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 Warren[1,2]*, Evangelos Evangelou[3,4]*, Claudia P Cabrera[1,2]*, He Gao[3,5]*, Meixia Ren[1,6]*, Borbala Mifsud[1]*, Ioanna Ntalla[1], Praveen Surendran[7], Chunyu Liu[8-10], James P Cook[11], Aldi Kraja[12], Fotios Drenos[13], Marie Loh[3,14], Niek Verweij[15], Jonathan Marten[16], Ibrahim Karaman[3,5], Marcelo P Segura Lepe[3,17], Paul O'Reilly[18], Joanne Knight[19], Harold Snieder[20], Norihiro Kato[21], Jiang He[22], E Shyong Tai[23,24], Abdullah M Said[15], David Porteous[25], Maris Alver[26], Neil Poulter[27], Martin Farrall[28], Ron T Gansevoort[29], Sandosh Padmanabhan[30], Reedik Mägi[26], Alice Stanton[31], John Connell[32], Stephan J L Bakker[33], Andres Metspalu[26], Denis Shields[34], Simon Thom[35], Morris Brown[1,2], Peter Sever[36], Tõnu Esko[26,37], Caroline Hayward[16], Pim van der Harst[15], Danish Saleheen[38], Rajiv Chowdhury[7], John C Chambers[3,39-41], Daniel I Chasman[42,43], Aravinda Chakravarti[44], Christopher Newton-Cheh[37,45,46], Cecilia M Lindgren[47], Daniel Levy[8,10], Jaspal S Kooner[40,48,49], Bernard Keavney[50], Maciej Tomaszewski[50], Nilesh J Samani[51,52], Joanna M M Howson[7] , Martin D Tobin[53], Patricia B Munroe[1,2], Georg B Ehret[44,54], Louise V Wain[53], Michael R Barnes[1,2]*, Ioanna Tzoulaki[3-5]*, Mark J Caulfield[1,2]*†, Paul Elliott[3,5]*†

On behalf of the UK Biobank CardioMetabolic Consortium BP working group;
in collaboration with The International Consortium of Blood Pressure (ICBP) 1000G Analyses, The CHD Exome+ Consortium, The ExomeBP Consortium, The T2D-GENES Consortium, The GoT2DGenes Consortium, The Cohorts for Heart and Ageing Research in Genome Epidemiology (CHARGE) BP Exome Consortium, and The International Genomics of Blood Pressure (iGEN-BP) Consortium

* Equal contribution
† Corresponding author

1 William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.
2 National Institute for Health Research Cardiovascular Biomedical Research Unit at Barts, Barts Heart Centre, Queen Mary University of London, London, UK.
3 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, UK.
4 Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece.
5 MRC-PHE Centre for Environment and Health, Imperial College London, London, UK.
6 Department of Cardiology, Fujian Provincial Hospital, Fujian Medical University, Fuzhou, China.
7 Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
8 Population Sciences Branch, National Heart Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA.
9 National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA.

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, Lancester 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, 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[5-8]. 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 $P < 5x10^{-8}$, 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 *PAX2* ($P = 3.6 \times 10^{-18}$); for diastolic pressure, which plateaus in middle age, with rs76326501 near *METTL21A- ACO16735.1* ($P = 3.6 \times 10^{-18}$); and rs3889199 near *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 (*NADK-CPSF3L*, *GTF2B*, *METTL21A*-AC079767.3 and *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 *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 ~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 (*ACE*) locus ($P = 6.8 \times 10^{-14}$), from the renin-angiotensin system, a pathway which is targeted by current blood pressure treatments (ACE-inhibitors), as well as several other signals at known hypertension drug targets. These include *CACNA2D2* (rs743757, $P = 2.4 \times 10^{-10}$) targeted by calcium channel blockers, *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[12,13], 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[19] 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[20].

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. 2 a-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[21] 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[22] 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 ≥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[24].

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 ^{1}H 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[25,26](**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[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[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[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 have previously been implicated in the genetics of blood pressure. Several of our novel loci 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[28-30] (**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; $r^2 = 0.17$)[31]. 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[31]. 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[32] 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[33].

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[34-36]. 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[37,38].

We identify several loci containing genes involved in vascular signalling and second messenger systems such as *PDE5A* and *PDE10A*[39-41]. The phosphodiesterase PDE5A hydrolyses cyclic GMP and is inhibited by sildenafil which leads to vasodilatation[42]. 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[21,43,44]. 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[45] and blood pressure response to nifedipine, a calcium channel blocker antihypertensive drug[46]. *SLC8A1* encodes a sodium calcium exchanger expressed in cardiomyocytes which alters cardiac contractility and hypertrophy and shows abnormal blood pressure in SLC8A1 transgenic mice[47]. Variants at *SLC35F1* have been previously associated with resting heart rate and ventricular dimensions which could contribute to blood pressure elevation[48].

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[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[53-55].

Our analysis validates loci containing genes with prior physiological connection to blood pressure such as *BDNF, FAM208A,* and *CACNA2D2*[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[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[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[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 ~200kb 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[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[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)[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[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

1. Forouzanfar, M.H. *et al.* Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **386**, 2287-323 (2015).
2. Sundstrom, J. *et al.* Blood pressure-lowering treatment based on cardiovascular risk: a meta-analysis of individual patient data. *Lancet* **384**, 591-8 (2014).
3. Joffres, M. *et al.* Hypertension prevalence, awareness, treatment and control in national surveys from England, the USA and Canada, and correlation with stroke and ischaemic heart disease mortality: a cross-sectional study. *BMJ Open* **3**, e003423 (2013).
4. Falaschetti, E., Mindell, J., Knott, C. & Poulter, N. Hypertension management in England: a serial cross-sectional study from 1994 to 2011. *Lancet* **383**, 1912-9 (2014).
5. Poulter, N.R., Prabhakaran, D. & Caulfield, M. Hypertension. *Lancet* **386**, 801-12 (2015).
6. Feinleib, M. *et al.* The NHLBI twin study of cardiovascular disease risk factors: methodology and summary of results. *Am J Epidemiol* **106**, 284-5 (1977).
7. Mongeau, J.G., Biron, P. & Sing, C.F. The influence of genetics and household environment upon the variability of normal blood pressure: the Montreal Adoption Survey. *Clin Exp Hypertens A* **8**, 653-60 (1986).
8. Munoz, M. *et al.* Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat Genet* (2016).
9. Cabrera, C.P. *et al.* Exploring hypertension genome-wide association studies findings and impact on pathophysiology, pathways, and pharmacogenetics. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **7**, 73-90 (2015).
10. Ehret, G.B. *et al.* The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nature genetics* **in press** (2016).
11. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nature genetics* **in press** (2016).
12. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nature genetics* **in press** (2016).
13. Kato, N. *et al.* Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat Genet* **47**, 1282-93 (2015).
14. Elliott, P. & Peakman, T.C. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* **37**, 234-44 (2008).
15. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
16. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun* **6**, 8111 (2015).
17. The UK Biobank. Genotype imputation and genetic association studies of UKBiobank. (2015).
18. Wain, L.V. *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* **43**, 1005-11 (2011).
19. Staley, J.R. *et al.* PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* (2016).
20. Ettehad, D. *et al.* Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. *Lancet* **387**, 957-67 (2016).
21. Kato, N. *et al.* Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet* **43**, 531-8 (2011).
22. Munroe, P.B., Barnes, M.R. & Caulfield, M.J. Advances in Blood Pressure Genomics. *Circulation Research* **112**, 1365-1379 (2013).

23. den Hoed, M. *et al.* Identification of heart rate–associated loci and their effects on cardiac conduction and rhythm disorders. *Nature genetics* **45**, 621-631 (2013).
24. Hamilton, C.A., Brosnan, M.J., McIntyre, M., Graham, D. & Dominiczak, A.F. Superoxide excess in hypertension and aging: a common cause of endothelial dysfunction. *Hypertension* **37**, 529-34 (2001).
25. Shin, S.Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat Genet* **46**, 543-50 (2014).
26. Raffler, J. *et al.* Genome-Wide Association Study with Targeted and Non-targeted NMR Metabolomics Identifies 15 Novel Loci of Urinary Human Metabolic Individuality. *PLoS Genet* **11**, e1005487 (2015).
27. Ehret, G.B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
28. van Setten, J. *et al.* Genome-wide association study of coronary and aortic calcification implicates risk loci for coronary artery disease and myocardial infarction. *Atherosclerosis* **228**, 400-5 (2013).
29. McCarthy, J.J. *et al.* Large scale association analysis for identification of genes underlying premature coronary heart disease: cumulative perspective from analysis of 111 candidate genes. *J Med Genet* **41**, 334-41 (2004).
30. van Meurs, J.B. *et al.* Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. *Am J Clin Nutr* **98**, 668-76 (2013).
31. Pu, X. *et al.* ADAMTS7 cleavage and vascular smooth muscle cell migration is affected by a coronary-artery-disease-associated variant. *Am J Hum Genet* **92**, 366-74 (2013).
32. Bauer, R.C. *et al.* Knockout of Adamts7, a novel coronary artery disease locus in humans, reduces atherosclerosis in mice. *Circulation* **131**, 1202-13 (2015).
33. Boardman-Pretty, F. *et al.* Functional Analysis of a Carotid Intima-Media Thickness Locus Implicates BCAR1 and Suggests a Causal Variant. *Circ Cardiovasc Genet* **8**, 696-706 (2015).
34. Touyz, R.M. & Montezano, A.C. Vascular Nox4: a multifarious NADPH oxidase. *Circ Res* **110**, 1159-61 (2012).
35. Steppan, J., Barodka, V., Berkowitz, D.E. & Nyhan, D. Vascular stiffness and increased pulse pressure in the aging cardiovascular system. *Cardiol Res Pract* **2011**, 263585 (2011).
36. Yan, F. *et al.* Nox4 and redox signaling mediate TGF-beta-induced endothelial cell apoptosis and phenotypic switch. *Cell Death Dis* **5**, e1010 (2014).
37. Chan, E.C. *et al.* Nox4 modulates collagen production stimulated by transforming growth factor beta1 in vivo and in vitro. *Biochem Biophys Res Commun* **430**, 918-25 (2013).
38. Vasa-Nicotera, M. *et al.* miR-146a is modulated in human endothelial cell with aging. *Atherosclerosis* **217**, 326-30 (2011).
39. Tian, X. *et al.* Phosphodiesterase 10A upregulation contributes to pulmonary vascular remodeling. *PLoS One* **6**, e18136 (2011).
40. Takimoto, E. *et al.* Chronic inhibition of cyclic GMP phosphodiesterase 5A prevents and reverses cardiac hypertrophy. *Nat Med* **11**, 214-22 (2005).
41. Perez, N.G. *et al.* Phosphodiesterase 5A inhibition induces Na+/H+ exchanger blockade and protection against myocardial infarction. *Hypertension* **49**, 1095-103 (2007).
42. Oliver, J.J., Melville, V.P. & Webb, D.J. Effect of regular phosphodiesterase type 5 inhibition in hypertension. *Hypertension* **48**, 622-7 (2006).
43. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat Genet* **41**, 677-87 (2009).
44. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet* **41**, 666-76 (2009).
45. DeStefano, A.L. *et al.* Autosomal dominant orthostatic hypotensive disorder maps to chromosome 18q. *Am J Hum Genet* **63**, 1425-30 (1998).

46. Hong, X. *et al.* Genetic polymorphisms of the urea transporter gene are associated with antihypertensive response to nifedipine GITS. *Methods Find Exp Clin Pharmacol* **29**, 3-10 (2007).
47. Takimoto, E. *et al.* Sodium calcium exchanger plays a key role in alteration of cardiac function in response to pressure overload. *FASEB J* **16**, 373-8 (2002).
48. Ronaldson, P.T. & Davis, T.P. Targeting transporters: promoting blood-brain barrier repair in response to oxidative stress injury. *Brain Res* **1623**, 39-52 (2015).
49. Carta, L. *et al.* Fibrillins 1 and 2 perform partially overlapping functions during aortic development. *J Biol Chem* **281**, 8016-23 (2006).
50. Kazenwadel, J. *et al.* Loss-of-function germline GATA2 mutations in patients with MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature. *Blood* **119**, 1283-91 (2012).
51. Akashi, M., Higashi, T., Masuda, S., Komori, T. & Furuse, M. A coronary artery disease-associated gene product, JCAD/KIAA1462, is a novel component of endothelial cell-cell junctions. *Biochem Biophys Res Commun* **413**, 224-9 (2011).
52. Cakstina, I. *et al.* Primary culture of avian embryonic heart forming region cells to study the regulation of vertebrate early heart morphogenesis by vitamin A. *BMC Dev Biol* **14**, 10 (2014).
53. Wang, J., Karra, R., Dickson, A.L. & Poss, K.D. Fibronectin is deposited by injury-activated epicardial cells and is necessary for zebrafish heart regeneration. *Dev Biol* **382**, 427-35 (2013).
54. Dietrich, T. *et al.* ED-B fibronectin (ED-B) can be targeted using a novel single chain antibody conjugate and is associated with macrophage accumulation in atherosclerotic lesions. *Basic Res Cardiol* **102**, 298-307 (2007).
55. Stoynev, N. *et al.* Gene expression in peripheral blood of patients with hypertension and patients with type 2 diabetes. *J Cardiovasc Med (Hagerstown)* **15**, 702-9 (2014).
56. Erdos, B., Backes, I., McCowan, M.L., Hayward, L.F. & Scheuer, D.A. Brain-derived neurotrophic factor modulates angiotensin signaling in the hypothalamus to increase blood pressure in rats. *Am J Physiol Heart Circ Physiol* **308**, H612-22 (2015).
57. Chan, S.H., Wu, C.W., Chang, A.Y., Hsu, K.S. & Chan, J.Y. Transcriptional upregulation of brain-derived neurotrophic factor in rostral ventrolateral medulla by angiotensin II: significance in superoxide homeostasis and neural regulation of arterial pressure. *Circ Res* **107**, 1127-39 (2010).
58. Crespo, K., Menard, A. & Deng, A.Y. Retinoblastoma-associated protein 140 as a candidate for a novel etiological gene to hypertension. *Clin Exp Hypertens* **38**, 533-40 (2016).
59. Watanabe, Y. *et al.* Accumulation of common polymorphisms is associated with development of hypertension: a 12-year follow-up from the Ohasama study. *Hypertens Res* **33**, 129-34 (2010).
60. Gale, D.P., Harten, S.K., Reid, C.D., Tuddenham, E.G. & Maxwell, P.H. Autosomal dominant erythrocytosis and pulmonary arterial hypertension associated with an activating HIF2 alpha mutation. *Blood* **112**, 919-21 (2008).
61. Yndestad, A. *et al.* Elevated levels of activin A in clinical and experimental pulmonary hypertension. *J Appl Physiol (1985)* **106**, 1356-64 (2009).
62. Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. Intersalt Cooperative Research Group. *BMJ* **297**, 319-28 (1988).
63. Sacks, F.M. *et al.* Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. DASH-Sodium Collaborative Research Group. *N Engl J Med* **344**, 3-10 (2001).
64. Whelton, P.K. *et al.* Primary prevention of hypertension: clinical and public health advisory from The National High Blood Pressure Education Program. *JAMA* **288**, 1882-8 (2002).

65. Chan, Q. *et al.* An Update on Nutrients and Blood Pressure. *J Atheroscler Thromb* **23**, 276-89 (2016).

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).
BK holds a British Heart Foundation Personal Chair.
NJS holds a chair funded by the British Heart Foundation and is a NIHR Senior Investigator.
FD was funded by the MRC Unit at the University of Bristol (MC_UU_12013/1-9).
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 NHS 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.
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).
Replication consortium contributor: [ICBP-1000G] GBE, LVW, DL, AC, MJC, MDT, POR, JK, HS; [CHD Exome+ Consortium] PSu, RC, DSa, JMMH [ExomeBP Consortium] JPC, FD, PBM [T2D-GENES Consortium and GoT2DGenes Consortium] CML; [CHARGE] GBE, CL, AK, DL, CNC, DIC; [iGEN-BP] ML, JCC, NK, JH, EST, PE, JSK, PVDH.
Replication study contributor: [Lifelines] NV, PVDH, HS, AMS; [GS:SFHS] JM, CH, DP, SP; [EGCUT] TE, MA, RM, AM; [PREVEND] PVDH, NV, RTG, SJLB; [ASCOT] HRW, MJC, PBM, PS, NP, AS, DS, ST; [BRIGHT] HRW, MJC, PBM, MB, MF, JC; [Airwave] HG, EE, MPSL, IK, IT, PE.
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.

(a) UK Biobank GWAS																
Sentinel SNV in the locus					UK Biobank discovery					Replication			Combined			
Locus	Chr	Pos	rsID	EA	INFO	EAF	Beta	SE	*P*	Beta	SE	*P*	N	Beta	SE	*P*
Systolic blood pressure																
NADK-CPSF3L	1	1685921	rs139385870	D	0.99	0.50	-0.394	0.07	$1.9x10^{-8}$	-0.310	0.07	$1.0x10^{-5}$	281,890	-0.352	0.05	$1.3x10^{-12}$
CELA2A	1	15798197	rs3820068	A	0.99	0.81	0.497	0.09	$2.4x10^{-8}$	0.367	0.08	$5.3x10^{-6}$	310,776	0.425	0.06	$1.1x10^{-12}$
GTF2B	1	89360158	rs10922502	A	0.99	0.62	-0.475	0.07	$4.7x10^{-11}$	-0.307	0.06	$2.0x10^{-6}$	323,666	-0.382	0.05	$2.2x10^{-15}$
FOSL2	2	28635740	rs7562	T	0.98	0.52	0.365	0.07	$2.2x10^{-7}$	0.182	0.06	$3.7x10^{-3}$	319,942	0.263	0.05	$1.9x10^{-8}$
PRKD3	2	37517566	rs13420463	A	1.00	0.77	0.504	0.08	$1.4x10^{-9}$	0.244	0.07	$7.3x10^{-4}$	330,307	0.356	0.05	$7.0x10^{-11}$
METTL21A-AC079767.3	2	208526140	rs55780018	T	0.97	0.54	-0.426	0.07	$1.7x10^{-9}$	-0.360	0.07	$5.1x10^{-8}$	304,567	-0.391	0.05	$5.9x10^{-16}$
RYK	3	134000025	rs9859176	T	0.97	0.40	0.419	0.07	$6.4x10^{-9}$	0.248	0.06	$9.6x10^{-5}$	322,428	0.322	0.05	$1.3x10^{-11}$
NPNT	4	106911742	rs13112725	C	1.00	0.76	0.418	0.08	$3.1x10^{-7}$	0.450	0.08	$9.4x10^{-9}$	306,370	0.435	0.06	$1.5x10^{-14}$
TMEM161B	5	87514515	rs10059921	T	0.98	0.08	-0.644	0.13	$5.9x10^{-7}$	-0.417	0.12	$7.9x10^{-4}$	298,543	-0.526	0.09	$4.0x10^{-9}$
FBN2	5	127868199	rs6595838	A	0.93	0.30	0.483	0.08	$2.0x10^{-10}$	0.236	0.07	$4.5x10^{-4}$	328,401	0.344	0.05	$7.6x10^{-12}$
CASC15	6	22130601	rs6911827	T	0.99	0.45	0.433	0.07	$8.2x10^{-10}$	0.190	0.06	$2.1x10^{-3}$	326,471	0.296	0.05	$2.0x10^{-10}$
TFAP2D	6	50683009	rs78648104	T	1.00	0.92	-0.664	0.13	$1.2x10^{-7}$	-0.329	0.11	$4.0x10^{-3}$	305,426	-0.481	0.08	$1.3x10^{-8}$
MKLN1	7	131059056	rs13238550	A	1.00	0.40	0.486	0.07	$9.4x10^{-12}$	0.212	0.06	$7.1x10^{-4}$	325,647	0.331	0.05	$1.9x10^{-12}$
HIPK2	7	139463264	rs1011018	A	0.98	0.20	-0.441	0.09	$6.1x10^{-7}$	-0.244	0.08	$1.6x10^{-3}$	325,110	-0.329	0.06	$1.5x10^{-8}$
ZFAT	8	135612745	rs894344	A	1.00	0.60	-0.384	0.07	$6.8x10^{-8}$	-0.163	0.06	$8.2x10^{-3}$	329,834	-0.258	0.05	$3.2x10^{-8}$
PAX2	10	102604514	rs112184198	A	0.99	0.10	-0.826	0.12	$7.8x10^{-13}$	-0.532	0.10	$1.3x10^{-7}$	323,791	-0.659	0.08	$3.6x10^{-18}$
MCF2L	13	113636156	rs9549328	T	1.00	0.23	0.440	0.08	$1.5x10^{-7}$	0.218	0.08	$3.9x10^{-3}$	313,787	0.318	0.06	$1.5x10^{-8}$
FERMT2	14	53377540	rs9888615	T	0.99	0.29	-0.427	0.08	$3.5x10^{-8}$	-0.236	0.07	$4.3x10^{-4}$	326,235	-0.318	0.05	$3.5x10^{-10}$
PPP2R5E	14	63928546	rs8016306	A	0.99	0.80	0.454	0.09	$2.5x10^{-7}$	0.250	0.07	$7.9x10^{-4}$	329,869	0.335	0.06	$3.7x10^{-9}$
ABHD17C	15	81013037	rs35199222	A	0.99	0.45	0.353	0.07	$5.7x10^{-7}$	0.298	0.06	$1.7x10^{-6}$	323,407	0.322	0.05	$5.2x10^{-12}$
CFDP1	16	75331044	rs11643209	T	0.98	0.42	-0.481	0.07	$1.8x10^{-11}$	-0.222	0.06	$6.3x10^{-4}$	309,242	-0.339	0.05	$1.8x10^{-12}$
CRK	17	1333598	rs12941318	T	0.99	0.49	-0.317	0.07	$6.2x10^{-6}$	-0.226	0.07	$6.9x10^{-4}$	299,739	-0.269	0.05	$2.5x10^{-8}$
ACOX1	17	73949045	rs2467099	T	1.00	0.22	-0.423	0.08	$4.5x10^{-7}$	-0.216	0.07	$3.6x10^{-3}$	326,401	-0.307	0.06	$3.3x10^{-8}$
Diastolic blood pressure																
chr1mb25	1	25030470	rs6686889	T	0.99	0.25	0.231	0.05	$3.7x10^{-7}$	0.143	0.04	$9.1x10^{-4}$	322,575	0.185	0.03	$3.6x10^{-9}$
DNM3	1	172357441	rs12405515	T	0.98	0.56	-0.219	0.04	$4.1x10^{-8}$	-0.118	0.04	$1.6x10^{-3}$	328,543	-0.165	0.03	$1.4x10^{-9}$

GPATCH2	1	217718789	rs12408022	T	0.97	0.26	0.226	0.05	$5.9x10^{-7}$	0.172	0.04	$6.7x10^{-5}$	320,983	0.198	0.03	$2.4x10^{-10}$
CDC42BPA	1	227252626	rs10916082	A	1.00	0.73	-0.222	0.04	$5.3x10^{-7}$	-0.135	0.04	$1.5x10^{-3}$	327,636	-0.177	0.03	$8.4x10^{-9}$
WNT3A	1	228191075	rs2760061	A	0.98	0.47	0.235	0.04	$3.7x10^{-9}$	0.225	0.04	$1.1x10^{-8}$	312,761	0.230	0.03	$2.1x10^{-16}$
SDCCAG8	1	243471192	rs953492	A	0.99	0.46	0.293	0.04	$1.2x10^{-13}$	0.153	0.04	$4.6x10^{-5}$	325,253	0.220	0.03	$7.4x10^{-16}$
ADCY3	2	25139596	rs55701159	T	0.98	0.89	0.382	0.06	$1.1x10^{-9}$	0.193	0.06	$1.6x10^{-3}$	321,052	0.285	0.04	$7.2x10^{-11}$
SLC8A1	2	40567743	rs4952611	T	0.95	0.58	-0.200	0.04	$8.0x10^{-7}$	-0.114	0.04	$4.6x10^{-3}$	309,395	-0.157	0.03	$4.0x10^{-8}$
AC016735.1	2	43167878	rs76326501	A	0.98	0.91	0.426	0.07	$4.3x10^{-10}$	0.413	0.07	$1.5x10^{-9}$	318,127	0.419	0.05	$3.6x10^{-18}$
GPAT2-FAHD2CP	2	96675166	rs2579519	T	1.00	0.63	-0.259	0.04	$1.7x10^{-10}$	-0.137	0.04	$6.7x10^{-4}$	311,557	-0.197	0.03	$4.8x10^{-12}$
TEX41	2	145646072	rs1438896	T	1.00	0.30	0.288	0.04	$2.1x10^{-11}$	0.187	0.04	$4.3x10^{-6}$	329,278	0.234	0.03	$2.0x10^{-15}$
CCDC141	2	179786068	rs79146658	T	1.00	0.91	-0.375	0.07	$5.8x10^{-8}$	-0.245	0.07	$4.2x10^{-4}$	321,318	-0.311	0.05	$2.4x10^{-10}$
TMEM194B	2	191439591	rs7592578	T	0.99	0.19	-0.271	0.05	$8.9x10^{-8}$	-0.212	0.05	$1.7x10^{-5}$	304,672	-0.240	0.04	$9.5x10^{-12}$
TNS1	2	218668732	rs1063281	T	0.98	0.60	-0.231	0.04	$1.2x10^{-8}$	-0.172	0.04	$1.4x10^{-5}$	315,354	-0.200	0.03	$1.3x10^{-12}$
CAMKV-ACTBP13	3	49913705	rs36022378	T	0.99	0.80	-0.265	0.05	$6.3x10^{-8}$	-0.140	0.05	$3.9x10^{-3}$	319,983	-0.202	0.03	$4.7x10^{-9}$
CACNA2D2	3	50476378	rs743757	C	0.99	0.14	0.313	0.06	$2.9x10^{-8}$	0.184	0.05	$5.1x10^{-4}$	328,836	0.245	0.04	$2.4x10^{-10}$
FAM208A	3	56726646	rs9827472	T	1.00	0.37	-0.207	0.04	$3.6x10^{-7}$	-0.148	0.04	$1.7x10^{-4}$	323,058	-0.177	0.03	$4.3x10^{-10}$
RP11-439C8.2	3	154707967	rs143112823	A	0.97	0.09	-0.484	0.07	$2.9x10^{-12}$	-0.295	0.08	$2.3x10^{-4}$	297,343	-0.403	0.05	$1.4x10^{-14}$
SENP2	3	185317674	rs12374077	C	1.00	0.35	0.203	0.04	$8.3x10^{-7}$	0.127	0.04	$1.2x10^{-3}$	327,513	0.163	0.03	$9.2x10^{-9}$
PDE5A	4	120509279	rs66887589	T	1.00	0.52	-0.296	0.04	$5.7x10^{-14}$	-0.140	0.04	$2.1x10^{-4}$	324,397	-0.215	0.03	$3.4x10^{-15}$
POC5	5	75038431	rs10078021	T	0.99	0.63	-0.223	0.04	$4.7x10^{-8}$	-0.105	0.04	$9.2x10^{-3}$	314,172	-0.164	0.03	$1.3x10^{-8}$
CPEB4	5	173377636	rs72812846	A	0.97	0.28	-0.232	0.04	$1.6x10^{-7}$	-0.186	0.04	$2.4x10^{-5}$	312,601	-0.209	0.03	$2.2x10^{-11}$
PKHD1	6	51832494	rs13205180	T	0.97	0.49	0.218	0.04	$3.7x10^{-8}$	0.123	0.04	$1.1x10^{-3}$	325,419	0.168	0.03	$7.0x10^{-10}$
PDE10A	6	166178451	rs147212971	T	0.98	0.06	-0.421	0.08	$2.3x10^{-7}$	-0.289	0.09	$9.4x10^{-4}$	296,010	-0.360	0.06	$1.6x10^{-9}$
SLC35F1	6	118572486	rs9372498	A	0.98	0.08	0.459	0.07	$5.4x10^{-10}$	0.231	0.07	$5.6x10^{-4}$	330,625	0.334	0.05	$1.8x10^{-11}$
SNX31	8	101676675	rs2978098	A	0.99	0.54	0.212	0.04	$6.9x10^{-8}$	0.122	0.04	$1.4x10^{-3}$	324,424	0.165	0.03	$1.5x10^{-9}$
RP11-273G15.2	8	144060955	rs62524579	A	1.00	0.53	-0.202	0.04	$2.8x10^{-7}$	-0.140	0.05	$2.2x10^{-3}$	268,645	-0.175	0.03	$3.8x10^{-9}$
MTAP	9	21801530	rs4364717	A	0.99	0.55	-0.218	0.04	$3.5x10^{-8}$	-0.136	0.04	$2.9x10^{-4}$	327,173	-0.175	0.03	$1.3x10^{-10}$
BDNF	11	27728102	rs11030119	A	1.00	0.31	-0.211	0.04	$7.0x10^{-7}$	-0.119	0.04	$3.3x10^{-3}$	330,002	-0.163	0.03	$2.9x10^{-8}$
MYEOV	11	69079707	rs67330701	T	0.89	0.09	-0.415	0.07	$7.8x10^{-9}$	-0.314	0.08	$3.8x10^{-5}$	276,760	-0.367	0.05	$2.1x10^{-12}$
RP11-321F6.1	15	66869072	rs7178615	A	1.00	0.37	-0.207	0.04	$3.8x10^{-7}$	-0.152	0.04	$1.0x10^{-4}$	318,076	-0.179	0.03	$2.6x10^{-10}$
ADAMTS7	15	79070000	rs62012628	T	0.97	0.29	-0.295	0.04	$2.1x10^{-11}$	-0.147	0.06	$7.7x10^{-3}$	244,143	-0.238	0.03	$5.1x10^{-12}$
chr15mb95	15	95312071	rs12906962	T	0.98	0.68	-0.292	0.04	$5.3x10^{-12}$	-0.155	0.04	$1.5x10^{-4}$	319,952	-0.221	0.03	$5.6x10^{-14}$
PPL	16	4943019	rs12921187	T	1.00	0.43	-0.203	0.04	$3.0x10^{-7}$	-0.147	0.04	$1.2x10^{-4}$	326,469	-0.174	0.03	$2.5x10^{-10}$
FBXL19	16	30936743	rs72799341	A	1.00	0.24	0.235	0.05	$3.0x10^{-7}$	0.139	0.04	$1.6x10^{-3}$	324,502	0.185	0.03	$5.8x10^{-9}$
CMIP	16	81574197	rs8059962	T	0.98	0.42	-0.241	0.04	$2.0x10^{-9}$	-0.103	0.04	$8.5x10^{-3}$	319,839	-0.170	0.03	$1.3x10^{-9}$
ACE	17	61559625	rs4308	A	0.98	0.37	0.242	0.04	$3.2x10^{-9}$	0.186	0.04	$2.7x10^{-6}$	319,394	0.213	0.03	$6.8x10^{-14}$

MAPK4	18	48142854	rs745821	T	0.99	0.76	0.236	0.05	$3.2x10^{-7}$	0.150	0.04	$4.2x10^{-4}$	330,954	0.189	0.03	$1.4x10^{-9}$
CCNE1	19	30294991	rs62104477	T	0.99	0.33	0.209	0.04	$7.1x10^{-7}$	0.148	0.04	$2.4x10^{-4}$	320,347	0.177	0.03	$1.2x10^{-9}$
PLCB1	20	8626271	rs6108168	A	0.99	0.25	-0.305	0.05	$1.5x10^{-11}$	-0.127	0.04	$2.9x10^{-3}$	327,368	-0.211	0.03	$1.1x10^{-11}$
Pulse pressure																
chr1mb9	1	9441949	rs9662255	A	0.99	0.43	-0.303	0.05	$4.7x10^{-10}$	-0.130	0.04	$3.0x10^{-3}$	310,618	-0.207	0.03	$1.9x10^{-10}$
SF3A3	1	38455891	rs4360494	C	0.99	0.55	0.332	0.05	$5.7x10^{-12}$	0.224	0.05	$3.6x10^{-6}$	282,851	0.278	0.03	$3.7x10^{-16}$
RP4-710M16.1-PPAP2B	1	56576924	rs112557609	A	0.99	0.35	0.280	0.05	$3.2x10^{-8}$	0.187	0.04	$1.8x10^{-5}$	325,952	0.227	0.03	$6.8x10^{-12}$
FGGY	1	59653742	rs3889199	A	0.99	0.71	0.462	0.05	$3.3x10^{-18}$	0.271	0.05	$1.9x10^{-9}$	329,486	0.351	0.03	$1.8x10^{-24}$
C2orf43	2	20881840	rs2289081	C	0.99	0.36	-0.251	0.05	$5.3x10^{-7}$	-0.203	0.04	$1.7x10^{-6}$	329,140	-0.223	0.03	$5.5x10^{-12}$
PRKCE	2	46363336	rs11690961	A	1.00	0.88	0.437	0.07	$4.2x10^{-9}$	0.266	0.07	$4.6x10^{-5}$	327,847	0.340	0.05	$3.9x10^{-12}$
CEP68	2	65283972	rs74181299	T	0.99	0.62	0.296	0.05	$2.1x10^{-9}$	0.181	0.04	$2.0x10^{-5}$	324,224	0.230	0.03	$9.6x10^{-13}$
TCF7L1	2	85491365	rs11689667	T	0.99	0.54	0.256	0.05	$1.1x10^{-7}$	0.118	0.04	$3.8x10^{-3}$	330,634	0.176	0.03	$1.7x10^{-8}$
FN1	2	216300482	rs1250259	A	0.99	0.74	-0.457	0.05	$5.5x10^{-17}$	-0.210	0.05	$7.7x10^{-6}$	325,485	-0.314	0.04	$8.7x10^{-19}$
GATA2	3	128201889	rs62270945	T	1.00	0.03	0.861	0.14	$2.6x10^{-9}$	0.366	0.14	$9.5x10^{-3}$	279,925	0.607	0.10	$1.8x10^{-9}$
PALLD	4	169717148	rs1566497	A	0.98	0.42	0.320	0.05	$6.6x10^{-11}$	0.173	0.04	$4.8x10^{-5}$	320,948	0.236	0.03	$1.9x10^{-13}$
chr4mb174	4	174584663	rs17059668	C	0.98	0.92	-0.442	0.09	$9.0x10^{-7}$	-0.245	0.08	$2.2x10^{-3}$	313,277	-0.332	0.06	$2.8x10^{-8}$
LHFPL2	5	77837789	rs10057188	A	0.99	0.46	-0.280	0.05	$5.5x10^{-9}$	-0.149	0.04	$3.3x10^{-4}$	325,985	-0.205	0.03	$6.7x10^{-11}$
GJA1	6	121781390	rs11154027	T	0.99	0.47	0.311	0.05	$1.1x10^{-10}$	0.125	0.04	$3.7x10^{-3}$	316,708	0.207	0.03	$1.1x10^{-10}$
ESR1	6	152397912	rs36083386	I	1.00	0.11	0.651	0.08	$4.6x10^{-17}$	0.289	0.07	$1.0x10^{-5}$	323,303	0.439	0.05	$1.5x10^{-18}$
FNDC1	6	159699125	rs449789	C	1.00	0.14	0.480	0.07	$2.2x10^{-12}$	0.264	0.06	$1.3x10^{-5}$	325,584	0.359	0.05	$2.4x10^{-15}$
THBS2	6	169587103	rs1322639	A	1.00	0.78	0.433	0.06	$7.7x10^{-14}$	0.230	0.05	$3.4x10^{-6}$	319,866	0.316	0.04	$4.8x10^{-17}$
SUGCT	7	40447971	rs76206723	A	0.99	0.10	-0.405	0.08	$2.6x10^{-7}$	-0.305	0.07	$3.8x10^{-6}$	328,162	-0.346	0.05	$7.4x10^{-12}$
SLC20A2	8	42324765	rs2978456	T	1.00	0.55	-0.253	0.05	$1.3x10^{-7}$	-0.130	0.05	$4.4x10^{-3}$	304,964	-0.188	0.03	$1.2x10^{-8}$
TRAPPC9	8	141060027	rs4454254	A	1.00	0.63	-0.320	0.05	$9.4x10^{-11}$	-0.217	0.04	$2.9x10^{-7}$	330,022	-0.261	0.03	$5.1x10^{-16}$
SCAI	9	127900996	rs72765298	T	0.98	0.87	-0.392	0.07	$5.9x10^{-8}$	-0.358	0.07	$8.6x10^{-8}$	316,271	-0.374	0.05	$2.7x10^{-14}$
KIAA1462	10	30317073	rs9337951	A	0.94	0.34	0.301	0.05	$7.6x10^{-9}$	0.262	0.05	$5.5x10^{-8}$	299,646	0.280	0.04	$2.5x10^{-15}$
ARHGAP12	10	32082658	rs10826995	T	0.99	0.71	-0.317	0.05	$2.2x10^{-9}$	-0.133	0.05	$3.9x10^{-3}$	327,373	-0.212	0.03	$1.1x10^{-9}$
PRDM11	11	45208141	rs11442819	I	1.00	0.11	-0.412	0.07	$3.8x10^{-8}$	-0.185	0.06	$3.3x10^{-3}$	326,483	-0.279	0.05	$7.1x10^{-9}$
NOX4	11	89224453	rs2289125	A	0.98	0.21	-0.481	0.06	$3.1x10^{-16}$	-0.293	0.05	$2.9x10^{-8}$	307,682	-0.377	0.04	$9.1x10^{-22}$
CEP164	11	117283676	rs8258	T	1.00	0.38	0.341	0.05	$5.3x10^{-12}$	0.157	0.04	$2.4x10^{-4}$	327,038	0.236	0.03	$2.9x10^{-13}$
CCDC41	12	94880742	rs139236208	A	0.97	0.10	-0.442	0.08	$5.7x10^{-8}$	-0.288	0.08	$2.8x10^{-4}$	291,244	-0.363	0.06	$1.6x10^{-10}$
RP11-61O1.1	14	98587630	rs9323988	T	0.98	0.63	-0.291	0.05	$5.6x10^{-9}$	-0.156	0.04	$2.0x10^{-4}$	327,551	-0.212	0.03	$4.1x10^{-11}$
VAC14	16	70755610	rs117006983	A	0.46	0.01	1.448	0.30	$9.4x10^{-7}$	0.847	0.16	$1.8x10^{-7}$	250,766	0.986	0.14	$4.1x10^{-12}$
CDH13	16	83045790	rs7500448	A	0.98	0.75	0.386	0.06	$4.2x10^{-12}$	0.288	0.05	$1.8x10^{-9}$	321,958	0.329	0.04	$1.1x10^{-19}$
KIAA0753	17	6473828	rs7226020	T	0.96	0.56	-0.348	0.05	$1.3x10^{-12}$	-0.175	0.05	$1.4x10^{-4}$	303,389	-0.256	0.03	$2.3x10^{-14}$

TP53-SLC2A4	17	7571752	rs78378222	T	0.95	0.99	1.530	0.22	8.9×10^{-12}	0.487	0.18	7.9×10^{-3}	294,053	0.904	0.14	1.8×10^{-10}
KCNH4-HSD17B1	17	40317241	rs79089478	T	0.99	0.97	0.842	0.15	1.2×10^{-8}	0.377	0.13	4.4×10^{-3}	318,326	0.584	0.10	3.1×10^{-9}
PYY	17	42060631	rs62080325	A	0.98	0.66	-0.260	0.05	3.6×10^{-7}	-0.128	0.05	4.8×10^{-3}	315,689	-0.186	0.03	4.0×10^{-8}
MRC2	17	60767151	rs740698	T	0.99	0.56	-0.307	0.05	2.1×10^{-10}	-0.161	0.04	2.8×10^{-4}	311,450	-0.228	0.03	3.1×10^{-12}
SLC14A2	18	43097750	rs7236548	A	0.99	0.18	0.462	0.06	1.1×10^{-13}	0.273	0.05	2.2×10^{-7}	330,075	0.352	0.04	2.0×10^{-18}
SLC24A3	20	19465907	rs6081613	A	0.99	0.28	0.326	0.05	1.2×10^{-9}	0.213	0.05	8.1×10^{-6}	315,546	0.263	0.04	1.6×10^{-13}
ARVCF	22	19967980	rs12628032	T	0.99	0.30	0.269	0.05	2.4×10^{-7}	0.216	0.05	3.8×10^{-6}	310,292	0.240	0.03	5.5×10^{-12}
XRCC6	22	42038786	rs73161324	T	1.00	0.05	0.611	0.11	6.5×10^{-9}	0.380	0.11	3.1×10^{-4}	267,722	0.496	0.07	2.8×10^{-11}
(b) UK Biobank exome																
Systolic blood pressure																
SSPN	12	26438189	rs6487543	A	0.94	0.77	0.345	0.09	5.9×10^{-5}	0.279	0.06	2.1×10^{-6}	244,842	0.300	0.05	6.3×10^{-10}
Diastolic blood pressure																
MRAS	3	138119952	rs2306374	T	1.00	0.84	-0.237	0.05	9.3×10^{-6}	-0.155	0.04	9.3×10^{-5}	281,715	-0.184	0.03	7.4×10^{-9}
Pulse pressure																
CD34	1	208024820	rs12731740	T	1.00	0.10	-0.360	0.08	5.8×10^{-6}	-0.202	0.05	1.1×10^{-4}	279,078	-0.249	0.04	1.1×10^{-8}
ZNF638	2	71627539	rs3771371	T	1.00	0.57	-0.223	0.05	4.1×10^{-6}	-0.130	0.03	9.6×10^{-5}	280,285	-0.160	0.03	5.8×10^{-9}
CRACR2B	11	828916	rs7126805	A	1.00	0.73	0.262	0.05	1.1×10^{-6}	0.184	0.05	4.6×10^{-4}	145,162	0.222	0.04	3.3×10^{-9}

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; 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).

Secondary SNV in the locus						UK Biobank discovery					Replication			Combined			
Locus	**Chr**	**Pos**	**rsID**	**EA**	**Trait**	**INFO**	**EAF**	**Beta**	**SE**	***P***	**Beta**	**SE**	***P***	**N**	**Beta**	**SE**	***P***
(a) Novel loci from UK Biobank GWAS																	
NADK-CPSF3L	1	1254436	rs1886773	A	PP	0.99	0.03	-0.743	0.13	2.0×10^{-8}	-0.481	0.15	1.0×10^{-3}	233,789	-0.625	0.10	1.9×10^{-10}
RP4-710M16.1-PPAP2B	1	56938218	rs6588634	T	PP	0.99	0.89	0.403	0.08	2.1×10^{-7}	0.270	0.07	4.7×10^{-5}	329,029	0.326	0.05	1.0×10^{-10}
FN1	2	216245694	rs34923683	A	PP	1.00	0.02	0.837	0.15	4.8×10^{-8}	0.432	0.16	7.7×10^{-3}	285,653	0.646	0.11	6.8×10^{-9}
TP53-SLC2A4	17	7185062	rs5417	A	DBP	0.99	0.57	0.207	0.04	2.1×10^{-7}	0.207	0.04	1.1×10^{-7}	319,299	0.207	0.03	1.1×10^{-13}
KCNH4-HSD17B1	17	40709867	rs138643143	A	PP	0.85	0.07	0.539	0.10	1.4×10^{-7}	0.420	0.15	5.8×10^{-3}	229,161	0.502	0.08	3.3×10^{-9}
(b) Previously reported loci																	
UK Biobank GWAS																	
RNF207	1	6683240	rs14057	A	SBP	0.99	0.35	-0.394	0.07	7.5×10^{-8}	-0.235	0.06	2.0×10^{-4}	329,584	-0.303	0.05	2.5×10^{-10}
FIGN-GRB14	2	165513065	rs34271465	D	SBP	1.00	0.41	-0.370	0.07	1.9×10^{-7}	-0.277	0.06	6.9×10^{-6}	328,486	-0.317	0.05	9.9×10^{-12}
ENPEP	4	111431444	rs33966350	A	SBP	1.00	0.01	1.742	0.31	2.6×10^{-8}	1.525	0.41	1.8×10^{-4}	216,630	1.661	0.25	2.1×10^{-11}
GUCY1A3-GUCY1B3	4	156406054	rs146853253	D	PP	0.99	0.16	0.457	0.06	1.7×10^{-12}	0.212	0.06	1.4×10^{-4}	322,302	0.316	0.04	6.9×10^{-14}
EBF1	5	158220193	rs31864	A	PP	0.99	0.55	0.307	0.05	1.9×10^{-10}	0.132	0.04	1.5×10^{-3}	326,557	0.206	0.03	5.5×10^{-11}
EBF1	5	158448401	rs888987	C	DBP	0.96	0.37	0.208	0.04	4.4×10^{-7}	0.111	0.04	7.1×10^{-3}	311,814	0.160	0.03	4.3×10^{-8}
PDE3A	12	19979881	rs10841376	C	SBP	0.99	0.76	0.261	0.08	1.6×10^{-3}	0.362	0.07	5.1×10^{-7}	327,370	0.319	0.05	4.5×10^{-9}
PDE3A	12	20230639	rs10770612	A	PP	1.00	0.80	0.378	0.06	2.5×10^{-10}	0.259	0.05	1.8×10^{-6}	311,586	0.313	0.04	6.9×10^{-15}
PDE3A	12	20368269	rs60691990	T	DBP	0.98	0.65	0.344	0.04	1.4×10^{-16}	0.223	0.04	7.4×10^{-8}	323,722	0.283	0.03	5.0×10^{-22}
TBX5-TBX3	12	115928440	rs10850519	C	DBP	0.99	0.30	-0.244	0.04	1.4×10^{-8}	-0.188	0.04	4.7×10^{-6}	327,837	-0.214	0.03	5.1×10^{-13}
MYH6	14	23761094	rs12050260	T	PP	0.97	0.35	0.261	0.05	2.9×10^{-7}	0.132	0.05	4.1×10^{-3}	304,390	0.190	0.03	2.6×10^{-8}
FURIN-FES	15	91427692	rs138682554	A	SBP	0.85	0.03	1.274	0.23	5.1×10^{-8}	0.695	0.21	8.8×10^{-4}	279,876	0.952	0.16	9.8×10^{-10}
HOXB7	17	46874272	rs585736	A	PP	1.00	0.03	0.712	0.13	7.8×10^{-8}	0.517	0.13	4.1×10^{-5}	301,845	0.609	0.09	2.5×10^{-11}
INSR	19	7258405	rs11671314	C	SBP	0.94	0.13	0.532	0.11	8.3×10^{-7}	0.344	0.13	6.2×10^{-3}	253,103	0.452	0.08	3.4×10^{-8}
JAG1	20	10669188	rs2206815	A	PP	0.98	0.50	-0.432	0.05	3.9×10^{-19}	-0.247	0.04	2.7×10^{-9}	324,088	-0.326	0.03	4.7×10^{-25}
JAG1	20	10767811	rs1040922	T	DBP	0.99	0.28	-0.344	0.04	3.8×10^{-15}	-0.156	0.04	1.8×10^{-4}	325,879	-0.245	0.03	4.2×10^{-16}
PREX1	20	47411149	rs80346118	A	DBP	0.99	0.15	-0.305	0.06	3.1×10^{-8}	-0.243	0.05	5.6×10^{-6}	327,614	-0.273	0.04	1.1×10^{-12}

CRYAA-SIK1	21	44720890	rs79094191	T	DBP	0.98	0.96	-0.691	0.10	$3.9x10^{-11}$	-0.408	0.12	$4.4x10^{-4}$	284,734	-0.564	0.08	$3.8x10^{-13}$
UK Biobank exome																	
ST7L-CAPZA1-MOV10	1	113456546	rs1049434	A	DBP	1.00	0.44	-0.175	0.04	$9.7x10^{-6}$	-0.131	0.03	$1.1x10^{-5}$	264,717	-0.147	0.02	$6.6x10^{-10}$
CDH17	8	95264265	rs138582164	A	PP	0.78	0.001	5.199	0.99	$1.3x10^{-7}$	2.620	0.73	$3.2x10^{-4}$	226,592	3.529	0.59	$1.7x10^{-9}$

Locus: For (a) the locus name from Table 1 for the nearest annotated gene, (b) the name of the previously reported blood pressure locus; Pos: build 37; EA: effect allele; Trait: the validated trait with most significant association in the combined meta-analysis; INFO: imputation quality score; 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; (Note: within the UK Biobank discovery analysis the sample size was N=140,882/140,886 for systolic and pulse pressure / diastolic pressure.)

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.

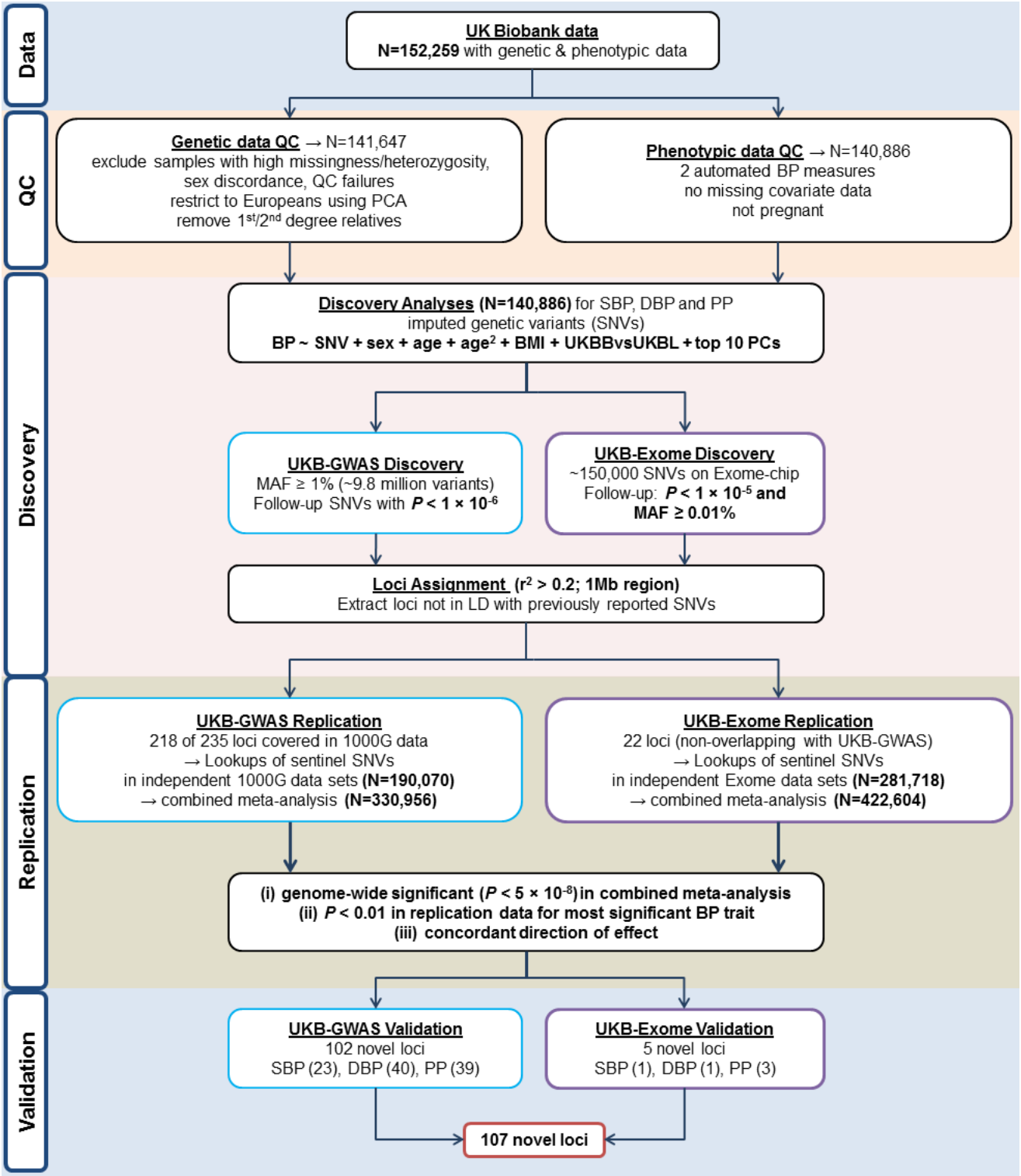
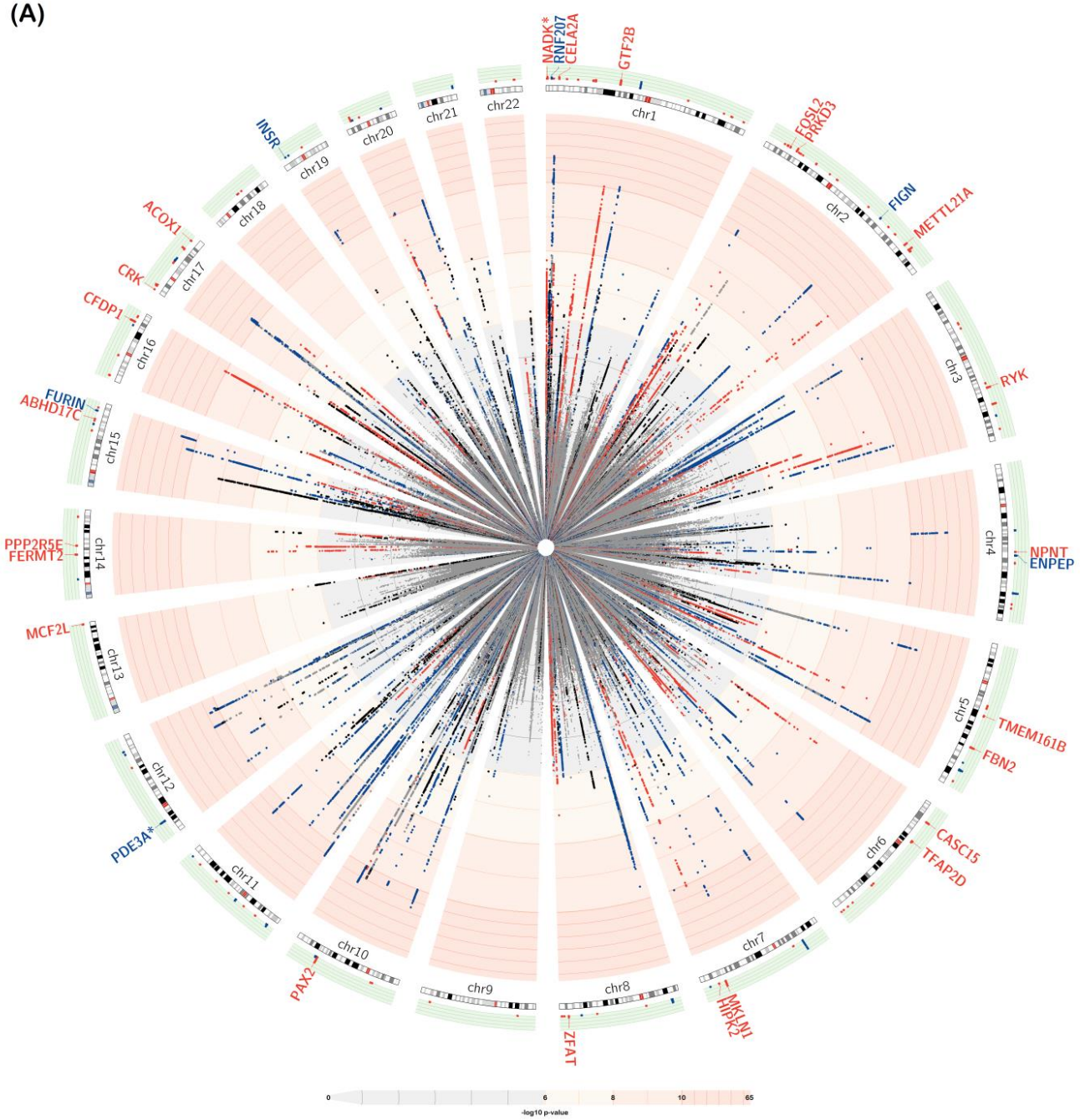


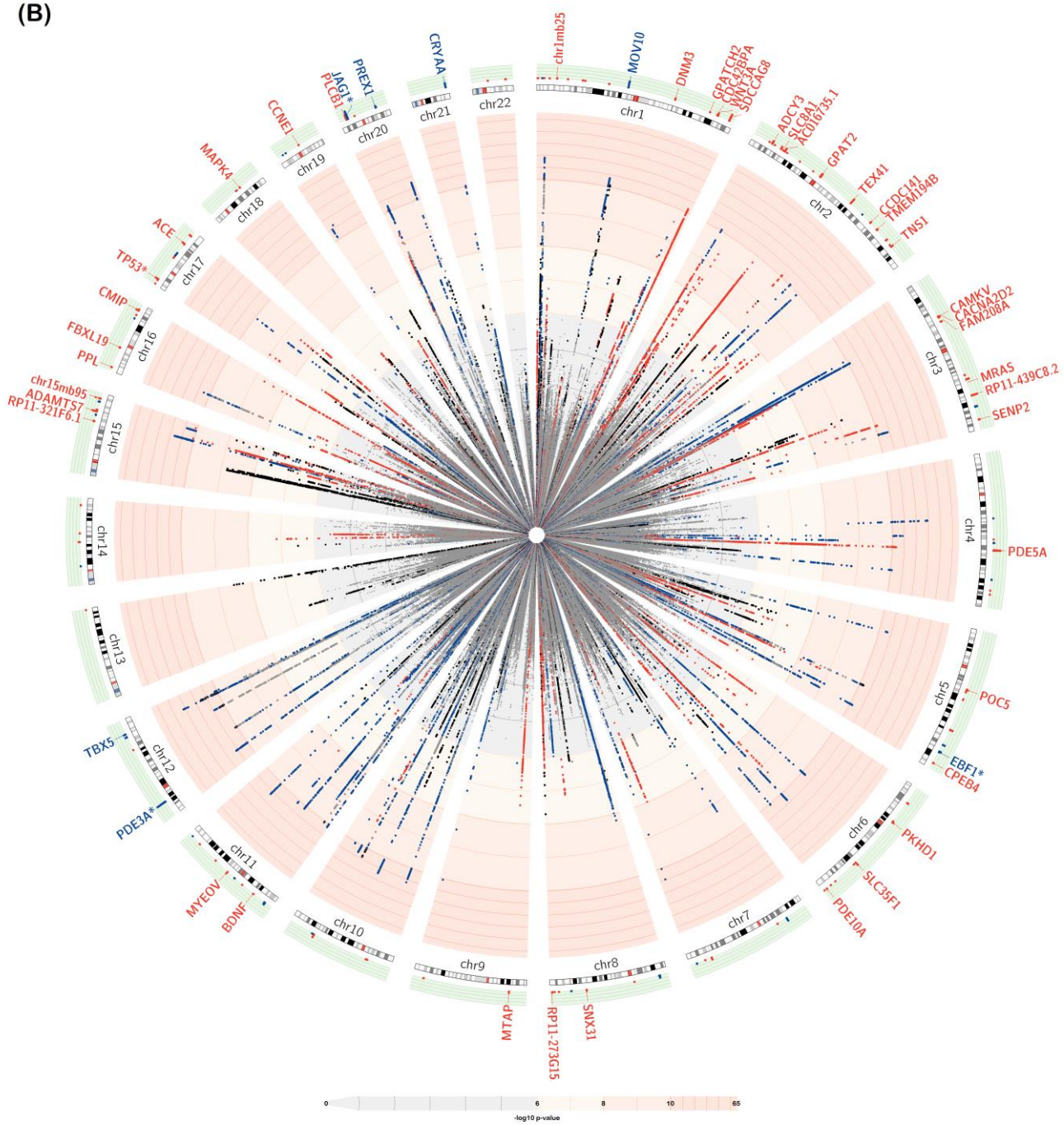
Figure 2: UK Biobank GWAS discovery Manhattan plots and Venn diagram of 107 novel validated loci. Plots (A), (B) and (C) show the UK Biobank GWAS discovery circos Manhattan plots for systolic blood pressure (SBP), diastolic blood pressure (DBP) and pulse pressure (PP) respectively. *P*-value results are plotted on a –log10 scale (see legend) for all ~9.8 million variants with Minor Allele Frequency (MAF) ≥ 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.

(A)


NADK*
RNF207
CELA2A
GTF2B
chr1
chr22
chr21
chr20
chr19
INSR
chr18
ACOX1
chr17
CRK
CFDP1
chr16
FURIN
ABHD17C
chr15
PPP2R5E
FERMT2
chr14
MCF2L
chr13
chr12
PDE3A*
chr11
chr10
PAX2
chr9
chr8
ZFAT
chr7
MKLN1
HIPK2
chr6
CASC15
TFAP2D
chr5
TMEM161B
FBN2
chr4
NPNT
ENPEP
chr3
RYK
chr2
FIGN
METTL21A
EOGL2
PRKD3
0
6
8
10
65
-log10 p-value

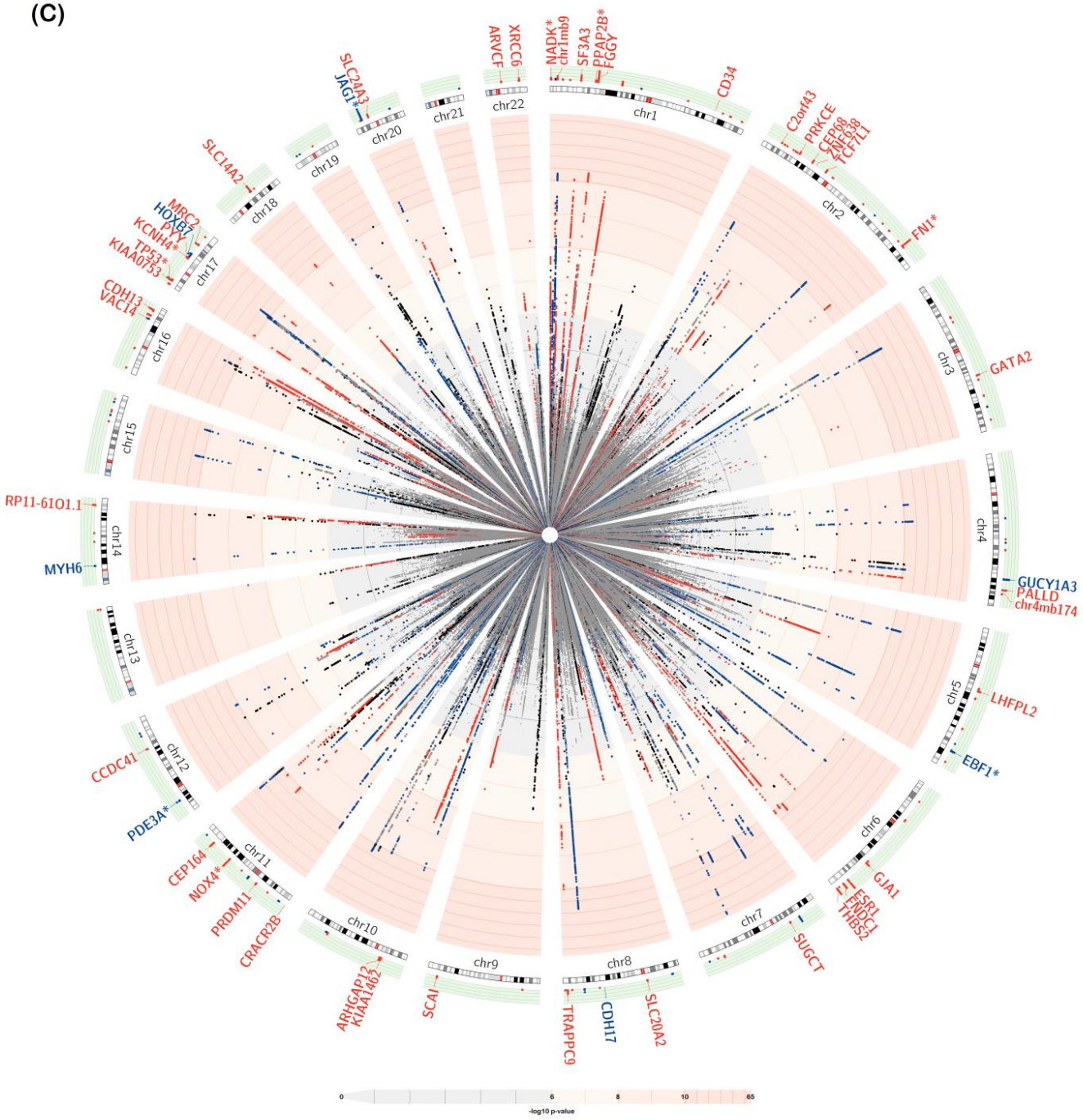
774

(B)


chr1mb25
MOV10
DNM3
GPATCH2
CDC42BPA
WNK3A
SDCCAG8
ADCY3
SLC8A1
AC016735.1
GPAT2
TEX41
CCDC141
TMEM194B
TNS1
CAMKV
CACNA2D2
FAM208A
MRAS
RP11-439C8.2
SENP2
PDE5A
POC5
EBF1*
CPEB4
PKHD1
SLC35F1
PDE10A
SNX31
RP11-273G15
MTAP
BDNF
MYEOV
PDE3A*
TBX5
PPL
FBXL19
CMIP
TP53*
ACE
MAPK4
CCNE1
JAG1*
PLCB1
PREX1
CRYAA
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
0
6
8
10
65
-log10 p-value

775

(C)


chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
NADK*
chr1mb9
SF3A3
PPAP2B*
FGGY
CD34
C2orf43
PRKCE
CER68
ZNF638
FCF1L1
FN1*
GATA2
GUCY1A3
PALLD
chr4mb174
LHFPL2
EBF1*
GJA1
ESR1
FAIM2
THBS2
SUGCT
SLC20A2
CDH17
TRAPPC9
SCAI
ARHGAP12
KIAA1462
CRACR2B
PRDM11
NOX4*
CEP164
PDE3A*
CCDC41
MYH6
RP11-61O1.1
CDH13
VAC14
KIAA0753
TP53*
KCNH4*
PXY
HOXB7
MRC2
SLC14A2
SLC2A3
JAG1*
ARVCF
XRCC6
0
6
8
10
65
-log10 p-value

776

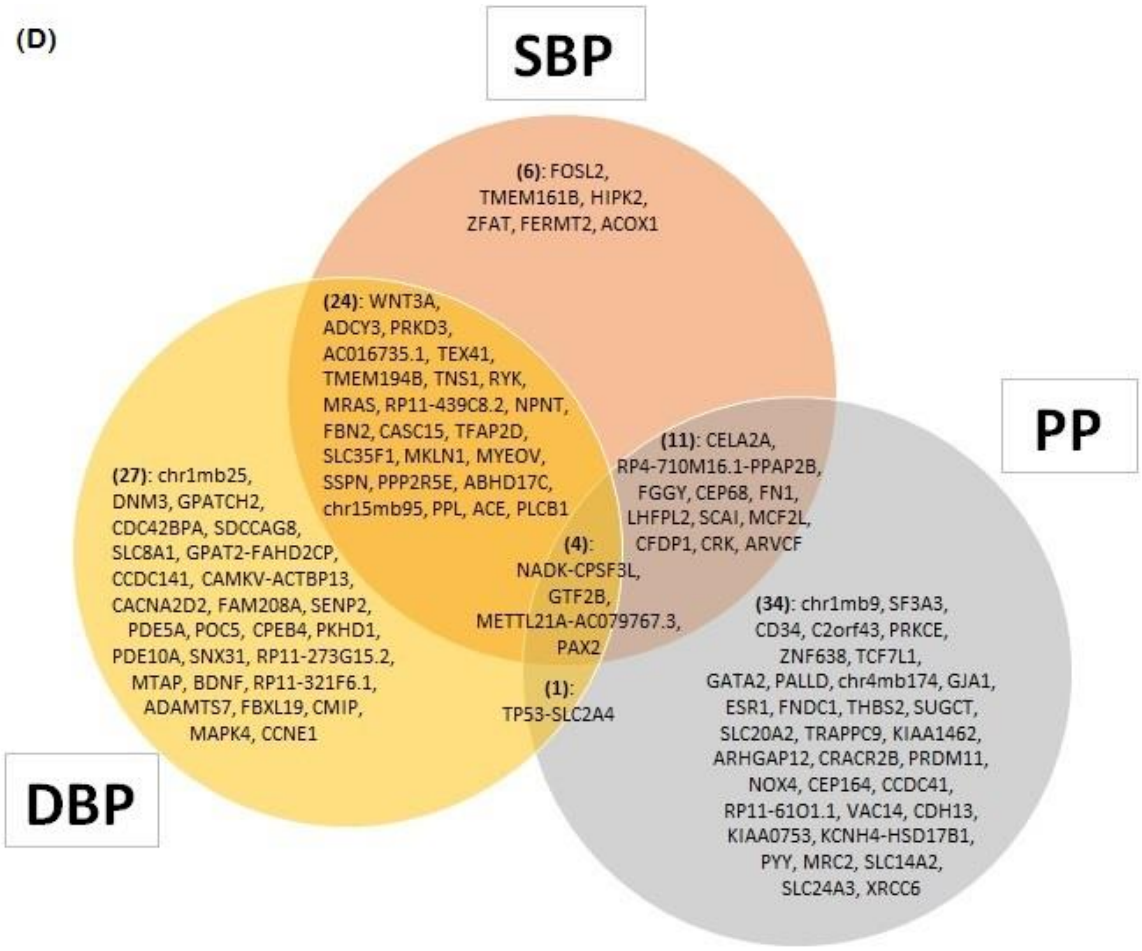

(D)
SBP
(6): FOSL2, TMEM161B, HIPK2, ZFAT, FERMT2, ACOX1
(24): WNT3A, ADCY3, PRKD3, AC016735.1, TEX41, TMEM194B, TNS1, RYK, MRAS, RP11-439C8.2, NPNT, FBN2, CASC15, TFAP2D, SLC35F1, MKLN1, MYEOV, SSPN, PPP2R5E, ABHD17C, chr15mb95, PPL, ACE, PLCB1
PP
(11): CELA2A, RP4-710M16.1-PPAP2B, FGGY, CEP68, FN1, LHFPL2, SCAI, MCF2L, CFDP1, CRK, ARVCF
(27): chr1mb25, DNM3, GPATCH2, CDC42BPA, SDCCAG8, SLC8A1, GPAT2-FAHD2CP, CCDC141, CAMKV-ACTBP13, CACNA2D2, FAM208A, SENP2, PDE5A, POC5, CPEB4, PKHD1, PDE10A, SNX31, RP11-273G15.2, MTAP, BDNF, RP11-321F6.1, ADAMTS7, FBXL19, CMIP, MAPK4, CCNE1
(4): NADK-CPSF3L, GTF2B, METTL21A-AC079767.3, PAX2
(34): chr1mb9, SF3A3, CD34, C2orf43, PRKCE, ZNF638, TCF7L1, GATA2, PALLD, chr4mb174, GJA1, ESR1, FNDC1, THBS2, SUGCT, SLC20A2, TRAPPC9, KIAA1462, ARHGAP12, CRACR2B, PRDM11, NOX4, CEP164, CCDC41, RP11-61O1.1, VAC14, CDH13, KIAA0753, KCNH4-HSD17B1, PYY, MRC2, SLC14A2, SLC24A3, XRCC6
(1): TP53-SLC2A4
DBP

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 ($P < 5\times10^{-8}$). 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.

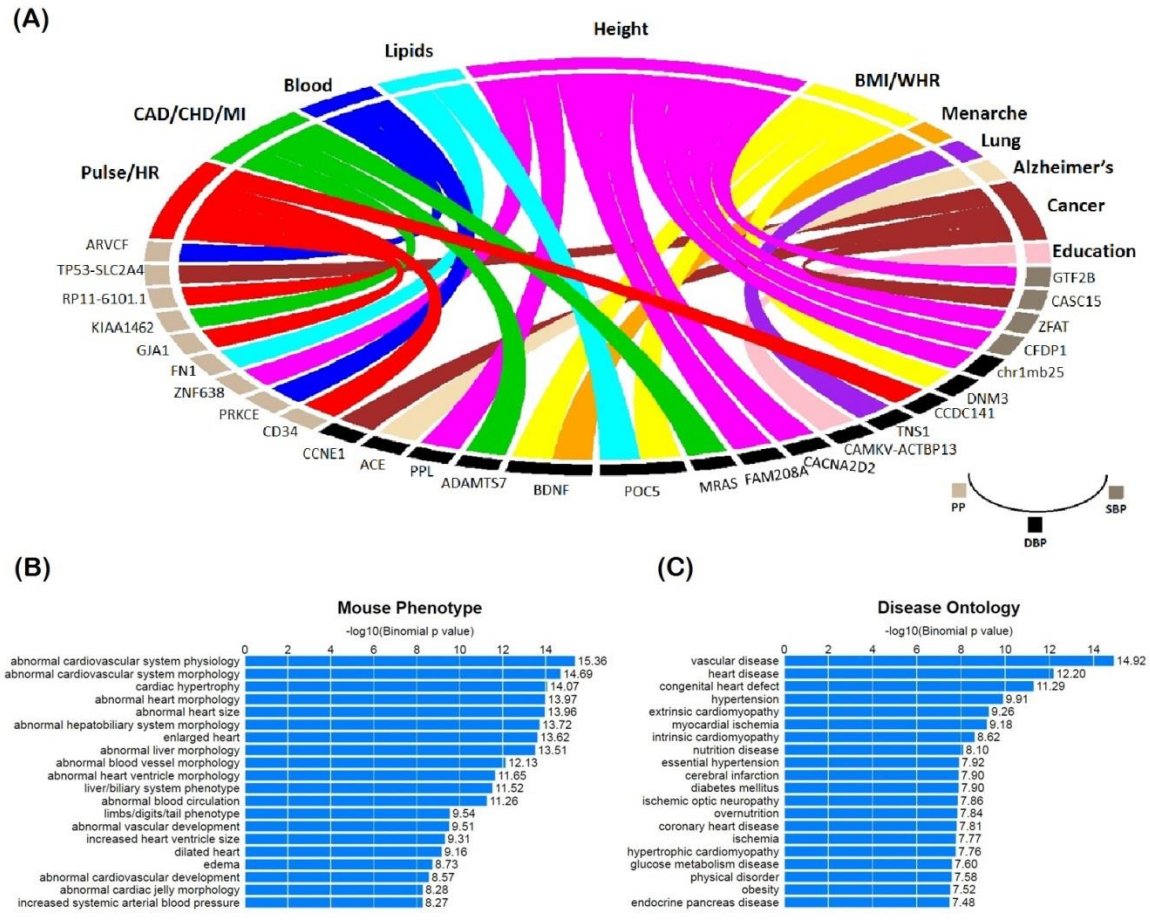
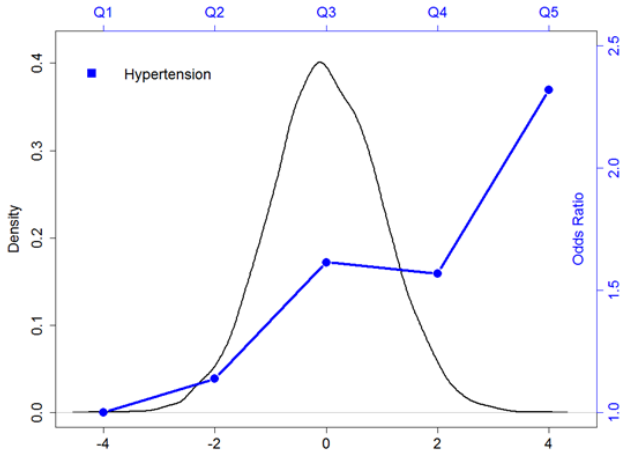
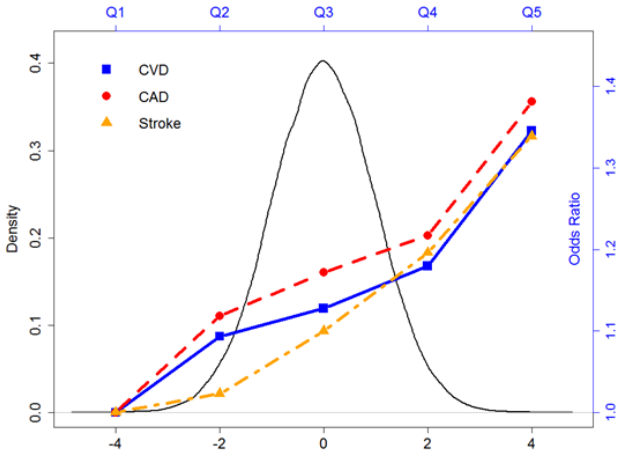


Figure 4: Distribution of a Genetic Risk Score (GRS) based on novel and previously reported blood pressure variants and its relationship with blood pressure levels, hypertension and cardiovascular disease (CVD) outcomes. (A): Distribution of GRS in the independent Airwave study and odds ratio of hypertension 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.

(A)



(C)



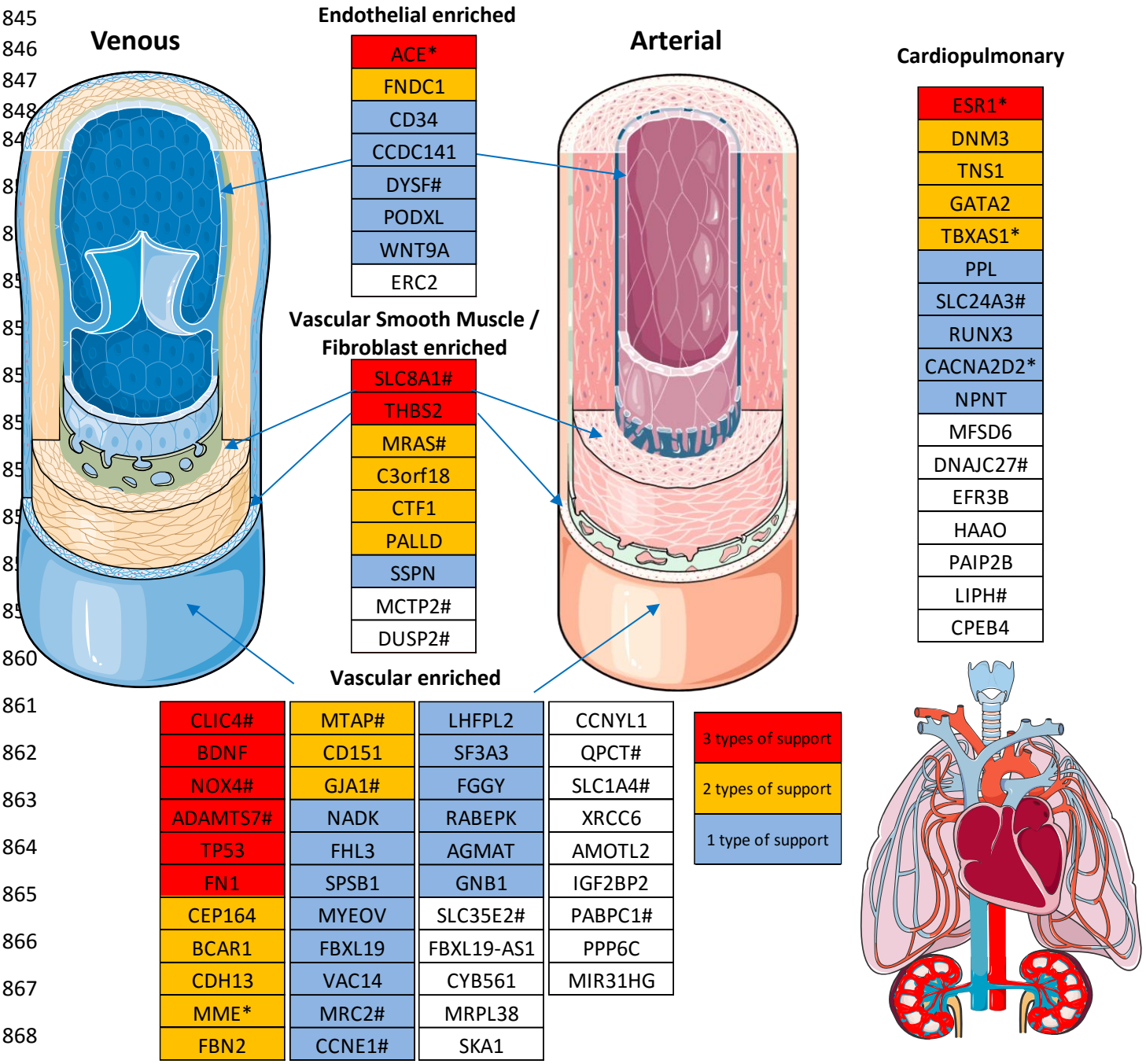
(B)

(mmHg)	Q1	Q2	Q3	Q4	Q5
SBP	136.4	138.7	142.5	142.5	145.7
DBP	82.9	84.3	85.9	86.0	87.6
PP	53.4	54.4	56.7	56.5	58.2

(D)

(count)	Q1	Q2	Q3	Q4	Q5
CVD	2,462	2,689	2,759	2,874	3,229
CAD	1,783	1,996	2,078	2,154	2,417
Stroke	581	597	640	695	776

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 *kmeans* clustering to the Principal Component Analysis (PCA) data a total of N=145,315 Europeans remain. Then we use the kinship data to exclude 1st and 2nd 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[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 $P < 1x10^{-6}$ 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 $P < 1x10^{-5}$. 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 ≥ 1%, INFO > 0.1 and $P<1x10^{-6}$, 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 ≥ 0.01%, INFO > 0.1 and $P < 1x10^{-5}$ 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[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 –log10 scale, i.e. if –log10(*P*) / -log10(*P*_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 (ENCODE) 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' un-translated 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 Chembl (www.ebi.ac.uk/chembl/) and Drug Gene Interaction database (dgidb.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

Fantom5 data (accessed 16 Aug 2016), http://fantom.gsc.riken.jp/5/

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

66. The UK Biobank Array Design Group. UK Biobank Axiom Array Content Summary. (2014).
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 Biobank. *Lancet Respir Med* **3**, 769-81 (2015).
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* **24**, 2911-35 (2005).
69. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-913 (2007).
70. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
71. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
72. Barnes, M.R. Exploring the landscape of the genome. *Methods Mol Biol* **628**, 21-38 (2010).
73. Gong, J. *et al.* Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. *Hum Mutat* **33**, 254-63 (2012).
74. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**(2015).
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 exploration. *Bioinformatics* **28**, 419-20 (2012).

77. McLean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501 (2010).
78. Elliott, P. *et al.* The Airwave Health Monitoring Study of police officers and staff in Great Britain: rationale, design and methods. *Environ Res* **134**, 280-5 (2014).
79. Petersen, M. *et al.* Quantification of lipoprotein subclasses by proton nuclear magnetic resonance-based partial least-squares regression models. *Clin Chem* **51**, 1457-61 (2005).
80. Chadeau-Hyam, M. *et al.* Metabolic profiling and the metabolome-wide association study: significance level for biomarker identification. *J Proteome Res* **9**, 4620-7 (2010).