QC of the genetic data, were of European ancestry, and met our phenotypic data QC requirements for availability of blood pressure data and covariates.
   To maximize sample size in the discovery, we recruited an additional 148,890 samples from 23 new cohorts, in addition to the 150,134 samples already existing from the 54 cohorts within the published ICBP-1000G project, giving a total sample size of 299,024 in the ICBP meta-analysis.
   Hence a total discovery sample size of N=757,601.
   Our combined meta-analysis sample size was N=1,006,863 after combining with the data from the replication cohorts (N=220,520 from MVP and N=28,742 from EGCLUT).

2. Data exclusions
   Describe any data exclusions.
   Within UK Biobank, we excluded samples according to both genetic data quality control (QC) and phenotypic data QC. From genetic data QC, we excluded 968 subjects listed as QC outliers for heterozygosity or missingness within the centrally provided UK Biobank sample QC files, and 378 individuals with sex discordance between the phenotypic and genetically inferred sex. We also restricted to subjects of European ancestry, according to both self-reported ethnicity status and ancestry clustering using PCA data. For phenotypic QC, we excluded any subjects with no BP measurements, missing BMI covariate data, pregnant (N=372) and those individuals who had withdrawn consent (N=36).
   Similar sample QC was performed at study level within each of the ICBP and replication cohorts.

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   Novel loci identified from our 2-stage approach were robustly replicated using independent replication datasets. Novel loci identified from our 1-stage approach met our criteria for internal replication by showing significant support within each of the UKB and ICBP GWAS datasets separately.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   N/A for GWAS

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   N/A for GWAS
   (Note data collection of UK Biobank was done centrally, not performed by us)

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.
6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

<table>
<thead>
<tr>
<th>n/a</th>
<th>Confirmed</th>
</tr>
</thead>
</table>

- [ ] The exact sample size \((n)\) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- [ ] A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- [ ] A statement indicating how many times each experiment was replicated
- [ ] The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- [ ] A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- [ ] The test results (e.g. \(P\) values) given as exact values whenever possible and with confidence intervals noted
- [ ] A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- [ ] Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

7. Software

Policy information about availability of computer code

Describe the software used to analyze the data in this study.

For the Primary GWAS analysis, BOLT-LMM software v2.3 was used for running an association analysis using linear mixed modelling; then METAL software was used for all meta-analyses with a fixed effects inverse variance weighted meta-analysis approach. We used R software for any general statistical analyses, for secondary analyses (e.g. variance explained analyses, risk score analyses) and for producing plots in the figures. We used PLINK software for LD calculations of variants. For the bioinformatics analyses, specific software was used for each different analysis. Each method and the software used is described in the Online Methods, the Supplementary Methods and also summarised in Supplementary Figure 3. For example, the Variant Effect Predictor (VEP) tool is used for variant annotation; DEPICT software is used for enrichment testing, etc.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

8. Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company:

N/A (not labwork)

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species):

N/A
10. Eukaryotic cell lines
   a. State the source of each eukaryotic cell line used. N/A
   b. Describe the method of cell line authentication used. N/A
   c. Report whether the cell lines were tested for mycoplasma contamination. N/A
   d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals
   Provide details on animals and/or animal-derived materials used in the study.
   N/A

Policy information about studies involving human research participants

12. Description of human research participants
   Describe the covariate-relevant population characteristics of the human research participants.
   Summary descriptives of UKB, ICBP and MVP/EGCUT individuals are provided in Sup Tables 1a, 1b and 1c, respectively, showing: blood pressure measurements as the phenotype; age, sex and BMI values as covariates; hypertension status and the use of BP-lowering medication.