Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nature genetics, 49 (3). pp. 403-415. ISSN 1061-4036 DOI: https://doi.org/10.1038/ng.3768

Downloaded from: http://researchonline.lshtm.ac.uk/3429633/

DOI: https://doi.org/10.1038/ng.3768
Usage Guidelines:

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by-nc-nd/2.5/
Discovery and validation of 107 blood pressure loci from UK Biobank offers novel biological insights into cardiovascular risk

Short title: Novel blood pressure loci in UK Biobank

The UK Biobank Cardio-metabolic Traits Consortium Blood Pressure Working Group.

Helen R Warren1,2*, Evangelos Evangelou3,4*, Claudia P Cabrera1,2*, He Gao3,5*, Meixia Ren1,6*, Borbala Mifsud14*, Ioanna Ntalla1, Praveen Surendran7, Chunyu Liu8-10, James P Cook11, Aldi Kraja12, Fotios Drenos13, Marie Loh3,14, Niek Verweij11, Ioanna Tzoulaki3-5*, Mark J Caulfield1,2*†, Paul Elliott3,5*†, Pierre Tuck11, John Connell32, Stephan J L Bakker33, Andres Metspalu26, Nilesh J Samani51,52, Joanna M M Howson7, Martin Farrall12, Sandosh Padmanabhan30, Reek H Nyman34, Simon Tong55, Morris Brown1,2, Peter Sever18, Tõnu Esko26,37, Caroline Hayward16, Pim van der Harst15, Danish Saleheen38, Rajiv Chowdhury7, John C Chambers3,39-41, Daniel I Chasman42,43, Aravinda Chakravarti44, Christopher Newton-Cheh37,45,46, Cecilia M Lindgren47, Daniel Levy8,10, Jaspal S Koone48,49, Bernard Keavney50, Maciej Tomaszewski50, Niles J Samani51,52, Joanna M M Howson7, Martin D Tobin53, Patricia B Munroe1,2, Georg B Ehret44,54, Louise V Wain55, Michael R Barnes1,2*, Ioanna Tzoulaki3-5*, Mark J Caulfield1,2*, Paul Elliott3,5*†

On behalf of the UK Biobank CardioMetabolic Consortium BP working group;

* Equal contribution
† Corresponding author
10 Boston University School of Public Health, Boston, MA, USA.
11 Department of Biostatistics, University of Liverpool, Liverpool, UK.
12 Division of Statistical Genomics, Center for Genome Sciences, Washington University School of Medicine, St. Louis MO, USA.
13 MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK.
14 Translational Laboratory in Genetic Medicine, Agency for Science, Technology and Research (A*STAR), Singapore.
15 Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.
16 MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK.
17 Bayer Pharma AG, Berlin, Germany.
18 Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK.
19 Data Science Institute, Lancaster University, Lancaster, UK.
20 Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.
21 Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan.
22 Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana, USA.
23 Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore.
24 Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.
25 Centre for Genomic & Experimental Medicine, Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, UK.
26 Estonian Genome Center, University of Tartu, Tartu, Estonia.
27 Imperial Clinical Trials Unit, School of Public Health, Imperial College London, London, UK.
28 Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, Oxford, UK.
29 Department of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.
30 Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK.
31 Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland.
32 Ninewells Hospital & Medical School, University of Dundee, Dundee, UK.
33 Department of Internal Medicine, University of Groningen, University Medical Center Groningen, Groningen, Groningen, the Netherlands.
34 School of Medicine, Conway Institute, Dublin, Ireland.
35 International Centre for Circulatory Health, Imperial College London, London, UK.
36 National Heart and Lung Institute, Imperial College London, London, UK.
37 Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, USA.
38 Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, USA.
39 Ealing Hospital National Health Service (NHS) Trust, Middlesex, UK.
40 Imperial College Healthcare NHS Trust, London, UK.
41 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.
42 Division of Preventive Medicine, Brigham and Women's Hospital, Boston MA, USA.
43 Harvard Medical School, Boston, MA, USA.
Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA.
Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA.
Wellcome Trust Center for Human Genetics, University of Oxford, Oxford, UK.
Department of Cardiology, Ealing Hospital NHS Trust, Southall, Middlesex, UK.
National Heart and Lung Institute, Cardiovascular Sciences, Hammersmith Campus, Imperial College London, London, UK.
Division of Cardiovascular Sciences, The University of Manchester, Manchester, UK.
Department of Cardiovascular Sciences, University of Leicester, BHF Cardiovascular Research Centre, Glenfield Hospital, Leicester, UK.
NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK.
Department of Health Sciences, University of Leicester, Leicester, UK.
Cardiology, Department of Medicine, Geneva University Hospital, Geneva, Switzerland.

Corresponding authors: Paul Elliott (p.elliott@imperial.ac.uk) and Mark Caulfield (m.j.caulfield@qmul.ac.uk)
Abstract:

Elevated blood pressure is the leading heritable risk factor for cardiovascular disease worldwide. We report genetic association of blood pressure (systolic, diastolic, pulse pressure) among UK Biobank participants of European ancestry with independent replication in other cohorts, leading to discovery and validation of 107 novel loci. We also identify new independent variants at 16 previously reported blood pressure loci. Combined with results from a range of in-silico functional analyses and wet bench experiments, our findings highlight new biological pathways for blood pressure regulation enriched for genes expressed in vascular tissues and identify potential therapeutic targets for hypertension. Results from genetic risk score models raise the possibility of a precision medicine approach through early lifestyle intervention to offset the impact of blood pressure raising variants on future cardiovascular disease risk.
Elevated blood pressure is a strong, heritable and modifiable driver of risk for stroke and coronary artery disease and a leading cause of global mortality and morbidity\(^1\)\(^2\). In most populations blood pressure rises with age and by older ages over 50% of the population has hypertension\(^3\)\(^4\). Raised blood pressure is heritable and arises from a complex interplay of lifestyle exposures and genetic background\(^5\)\(^6\). To date, studies including genome-wide meta-analyses of up to 2.5 million HapMap imputed variants across multiple studies, and analyses of bespoke or exome content, have identified 163 genetic variants of mostly modest or weak effect on blood pressure at 122 loci\(^9\)\(^13\). Here, we report association analyses between three blood pressure traits (systolic, diastolic and pulse pressure) and genetic variants among the first \(~150,000\) UK Biobank participants, with independent replication in large international consortia and other cohorts, providing new biological insights into blood pressure regulation.

UK Biobank is a prospective cohort study of 500,000 men and women aged 40-69 years with extensive baseline phenotypic measurements according to a standardized protocol (including blood pressure by a semi-automated device), stored biological samples (including DNA)\(^14\), and follow-up by electronic health record linkage\(^15\). Participants were genotyped using a customised array (including GWAS and exome content) and with genome-wide imputation based on 1000 Genomes and UK10K sequencing data\(^16\)\(^17\).

Our study design is summarised in Fig. 1. Briefly, of the 152,249 UK Biobank participants with genotype data, after quality measures and exclusions (see Methods Online), we study 140,886 unrelated individuals of European ancestry with two seated clinic blood pressure measurements (Supplementary Table 1). We carry out genome-wide association study (GWAS) analyses of systolic, diastolic and pulse pressure using single-variant linear regression under an additive model, based on \(~9.8\) million single nucleotide variants (SNVs) with minor allele frequency (MAF) \(\geq 1\%\) and imputation quality score (INFO) \(>0.1\). We then consider for replication SNVs with \(P < 1 \times 10^{-6}\) and take forward the sentinel SNV (i.e. with lowest \(P\)-value) at each locus, with a locus being defined by linkage disequilibrium (LD) \(r^2 < 0.2\), within a 1Mb interval. We similarly analyse exome content for variants with MAF \(\geq 0.01\%\), including rare variants, taking into replication the sentinel SNV (\(P < 1 \times 10^{-5}\)) from loci that are non-overlapping \(r^2 < 0.2\) with the GWAS findings. Overall we took the sentinel SNVs from 240 loci into replication (\(r^2 < 0.2\) and >500kb from previously reported blood pressure SNVs and not annotated to previously reported blood pressure genes: 218 from GWAS and 22 from the exome analysis (GWAS variants from an additional 17 novel loci could not be taken into replication due to the absence of the variant or a proxy in the replication resources (Supplementary Tables 2 and 3).

The replication resources comprise a large BP meta-analysis consortium and further cohorts with 1000 Genomes data for the GWAS findings (Supplementary Table 4), and large blood pressure exome consortia meta-analyses, both with individuals of European ancestry. We use \(P < 5 \times 10^{-8}\) to denote genome-wide significance in the combined (discovery and replication) meta-analyses, also requiring significant association (\(P < 0.01\)) in the replication data alone and concordant direction of effect. Additionally, we take forward for replication potential secondary signals at 51 previously reported blood pressure loci (excluding the HLA region).

To better understand the functional consequences of our new discoveries as well as previously reported variants, we carry out a series of in silico investigations including
expression Quantitative Trait Locus (eQTL) analyses, tissue and DNASE hypersensitivity site enrichment and pathway analyses (Supplementary Fig. 1). We also test for long-range regulatory interactions (Hi-C) and investigate metabolomics signatures associated with our sentinel SNVs. Finally, we undertake experimental analysis of gene expression in relevant vascular tissue for selected putative functional SNVs.

RESULTS

Discovery and validation of genetic variants at novel loci

Of the 240 not previously reported loci taken forward to replication, we validate 107 novel loci at \( P < 5 \times 10^{-8} \), of which 102 derive from the GWAS analysis replicated and meta-analysed in a total of 330,956 individuals (Table 1a; Fig. 2a-c; Supplementary Fig. 2a), and a further five are from the exome analysis validated in a total of 422,604 individuals from the combined meta-analysis (Table 1b and Supplementary Fig. 2b; Supplementary Tables 5 and 6). Most SNVs also show association with hypertension in the UK Biobank data, for example 93 of the 107 novel sentinel SNVs are nominally significant \( (P < 0.01) \) (Supplementary Table 7).

Our results for systolic, diastolic and pulse pressure are shown in Figs. 2a,b,c respectively. The most significant association signal for systolic pressure, which rises with age is with rs112184198 near \( \text{PAX2} \) \( (P = 3.6 \times 10^{-18}) \); for diastolic pressure, which plateaus in middle age, with rs76326501 near \( \text{METTL21A} - \text{ACO16735.1} \) \( (P = 3.6 \times 10^{-18}) \); and rs3889199 near \( \text{FGGY} \) \( (P = 1.8 \times 10^{-24}) \) for pulse pressure, which increases with age and arterial stiffening \( ^{18} \). However, as blood pressure traits are highly correlated, we unsurprisingly report considerable overlap in these findings (Supplementary Fig. 3). Many loci are associated with more than one blood pressure trait at genome-wide significance. For example, in the combined meta-analysis, 24 novel loci are associated with both systolic and diastolic pressure, 11 with both systolic and pulse pressure, one locus with both diastolic and pulse pressure and four loci (\( \text{NAKD-CPSF3L}, \text{GTF2B}, \text{METTL21A-AC079767.3} \) and \( \text{PAX2} \)) are associated with all three traits (Fig. 1d). We further note that many of the pulse pressure associated SNVs have opposing directions of effect for systolic and diastolic pressure, and are less likely to have strong associations with hypertension.

After conditional analysis on the sentinel SNV we identify five validated secondary SNVs in novel regions that are independently associated with blood pressure traits (Table 2a; Supplementary Table 8). We also note the existence of a rare validated potential secondary variant at the \( \text{NOX4} \) locus (rs56061986, MAF = 0.3%); although we do not claim this as an independent signal after conditioning on the sentinel variant, its relatively large effect on blood pressure remains (Supplementary Table 8). The contribution of our novel loci increases the percentage trait variance explained by \( \sim 1\% \), e.g. compared with 2.59% for previously reported SNVs alone, taken together, the novel and previously reported SNVs explain 3.56% of variance for systolic blood pressure, in an independent population.

For the first time in GWAS we report a signal at the angiotensin converting enzyme (\( \text{ACE} \)) locus \( (P = 6.8 \times 10^{-14}) \), from the renin-angiotensin system, a pathway which is targeted by current blood pressure treatments (\( \text{ACE-inhibitors} \)), as well as several other signals at known hypertension drug targets. These include \( \text{CACNA2D2} \) (rs743757, \( P = 2.4 \times 10^{-10}) \) targeted by calcium channel blockers, \( \text{MME} \) (rs143112823 in the RP11-439C8.2 locus, \( P = 1.4 \times 10^{-14}) \)
targeted by omapatrilat for treating hypertension, ADRA2B (rs2579519 in the GPAT2-FAHD2CP locus, \( P = 4.8 \times 10^{-12} \)) targeted by beta blockers, SLC14A2 (rs7236548, \( P = 2.0 \times 10^{-18} \)) targeted by the hypertension drug nifedipine, and phosphodiesterase 5A (PDE5A; rs66887589, \( P = 3.4 \times 10^{-15} \)) targeted by sildenafil for treating pulmonary hypertension.

Additionally, we evaluate our novel SNVs, where available, in cohorts of non-European ancestry, while recognising that these analyses are likely underpowered (Supplementary Table 9). For the GWAS SNVs, we find concordance in direction of effect (\( P < 0.05 \)) for all three blood pressure traits for individuals of East Asian ancestry, and for diastolic pressure for South Asian ancestry. For the exome analyses, we find concordance in direction of effect among individuals of Hispanic ancestry. Despite small numbers, these findings point to cosmopolitan effects for many of the blood pressure associated variants.

A PhenoScanner search revealed that 27 of our 107 novel sentinel SNVs (or proxies; \( r^2 \geq 0.8 \)) exhibit genome-wide significant associations (Fig. 3a) with other traits, including cardiovascular outcomes (e.g. coronary artery disease, myocardial infarction), cardiovascular risk factors (e.g. lipids, height, body mass index) and non-cardiovascular traits (e.g. lung function, cancer, Alzheimer’s). While some of these associations may reflect pleiotropy, for others such as coronary artery disease it is likely from evidence from trials that elevated blood pressure lies on the causal pathway.

**Associations at previously reported loci**

In the conditional analyses, we identify 22 secondary SNVs (17 common, one rare and four low-frequency variants) that are conditionally independent of the blood pressure associated SNVs at 16 previously reported loci (Table 2b; Supplementary Tables 10 and 11). One rare variant (rs138582164, MAF=0.1%) in the CDH17 locus anticipated to act as an exonic stop/gain mutation at the GEM gene is associated with a relatively large effect on pulse pressure (3.5 mm Hg per allele copy, Table 2b). At three previously reported loci (EBF1, PDE3A, JAG1) we identify multiple independent secondary SNVs in addition to the previously reported SNVs (Supplementary Table 10).

We confirm association (\( P < 0.01 \)) in the UK Biobank data for 119 of the 122 previously reported blood pressure loci (160 of 163 SNVs) for one or more blood pressure traits (Fig. 2a-c; Supplementary Table 12). Only three previously reported SNVs do not replicate in UK Biobank, one of which (rs11066280, RPL6-ALDH1) was identified from a GWAS of East Asian ancestry and may have ancestry-specific effects.

We also examine findings for low-frequency and rare gene mutations previously reported to be associated with monogenic hypertension disorders and included on the UK Biobank gene array. Due to a lack of power for testing rare variants, even within a large single study, only one monogenic mutation (rs199469624; KLH3; MAF=0.02%) shows nominally significant association (\( P < 0.05 \); Supplementary Table 13). However, we do detect a large effect of this rare variant (8.2 mm Hg and 5.6 mm Hg per allele for systolic and pulse pressure respectively) within the UK Biobank data.
Functional analyses

We annotate the 107 novel loci to 212 genes (based on LD $r^2 \geq 0.8$) and seek putative function from in silico analyses of our novel and previously reported loci, as well as undertaking gene expression experiments for selected SNVs in relevant vascular tissue. Of the 107 novel sentinel SNVs only three are Indels, all other variants are single nucleotide polymorphisms (SNPs). We identify non-synonymous SNVs at 13 of the 107 novel loci, including three non-synonymous novel sentinel SNVs (rs1250259 at FN1 locus, rs78648104 at TFAP2D and rs7127805 at CRACR2B locus) (Supplementary Table 14). Furthermore three of the 13 novel loci contain non-synonymous SNVs that are predicted to be damaging in TFAP2D (rs78648104), NOX4 (rs56061986, see above) and CCDC141 (rs17362588, reported to be associated with heart rate$^{23}$) (Fig. 3a). Beyond the coding regions we identify 29 novel associated SNVs in 3′UTRs which are predicted to significantly weaken or cause loss of miRNA regulation by altering the recognition motif in seven genes, and strengthen or create target sites for miRNA binding in 13 genes (Supplementary Table 14).

Our expression Quantitative Trait locus (eQTL) analysis shows that many novel loci contain variants with eQTLs across a range of different tissues (Supplementary Table 15). Of the 107 novel loci, 59 contain variants with eQTLs in at least one tissue. We observe arterial tissue as the tissue having the largest number of loci with eQTLs (Supplementary Fig. 4). Our follow-up targeted in-silico analysis reveals six novel loci with eQTLs in arterial tissue (Supplementary Table 14). For example, the GTEx tibial artery eQTL in SF3A3 (rs4360494) shows strong in silico supporting evidence, including an arterial DNase I site within which the major C allele removes a predicted AP-2 binding site (Supplementary Fig. 5). Hence we prioritised this gene for in vitro functional analysis (see below).

By considering all loci together from both novel and previously reported loci, our analysis using DEPICT identifies enrichment of expression across 31 tissues and cells (Supplementary Fig.6; Supplementary Table 16), with greatest enrichment in the arteries ($P = 1.9 \times 10^{-6}$, false discovery rate (FDR) $< 1\%$). We use FORGE to investigate and identify significant (FDR, $P < 0.05$) cell type specific enrichment within DNase I hypersensitive sites in a range of tissues including dermal and lung microvascular endothelial cell types, and cardiac fibroblasts (Supplementary Fig. 7). For a set of curated candidate regulatory SNVs from novel loci (see Supplementary Methods), widespread enrichment is found in microvascular endothelium, aortic smooth muscle, aortic fibroblasts, vascular epithelium, heart and skin (Supplementary Fig. 7). In addition, we identify significant enrichment of histone marks in a wide range of cell types, including strong enrichment seen for H3K4Me3 (an activating modification found near promoters) marks in umbilical vein endothelial cells (HUVEC) (Supplementary Fig. 8). To explore expression at the level of cardiovascular cell types specifically, we use Fantom5 reference transcript expression data (see Methods Online) to cluster the 212 genes annotated to our 107 novel loci according to tissue specificity (Supplementary Fig. 9), with the significantly clustered genes forming four tissue-specific clusters, including a vascular smooth muscle cell (VSMC) and fibroblast cluster, an endothelial cell cluster (including probable endothelial cells in highly vascularised tissues), and a combined vascular cell cluster.
Additionally, Ingenuity pathway analysis and upstream transcriptional analysis show enrichment of canonical pathways implicated in cardiovascular disease, including those targeted by antihypertensive drugs, such as the alpha-adrenergic, CXCR4, endothelin signalling and angiotensin receptor pathways (Supplementary Table 17). In keeping with vascular mediation of genetic influence we identify diphenyleneiodonium, an inhibitor of flavin-containing oxidases, including NAD(P)H oxidase, which is reported to reverse endothelial dysfunction (and hypertension) in a rat model. In order to identify long range target genes of non-coding variants, we use chromatin interaction (Hi-C) data from HUVEC, as enhancers and silencers often form chromatin loops with their target promoter. In most loci the strongest promoter interaction involves a gene in high LD with the SNV but for 21 loci we find a distal potential target gene (Supplementary Table 14). Ingenuity pathway analysis of the distal genes shows the greatest enrichment in regulators of cardiac hypertrophy.

We further evaluate pleiotropy using the Genomic Regions Enrichment of Annotations Tool (GREAT) to study enrichment of mouse phenotype and human disease ontology terms across all our novel and previously reported loci. These highlight cardiovascular system abnormalities and vascular disease as the most highly enriched terms (Fig. 3b & 3c).

Collectively evidence from eQTLs, DEPICT, DNase I sites, histone marks, Hi-C data and ontological analyses indicates predominant vascular and cardiovascular tissue involvement for genes within the blood pressure associated loci. For example, aggregating all loci together in the DEPICT analysis, we observe greatest enrichment in arterial tissue, which has the largest proportion of novel loci having variants with eQTLs.

We also look for association of our validated sentinel SNVs with metabolomic signatures. Three novel SNVs within the NOX4, KCNH4 and LHFPL2 loci show significant associations (family-wise error rate < 5%) with lipoprotein sub-fractions from 1H Nuclear Magnetic Resonance (NMR) spectroscopy analysis of 2,000 Airwave study samples (Supplementary Tables 18 and 19). The results for these variants suggest a link between blood pressure regulation and lipid metabolism. Eleven SNVs (including at LHFPL2 locus) show association (family wise error rate < 5%) with metabolites in blood or urine from the publicly available “Metabolomics GWAS Server” resource based on mass spectrometry (Supplementary Table 19), including sugar acids, sphingolipids, fatty acids, glycerophospholipids, organic acids and benzene derivatives.

Several genes and variants with putative function are highlighted in our in silico analysis as having biological support (e.g. eQTLs or nsSNVs) and those with novelty and tractability to laboratory investigation (e.g. expression in available tissue models) are prioritized. Variants in three genes are selected for experimental testing and successfully genotyped, each for at least 100 samples. We select ADAMTS7 due to strong biological support (e.g. mouse knockout phenotype), SF3A3 due to eQTLs and NOX4 as it contains a rare nsSNV in addition to common variant associations. We use quantitative polymerase chain reaction (qPCR) to study the impact of these sentinel variants on gene expression in human vascular smooth muscle (VSMCs) and endothelial cells (ECs) (see Methods Online). For SF3A3, the major C allele of
sentinel variant rs4360494 associated with increased pulse pressure is also associated with
SF3A3 expression in human VSMCs, although this SNV is not related to expression in
endothelial cells (Supplementary Fig. 10a); and the T allele of SNV rs62012628 in ADAMTS7,
associated with lower diastolic pressure, is associated with reduced ADAMTS7 expression in
human VSMCs (Supplementary Fig. 10b). Moreover, we find that the minor A allele of
sentinel SNV rs2289125 at the NOX4 locus correlates with increased NOX4 expression in ECs
though not VSMCs (Supplementary Fig. 10c). Our study thus finds evidence for novel cis-
eQTLs in ADAMTS7 and NOX4 in addition to validating the previously reported GTEx eQTL in
SF3A3, and supports the vascular expression of these genes.

Genetic risk of increased blood pressure, hypertension and cardiovascular outcomes

We create an unbiased genetic risk score (GRS) (Supplementary Table 20) to evaluate, in an
independent cohort (Airwave, see Methods Online), the impact of the combination of our
validated novel and previously reported loci on blood pressure levels and risk of hypertension.
The combination of these blood pressure influencing variants is associated with sex-adjusted
mean systolic pressure that is 9.3 mm Hg (95% CI 6.9 to 11.7 mm Hg, \( P = 1.0 \times 10^{-13} \)) higher at
ages 50 years and over, comparing the upper and lower fifths of the GRS distribution; and an
over two-fold higher risk of hypertension (OR 2.32 95% CI 1.76 to 3.06; \( P = 2.8 \times 10^{-9} \)) (Fig. 4;
Supplementary Table 21). Similar results were obtained from GRS associations with blood
pressure and hypertension within UK Biobank (Supplementary Table 22). In UK Biobank –
based on self-reported health data, record linkage to Hospital Episode Statistics and mortality
follow-up data (Supplementary Table 23) – we show that the GRS is associated with increased
risk of stroke, coronary heart disease and all cardiovascular outcomes, comparing the upper
and lower fifths of the GRS distribution, with sex-adjusted odds ratios of 1.34 (95% CI 1.20 to
1.49, \( P = 1.5 \times 10^{-7} \)), 1.38 (95% CI 1.30 to 1.47, \( P = 4.3 \times 10^{-23} \)) and 1.35 (95% CI 1.27 to 1.42,
\( P = 1.3 \times 10^{-25} \)) respectively (Fig. 4; Supplementary Table 24).

DISCUSSION

A key attribute of this study is the combination of a large, single discovery sample with
standardized blood pressure measurement and a dense 1000 Genomes imputation strategy
(UK 10K enhanced 1000G imputation), yielding a high quality dataset of ~9.8 million variants
for study\textsuperscript{16}. This is the largest genetic association analysis for blood pressure to date taking
advantage of major international consortia for parallel replication of common and low-
frequency variants, based in total on data from 330,956 individuals and exonic SNVs in a total
of 422,604 individuals\textsuperscript{27}. This strategy resulted in the discovery of 107 robustly validated novel
loci for blood pressure traits. In previous large-scale blood pressure genome-wide association
scans we estimated that an effective doubling of sample size from a discovery cohort of
70,000 to 140,000 individuals with ~2.5 million imputed variants would double the number
of validated loci, resulting in an estimated ~30 additional loci for blood pressure traits\textsuperscript{27}. Here
we find over three times that number, taking advantage of UK Biobank’s standardized
approach to data collection, biobanking, genotyping and enhanced imputation strategy.
Nonetheless, despite its size, our study is still under-powered to find rare variants - the vast
majority of our findings are common variants, with similarly modest or small effect sizes as
for previously reported variants (Supplementary Fig. 11). There may be greater potential for
identifying rare variants from the future release of genetic data for all 500,000 UK Biobank participants.

Our findings point to new biology as well as highlighting novel gene regions in systems that affect atherosclerosis or vascular remodelling (ADAMTS7, THBS2, CFDP1) and exhibit locus pleiotropy in prior genome-wide association studies for coronary artery disease or carotid intimal-media thickness (Fig. 3a and Fig. 5). In previous work we have shown that expression of ADAMTS7 is upregulated and increases vascular smooth muscle cell migration in response to vascular injury in relation to a distinct coronary artery variant (rs3825807 which is not in strong LD with our sentinel SNV; $r^2 = 0.17$). In endothelial cells ADAMTS7 acts as a metalloproteinase to cleave thrombospondin-1 encoded by THBS2 which leads to reduced endothelial cell migration and plays a role in neo-intimal repair in the vessel wall. Our functional work indicates that the allele associated with lower diastolic pressure is also associated with lower ADAMTS7 expression in human vascular smooth muscle cells; this fits with the murine knockout that exhibits reduced atherosclerosis. At the CFDP1 locus our sentinel SNV is in high LD ($r^2 = 0.95$) with a variant previously associated with carotid intimal-medial thickness.

We identify both common and rare variant associations at the novel NADPH oxidase 4 (NOX4) locus. This oxidase generates reactive oxygen species in the endothelium and may contribute to salt sensitive hypertension in the kidney and the vasculature. We found that the allele of the common variant at NOX4 locus correlates with increased tissue specific NOX4 expression in endothelial cells rather than vascular smooth muscle cells (Supplementary Figure 10c). NOX4 mediates endothelial cell apoptosis and facilitates vascular collagen synthesis contributing to endothelial dysfunction and arterial stiffness, and may explain the association with pulse pressure.

We identify several loci containing genes involved in vascular signalling and second messenger systems such as PDE5A and PDE10A. The phosphodiesterase PDE5A hydrolyses cyclic GMP and is inhibited by sildenafil which leads to vasodilatation. This finding fits with our previous discoveries of a role for gene loci encoding elements of natriuretic peptide-nitric oxide pathway and guanylate cyclase signalling systems in blood pressure regulation. Our findings strengthen the case for evaluating the opportunity to repurpose PDE5A inhibitors for use in hypertension.

The importance of microvascular function is emphasised by the solute carrier transporters such as SLC14A2 encoding a urea transporter, which has previously been linked to autosomal dominant Streeten type orthostatic hypotensive disorder and blood pressure response to nifedipine, a calcium channel blocker antihypertensive drug. SLC8A1 encodes a sodium calcium exchanger expressed in cardiomyocytes which alters cardiac contractility and hypertrophy and shows abnormal blood pressure in SLC8A1 transgenic mice. Variants at SLC35F1 have been previously associated with resting heart rate and ventricular dimensions which could contribute to blood pressure elevation.
We also identify loci that are involved in cardiovascular development (GATA2, KIAA1462, FBN2, FN1 and HAND2) such as fibrillin 2 (FBN2) which overlaps in action with fibrillin 1 in development of the aortic matrix\textsuperscript{49-53}. In addition, fibronectin expression is increased in hypertension and in atherosclerosis but it may also play a role in the development of the heart\textsuperscript{53-55}.

Our analysis validates loci containing genes with prior physiological connection to blood pressure such as BDNF, FAM208A, and CACNA2D2\textsuperscript{56-58}. The neurotrophin Brain Derived Neurotrophic Factor modulates angiotensin 11 in the brain to elevate blood pressure in experimental models and higher serum levels correlate with reduced risk of cardiovascular disease and mortality\textsuperscript{56}. In experimental models FAM208A, which is thought to be a transcription factor, is a strong candidate for a quantitative trait locus for blood pressure\textsuperscript{58}. The gene CACNA2D2 encodes a subunit of the L-type calcium channel that is most abundantly expressed in the atrium and in neurones and may be a target for negatively chronotropic and inotropic calcium channel antagonists which reduce blood pressure\textsuperscript{59}.

This is the first time long range genomic interactions have been sought using Hi-C for blood pressure, where the promoter region has a strong chromatin interaction with a novel SNV. One such gene is EPAS1, which is \textasciitilde200kb away from the SNV (rs11690961). It encodes hypoxia-inducible factor 2alpha, which affects catecholamine homeostasis, protects against heart failure and mutations in the gene are associated with pulmonary hypertension\textsuperscript{60}. Another gene is INHBA, 1.3Mb away from the SNV (rs12531683), which is elevated in pulmonary hypertension and contributes to vascular remodelling by inducing expression of endothelin-1 and plasminogen activator inhibitor-1 in pulmonary smooth muscle cells\textsuperscript{61}.

Our observation that the blood pressure genetic risk score is associated with 9-10 mm Hg higher blood pressure at age 50+ years when comparing the top vs bottom fifths of the distribution in an independent population has potential clinical and public health implications. Were the genetic risk score to be measured at birth or in childhood, there would be the possibility of adopting an early precision medicine approach to risk management through lifestyle intervention (i.e. reduced sodium intake, increased potassium intake, maintenance of optimal weight, low adult alcohol consumption and regular exercise)\textsuperscript{62-64}. Studies of non-pharmacologic approaches to blood pressure control indicate that we could achieve 10 mm Hg or more reduction in systolic blood pressure through lifestyle measures alone\textsuperscript{65}. This would be sufficient to offset the genetic influence on the rise of blood pressure from young adulthood to middle age and reduce the resultant high prevalence of hypertension at older ages. Such a precision medicine approach could thus mitigate the risk of future cardiovascular disease among people at high genetic risk of raised blood pressure.

We describe 107 novel validated loci for blood pressure offering new biology, identifying potential new therapeutic targets and raising the possibility of a precision medicine approach to modify risk of hypertension and cardiovascular outcomes. In total this brings the number of combined novel and previously reported loci for blood pressure traits to 229, representing a major advance in our understanding of the genetic architecture of blood pressure.
References


Acknowledgements

HRW, CPC, MR, MRB, PBM, MB and MJC were funded by the National Institutes for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Biomedical Research Unit in Cardiovascular Disease at Barts and The London School of Medicine and Dentistry. MJC is an NIHR Senior Investigator.

HG was funded by the NIHR Imperial College Health Care NHS Trust and Imperial College London Biomedical Research Centre. MR was a recipient from China Scholarship Council (No. 2011632047). BM holds an MRC eMedLab Medical Bioinformatics Career Development Fellowship, funded from award MR/L016311/1. JMMH was funded by the UK Medical Research Council (G0800270), British Heart Foundation (SP/09/002), UK National Institute for Health Research Cambridge Biomedical Research Centre, European Research Council (268834), European Commission Framework Programme 7 (HEALTH-F2-2012-279233). BH holds a British Heart Foundation Personal Chair. NJS holds a chair funded by the British Heart Foundation and is a NIHR Senior Investigator. PSu was funded by the UK Medical Research Council (G0800270). CL and AK were funded by the NHLBI intramural funding. CNC was funded by the National Institutes of Health (HL113933, HL124262). PVDH was funded by the ZonMw grant 90.700.441, Marie Sklodowska-Curie GF (call: H2020-MSCA-IF-2014, Project ID: 661395). NV was supported by Marie Sklodowska-Curie GF grant (661395) and ICIN-NHI. NP received funding from the UK National Institute for Health Research Biomedical Research Centre funding scheme and also from his Senior Investigator Award. PE is an NIHR Senior Investigator and was funded by the NIHR Imperial College Health Care Trust and Imperial College London Biomedical Research Centre, the UK Medical Research Council and Public Health England as Director of the MRC-PHE Centre for Environment and Health, and the NIHR Health Protection Research Unit on the Health Effects of Environmental Hazards. Effects of Environmental Hazards. Part of this work used computing resources provided by the MRC-funded UK MEDical Bioinformatics partnership programme (UK MED-BIO) (MR/L01632X/1). Some of this work used the ALICE and SPECTRE High Performance Computing Facilities at the University of Leicester. This research has been conducted using the UK Biobank Resource.

Conflicts/Disclosures

MJC is Chief Scientist for Genomics England, a wholly owned UK government company, to deliver the 100,000 Genomes Project.
Author Contributions

Central analysis: HRW, CPC, HG, MRB, MPSL, MR, IT, BM, IK, EE.

Writing of the paper: HRW*, MRB, EE, CPC, HG, IT, BM, MR, MJC*, PE* (*Writing group leads).

Working group membership: MJC*, HRW, EE, IT, PBM, LV, NJS, MT, JMMH, MDT, IN, BK, HG, MRB, CPC, JSK, PE* (*Co-Chairs).


All authors critically reviewed and approved the final version of the manuscript.
Table 1: Association results for the sentinel variant from each novel validated locus from (a) UK Biobank GWAS discovery and (b) UK Biobank exome discovery. Results are shown for the primary blood pressure trait with most significant association from the combined meta-analysis.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chr</th>
<th>Pos</th>
<th>rsID</th>
<th>EA</th>
<th>INFO</th>
<th>EAF</th>
<th>Beta</th>
<th>SE</th>
<th>P</th>
<th>Beta</th>
<th>SE</th>
<th>P</th>
<th>N</th>
<th>Beta</th>
<th>SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Systolic blood pressure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NADK-CPSF3L</td>
<td>1</td>
<td>1685921</td>
<td>rs139385870</td>
<td>D</td>
<td>0.99</td>
<td>0.50</td>
<td>-0.394</td>
<td>0.07</td>
<td>1.9x10^{-8}</td>
<td>-0.310</td>
<td>0.07</td>
<td>1.0x10^{-5}</td>
<td>281,890</td>
<td>-0.352</td>
<td>0.05</td>
<td>1.3x10^{-12}</td>
</tr>
<tr>
<td>CELA2A</td>
<td>1</td>
<td>15798197</td>
<td>rs3820068</td>
<td>A</td>
<td>0.99</td>
<td>0.81</td>
<td>0.497</td>
<td>0.09</td>
<td>2.4x10^{-8}</td>
<td>0.367</td>
<td>0.08</td>
<td>5.3x10^{-6}</td>
<td>310,776</td>
<td>0.425</td>
<td>0.06</td>
<td>1.1x10^{-12}</td>
</tr>
<tr>
<td>GTF2B</td>
<td>1</td>
<td>89360158</td>
<td>rs10922502</td>
<td>A</td>
<td>0.99</td>
<td>0.62</td>
<td>-0.475</td>
<td>0.07</td>
<td>4.7x10^{-11}</td>
<td>-0.307</td>
<td>0.06</td>
<td>2.0x10^{-6}</td>
<td>323,666</td>
<td>-0.382</td>
<td>0.05</td>
<td>2.2x10^{-15}</td>
</tr>
<tr>
<td>FOSL2</td>
<td>2</td>
<td>28635740</td>
<td>rs7562</td>
<td>T</td>
<td>0.98</td>
<td>0.52</td>
<td>0.365</td>
<td>0.07</td>
<td>2.2x10^{-7}</td>
<td>0.182</td>
<td>0.06</td>
<td>3.7x10^{-3}</td>
<td>319,942</td>
<td>0.263</td>
<td>0.05</td>
<td>1.9x10^{-8}</td>
</tr>
<tr>
<td>PRKD3</td>
<td>2</td>
<td>37517566</td>
<td>rs13420463</td>
<td>A</td>
<td>1.00</td>
<td>0.77</td>
<td>0.504</td>
<td>0.08</td>
<td>1.4x10^{-8}</td>
<td>0.244</td>
<td>0.07</td>
<td>7.3x10^{-4}</td>
<td>330,307</td>
<td>0.356</td>
<td>0.05</td>
<td>7.0x10^{-11}</td>
</tr>
<tr>
<td>METTL21A-ACD797673</td>
<td>2</td>
<td>208526140</td>
<td>rs55780018</td>
<td>T</td>
<td>0.97</td>
<td>0.54</td>
<td>-0.426</td>
<td>0.07</td>
<td>1.7x10^{-9}</td>
<td>-0.360</td>
<td>0.07</td>
<td>5.1x10^{-8}</td>
<td>304,567</td>
<td>-0.391</td>
<td>0.05</td>
<td>5.9x10^{-16}</td>
</tr>
<tr>
<td><strong>Diastolic blood pressure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr1nb25</td>
<td>1</td>
<td>25030470</td>
<td>rs6668889</td>
<td>T</td>
<td>0.99</td>
<td>0.25</td>
<td>0.231</td>
<td>0.05</td>
<td>3.7x10^{-7}</td>
<td>0.143</td>
<td>0.04</td>
<td>9.1x10^{-4}</td>
<td>322,575</td>
<td>0.185</td>
<td>0.03</td>
<td>3.6x10^{-9}</td>
</tr>
<tr>
<td>DNM3</td>
<td>1</td>
<td>172357441</td>
<td>rs12405515</td>
<td>T</td>
<td>0.98</td>
<td>0.56</td>
<td>-0.219</td>
<td>0.04</td>
<td>4.1x10^{-8}</td>
<td>-0.118</td>
<td>0.04</td>
<td>1.6x10^{-3}</td>
<td>328,543</td>
<td>-0.165</td>
<td>0.03</td>
<td>1.4x10^{-9}</td>
</tr>
<tr>
<td>Gene</td>
<td>ID</td>
<td>Chromosome Start</td>
<td>Chromosome End</td>
<td>p-Value (Gene)</td>
<td>p-Value (TA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>-----------------</td>
<td>----------------</td>
<td>---------------</td>
<td>--------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPATCH2</td>
<td>rs12408022</td>
<td>1</td>
<td>99063282</td>
<td>0.97</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDE4A</td>
<td>rs10078021</td>
<td>5</td>
<td>100782386</td>
<td>0.99</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PKHD1</td>
<td>rs1302180</td>
<td>6</td>
<td>13021807</td>
<td>0.99</td>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLC35F1</td>
<td>rs937246</td>
<td>11</td>
<td>93724696</td>
<td>0.98</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNX31</td>
<td>rs2978098</td>
<td>10</td>
<td>29780957</td>
<td>0.99</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP11-273G15.2</td>
<td>rs62524579</td>
<td>8</td>
<td>62524529</td>
<td>1.00</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTAP</td>
<td>rs4364717</td>
<td>9</td>
<td>4364717</td>
<td>0.99</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDNF</td>
<td>rs11030119</td>
<td>11</td>
<td>11030126</td>
<td>1.00</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYEOV</td>
<td>rs6733070</td>
<td>11</td>
<td>6733070</td>
<td>0.89</td>
<td>-0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP11-32I6.1</td>
<td>rs1209662</td>
<td>15</td>
<td>12096626</td>
<td>0.98</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADAMTS7</td>
<td>rs2012628</td>
<td>15</td>
<td>2012628</td>
<td>0.97</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr15mb95</td>
<td>rs1290662</td>
<td>15</td>
<td>1290662</td>
<td>0.98</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPL</td>
<td>rs12921187</td>
<td>16</td>
<td>12921187</td>
<td>0.10</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBX19</td>
<td>rs72799341</td>
<td>16</td>
<td>72799341</td>
<td>0.98</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMIP</td>
<td>rs8059962</td>
<td>16</td>
<td>8059962</td>
<td>0.98</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACE</td>
<td>rs4308</td>
<td>17</td>
<td>4308</td>
<td>0.98</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>RSID</td>
<td>Chromosome</td>
<td>Position</td>
<td>Haplotype</td>
<td>MAF</td>
<td>p-value</td>
<td>OR</td>
<td>95% CI Low</td>
<td>95% CI High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>------------</td>
<td>----------</td>
<td>-----------</td>
<td>-----</td>
<td>----------</td>
<td>--------</td>
<td>------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPK4</td>
<td>rs48142854</td>
<td>chr4</td>
<td>745821</td>
<td>A</td>
<td>0.99</td>
<td>0.76</td>
<td>0.236</td>
<td>0.05</td>
<td>3.2x10^-7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCNE1</td>
<td>rs30294991</td>
<td>chr4</td>
<td>62104477</td>
<td>T</td>
<td>0.99</td>
<td>0.33</td>
<td>0.209</td>
<td>0.04</td>
<td>7.1x10^-7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLCB1</td>
<td>rs8626271</td>
<td>chr4</td>
<td>6108168</td>
<td>A</td>
<td>0.99</td>
<td>0.25</td>
<td>-0.305</td>
<td>0.05</td>
<td>1.5x10^-11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulse Pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr1m8b</td>
<td>rs112557609</td>
<td>chr1</td>
<td>463574</td>
<td>A</td>
<td>0.99</td>
<td>0.71</td>
<td>0.462</td>
<td>0.05</td>
<td>3.3x10^-18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3A3</td>
<td>rs36083386</td>
<td>chr1</td>
<td>4365891</td>
<td>C</td>
<td>0.99</td>
<td>0.55</td>
<td>0.332</td>
<td>0.05</td>
<td>5.7x10^-12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP4-710M16.1-PPAP2B</td>
<td></td>
<td></td>
<td>5657692</td>
<td>A</td>
<td>0.99</td>
<td>0.35</td>
<td>0.280</td>
<td>0.05</td>
<td>3.2x10^-8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGGY</td>
<td>rs62270945</td>
<td>chr1</td>
<td>8625778</td>
<td>T</td>
<td>0.99</td>
<td>0.62</td>
<td>0.296</td>
<td>0.05</td>
<td>2.1x10^-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2orf43</td>
<td>rs4365336</td>
<td>chr1</td>
<td>1205259</td>
<td>A</td>
<td>0.99</td>
<td>0.74</td>
<td>-0.457</td>
<td>0.05</td>
<td>5.5x10^-17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRKCE</td>
<td>rs1690961</td>
<td>chr1</td>
<td>4636336</td>
<td>A</td>
<td>1.00</td>
<td>0.88</td>
<td>0.437</td>
<td>0.07</td>
<td>4.2x10^-9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEP68</td>
<td>rs74181299</td>
<td>chr1</td>
<td>5283972</td>
<td>T</td>
<td>0.99</td>
<td>0.54</td>
<td>0.256</td>
<td>0.05</td>
<td>1.1x10^-7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCF7L1</td>
<td>rs85491365</td>
<td>chr1</td>
<td>11689667</td>
<td>T</td>
<td>0.99</td>
<td>0.42</td>
<td>-0.44</td>
<td>0.09</td>
<td>3.8x10^-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FN1</td>
<td>rs1250259</td>
<td>chr1</td>
<td>21630048</td>
<td>A</td>
<td>0.99</td>
<td>0.74</td>
<td>-0.457</td>
<td>0.05</td>
<td>5.5x10^-17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATA2</td>
<td>rs62270945</td>
<td>chr1</td>
<td>1280188</td>
<td>T</td>
<td>1.00</td>
<td>0.03</td>
<td>0.861</td>
<td>0.14</td>
<td>2.6x10^-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PALLD</td>
<td>rs16971148</td>
<td>chr1</td>
<td>5654697</td>
<td>A</td>
<td>0.98</td>
<td>0.42</td>
<td>0.320</td>
<td>0.05</td>
<td>6.6x10^-11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr4mb174</td>
<td>rs17059668</td>
<td>chr1</td>
<td>17458466</td>
<td>C</td>
<td>0.98</td>
<td>0.92</td>
<td>-0.44</td>
<td>0.09</td>
<td>3.8x10^-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LHFPL2</td>
<td>rs10057188</td>
<td>chr1</td>
<td>7783778</td>
<td>A</td>
<td>0.99</td>
<td>0.46</td>
<td>-0.280</td>
<td>0.05</td>
<td>5.5x10^-9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIA1</td>
<td>rs11154027</td>
<td>chr1</td>
<td>12178139</td>
<td>T</td>
<td>0.99</td>
<td>0.47</td>
<td>0.311</td>
<td>0.05</td>
<td>1.1x10^-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESR1</td>
<td>rs36083386</td>
<td>chr1</td>
<td>15239712</td>
<td>I</td>
<td>1.00</td>
<td>0.11</td>
<td>0.651</td>
<td>0.08</td>
<td>4.6x10^-17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FNDC1</td>
<td>rs449789</td>
<td>chr1</td>
<td>15969915</td>
<td>C</td>
<td>1.00</td>
<td>0.14</td>
<td>0.480</td>
<td>0.07</td>
<td>2.2x10^-12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THBS2</td>
<td>rs1322639</td>
<td>chr1</td>
<td>16958713</td>
<td>A</td>
<td>1.00</td>
<td>0.78</td>
<td>0.433</td>
<td>0.06</td>
<td>7.7x10^-14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUGCT</td>
<td>rs76206723</td>
<td>chr1</td>
<td>40447971</td>
<td>T</td>
<td>0.99</td>
<td>0.10</td>
<td>-0.405</td>
<td>0.08</td>
<td>2.6x10^-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLC20A2</td>
<td>rs2978456</td>
<td>chr1</td>
<td>4324765</td>
<td>T</td>
<td>1.00</td>
<td>0.55</td>
<td>-0.253</td>
<td>0.05</td>
<td>1.3x10^-7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRAPPC9</td>
<td>rs4454254</td>
<td>chr1</td>
<td>14106002</td>
<td>A</td>
<td>1.00</td>
<td>0.63</td>
<td>-0.320</td>
<td>0.05</td>
<td>9.4x10^-11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCAI</td>
<td>rs72765298</td>
<td>chr1</td>
<td>12790096</td>
<td>T</td>
<td>0.98</td>
<td>0.87</td>
<td>-0.392</td>
<td>0.07</td>
<td>5.9x10^-8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIAA1462</td>
<td>rs9337951</td>
<td>chr1</td>
<td>30317073</td>
<td>A</td>
<td>0.94</td>
<td>0.34</td>
<td>0.301</td>
<td>0.05</td>
<td>7.6x10^-9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARHGAP12</td>
<td>rs10826995</td>
<td>chr1</td>
<td>30286258</td>
<td>T</td>
<td>0.99</td>
<td>0.71</td>
<td>-0.317</td>
<td>0.05</td>
<td>2.2x10^-9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PADM11</td>
<td>rs11442819</td>
<td>chr1</td>
<td>45208141</td>
<td>I</td>
<td>1.00</td>
<td>0.11</td>
<td>-0.712</td>
<td>0.07</td>
<td>3.8x10^-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOX4</td>
<td>rs2289125</td>
<td>chr1</td>
<td>89224453</td>
<td>A</td>
<td>0.98</td>
<td>0.21</td>
<td>-0.481</td>
<td>0.06</td>
<td>3.1x10^-16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEP164</td>
<td>rs8258</td>
<td>chr1</td>
<td>11728367</td>
<td>T</td>
<td>1.00</td>
<td>0.38</td>
<td>0.341</td>
<td>0.05</td>
<td>5.3x10^-12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDDC41</td>
<td>rs13923620</td>
<td>chr1</td>
<td>94880742</td>
<td>A</td>
<td>0.97</td>
<td>0.10</td>
<td>-0.442</td>
<td>0.08</td>
<td>5.7x10^-8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP11-6101.1</td>
<td>rs9323988</td>
<td>chr1</td>
<td>98587630</td>
<td>T</td>
<td>0.98</td>
<td>0.63</td>
<td>-0.291</td>
<td>0.05</td>
<td>5.6x10^-9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAC14</td>
<td>rs7055610</td>
<td>chr1</td>
<td>117006983</td>
<td>A</td>
<td>0.46</td>
<td>0.01</td>
<td>1.448</td>
<td>0.30</td>
<td>9.4x10^-7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDH13</td>
<td>rs7500448</td>
<td>chr1</td>
<td>83045790</td>
<td>A</td>
<td>0.98</td>
<td>0.75</td>
<td>0.386</td>
<td>0.06</td>
<td>4.2x10^-12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIAA0753</td>
<td>rs7226020</td>
<td>chr1</td>
<td>6473828</td>
<td>T</td>
<td>0.96</td>
<td>0.56</td>
<td>-0.348</td>
<td>0.05</td>
<td>1.3x10^-12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locus</td>
<td>TP53-SLC2A4</td>
<td>rs78378222</td>
<td>17</td>
<td>7571752</td>
<td>1</td>
<td>0.95</td>
<td>0.99</td>
<td>1.530</td>
<td>0.22</td>
<td>8.9x10^-12</td>
<td>0.487</td>
<td>0.18</td>
<td>7.9x10^-3</td>
<td>294,053</td>
<td>0.904</td>
<td>0.14</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-------------</td>
<td>-----</td>
<td>--------</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
<td>------------</td>
<td>------</td>
<td>-----</td>
<td>---------</td>
<td>----------</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>KCNH4-HSD17B1</td>
<td>rs79089478</td>
<td>17</td>
<td>40317241</td>
<td>1</td>
<td>0.99</td>
<td>0.97</td>
<td>0.842</td>
<td>0.15</td>
<td>1.2x10^-4</td>
<td>0.377</td>
<td>0.13</td>
<td>4.4x10^-3</td>
<td>318,326</td>
<td>0.584</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>PYY</td>
<td>rs62080325</td>
<td>17</td>
<td>42060631</td>
<td>A</td>
<td>0.98</td>
<td>0.66</td>
<td>-0.260</td>
<td>0.05</td>
<td>3.6x10^-7</td>
<td>-0.128</td>
<td>0.05</td>
<td>4.8x10^-3</td>
<td>315,689</td>
<td>-0.186</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>MRC2</td>
<td>rs740698</td>
<td>17</td>
<td>60767151</td>
<td>T</td>
<td>0.99</td>
<td>0.56</td>
<td>-0.307</td>
<td>0.05</td>
<td>2.1x10^-10</td>
<td>-0.161</td>
<td>0.04</td>
<td>2.8x10^-4</td>
<td>311,450</td>
<td>-0.228</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>SLC14A2</td>
<td>rs7236548</td>
<td>18</td>
<td>43097750</td>
<td>A</td>
<td>0.99</td>
<td>0.18</td>
<td>0.462</td>
<td>0.06</td>
<td>1.1x10^-13</td>
<td>0.273</td>
<td>0.05</td>
<td>2.2x10^-7</td>
<td>330,075</td>
<td>0.352</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>SLC24A3</td>
<td>rs6081613</td>
<td>20</td>
<td>19465907</td>
<td>A</td>
<td>0.99</td>
<td>0.28</td>
<td>0.326</td>
<td>0.05</td>
<td>1.2x10^-9</td>
<td>0.213</td>
<td>0.05</td>
<td>8.1x10^-6</td>
<td>315,546</td>
<td>0.263</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>ARVC</td>
<td>rs12628032</td>
<td>22</td>
<td>19967980</td>
<td>T</td>
<td>0.99</td>
<td>0.30</td>
<td>0.269</td>
<td>0.05</td>
<td>2.4x10^-7</td>
<td>0.216</td>
<td>0.05</td>
<td>3.8x10^-6</td>
<td>310,292</td>
<td>0.240</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>XRCC6</td>
<td>rs73161324</td>
<td>22</td>
<td>42038786</td>
<td>T</td>
<td>1.00</td>
<td>0.05</td>
<td>0.611</td>
<td>0.11</td>
<td>6.5x10^-9</td>
<td>0.380</td>
<td>0.11</td>
<td>3.1x10^-4</td>
<td>267,722</td>
<td>0.496</td>
<td>0.07</td>
</tr>
</tbody>
</table>

| (b) UK Biobank exome |
|---------------------|----------------|-----|--------|-----|-----|-------|------|-----|---------|------|-----|---------|----------|-------|-----|----------|
| **Systolic blood pressure** |
| SSPN | 12 | 26438189 | rs6487543 | A | 0.94 | 0.77 | 0.345 | 0.09 | 5.9x10^-5 | 0.279 | 0.06 | 2.1x10^-6 | 244,842 | 0.300 | 0.05 | 6.3x10^-10 |
| **Diastolic blood pressure** |
| MRA5 | 3 | 138119952 | rs2306374 | T | 1.00 | 0.84 | -0.237 | 0.05 | 9.3x10^-6 | -0.155 | 0.04 | 9.3x10^-5 | 281,715 | -0.184 | 0.03 | 7.4x10^-9 |
| **Pulse pressure** |
| CD34 | 1 | 208024820 | rs12731740 | T | 1.00 | 0.10 | -0.360 | 0.08 | 5.8x10^-6 | -0.202 | 0.05 | 1.1x10^-4 | 279,078 | -0.249 | 0.04 | 1.1x10^-8 |
| ZNF638 | 2 | 71627539 | rs3771371 | T | 1.00 | 0.57 | -0.223 | 0.05 | 4.1x10^-6 | -0.130 | 0.03 | 9.6x10^-5 | 280,285 | -0.160 | 0.03 | 5.8x10^-9 |
| CRACR2B | 11 | 828916 | rs7126805 | A | 1.00 | 0.73 | 0.262 | 0.05 | 1.1x10^-6 | 0.184 | 0.05 | 4.6x10^-4 | 145,162 | 0.222 | 0.04 | 3.3x10^-9 |

720 Locus: named according to the nearest annotated gene(s); Pos: build 37; EA: effect allele; INFO: imputation quality score from SNPTEST; EAF: effect allele frequency from discovery data in UK Biobank; Beta: effect estimate from linear regression; SE: Standard Error of effect estimate; P: P-value of association; N: total sample size analysed;
721 Note: within the UK Biobank discovery analysis sample size was N=140,882/140,886 for systolic and pulse pressure / diastolic pressure.
Table 2: Association results for new independent secondary variants identified at (a) novel loci and (b) previously reported blood pressure loci from either UK Biobank-GWAS or exome discovery. All listed secondary variants were validated in the replication meta-analyses and passed the conditional test for independence from the (a) sentinel novel variant from Table 1, or (b) previously reported SNVs (see Supplementary Tables 8 and 10).

<table>
<thead>
<tr>
<th>Secondary SNV in the locus</th>
<th>UK Biobank discovery</th>
<th>Replication</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locus Chr Pos rsID EA Trait</td>
<td>INFO EAF Beta SE P</td>
<td>Beta SE P</td>
<td>N Beta SE P</td>
</tr>
<tr>
<td>(a) Novel loci from UK Biobank GWAS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NADK-CPSF3L 1 1254436 rs1886773 A PP</td>
<td>0.99 0.03 -0.743 0.13 2.0x10^-8</td>
<td>-0.481 0.15 1.0x10^-3</td>
<td>233,789 -0.625 0.10 1.9x10^-10</td>
</tr>
<tr>
<td>RP4-710M16.1-PPAP2B 1 56938218 rs6588634 T PP</td>
<td>0.99 0.89 0.403 0.08 2.1x10^-7</td>
<td>0.270 0.07 4.7x10^-5</td>
<td>329,029 0.326 0.05 1.0x10^-10</td>
</tr>
<tr>
<td>FN1 2 216245694 rs34923683 A PP</td>
<td>1.00 0.02 0.837 0.15 4.8x10^-8</td>
<td>0.432 0.16 7.7x10^-3</td>
<td>285,653 0.646 0.11 6.8x10^-9</td>
</tr>
<tr>
<td>TPS3-SLC2A4 17 7185062 rs5417 A DBP</td>
<td>0.99 0.57 0.207 0.04 2.1x10^-7</td>
<td>0.207 0.04 1.1x10^-7</td>
<td>319,299 0.207 0.03 1.1x10^-13</td>
</tr>
<tr>
<td>KCNH4-HSD17B1 17 40709867 rs138643143 A PP</td>
<td>0.85 0.07 0.539 0.10 1.4x10^-7</td>
<td>0.420 0.15 5.8x10^-3</td>
<td>229,161 0.502 0.08 3.3x10^-9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) Previously reported GWAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK Biobank GWAS</td>
</tr>
<tr>
<td>Locus Chr Pos rsID EA Trait</td>
</tr>
<tr>
<td>RNF207 1 6683240 rs14057 A SBP</td>
</tr>
<tr>
<td>FIGN-GRB14 2 165513065 rs34271465 D SBP</td>
</tr>
<tr>
<td>ENPEP 4 111431444 rs3366350 A SBP</td>
</tr>
<tr>
<td>GUCY1A3-GUCY1B3 4 1564606054 rs146853253 D PP</td>
</tr>
<tr>
<td>EBF1 5 158220193 rs31864 A PP</td>
</tr>
<tr>
<td>EBF1 5 158448401 rs888987 C DBP</td>
</tr>
<tr>
<td>PDE3A 12 19979881 rs10841376 C SBP</td>
</tr>
<tr>
<td>PDE3A 12 20230639 rs10707612 A PP</td>
</tr>
<tr>
<td>PDE3A 12 20368269 rs60691990 T DBP</td>
</tr>
<tr>
<td>TBX5-TBX3 12 115928440 rs10850519 C DBP</td>
</tr>
<tr>
<td>MVH6 12 23761094 rs12050260 T PP</td>
</tr>
<tr>
<td>FURIN-FES 15 91427692 rs138682554 A SBP</td>
</tr>
<tr>
<td>HOXB7 17 46874272 rs585736 A PP</td>
</tr>
<tr>
<td>INSR 19 7258405 rs11671314 C SBP</td>
</tr>
<tr>
<td>JAG1 20 10669188 rs2206815 A PP</td>
</tr>
<tr>
<td>JAG1 20 10767811 rs1040922 T DBP</td>
</tr>
<tr>
<td>PREX1 20 47411149 rs80346118 A DBP</td>
</tr>
<tr>
<td>Locus</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td><strong>CRYAA-SIK1</strong></td>
</tr>
<tr>
<td><strong>ST7L-CAPZA1-MOV10</strong></td>
</tr>
<tr>
<td><strong>CDH17</strong></td>
</tr>
</tbody>
</table>
**Figure 1:** Study design schematic for discovery and validation of novel loci. N: sample size; QC: Quality Control; PCA: Principal Component Analysis; BP: blood pressure; SBP: systolic BP; DBP: diastolic BP; PP: pulse pressure; SNVs: single nucleotide variants; BMI: body mass index; UKB: UK Biobank; UKBL: UK BiLEVE; GWAS: Genome-wide association study; MAF: Minor Allele Frequency; P: P-value; LD: Linkage Disequilibrium; 1000G: 1000 Genomes.
Figure 2: UK Biobank GWAS discovery Manhattan plots and Venn diagram of 107 novel validated loci. Plots (A), (B) and (C) show the UK Biobank GWAS discovery circos Manhattan plots for systolic blood pressure (SBP), diastolic blood pressure (DBP) and pulse pressure (PP) respectively. P-value results are plotted on a $-\log_{10}$ scale (see legend) for all $\sim$9.8 million variants with Minor Allele Frequency (MAF) $\geq$ 1% and imputation quality INFO $> 0.1$ analysed within the GWAS discovery. Associations are plotted in red for all variants within validated novel loci, in black for variants within novel loci which were looked-up ($P<1\times10^{-6}$) in replication data but did not replicate, in blue for all variants within previously reported blood pressure loci, and grey otherwise. Loci names labelled around the edge are specific to each blood pressure trait, with red labels corresponding to novel loci validated for the given trait (102 novel loci from Table 1a in total across plots (A-C) from GWAS), and blue labels corresponding to previously reported loci within which new independent secondary variants were identified (20 GWAS variants in total from Table 2b). Plot (D) presents a Venn diagram, showing concordance of significant associations across the three blood pressure phenotypes for the 107 novel sentinel variants (Table 1) from both the GWAS and exome analyses.
Figure 3: Association of blood pressure loci with other traits. Plot (A) shows results for associations with other traits which were extracted from the PhenoScanner database for the sentinel novel variants from Table 1, including proxies in Linkage Disequilibrium ($r^2 \geq 0.8$), with genome-wide significant associations ($P < 5 \times 10^{-8}$). The loci are grouped by blood pressure traits ordered right to left according to the loci in Table 1. There are four systolic blood pressure associated loci, 14 diastolic blood pressure associated loci and nine pulse pressure associated loci with associations with other traits reported in the literature. Traits are grouped into different disease categories: “Pulse/HR” includes pulse, heart rate, pulse wave velocity and aortic stiffness traits; “CAD/CHD/MI”: Coronary Artery Disease / Coronary Heart Disease / Myocardial Infarction; “Blood” traits: Haemoglobin levels and platelet counts; “Lipids”: LDL and Total Cholesterol; “BMI/WHR” includes Body Mass Index, weight, obesity, waist or hip circumference, Waist-Hip-Ratio; “Menarche”: age at menarche; “Lung”: lung function (FEV1); “Alzheimer’s” traits refers to Cerebrospinal fluid levels of Alzheimer’s disease related proteins; “Cancer” includes carcinomas, neuroblastomas, bladder cancer; “Education”: years of educational attainment. Plots (B) and (C) show mouse phenotype enrichment and disease ontology enrichment, respectively, of novel and previously reported variants. Enrichment was performed using the GREAT tool (http://bejerano.stanford.edu/great) with the sentinel SNVs as query.
Figure 4: Distribution of a Genetic Risk Score (GRS) based on novel and previously reported blood pressure variants and its relationship with blood pressure levels, hypertension and cardiovascular disease (CVD) outcomes. (A): Distribution of GRS in the independent Airwave study and odds ratio of hypertension at age 50+ comparing each of the upper four GRS quintiles with the lowest quintile. (B): Mean blood pressures in Airwave study age 50+ across GRS quintiles. (C): Distribution of GRS in UK Biobank and odds ratio of CVD, Coronary Artery Disease (CAD) and stroke comparing each of the upper four GRS quintiles with the lowest quintile. (D) Number of CVD, CAD and stroke outcomes (self-reports, events and deaths) across GRS quintiles in UK Biobank participants.

<table>
<thead>
<tr>
<th>(mmHg)</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBP</td>
<td>136.4</td>
<td>138.7</td>
<td>142.5</td>
<td>142.5</td>
<td>145.7</td>
</tr>
<tr>
<td>DBP</td>
<td>82.9</td>
<td>84.3</td>
<td>85.9</td>
<td>86.0</td>
<td>87.6</td>
</tr>
<tr>
<td>PP</td>
<td>53.4</td>
<td>54.4</td>
<td>56.7</td>
<td>56.5</td>
<td>58.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(count)</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVD</td>
<td>2,462</td>
<td>2,689</td>
<td>2,759</td>
<td>2,874</td>
<td>3,229</td>
</tr>
<tr>
<td>CAD</td>
<td>1,783</td>
<td>1,996</td>
<td>2,078</td>
<td>2,154</td>
<td>2,417</td>
</tr>
<tr>
<td>Stroke</td>
<td>581</td>
<td>597</td>
<td>640</td>
<td>695</td>
<td>776</td>
</tr>
</tbody>
</table>
**Figure 5:** Summary of novel gene cardiovascular expression. Genes are shown on the basis of their tissue expression and supporting evidence summarised in Supplementary Table 14, based on Knockout (KO) phenotype, previously reported blood pressure biology or a strong functional rationale: eQTL (expression Quantitative Trait Loci), nsSNV (non-synonymous SNV), Hi-C. Multiple lines of evidence indicate the central importance of the vasculature in blood pressure regulation and we thus highlight existing drugged (*) and druggable (#) targets among these genes. Illustrations used elements with permission from Servier Medical Art: www.servier.fr/servier-medical-art.
Online Methods

**UK Biobank data**

Our Genome Wide Association Study (GWAS) analysis is performed using data from the interim release of the first ~150k UK Biobank participants (Supplementary Methods)\(^{17}\). These consist of ~100k individuals from UK Biobank genotyped at ~800,000 single nucleotide variants (SNVs) with a custom Affymetrix UK Biobank Axiom Array chip\(^{66}\) and ~50k individuals genotyped with a custom Affymetrix UK BiLEVE Axiom Array chip from the UK BiLEVE study\(^ {67}\), which is a subset of UK Biobank. SNVs were imputed centrally by UK Biobank using a merged UK10K sequencing + 1000 Genomes imputation reference panel.

**Quality control**

Following quality control (QC) procedures already carried out centrally by UK Biobank, we exclude discordant SNVs and samples with QC failures, gender discordance and high heterozygosity/missingness. We further restrict our data to a subset of individuals of European ancestry. By applying *k*-means clustering to the Principal Component Analysis (PCA) data a total of N=145,315 Europeans remain. Then we use the kinship data to exclude 1\(^{st}\) and 2\(^{nd}\) degree relatives, with N=141,647 unrelated individuals remaining. Finally we restrict our data to non-pregnant individuals with two automated BP measurements available, resulting in a maximum of N=140,886 individuals for analysis (Supplementary Methods).

**Phenotypic data**

After calculating the mean systolic and diastolic pressure values from the two blood pressure measurements, we adjust for medication use by adding 15 and 10 mmHg to systolic and diastolic pressure, respectively, for individuals reported to be taking blood pressure-lowering medication (21.4% of individuals)\(^ {68}\). Pulse Pressure is calculated as systolic minus diastolic pressure, according to the medication-adjusted traits. Hypertension, used in secondary analyses, is defined as: (i) systolic pressure ≥ 140 mmHg, or (ii) diastolic pressure ≥ 90 mmHg, (iii) or taking blood pressure-lowering medication; otherwise individuals are classified as non-hypertensive. Descriptive summary statistics are provided for all individuals, and stratified by UK Biobank vs UK BiLEVE participants (Supplementary Table 1).

**Analysis models**

For the GWAS, we perform linear regression analyses of the three (untransformed) continuous, medication-adjusted BP traits (systolic, diastolic and pulse pressure) for all measured and imputed genetic variants in dosage format using SNPTEST software\(^ {69}\) under an additive genetic model. We carry out a similar analysis for the exome content. Each analysis includes the following covariates: sex, age, age\(^2\), body mass index, top ten PCs and a binary indicator variable for UK Biobank vs UK BiLEVE to adjust for the different genotyping chips. We also run an association analysis within UK Biobank for validated novel blood pressure SNVs and hypertension using logistic regression under an additive model with adjustments as above. There are 76,554 hypertensive cases and the 64,384 remaining participants are treated as non-hypertensive controls. This sample size is slightly larger than the N=140,866...
used in the main analyses, since participants with only one blood pressure measurement, but with reported blood pressure-lowering medication, could be included as hypertensive.

**Previously reported variants**

We compile a list of all SNVs previously reported to be associated with blood pressure (Supplementary Table 12). This list includes all published SNVs which have been identified and validated from previous GWAS, CardioMetabochip and exome chip projects\textsuperscript{10-12}. We augment this list to include all 34,459 SNVs in Linkage Disequilibrium (LD) with the previously reported SNVs, according to a threshold of $r^2 \geq 0.2$. Results for all these variants are extracted for each of the three blood pressure traits, to check previously reported blood pressure associations in the UK Biobank data, according to whether the sentinel SNV or a variant at the locus in LD ($r^2 \geq 0.2$) with it reached nominal significance ($P < 0.01$) for association with at least one of the three BP traits.

**Replication strategy**

We use three independent external data sets for replication (Supplementary Methods). First, for the GWAS analysis based on advanced 1000 Genomes imputation enhanced by UK10K data we consider SNVs with MAF ≥ 1% and perform a reciprocal replication exchange with the International Consortium of Blood Pressure (ICBP) 1000 Genomes meta-analysis (max N = 150,134). The imputation strategy for ICBP 1000 Genomes meta-analysis is based on an earlier imputation grid for the 1000 Genomes project. In addition, we recruit further cohorts with 1000 Genomes data which had not contributed to the ICBP-1000 Genomes discovery meta-analysis: ASCOT-UK (N = 3,803), ASCOT-SC (N = 2,462), BRIGHT (N = 1,791), Generation Scotland (GS) (N = 9,749), EGCUT (N = 5,468), Lifelines (N = 13,292) and PREVEND (N = 3,619). This gives a total of N = 190,318 independent replication samples for the GWAS discovery.

Second, because the UK Biobank and UK BiLEVE genotyping chips contain exome content, we sought replication from two blood pressure exome consortia (European exome consortium and the Cohorts for Heart and Ageing research in Genome Epidemiology – CHARGE BP exome consortium), to allow validation of coding variants and variants with lower frequency. The European exome consortium (N = 161,926) and CHARGE consortium (N = 119,792) give a total of N = 281,718 independent replication samples for the UK Biobank exome discovery.

Note that the lookups for GWAS and exome discovery are distinct sets of SNVs. Loci are assigned sequentially, prioritising the primary GWAS discovery first, then considering any remaining loci with non-overlapping exome content for replication in the independent exome replication resources.

**Statistical criteria for replication**

For the GWAS discovery, there are ~9.8 million SNVs with MAF ≥ 1% and INFO > 0.1. We consider for follow-up any SNVs with $P < 1 \times 10^{-5}$ for any of the three blood pressure traits. For the exome discovery, there are 149,026 exome SNVs (Supplementary Methods) which were polymorphic with INFO > 0.1; for follow-up we consider all SNVs with MAF ≥ 0.01% and $P < 1 \times 10^{-5}$. All such SNVs are annotated to loci according to both an LD threshold of $r^2 \geq 0.2$ and a
1Mb interval region (see Supplementary Methods), and signals are classified either as belonging to novel loci, or being potential secondary signals at previously reported loci.

**Selection of variants for follow-up**

The sentinel (most significant) SNV from each association signal is selected for follow-up, all of which are pairwise-independent by LD ($r^2 < 0.2$). For the GWAS discovery, we check that potential lookup SNVs are covered within the ICBP-1000G replication data (Supplementary Methods). Of the 235 novel loci containing previously unreported SNVs with MAF $\geq 1\%$, INFO $> 0.1$ and $P<1\times10^{-6}$, 218 are covered, and similarly 100 of the 123 potential secondary SNVs at 51 of the 54 previously reported BP loci are available for follow-up. For the exome discovery, by following up SNVs with MAF $\geq 0.01\%$, INFO $> 0.1$ and $P < 1\times10^{-5}$ across the three blood pressure traits, we carry forward for replication sentinel SNVs at 22 novel loci, and potential secondary SNVs at three previously reported loci. We produce locus zoom plots for each of the lookup variants.

**Replication meta-analyses**

The replication and combined meta-analyses were perform within METAL software\textsuperscript{70} using fixed effects inverse variance weighted meta-analysis (Supplementary Methods). The combined meta-analysis of both the UK Biobank discovery (N = 140,886) and GWAS replication meta-analysis (max N = 190,070) include a total maximum sample size of N = 330,956. For the exome combined meta-analysis, we synthesize data from the UK Biobank discovery exome content (max N=140,866), with the replication dataset from both exome consortia (total max N=281,718), giving a maximum sample size of N=422,604.

**Validation Criteria**

In our study a signal is declared validated if it satisfies ALL of the following three criteria:

(i) the sentinel SNV is genome-wide significant ($P < 5\times10^{-8}$) in the combined meta-analysis for any of the three blood pressure traits;

(ii) the sentinel SNV is significant ($P < 0.01$) in the replication meta-analysis alone for association with the most significantly associated blood pressure trait from the combined meta-analysis;

(iii) the sentinel SNV has concordant direction of effect between the UK Biobank discovery and the replication meta-analysis for the most significantly associated blood pressure trait from the combined meta-analysis.

**Secondary signals**

By conditional analysis within UK Biobank data we assess all validated secondary signals from novel and previously reported loci for independence from the sentinel or previously reported SNV, respectively (Supplementary Methods). We declare a secondary signal to be independent of the previously reported SNV if there is less than a 1.5 fold difference between the main association and conditional association $P$-values on a $-\log_{10}$ scale, i.e. if $-\log_{10}(P) / -\log_{10}(P_{\text{cond}}) < 1.5$. Note that the lookup criteria already ensure that the secondary variant
is not in LD ($r^2 < 0.2$) with the previously reported SNV. If more than one SNV in a region is found to be independent we undertake further rounds of iterative conditional analysis.

**Lookups in non-European ancestries**

As a secondary analysis, we look up 102 and 5 novel validated SNVs from the UK Biobank-GWAS and exome analyses, respectively, in non-European ancestry samples. These comprise analysis of East Asian (N = 31,513) and South Asian (N = 33,115) ancestry data from the iGEN-BP consortium\(^{13}\) for the GWAS lookups, and South Asian (N = 25,937), African American (N = 21,488) and Hispanic (N = 4,581) ancestry data from the CHARGE BP exome consortium\(^{12}\) and CHD+ Exome consortium\(^{11}\), for the exome content lookups (Supplementary Methods). We carry out a binomial (sign) test based on the number of SNVs with consistent directions of effect between UK Biobank and each of the non-European ancestry samples.

**Monogenic blood pressure gene lookups**

The UK Biobank and UK BiLEVE arrays include some rare coding variants for monogenic disorders. We collate a list of all specific mutation variants within genes known to be associated with monogenic blood pressure disorders\(^{22}\). Results from the UKB discovery association analyses for all three blood pressure traits are extracted for any of these SNVs directly covered within the UK Biobank dataset (Supplementary Table 13). Note that a search of proxies did not augment the list of available variants, so results are reported for the specific variants only.

**Functional analyses**

In order to prioritise associated SNVs, we use an integrative bioinformatics approach to collate functional annotation at both the variant and gene level for each SNV within the blood pressure loci (all SNVs in LD $r^2 \geq 0.8$ with the blood pressure-associated SNVs). At the variant level we use ANNOVAR\(^{71}\) to obtain comprehensive functional characterisation of variants, including gene location, conservation and amino acid substitution impact based on a range of prediction tools.

We use the University of California Santa Cruz (UCSC) genome browser to review sequence specific context of SNVs in relation to function, particularly in the Encyclopedia of DNA Elements (ENCODEx) dataset\(^{72}\). We use the UCSC table browser to annotate SNVs in ENCODE regulatory regions. We evaluate SNVs for impact on putative micro RNA target sites in the 3’ untranslated regions (3’UTR) of transcripts by a query of the miRNASNP database\(^{73}\). We evaluate all SNVs in LD ($r^2 \geq 0.8$) with our novel sentinel SNVs for evidence of mediation of expression quantitative trait loci (eQTL) in all 44 tissues using the Genotype-Tissue Expression (GTEx) database (www.gtexportal.org), in order to identify novel loci which are highly expressed, and to highlight specific tissue types which show eQTLs for a large proportion of novel loci. We further seek to identify novel loci with the strongest evidence of eQTL associations in arterial tissue, in particular.

At the gene level, we use Ingenuity Pathway Analysis (IPA) software (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) to review genes with prior links to blood pressure, based on annotation with the “Blood Pressure” Medline Subject Heading (MESH) term which is
annotated to 684 genes. We also use IPA to identify genes which interact with blood pressure MESH annotated genes, and evaluate genes for evidence of small molecule druggability based on queries of Chemb (www.ebi.ac.uk/chembl/) and Drug Gene Interaction database (dgideb.genome.wustl.edu).

We then perform overall enrichment testing across all loci. Firstly, we use DEPICT\(^74\) (Data-driven Expression Prioritized Integration for Complex Traits) to identify highly expressed tissues and cells within the blood pressure loci. DEPICT uses a large number of microarrays (~37k) to identify cells and tissues where the genes are highly expressed and uses precomputed GWAS phenotypes to adjust for co-founding sources. DEPICT provides a *P*-value of enrichment and false discovery rates adjusted *P*-values for each tissue/cells tested.

Furthermore, to investigate regulatory regions, we employ a two tiered approach to investigate cell type specific enrichment within DNase I sites using FORGE, which tests for enrichment of SNVs within DNase I sites in 123 cell types from the Epigenomics Roadmap Project and ENCODE\(^75\) (Supplementary Methods). Novel sentinel SNVs discovered in our study are analysed along with previously reported SNVs and secondary signals (with *P*-value < 1×10\(^{-4}\)) to evaluate the overall tissue specific enrichment of blood pressure associated variants. In a second analysis we use FORGE (with no LD filter) to investigate directly our curated candidate regulatory SNVs for overlap with cell-specific DNase I signals.

GenomeRunner\(^76\) is used to search for enrichment of novel and previously reported sentinel SNVs with histone modification mark genomic features (Supplementary Methods). Relevant cardiovascular tissue expression is investigated using Fantom5 reference transcript expression data (fantom.gsc.riken.jp/5) (Supplementary Methods).

We use IPA (IPA\(^®\), QIAGEN Redwood City, www.qiagen.com/ingenuity) to identify biological pathways and transcriptional upstream regulators enriched for genes within the blood pressure loci. The transcriptional upstream regulator analysis aims to identify transcription factors, compounds, drugs, kinases and other molecules, for which the target is one of the blood pressure genes under investigation.

We query SNVs against PhenoScanner\(^19\) to investigate trait pleiotropy, extracting all association results with nominal significance at *P* < 0.05 for full reporting (Supplementary Table 14), and then extract genome-wide significant results to highlight the novel loci with strongest evidence of association with other traits (Fig. 3a). We also use the Genomic Regions Enrichment of Annotations Tool (GREAT) to study gene set enrichment of mouse phenotype and disease ontology terms within our novel and previously reported loci, using default SNV to gene mapping settings\(^77\).

We carry out metabolomics analysis using two sets of data. First we use \(^1\)H NMR lipidomics data on plasma from a subset of 2,000 participants of the Airwave Health Monitoring Study\(^78,79\) (Supplementary Methods). For each replicated blood pressure-associated SNV we ran association tests with the lipidomics data using linear regression analyses, adjusted for age and sex. We computed significance thresholds using a permutation derived family wise error rate (5%) to account for the high correlation structure of these data (ENT=35)\(^80\). We also test each replicated SNV against published genome-wide vs metabolome-wide associations.
in plasma and urine using publicly available data from the “Metabolomics GWAS Server” to identify metabolites that have been associated with variants of interest at $P < 3.0 \times 10^{-4}$ (Bonferroni corrected $P$ for validated signals)$^{25,26}$.

**Experimental methods**

We prioritise novel genes for laboratory testing on the basis of evidence for SNV function (including coding variants, eQTLs and Hi-C interactions), biological support for relevance to blood pressure (from literature review) and transgenic phenotype. We perform genotyping and Quantitative Reverse-Transcription Polymerase Chain Reaction (q RT-PCR) for the selected sentinel variants of interest using human vascular smooth muscle cells and endothelial cells and test for expression levels (Supplementary Methods).

**Genetic risk scores**

First, by calculating genetic risk scores (GRS), we use the Airwave study$^{78}$ data to assess the effect in an independent cohort of the blood pressure-associated variants on blood pressure and risk of hypertension (Supplementary Methods). This provides an estimate of the combined effect of the blood pressure raising variants avoiding bias by “winners curse”. We create three trait-specific weighted GRSs (i.e. systolic, diastolic and pulse pressure), for all pairwise-independent, LD-filtered ($r^2 < 0.2$) previously reported variants and validated novel variants (sentinel and secondary SNVs) combined, using SNVs available in Airwave (Supplementary Table 20). For the previously reported variants, we weight blood pressure increasing alleles by the trait-specific beta coefficients from the UK Biobank discovery GWAS. For the novel variants, beta coefficients of the replication meta-analysis for each blood pressure trait are used as independent, unbiased weights.

For risk score analyses we derive an average blood pressure GRS, as the average of the systolic and diastolic pressure GRSs. We standardize the GRS to have mean of zero and standard deviation of one. We assess the association of the continuous GRS variable with corresponding blood pressure trait by simple linear regression. We also run a logistic regression to examine the association of each GRS with risk of hypertension. We perform each analysis both with and without adjustment for sex, for comparison. We compare blood pressure levels and risk of hypertension for individuals in the top and bottom 20% of the GRS distribution at ages 50 years and over using linear and logistic regression, respectively.

To calculate the percent of variance for each blood pressure trait explained by its corresponding trait-specific GRS, not accounted for by known factors, we generate the residuals from the regression model of each trait against covariates of age, age-square, sex and body mass index. We then fit a second linear model for the trait residuals with all the variants in the GRS plus the top 10 principal components. Within the Airwave study, these percentage variance explained results are calculated within an independent population.

We also assess the association of the GRSs with cardiovascular outcomes in the UK Biobank data, based on self-reported medical history, and linkage to hospitalization and mortality data. We include all pairwise-independent previously reported blood pressure variants and validated novel variants. We use logistic regression with binary outcome variables for
coronary heart disease, stroke and cardiovascular disease (see Supplementary Methods) and
GRS as explanatory variable (with and without sex adjustment).

URLs

FORGE (accessed 16 Aug 2016),
http://browser.1000genomes.org/Homo_sapiens/UserData/Forge?db=core


ENCODE DNase I data (wgEncodeAwgDnaseMasterSites; accessed 20 Aug 2016 using Table
browser)

ENCODE cell type data (accessed 20 Aug 2016),
http://genome.ucsc.edu/ENCODE/cellTypes.html.

Exome chip design:

http://genome.sph.umich.edu/wiki/Exome_Chip_Design

References


67. Wain, L.V. et al. Novel insights into the genetics of smoking behaviour, lung function, and
chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK

68. Tobin, M.D., Sheehan, N.A., Scurrah, K.J. & Burton, P.R. Adjusting for treatment effects in
studies of quantitative traits: antihypertensive therapy and systolic blood pressure. Stat Med


70. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide

71. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from


73. Gong, J. et al. Genome-wide identification of SNPs in microRNA genes and the SNP effects on

74. Pers, T.H. et al. Biological interpretation of genome-wide association studies using predicted

75. Dunham, I., Kulesha, E., Iotchkova, V., Morganella, S. & Birney, E. FORGE: A tool to discover
cell specific enrichments of GWAS associated SNPs in regulatory regions [version 1; referees:
2 approved with reservations], (2015).

76. Dozmorov, M.G., Cara, L.R., Giles, C.B. & Wren, J.D. GenomeRunner: automating genome

