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[1,2]*, Borbala Mifsud[1]*, Ioanna Ntalla[1], Praveen Surendran[6], Chunyu Liu[7-9], James P Cook[10], Aldi Kraja[11], Fotios Drenos[12,13], Marie Loh[3,14], Niek Verweij[15-18], Jonathan Marten[19], Ibrahim Karaman[3,5], Marcelo P Segura Lepe[3,20], Paul O'Reilly[21], Joanne Knight[22], Harold Snieder[23], Norihiro Kato[24], Jiang He[25], E Shyong Tai[26,27], Abdullah M Said[15], David Porteous[28], Maris Alver[29], Neil Poulter[30], Martin Farrall[31], Ron T Gansevoort[32], Sandosh Padmanabhan[33], Reedik Mägi[29], Alice Stanton[34], John Connell[35], Stephan J L Bakker[36], Andres Metspalu[29], Denis Shields[37], Simon Thom[38], Morris Brown[1,2], Peter Sever[39], Tõnu Esko[16,29], Caroline Hayward[19], Pim van der Harst[15], Danish Saleheen[40-42], Rajiv Chowdhury[6], John C Chambers[3,43-45], Daniel I Chasman[46,47], Aravinda Chakravarti[48], Christopher Newton-Cheh[16-18], Cecilia M Lindgren[16,49,50], Daniel Levy[7,9], Jaspal S Kooner[44,51,52], Bernard Keavney[53], Maciej Tomaszewski[53], Nilesh J Samani[54,55], Joanna M M Howson[6] , Martin D Tobin[56], Patricia B Munroe[1,2], Georg B Ehret[48,57], Louise V Wain[56], 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 Barts Cardiovascular Biomedical Research Unit, 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 Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
7 Population Sciences Branch, National Heart Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA.
8 Boston University School of Public Health, Boston, MA, USA.
9 National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA.
10 Department of Biostatistics, University of Liverpool, Liverpool, UK.
11 Division of Statistical Genomics, Center for Genome Sciences, Washington University School of Medicine, St. Louis MO, USA.

12 MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK.
13 Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, Rayne Building, University College London, London, WC1E 6JF, 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 Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, USA.
17 Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA.
18 Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA.
19 MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK.
20 Bayer Pharma AG, Berlin, Germany.
21 Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK.
22 Data Science Institute, Lancester University, Lancaster, UK.
23 Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.
24 Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan.
25 Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana, USA.
26 Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore.
27 Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.
28 Centre for Genomic & Experimental Medicine, Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, UK.
29 Estonian Genome Center, University of Tartu, Tartu, Estonia.
30 Imperial Clinical Trials Unit, School of Public Health, Imperial College London, London, UK.
31 Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, Oxford, UK.
32 Department of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.
33 Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK.
34 Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Ireland.
35 Ninewells Hospital & Medical School, University of Dundee, Dundee, UK.
36 Department of Internal Medicine, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.
37 School of Medicine, Conway Institute, Dublin, Ireland.
38 International Centre for Circulatory Health, Imperial College London, London, UK.
39 National Heart and Lung Institute, Imperial College London, London, UK.
40 Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, USA.
41 Centre for Non-Communicable Diseases, Karachi, Pakistan.
42 Department of Public Health and Primary Care, University of Cambridge, UK.
43 Ealing Hospital National Health Service (NHS) Trust, Middlesex, UK.
44 Imperial College Healthcare NHS Trust, London, UK.
45 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.
46 Division of Preventive Medicine, Brigham and Women's Hospital, Boston MA, USA.

47 Harvard Medical School, Boston, MA, USA.
48 Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
49 Wellcome Trust Center for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.
50 The Big Data Institute at the Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford OX3 7BN, UK.
51 Department of Cardiology, Ealing Hospital NHS Trust, Southall, Middlesex, UK.
52 National Heart and Lung Institute, Cardiovascular Sciences, Hammersmith Campus, Imperial College London, London, UK.
53 Division of Cardiovascular Sciences, The University of Manchester, Manchester, UK.
54 Department of Cardiovascular Sciences, University of Leicester, BHF Cardiovascular Research Centre, Glenfield Hospital, Leicester, UK.
55 NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK.
56 Department of Health Sciences, University of Leicester, Leicester, UK.
57 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 11 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: Omron HEM-7015IT digital blood pressure monitor), 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 $P < 1\times10^{-6}$ 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\times10^{-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 $P < 5\times10^{-8}$ to denote genome-wide significance in the combined (discovery and replication) meta-analyses, also requiring evidence of support ($P < 0.01$) 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). We note that the replication *P*-value threshold of $P < 0.01$ is more stringent than a range of

196 thresholds calculated according to False Discovery Rate (FDR) which gives FDR thresholds of
197 $0.03 < P < 0.04$ (see Supplementary Methods).

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

205 **RESULTS**

206 **Discovery and validation of genetic variants at novel loci**

207 Of the 240 not previously reported loci taken forward to replication, we validate 107 novel
208 loci at $P < 5x10^{-8}$, of which 102 derive from the GWAS analysis replicated and meta-analysed
209 in a total of 330,956 individuals (**Table 1a**; **Supplementary Fig. 2a-c; Supplementary Fig. 3a**),
210 and a further five are from the exome analysis validated in a total of 422,604 individuals from
211 the combined meta-analysis (**Table 1b and Supplementary Fig. 3b; Supplementary Tables 5**
212 **and 6**). Most SNVs also show association with hypertension in the UK Biobank data, for
213 example 93 of the 107 validated novel sentinel SNVs are nominally significant ($P < 0.01$)
214 (**Supplementary Table 7**).
215
216 Our results for systolic, diastolic and pulse pressure are shown in **Supplementary Figs. 2a,b,c**
217 respectively. The most significant association signal for systolic pressure, which rises with age
218 is with rs112184198 near *PAX2* ($P = 3.6 \times 10^{-18}$); for diastolic pressure, which plateaus in middle
219 age, with rs76326501 near *METTL21A- ACO16735.1* ($P = 3.6 \times 10^{-18}$); and rs3889199 near *FGGY*
220 ($P = 1.8 \times 10^{-24}$) for pulse pressure, which increases with age and arterial stiffening[18]. However,
221 as blood pressure traits are highly correlated, we unsurprisingly report considerable overlap
222 in these findings (**Supplementary Fig. 4**). Many loci are associated with more than one blood
223 pressure trait at genome-wide significance. For example, in the combined meta-analysis, 24
224 validated novel loci are associated with both systolic and diastolic pressure, 11 with both
225 systolic and pulse pressure, one locus with both diastolic and pulse pressure and four loci
226 (*NADK-CPSF3L*, *GTF2B*, *METTL21A*-AC079767.3 and *PAX2*) are associated with all three traits
227 (**Fig. 2**). We further note that many of the pulse pressure associated SNVs have opposing
228 directions of effect for systolic and diastolic pressure, and are less likely to have strong
229 associations with hypertension.
230
231 After conditional analysis on the sentinel SNV we identify five validated secondary SNVs in
232 novel regions that are independently associated with blood pressure traits (**Table 2a;**
233 **Supplementary Table 8**). We also note the existence of a rare validated potential secondary
234 variant at the *NOX4* locus (rs56061986, MAF = 0.3%); although we do not claim this as an
235 independent signal after conditioning on the sentinel variant, its relatively large effect on
236 blood pressure remains (**Supplementary Table 8**). The contribution of our validated novel loci
237 increases the percentage trait variance explained by ~1%, e.g. compared with 2.59% for
238 previously reported SNVs alone, taken together, the validated novel and previously reported
239 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 x 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 x 10^{-10}) targeted by calcium channel blockers, *MME* (rs143112823 in the RP11-439C8.2 locus, P = 1.4 x 10^{-14}) targeted by omapatrilat for treating hypertension, *ADRA2B* (rs2579519 in the *GPAT2-FAHD2CP* locus, P = 4.8 x 10^{-12}) targeted by beta blockers, *SLC14A2* (rs7236548, P = 2.0 x 10^{-18}) targeted by the hypertension drug nifedipine, and phosphodiesterase 5A (*PDE5A*; rs66887589, P = 3.4 x 10^{-15}) targeted by sildenafil for treating pulmonary hypertension.

Additionally, we evaluate our validated 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 validated 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). Some of these associations may reflect genuine pleiotropic effects of variants. In some cases, such as for coronary artery disease, association with blood pressure may either be due to pleiotropy, or reflect the fact 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**).

The UK Biobank data show support (P < 0.01) for 119 of the 122 previously reported blood pressure loci (159 of 163 SNVs) for one or more blood pressure traits (**Supplementary Fig. 2 a-c; Supplementary Table 12**). Thus we do not show support in UK Biobank for only four previously reported SNVs, one of which (rs11066280, *RPL6-ALDH1*) was identified from a GWAS of East Asian ancestry[21] and may indicate ancestry-specific effects. We compare the MAF and effect sizes in UK Biobank with the published results of previously reported variants (**Supplementary Figure 5**), indicating consistency of results between the two sources of data.

284 We also examine findings for low-frequency and rare gene mutations previously reported to
285 be associated with monogenic hypertension disorders[22] and included on the UK Biobank gene
286 array. Even within a large single study, there is still a lack of power for testing the impact of
287 rare variants and it remains inconclusive as to whether any monogenic mutations also affect
288 blood pressure levels within the general population. From the look-up results obtained within
289 the UK Biobank data (**Supplementary Table 13**), there is suggestion that the variant with the
290 lowest *P*-value (rs387907156; *KLH3*; MAF=0.02%) has a large effect on blood pressure (8.2
291 mm Hg per allele (SE=4.1); $P = 0.046$ and 5.6 mm Hg (SE=2.6); $P = 0.048$ for systolic and pulse
292 pressure respectively).

293

294 **Functional analyses**

295 We annotate the 107 validated novel loci to 212 genes (based on LD $r^2 \geq 0.8$) and seek putative
296 function from *in silico* analyses of our novel and previously reported loci, as well as
297 undertaking gene expression experiments for selected SNVs in relevant vascular tissue.
298 Candidate genes with the strongest supporting evidence are indicated in the last column of
299 Table 1 with an indication of the supporting data source. All genome wide-significant variants
300 in LD ($r^2>0.8$) for (a) validated novel loci and (b) previously reported loci, ranked by supporting
301 evidence are annotated in **Supplementary Table 14**. Of the 107 validated novel sentinel SNVs
302 only three are Indels, all other variants are single nucleotide polymorphisms (SNPs). We
303 identify non-synonymous SNVs at 13 of the 107 validated novel loci, including three non-
304 synonymous novel sentinel SNVs (rs1250259 at *FN1* locus, rs78648104 at *TFAP2D* and
305 rs7127805 at *CRACR2B* locus) (**Supplementary Table 15**). Furthermore three of the 13
306 validated novel loci contain non-synonymous SNVs that are predicted to be damaging
307 (ANNOVAR) in *TFAP2D* (rs78648104), *NOX4* (rs56061986, see above) and *CCDC141*
308 (rs17362588, reported to be associated with heart rate[23]) (**Fig. 3a**). Beyond the coding regions
309 we identify 29 novel associated SNVs in 3'UTRs which are predicted to significantly weaken
310 or cause loss of miRNA regulation by altering the recognition motif in seven genes, and
311 strengthen or create target sites for miRNA binding in 13 genes (based on miRNASNP db,
312 **Supplementary Table 15**).

313 Our expression Quantitative Trait locus (eQTL) analysis (based on GTEx data) shows that many
314 novel loci contain variants with eQTLs across a range of different tissues (**Supplementary**
315 **Table 16**). Of the 107 validated novel loci, 59 contain variants with eQTLs in at least one tissue.
316 We observe arterial tissue as the tissue having the largest number of loci with eQTLs
317 (**Supplementary Fig. 6**). Our follow-up targeted *in-silico* analysis reveals six novel loci with
318 eQTLs in arterial tissue (**Supplementary Table 15**). For example, the GTEx tibial artery eQTL
319 in *SF3A3* (rs4360494) shows strong *in silico* supporting evidence, including an arterial DNase
320 I site within which the major C allele removes a predicted AP-2 binding site **(Supplementary**
321 **Fig. 7**). Hence we prioritised this gene for *in vitro* functional analysis (see below).

322 By considering all loci together from both validated novel and previously reported loci, our
323 analysis using DEPICT identifies enrichment of expression across 31 tissues and cells
324 (**Supplementary Fig.8; Supplementary Table 17**), with greatest enrichment in the arteries (*P*

325 = 1.9 x 10^{-6}, false discovery rate (FDR) < 1%). We use FORGE to investigate and identify
326 significant (FDR, *P* <0.05) cell type specific enrichment within DNase I hypersensitive sites in
327 a range of tissues including dermal and lung microvascular endothelial cell types, and cardiac
328 fibroblasts (**Supplementary Fig. 9**). For a set of curated candidate regulatory SNVs from
329 validated novel loci (see Supplementary Methods), widespread enrichment is found in
330 microvascular endothelium, aortic smooth muscle, aortic fibroblasts, vascular epithelium,
331 heart and skin (**Supplementary Fig. 9**). In addition, we identify significant enrichment of
332 histone marks in a wide range of cell types, including strong enrichment seen for H3K4Me3
333 (an activating modification found near promoters) marks in umbilical vein endothelial cells
334 (HUVEC) (**Supplementary Fig. 10**). To explore expression at the level of cardiovascular cell
335 types specifically, we use Fantom5 reference transcript expression data (see Methods Online)
336 to cluster the 212 genes annotated to our 107 validated novel loci according to tissue
337 specificity **(Supplementary Fig. 11**), with the significantly clustered genes forming four tissue-
338 specific clusters, including a vascular smooth muscle cell (VSMC) and fibroblast cluster, an
339 endothelial cell cluster (including probable endothelial cells in highly vascularised tissues),
340 and a combined vascular cell cluster.

341 Additionally, Ingenuity pathway analysis and upstream transcriptional analysis show
342 enrichment of canonical pathways implicated in cardiovascular disease, including those
343 targeted by antihypertensive drugs, such as the alpha-adrenergic, CXCR4, endothelin
344 signalling and angiotensin receptor pathways (**Supplementary Table 18**). In keeping with
345 vascular mediation of genetic influence we identify diphenyleneiodonium, an inhibitor of
346 flavin-containing oxidases, including NAD(P)H oxidase, which is reported to reverse
347 endothelial dysfunction (and hypertension) in a rat model[24].

348 In order to identify long range target genes of non-coding variants, we use chromatin
349 interaction (Hi-C) data from HUVEC, as enhancers and silencers often form chromatin loops
350 with their target promoter. In most loci the strongest promoter interaction involves a gene in
351 high LD with the SNV but for 21 loci we find a distal potential target gene (**Supplementary**
352 **Table 15**). Ingenuity pathway analysis of the distal genes shows the greatest enrichment in
353 regulators of cardiac hypertrophy.

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

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

363 We also look for association of our validated sentinel SNVs with metabolomic signatures.
364 Three novel SNVs within the *NOX4*, *KCNH4* and *LHFPL2* loci show significant associations
365 (family-wise error rate < 5%) with lipoprotein sub-fractions from ^{1}H Nuclear Magnetic

366 Resonance (NMR) spectroscopy analysis of 2,000 Airwave study samples (**Supplementary**
367 **Tables 19 and 20**). The results for these variants suggest a link between blood pressure
368 regulation and lipid metabolism. Eleven SNVs (including at *LHFPL2* locus) show association
369 (family wise error rate < 5%) with metabolites in blood or urine from the publicly available
370 "Metabolomics GWAS Server" resource based on mass spectrometry[25,26] (**Supplementary**
371 **Table 20**), including sugar acids, sphingolipids, fatty acids, glycerophospholipids, organic acids
372 and benzene derivatives.

373 Several genes and variants with putative function are highlighted in our *in silico* analysis as
374 having biological support (e.g. eQTLs or nsSNVs) and those with novelty and tractability to
375 laboratory investigation (e.g. expression in available tissue models) are prioritized. Sentinel
376 variants in three genes are selected for experimental testing and successfully genotyped, each
377 for at least 100 samples. We select *ADAMTS7* due to strong biological support (e.g. mouse
378 knockout phenotype), *SF3A3* due to eQTLs and *NOX4* as it contains a rare nsSNV in addition
379 to common variant associations. All three SNVs reached highly significant levels of association
380 with blood pressure in the combined meta-analysis (**Table 1)**: rs62012628 at *ADAMTS7* for
381 diastolic pressure (0.238 mmHg per allele ±0.03, P=5.1x10^{-12}, N=244,143); rs4360494 at *SF3A3*
382 for pulse pressure (0.278 mmHg ±0.03, P=3.7x10^{-16}, N=307,682); rs2289125 at *NOX4* for pulse
383 pressure (-0.377 mmHg ±0.04, P=9.1x10^{-22}, N=282,851). We use quantitative polymerase
384 chain reaction (qPCR) to study the impact of these sentinel variants on gene expression in
385 human vascular smooth muscle (VSMCs) and endothelial cells (ECs) (see Methods Online). For
386 *SF3A3*, the major C allele of sentinel variant rs4360494 associated with increased pulse
387 pressure is also associated with *SF3A3* expression in human VSMCs, although this SNV is not
388 related to expression in endothelial cells (**Supplementary Fig. 12a**); and the T allele of SNV
389 rs62012628 in *ADAMTS7*, associated with lower diastolic pressure, is associated with reduced
390 *ADAMTS7* expression in human VSMCs (**Supplementary Fig. 12b**). Moreover, we find that the
391 minor A allele of sentinel SNV rs2289125 at the *NOX4* locus correlates with increased *NOX4*
392 expression in ECs though not VSMCs (**Supplementary Fig. 12c**). Our study thus finds evidence
393 for novel cis-eQTLs in *ADAMTS7* and *NOX4* in addition to validating the previously reported
394 GTEx eQTL in *SF3A3*, and supports the vascular expression of these genes.

395 **Genetic risk of increased blood pressure, hypertension and cardiovascular outcomes**

396 We create an unbiased genetic risk score (GRS) (**Supplementary Table 21**) to evaluate, in an
397 independent cohort (Airwave, see Methods Online), the impact of the combination of our
398 validated novel and previously reported loci on blood pressure levels and risk of hypertension.
399 When compared with the lowest quintile of the distribution of the GRS, individuals >50 years
400 in the highest quintile have sex-adjusted mean systolic pressure higher by 9.3 mm Hg (95% CI
401 6.9 to 11.7 mm Hg, P =1.0 x 10-13) and an over two-fold higher risk of hypertension (OR 2.32
402 95% CI 1.76 to 3.06; P=2.8 x 10-9) compared with individuals in the lowest quintile of the GRS
403 distribution (**Fig. 4; Supplementary Table 22**). Similar results were obtained from GRS
404 associations with blood pressure and hypertension within UK Biobank (**Supplementary Table**
405 **23**). In UK Biobank − based on self-reported health data, record linkage to Hospital Episode
406 Statistics and mortality follow-up data (**Supplementary Table 24**) − we show that the GRS is
407 associated with increased risk of stroke, coronary heart disease and all cardiovascular
408 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, $P = 1.5\times10^{-7}$), 1.38 (95% CI 1.30 to 1.47, $P = 4.3\times10^{-23}$) and 1.35 (95% CI 1.27 to 1.42, $P = 1.3\times10^{-25}$) respectively (**Fig. 4; Supplementary Table 25**). Results are also provided for incident-only cases (**Supplementary Table 26**).

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. Despite its size, our study is still under-powered to find low-frequency variants and the vast majority of our findings are common variants, with similarly modest or small effect sizes as previously reported validated variants (**Supplementary Fig. 13**). Our GWAS, which was restricted to MAF ≥ 1%, only identified four novel sentinel SNVs of low-frequency (1% ≤ MAF < 5%) and our Exome analysis, despite allowing for rare variant discovery, did not identify any rare novel sentinel SNVs. The only rare and low-frequency variants identified were secondary SNVs within previously reported loci. The lack of rare variant discovery could also be due to the challenge of detecting rare variants from imputed data, in contrast to the recent Exome-chip studies which identified some novel rare SNVs from genotyped data[11,12]. 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 validated 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 atherosclerosis. SF3A3 is a splicing

451 factor with no prior links to blood pressure other than our reported association and eQTL.
452 NOX4 has an established role in the endothelium where it enhances vasodilatation and
453 reduces blood pressure in vivo[32]. At the CFDP1 locus our sentinel SNV is in high LD (r^2 = 0.95)
454 with a variant previously associated with carotid intimal-medial thickness. Collectively our
455 findings highlight a potential common mechanism among these genes in vascular remodelling
456 that has previously been observed in small resistance arteries in essential hypertension[33].

457 We identify both common and rare variant associations at the novel NADPH oxidase 4 (*NOX4*)
458 locus. This oxidase generates reactive oxygen species in the endothelium and may contribute
459 to salt sensitive hypertension in the kidney and the vasculature[34-36]. We found that the allele
460 of the common variant at *NOX4* locus correlates with increased tissue specific *NOX4*
461 expression in endothelial cells rather than vascular smooth muscle cells (**Supplementary**
462 **Figure 12c**). NOX4 mediates endothelial cell apoptosis and facilitates vascular collagen
463 synthesis contributing to endothelial dysfunction and arterial stiffness, and may explain the
464 association with pulse pressure[37,38].

465 We identify several loci containing genes involved in vascular signalling and second
466 messenger systems such as *PDE5A* and *PDE10A*[39-41]. The phosphodiesterase PDE5A
467 hydrolyses cyclic GMP and is inhibited by sildenafil which leads to vasodilatation[42]. This
468 finding fits with our previous discoveries of a role for gene loci encoding elements of
469 natriuretic peptide-nitric oxide pathway and guanylate cyclase signalling systems in blood
470 pressure regulation[21,43,44]. Our findings strengthen the case for evaluating the opportunity to
471 repurpose PDE5A inhibitors for use in hypertension.

472 The importance of microvascular function is emphasised by the solute carrier transporters
473 such as *SLC14A2* encoding a urea transporter, which has previously been linked to autosomal
474 dominant Streeten type orthostatic hypotensive disorder[45] and blood pressure response to
475 nifedipine, a calcium channel blocker antihypertensive drug[46]. *SLC8A1* encodes a sodium
476 calcium exchanger expressed in cardiomyocytes which alters cardiac contractility and
477 hypertrophy and shows abnormal blood pressure in SLC8A1 transgenic mice[47]. Variants at
478 *SLC35F1* have been previously associated with resting heart rate and ventricular dimensions
479 which could contribute to blood pressure elevation[48].

480 We also identify loci that are involved in cardiovascular development (*GATA2, KIAA1462,*
481 *FBN2*, *FN1* and *HAND2)* such as fibrillin 2 (*FBN2*) which overlaps in action with fibrillin 1 in
482 development of the aortic matrix[49-53]. In addition, fibronectin expression is increased in
483 hypertension and in atherosclerosis but it may also play a role in the development of the
484 heart[53-55].

485 Our analysis validates loci containing genes with prior physiological connection to blood
486 pressure such as *BDNF, FAM208A,* and *CACNA2D2*[56-58]. The neurotrophin Brain Derived
487 Neurotrophic Factor modulates angiotensin 11 in the brain to elevate blood pressure in
488 experimental models and higher serum levels correlate with reduced risk of cardiovascular
489 disease and mortality[56]. In experimental models FAM208A, which is thought to be a
490 transcription factor, is a strong candidate for a quantitative trait locus for blood pressure[58].
491 The gene *CACNA2D2* encodes a subunit of the L-type calcium channel that is most abundantly

492 expressed in the atrium and in neurones and may be a target for negatively chronotropic and
493 inotropic calcium channel antagonists which reduce blood pressure[59].

494 This is the first time long range genomic interactions have been sought using Hi-C for blood
495 pressure, where the promoter region has a strong chromatin interaction with a novel SNV.
496 One such gene is *EPAS1*, which is ~200kb away from the SNV (rs11690961). It encodes
497 hypoxia-inducible factor 2alpha, which affects catecholamine homeostasis, protects against
498 heart failure and mutations in the gene are associated with pulmonary hypertension[60].
499 Another gene is *INHBA*, 1.3Mb away from the SNV (rs12531683**)**, which is elevated in
500 pulmonary hypertension and contributes to vascular remodelling by inducing expression of
501 endothelin-1 and plasminogen activator inhibitor-1 in pulmonary smooth muscle cells[61].

502 Our observation that the blood pressure genetic risk score is associated with 9-10 mm Hg
503 higher blood pressure at age 50+ years when comparing the top vs bottom fifths of the
504 distribution in an independent population has potential clinical and public health implications.
505 The results are particularly striking when stratified by age, due to a significant interaction of
506 the GRS with age (P ranging between 9.96×10^{-11} and 1.16×10^{-3} for interaction with continuous
507 blood pressure traits, and P = 0.012 for hypertension). Measuring the genetic risk score in
508 early life raises the possibility of adopting an early precision medicine approach to risk
509 management through lifestyle intervention (i.e. reduced sodium intake, increased potassium
510 intake, maintenance of optimal weight, low adult alcohol consumption and regular
511 exercise)[62-64]. Indeed, studies of non-pharmacologic approaches to blood pressure control
512 indicate that we might achieve 10 mm Hg or more reduction in systolic blood pressure
513 through lifestyle measures alone[65]. At the same time, recent evidence suggests that favorable
514 lifestyle may offset the cardiovascular sequelae associated with high genetic risk[66]. However,
515 as the above data are observational, it is not certain to what extent adherence to lifestyle
516 recommendations amongst high genetic risk individuals could result in favorable outcomes.
517 Given the substantial effect of genetic risk score on blood pressure by middle-age, the
518 potential for adopting early lifestyle intervention amongst individuals at high genetic risk,
519 along with population-wide measures to lower blood pressure, warrants further study.

520 Since the completion of our study, another blood pressure GWAS has been recently
521 published[67]. This used UK Biobank data within a larger single-stage combined meta-analysis,
522 reporting a total of 316 loci, including 241 loci identified from the meta-analysis involving UK
523 Biobank that were not tested for validation, as no replication resource was available. Our
524 study reports 107 validated novel loci, of which 32 are detected and validated for the first
525 time in our analysis of UK Biobank. In addition, 75 sentinel SNVs are in LD ($r^2 \geq 0.2$) with the
526 recently reported loci and we are able to validate at least 53 of these for the first time in our
527 study. Furthermore we note that 49 of the reported loci from this recent study[67] did not
528 validate in our large independent replication resource.

529 We describe 107 validated novel loci for blood pressure offering new biology, identifying
530 potential new therapeutic targets and raising the possibility of a precision medicine approach
531 to modify risk of hypertension and cardiovascular outcomes. In total this brings the number
532 of combined validated novel and previously reported loci for blood pressure traits to 229,

533 representing a major advance in our understanding of the genetic architecture of blood
534 pressure.

535 **References**

536 1. Forouzanfar, M.H. *et al.* Global, regional, and national comparative risk assessment
537 of 79 behavioural, environmental and occupational, and metabolic risks or clusters
538 of risks in 188 countries, 1990-2013: a systematic analysis for the Global Burden of
539 Disease Study 2013. *Lancet* **386**, 2287-323 (2015).
540 2. Sundstrom, J. *et al.* Blood pressure-lowering treatment based on cardiovascular risk:
541 a meta-analysis of individual patient data. *Lancet* **384**, 591-8 (2014).
542 3. Joffres, M. *et al.* Hypertension prevalence, awareness, treatment and control in
543 national surveys from England, the USA and Canada, and correlation with stroke and
544 ischaemic heart disease mortality: a cross-sectional study. *BMJ Open* **3**, e003423
545 (2013).
546 4. Falaschetti, E., Mindell, J., Knott, C. & Poulter, N. Hypertension management in
547 England: a serial cross-sectional study from 1994 to 2011. *Lancet* **383**, 1912-9 (2014).
548 5. Poulter, N.R., Prabhakaran, D. & Caulfield, M. Hypertension. *Lancet* **386**, 801-12
549 (2015).
550 6. Feinleib, M. *et al.* The NHLBI twin study of cardiovascular disease risk factors:
551 methodology and summary of results. *Am J Epidemiol* **106**, 284-5 (1977).
552 7. Mongeau, J.G., Biron, P. & Sing, C.F. The influence of genetics and household
553 environment upon the variability of normal blood pressure: the Montreal Adoption
554 Survey. *Clin Exp Hypertens A* **8**, 653-60 (1986).
555 8. Munoz, M. *et al.* Evaluating the contribution of genetics and familial shared
556 environment to common disease using the UK Biobank. *Nat Genet* (2016).
557 9. Cabrera, C.P. *et al.* Exploring hypertension genome-wide association studies findings
558 and impact on pathophysiology, pathways, and pharmacogenetics. *Wiley*
559 *Interdisciplinary Reviews: Systems Biology and Medicine* **7**, 73-90 (2015).
560 10. Ehret, G.B. *et al.* The genetics of blood pressure regulation and its target organs from
561 association studies in 342,415 individuals. *Nat Genet* **48**, 1171-1184 (2016).
562 11. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants
563 associated with blood pressure and hypertension. *Nat Genet* **48**, 1151-1161 (2016).
564 12. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood
565 pressure and overlapping with metabolic trait loci. *Nat Genet* **48**, 1162-1170 (2016).
566 13. Kato, N. *et al.* Trans-ancestry genome-wide association study identifies 12 genetic
567 loci influencing blood pressure and implicates a role for DNA methylation. *Nat Genet*
568 **47**, 1282-93 (2015).
569 14. Elliott, P. & Peakman, T.C. The UK Biobank sample handling and storage protocol for
570 the collection, processing and archiving of human blood and urine. *Int J Epidemiol*
571 **37**, 234-44 (2008).
572 15. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a
573 wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779
574 (2015).
575 16. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the
576 UK10K haplotype reference panel. *Nat Commun* **6**, 8111 (2015).
577 17. The UK Biobank. Genotype imputation and genetic association studies of UKBiobank.
578 (2015).

579 18. Wain, L.V. *et al.* Genome-wide association study identifies six new loci influencing
580 pulse pressure and mean arterial pressure. *Nat Genet* **43**, 1005-11 (2011).
581 19. Staley, J.R. *et al.* PhenoScanner: a database of human genotype-phenotype
582 associations. *Bioinformatics* (2016).
583 20. Ettehad, D. *et al.* Blood pressure lowering for prevention of cardiovascular disease
584 and death: a systematic review and meta-analysis. *Lancet* **387**, 957-67 (2016).
585 21. Kato, N. *et al.* Meta-analysis of genome-wide association studies identifies common
586 variants associated with blood pressure variation in east Asians. *Nat Genet* **43**, 531-8
587 (2011).
588 22. Munroe, P.B., Barnes, M.R. & Caulfield, M.J. Advances in Blood Pressure Genomics.
589 *Circulation Research* **112**, 1365-1379 (2013).
590 23. den Hoed, M. *et al.* Identification of heart rate–associated loci and their effects on
591 cardiac conduction and rhythm disorders. *Nature genetics* **45**, 621-631 (2013).
592 24. Hamilton, C.A., Brosnan, M.J., McIntyre, M., Graham, D. & Dominiczak, A.F.
593 Superoxide excess in hypertension and aging: a common cause of endothelial
594 dysfunction. *Hypertension* **37**, 529-34 (2001).
595 25. Shin, S.Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat*
596 *Genet* **46**, 543-50 (2014).
597 26. Raffler, J. *et al.* Genome-Wide Association Study with Targeted and Non-targeted
598 NMR Metabolomics Identifies 15 Novel Loci of Urinary Human Metabolic
599 Individuality. *PLoS Genet* **11**, e1005487 (2015).
600 27. Ehret, G.B. *et al.* Genetic variants in novel pathways influence blood pressure and
601 cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
602 28. van Setten, J. *et al.* Genome-wide association study of coronary and aortic
603 calcification implicates risk loci for coronary artery disease and myocardial infarction.
604 *Atherosclerosis* **228**, 400-5 (2013).
605 29. McCarthy, J.J. *et al.* Large scale association analysis for identification of genes
606 underlying premature coronary heart disease: cumulative perspective from analysis
607 of 111 candidate genes. *J Med Genet* **41**, 334-41 (2004).
608 30. van Meurs, J.B. *et al.* Common genetic loci influencing plasma homocysteine
609 concentrations and their effect on risk of coronary artery disease. *Am J Clin Nutr* **98**,
610 668-76 (2013).
611 31. Pu, X. *et al.* ADAMTS7 cleavage and vascular smooth muscle cell migration is affected
612 by a coronary-artery-disease-associated variant. *Am J Hum Genet* **92**, 366-74 (2013).
613 32. Ray, R. *et al.* Endothelial Nox4 NADPH oxidase enhances vasodilatation and reduces
614 blood pressure in vivo. *Arterioscler Thromb Vasc Biol* **31**, 1368-76 (2011).
615 33. Rizzoni, D. & Agabiti-Rosei, E. Structural abnormalities of small resistance arteries in
616 essential hypertension. *Intern Emerg Med* **7**, 205-12 (2012).
617 34. Touyz, R.M. & Montezano, A.C. Vascular Nox4: a multifarious NADPH oxidase. *Circ*
618 *Res* **110**, 1159-61 (2012).
619 35. Steppan, J., Barodka, V., Berkowitz, D.E. & Nyhan, D. Vascular stiffness and increased
620 pulse pressure in the aging cardiovascular system. *Cardiol Res Pract* **2011**, 263585
621 (2011).
622 36. Yan, F. *et al.* Nox4 and redox signaling mediate TGF-beta-induced endothelial cell
623 apoptosis and phenotypic switch. *Cell Death Dis* **5**, e1010 (2014).

624 37. Chan, E.C. *et al.* Nox4 modulates collagen production stimulated by transforming
625 growth factor beta1 in vivo and in vitro. *Biochem Biophys Res Commun* **430**, 918-25
626 (2013).
627 38. Vasa-Nicotera, M. *et al.* miR-146a is modulated in human endothelial cell with aging.
628 *Atherosclerosis* **217**, 326-30 (2011).
629 39. Tian, X. *et al.* Phosphodiesterase 10A upregulation contributes to pulmonary
630 vascular remodeling. *PLoS One* **6**, e18136 (2011).
631 40. Takimoto, E. *et al.* Chronic inhibition of cyclic GMP phosphodiesterase 5A prevents
632 and reverses cardiac hypertrophy. *Nat Med* **11**, 214-22 (2005).
633 41. Perez, N.G. *et al.* Phosphodiesterase 5A inhibition induces Na+/H+ exchanger
634 blockade and protection against myocardial infarction. *Hypertension* **49**, 1095-103
635 (2007).
636 42. Oliver, J.J., Melville, V.P. & Webb, D.J. Effect of regular phosphodiesterase type 5
637 inhibition in hypertension. *Hypertension* **48**, 622-7 (2006).
638 43. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension.
639 *Nat Genet* **41**, 677-87 (2009).
640 44. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci
641 associated with blood pressure. *Nat Genet* **41**, 666-76 (2009).
642 45. DeStefano, A.L. *et al.* Autosomal dominant orthostatic hypotensive disorder maps to
643 chromosome 18q. *Am J Hum Genet* **63**, 1425-30 (1998).
644 46. Hong, X. *et al.* Genetic polymorphisms of the urea transporter gene are associated
645 with antihypertensive response to nifedipine GITS. *Methods Find Exp Clin Pharmacol*
646 **29**, 3-10 (2007).
647 47. Takimoto, E. *et al.* Sodium calcium exchanger plays a key role in alteration of cardiac
648 function in response to pressure overload. *FASEB J* **16**, 373-8 (2002).
649 48. Ronaldson, P.T. & Davis, T.P. Targeting transporters: promoting blood-brain barrier
650 repair in response to oxidative stress injury. *Brain Res* **1623**, 39-52 (2015).
651 49. Carta, L. *et al.* Fibrillins 1 and 2 perform partially overlapping functions during aortic
652 development. *J Biol Chem* **281**, 8016-23 (2006).
653 50. Kazenwadel, J. *et al.* Loss-of-function germline GATA2 mutations in patients with
654 MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for
655 GATA2 in the lymphatic vasculature. *Blood* **119**, 1283-91 (2012).
656 51. Akashi, M., Higashi, T., Masuda, S., Komori, T. & Furuse, M. A coronary artery
657 disease-associated gene product, JCAD/KIAA1462, is a novel component of
658 endothelial cell-cell junctions. *Biochem Biophys Res Commun* **413**, 224-9 (2011).
659 52. Cakstina, I. *et al.* Primary culture of avian embryonic heart forming region cells to
660 study the regulation of vertebrate early heart morphogenesis by vitamin A. *BMC Dev*
661 *Biol* **14**, 10 (2014).
662 53. Wang, J., Karra, R., Dickson, A.L. & Poss, K.D. Fibronectin is deposited by injury-
663 activated epicardial cells and is necessary for zebrafish heart regeneration. *Dev Biol*
664 **382**, 427-35 (2013).
665 54. Dietrich, T. *et al.* ED-B fibronectin (ED-B) can be targeted using a novel single chain
666 antibody conjugate and is associated with macrophage accumulation in
667 atherosclerotic lesions. *Basic Res Cardiol* **102**, 298-307 (2007).
668 55. Stoynev, N. *et al.* Gene expression in peripheral blood of patients with hypertension
669 and patients with type 2 diabetes. *J Cardiovasc Med (Hagerstown)* **15**, 702-9 (2014).

670 56. Erdos, B., Backes, I., McCowan, M.L., Hayward, L.F. & Scheuer, D.A. Brain-derived
671 neurotrophic factor modulates angiotensin signaling in the hypothalamus to increase
672 blood pressure in rats. *Am J Physiol Heart Circ Physiol* **308**, H612-22 (2015).
673 57. Chan, S.H., Wu, C.W., Chang, A.Y., Hsu, K.S. & Chan, J.Y. Transcriptional upregulation
674 of brain-derived neurotrophic factor in rostral ventrolateral medulla by angiotensin
675 II: significance in superoxide homeostasis and neural regulation of arterial pressure.
676 *Circ Res* **107**, 1127-39 (2010).
677 58. Crespo, K., Menard, A. & Deng, A.Y. Retinoblastoma-associated protein 140 as a
678 candidate for a novel etiological gene to hypertension. *Clin Exp Hypertens* **38**, 533-40
679 (2016).
680 59. Watanabe, Y. *et al.* Accumulation of common polymorphisms is associated with
681 development of hypertension: a 12-year follow-up from the Ohasama study.
682 *Hypertens Res* **33**, 129-34 (2010).
683 60. Gale, D.P., Harten, S.K., Reid, C.D., Tuddenham, E.G. & Maxwell, P.H. Autosomal
684 dominant erythrocytosis and pulmonary arterial hypertension associated with an
685 activating HIF2 alpha mutation. *Blood* **112**, 919-21 (2008).
686 61. Yndestad, A. *et al.* Elevated levels of activin A in clinical and experimental pulmonary
687 hypertension. *J Appl Physiol (1985)* **106**, 1356-64 (2009).
688 62. Sacks, F.M. *et al.* Effects on blood pressure of reduced dietary sodium and the
689 Dietary Approaches to Stop Hypertension (DASH) diet. DASH-Sodium Collaborative
690 Research Group. *N Engl J Med* **344**, 3-10 (2001).
691 63. Intersalt: an international study of electrolyte excretion and blood pressure. Results
692 for 24 hour urinary sodium and potassium excretion. Intersalt Cooperative Research
693 Group. *BMJ* **297**, 319-28 (1988).
694 64. Whelton, P.K. *et al.* Primary prevention of hypertension: clinical and public health
695 advisory from The National High Blood Pressure Education Program. *JAMA* **288**,
696 1882-8 (2002).
697 65. Chan, Q. *et al.* An Update on Nutrients and Blood Pressure. *J Atheroscler Thromb* **23**,
698 276-89 (2016).
699 66. Khera, A.V. *et al.* Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary
700 Disease. *New England Journal of Medicine* **0**, null.
701 67. Hoffmann, T.J. *et al.* Genome-wide association analyses using electronic health
702 records identify new loci influencing blood pressure variation. *Nat Genet* (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 at Barts and The London School of Medicine and Dentistry.
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.
This research was supported by the British Heart Foundation (grant SP/13/2/30111).
Project title: Large-scale comprehensive genotyping of UK Biobank for cardiometabolic traits and diseases: UK CardioMetabolic Consortium (UKCMC).
PE is Director of the MRC-PHE Centre for Environment and Health and acknowledges support from the Medical Research Council and Public Health England. PE is a National Institute for Health Research (NIHR) senior investigator and acknowledges support from the NIHR Biomedical Research Centre at Imperial College Healthcare NHS Trust and Imperial College London, and the NIHR Health Protection Research Unit on Health Effects of Environmental Hazards.
This work used the computing resources of the UK MEDical BIOinformatics partnership - aggregation, integration, visualisation and analysis of large, complex data (UK MED-BIO) which is supported by the Medical Research Council (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 under Application Number 236.

Conflicts/Disclosures

MJC is Chief Scientist for Genomics England, a UK government company.

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.

765 **Table 1: Association results for the sentinel variant from each validated novel locus from (a) UK Biobank GWAS discovery and (b) UK**
766 **Biobank exome discovery.** Results are shown for the primary blood pressure trait with most significant association from the combined meta-
767 analysis.

(a) UK Biobank GWAS																	
Sentinel SNV in the locus					UK Biobank discovery				Replication			Combined					
Locus	**Chr**	**Pos**	**rsID**	**EA**	**EAF**	**Beta**	**SE**	***P***	**Beta**	**SE**	***P***	**N**	**Beta**	**SE**	***P***	**Traits**	**Candidate genes**
Systolic blood pressure																	
NADK-CPSF3L	1	1,685,921	rs139385870	D	0.50	-0.394	0.07	1.9×10^{-8}	-0.310	0.07	1.0×10^{-5}	281,890	-0.352	0.05	1.3×10^{-12}	DBP,PP	GNB1[df], NADK[efg]
CELA2A	1	15,798,197	rs3820068	A	0.81	0.497	0.09	2.4×10^{-8}	0.367	0.08	5.3×10^{-6}	310,776	0.425	0.06	1.1×10^{-12}	PP	AGMAT[df], CELA2A[efg]
GTF2B	1	89,360,158	rs10922502	A	0.62	-0.475	0.07	4.7×10^{-11}	-0.307	0.06	2.0×10^{-6}	323,666	-0.382	0.05	2.2×10^{-15}	DBP,PP	KYAT3[bf], GTF2B[defg]
FOSL2	2	28,635,740	rs7562	T	0.52	0.365	0.07	2.2×10^{-7}	0.182	0.06	3.7×10^{-3}	319,942	0.263	0.05	1.9×10^{-8}		FOSL2[efg]
PRKD3	2	37,517,566	rs13420463	A	0.77	0.504	0.08	1.4×10^{-9}	0.244	0.07	7.3×10^{-4}	330,307	0.356	0.05	7.0×10^{-11}	DBP	PRKD3[efg]
METTL21A-AC079767.3	2	208,526,140	rs55780018	T	0.54	-0.426	0.07	1.7×10^{-9}	-0.360	0.07	5.1×10^{-8}	304,567	-0.391	0.05	5.9×10^{-16}	DBP,PP	METTL21A[efg]
RYK	3	134,000,025	rs9859176	T	0.40	0.419	0.07	6.4×10^{-9}	0.248	0.06	9.6×10^{-5}	322,428	0.322	0.05	1.3×10^{-11}	DBP	RYK[efg]
NPNT	4	106,911,742	rs13112725	C	0.76	0.418	0.08	3.1×10^{-7}	0.450	0.08	9.4×10^{-9}	306,370	0.435	0.06	1.5×10^{-14}	DBP	NPNT[dfg]
TMEM161B	5	87,514,515	rs10059921	T	0.08	-0.644	0.13	5.9×10^{-7}	-0.417	0.12	7.9×10^{-4}	298,543	-0.526	0.09	4.0×10^{-9}		TMEM161B[g]
FBN2	5	127,868,199	rs6595838	A	0.30	0.483	0.08	2.0×10^{-10}	0.236	0.07	4.5×10^{-4}	328,401	0.344	0.05	7.6×10^{-12}	DBP	FBN2[defg]
CASC15	6	22,130,601	rs6911827	T	0.45	0.433	0.07	8.2×10^{-10}	0.190	0.06	2.1×10^{-3}	326,471	0.296	0.05	2.0×10^{-10}	DBP	CASC15[efg]
TFAP2D	6	50,683,009	rs78648104	T	0.92	-0.664	0.13	1.2×10^{-7}	-0.329	0.11	4.0×10^{-3}	305,426	-0.481	0.08	1.3×10^{-8}	DBP	TFAP2D[aefg]
MKLN1	7	131,059,056	rs13238550	A	0.40	0.486	0.07	9.4×10^{-12}	0.212	0.06	7.1×10^{-4}	325,647	0.331	0.05	1.9×10^{-12}	DBP	PODXL[cf], MKLN1[efg]
HIPK2	7	139,463,264	rs1011018	A	0.20	-0.441	0.09	6.1×10^{-7}	-0.244	0.08	1.6×10^{-3}	325,110	-0.329	0.06	1.5×10^{-8}		TBXAS1[cdf], HIPK2[cdefg]
ZFAT	8	135,612,745	rs894344	A	0.60	-0.384	0.07	6.8×10^{-8}	-0.163	0.06	8.2×10^{-3}	329,834	-0.258	0.05	3.2×10^{-8}		ZFAT[dfg]
PAX2	10	102,604,514	rs112184198	A	0.10	-0.826	0.12	7.8×10^{-13}	-0.532	0.10	1.3×10^{-7}	323,791	-0.659	0.08	3.6×10^{-18}	DBP,PP	PAX2[cefg]
MCF2L	13	113,636,156	rs9549328	T	0.23	0.440	0.08	1.5×10^{-7}	0.218	0.08	3.9×10^{-3}	313,787	0.318	0.06	1.5×10^{-8}	PP	MCF2L[defg]
FERMT2	14	53,377,540	rs9888615	T	0.29	-0.427	0.08	3.5×10^{-8}	-0.236	0.07	4.3×10^{-4}	326,235	-0.318	0.05	3.5×10^{-10}		FERMT2[defg]
PPP2R5E	14	63,928,546	rs8016306	A	0.80	0.454	0.09	2.5×10^{-7}	0.250	0.07	7.9×10^{-4}	329,869	0.335	0.06	3.7×10^{-9}	DBP	PPP2R5E[efg]
ABHD17C	15	81,013,037	rs35199222	A	0.45	0.353	0.07	5.7×10^{-7}	0.298	0.06	1.7×10^{-6}	323,407	0.322	0.05	5.2×10^{-12}	DBP	ABHD17C[efg]
CFDP1	16	75,331,044	rs11643209	T	0.42	-0.481	0.07	1.8×10^{-11}	-0.222	0.06	6.3×10^{-4}	309,242	-0.339	0.05	1.8×10^{-12}	PP	CFDP1[bfg], BCAR1[def]
CRK	17	1,333,598	rs12941318	T	0.49	-0.317	0.07	6.2×10^{-6}	-0.226	0.07	6.9×10^{-4}	299,739	-0.269	0.05	2.5×10^{-8}	PP	CRK[cdfg]

ACOX1	17	73,949,045	rs2467099	T	0.22	-0.423	0.08	$4.5x10^{-7}$	-0.216	0.07	$3.6x10^{-3}$	326,401	-0.307	0.06	$3.3x10^{-8}$		ACOX1[dfg], FBF1[acef]
Diastolic blood pressure																	
chr1mb25	1	25,030,470	rs6686889	T	0.25	0.231	0.05	$3.7x10^{-7}$	0.143	0.04	$9.1x10^{-4}$	322,575	0.185	0.03	$3.6x10^{-9}$		RUNX3[df], CLIC4[cdef], SRRM1[g]
DNM3	1	172,357,441	rs12405515	T	0.56	-0.219	0.04	$4.1x10^{-8}$	-0.118	0.04	$1.6x10^{-3}$	328,543	-0.165	0.03	$1.4x10^{-9}$		DNM3[defg]
GPATCH2	1	217,718,789	rs12408022	T	0.26	0.226	0.05	$5.9x10^{-7}$	0.172	0.04	$6.7x10^{-5}$	320,983	0.198	0.03	$2.4x10^{-10}$		GPATCH2[g]
CDC42BPA	1	227,252,626	rs10916082	A	0.73	-0.222	0.04	$5.3x10^{-7}$	-0.135	0.04	$1.5x10^{-3}$	327,636	-0.177	0.03	$8.4x10^{-9}$		CDC42BPA[efg]
WNT3A	1	228,191,075	rs2760061	A	0.47	0.235	0.04	$3.7x10^{-9}$	0.225	0.04	$1.1x10^{-8}$	312,761	0.230	0.03	$2.1x10^{-16}$	SBP	WNT9A[df], WNT3A[efg]
SDCCAG8	1	243,471,192	rs953492	A	0.46	0.293	0.04	$1.2x10^{-13}$	0.153	0.04	$4.6x10^{-5}$	325,253	0.220	0.03	$7.4x10^{-16}$		SDCCAG8[bcefg]
ADCY3	2	25,139,596	rs55701159	T	0.89	0.382	0.06	$1.1x10^{-9}$	0.193	0.06	$1.6x10^{-3}$	321,052	0.285	0.04	$7.2x10^{-11}$	SBP	ADCY3[efg]
SLC8A1	2	40,567,743	rs4952611	T	0.58	-0.200	0.04	$8.0x10^{-7}$	-0.114	0.04	$4.6x10^{-3}$	309,395	-0.157	0.03	$4.0x10^{-8}$		SLC8A1[cdefg]
AC016735.1	2	43,167,878	rs76326501	A	0.91	0.426	0.07	$4.3x10^{-10}$	0.413	0.07	$1.5x10^{-9}$	318,127	0.419	0.05	$3.6x10^{-18}$	SBP	PRKCE[df], HAAO[g]
GPAT2-FAHD2CP	2	96,675,166	rs2579519	T	0.63	-0.259	0.04	$1.7x10^{-10}$	-0.137	0.04	$6.7x10^{-4}$	311,557	-0.197	0.03	$4.8x10^{-12}$		ADRA2B[cdf], TCF7L1[cef], FAHD2CP[efg]
TEX41	2	145,646,072	rs1438896	T	0.30	0.288	0.04	$2.1x10^{-11}$	0.187	0.04	$4.3x10^{-6}$	329,278	0.234	0.03	$2.0x10^{-15}$	SBP	TEX41[g]
CCDC141	2	179,786,068	rs79146658	T	0.91	-0.375	0.07	$5.8x10^{-8}$	-0.245	0.07	$4.2x10^{-4}$	321,318	-0.311	0.05	$2.4x10^{-10}$		CCDC141[afg]
TMEM194B	2	191,439,591	rs7592578	T	0.19	-0.271	0.05	$8.9x10^{-8}$	-0.212	0.05	$1.7x10^{-5}$	304,672	-0.240	0.04	$9.5x10^{-12}$	SBP	NAB1[g]
TNS1	2	218,668,732	rs1063281	T	0.60	-0.231	0.04	$1.2x10^{-8}$	-0.172	0.04	$1.4x10^{-5}$	315,354	-0.200	0.03	$1.3x10^{-12}$	SBP	TNS1[cefg]
CAMKV-ACTBP13	3	49,913,705	rs36022378	T	0.80	-0.265	0.05	$6.3x10^{-8}$	-0.140	0.05	$3.9x10^{-3}$	319,983	-0.202	0.03	$4.7x10^{-9}$		CAMKV[efg]
CACNA2D2	3	50,476,378	rs743757	C	0.14	0.313	0.06	$2.9x10^{-8}$	0.184	0.05	$5.1x10^{-4}$	328,836	0.245	0.04	$2.4x10^{-10}$		CACNA2D2[dfg], C3orf18[def]
FAM208A	3	56,726,646	rs9827472	T	0.37	-0.207	0.04	$3.6x10^{-7}$	-0.148	0.04	$1.7x10^{-4}$	323,058	-0.177	0.03	$4.3x10^{-10}$		FAM208A[efg]
RP11-439C8.2	3	154,707,967	rs143112823	A	0.09	-0.484	0.07	$2.9x10^{-12}$	-0.295	0.08	$2.3x10^{-4}$	297,343	-0.403	0.05	$1.4x10^{-14}$	SBP	MME[defg]
SENP2	3	185,317,674	rs12374077	C	0.35	0.203	0.04	$8.3x10^{-7}$	0.127	0.04	$1.2x10^{-3}$	327,513	0.163	0.03	$9.2x10^{-9}$		SENP2[efg]
PDE5A	4	120,509,279	rs66887589	T	0.52	-0.296	0.04	$5.7x10^{-14}$	-0.140	0.04	$2.1x10^{-4}$	324,397	-0.215	0.03	$3.4x10^{-15}$		FABP2[cf], PDE5A[defg]
POC5	5	75,038,431	rs10078021	T	0.63	-0.223	0.04	$4.7x10^{-8}$	-0.105	0.04	$9.2x10^{-3}$	314,172	-0.164	0.03	$1.3x10^{-8}$		POC5[efg]
CPEB4	5	173,377,636	rs72812846	A	0.28	-0.232	0.04	$1.6x10^{-7}$	-0.186	0.04	$2.4x10^{-5}$	312,601	-0.209	0.03	$2.2x10^{-11}$		C5orf47[ef], CPEB4[g]
PKHD1	6	51,832,494	rs13205180	T	0.49	0.218	0.04	$3.7x10^{-8}$	0.123	0.04	$1.1x10^{-3}$	325,419	0.168	0.03	$7.0x10^{-10}$		PKHD1[cefg]
PDE10A	6	166,178,451	rs147212971	T	0.06	-0.421	0.08	$2.3x10^{-7}$	-0.289	0.09	$9.4x10^{-4}$	296,010	-0.360	0.06	$1.6x10^{-9}$		PDE10A[defg]
SLC35F1	6	118,572,486	rs9372498	A	0.08	0.459	0.07	$5.4x10^{-10}$	0.231	0.07	$5.6x10^{-4}$	330,625	0.334	0.05	$1.8x10^{-11}$	SBP	SLC35F1[dfg]
SNX31	8	101,676,675	rs2978098	A	0.54	0.212	0.04	$6.9x10^{-8}$	0.122	0.04	$1.4x10^{-3}$	324,424	0.165	0.03	$1.5x10^{-9}$		SNX31[efg]

RP11-273G15.2	8	144,060,955	rs62524579	A	0.53	-0.202	0.04	$2.8x10^{-7}$	-0.140	0.05	$2.2x10^{-3}$	268,645	-0.175	0.03	$3.8x10^{-9}$		CYP11B1[cdf], CYP11B2[cfg]
MTAP	9	21,801,530	rs4364717	A	0.55	-0.218	0.04	$3.5x10^{-8}$	-0.136	0.04	$2.9x10^{-4}$	327,173	-0.175	0.03	$1.3x10^{-10}$		MTAP[acefg]
BDNF	11	27,728,102	rs11030119	A	0.31	-0.211	0.04	$7.0x10^{-7}$	-0.119	0.04	$3.3x10^{-3}$	330,002	-0.163	0.03	$2.9x10^{-8}$		BDNF[cdefg]
MYEOV	11	69,079,707	rs67330701	T	0.09	-0.415	0.07	$7.8x10^{-9}$	-0.314	0.08	$3.8x10^{-5}$	276,760	-0.367	0.05	$2.1x10^{-12}$	SBP	MYEOV[efg]
RP11-321F6.1	15	66,869,072	rs7178615	A	0.37	-0.207	0.04	$3.8x10^{-7}$	-0.152	0.04	$1.0x10^{-4}$	318,076	-0.179	0.03	$2.6x10^{-10}$		LCTL[efg]
ADAMTS7	15	79,070,000	rs62012628	T	0.29	-0.295	0.04	$2.1x10^{-11}$	-0.147	0.06	$7.7x10^{-3}$	244,143	-0.238	0.03	$5.1x10^{-12}$		ADAMTS7[cdefg]
chr15mb95	15	95,312,071	rs12906962	T	0.68	-0.292	0.04	$5.3x10^{-12}$	-0.155	0.04	$1.5x10^{-4}$	319,952	-0.221	0.03	$5.6x10^{-14}$	SBP	LOC440311[ef], MCTP2[g]
PPL	16	4,943,019	rs12921187	T	0.43	-0.203	0.04	$3.0x10^{-7}$	-0.147	0.04	$1.2x10^{-4}$	326,469	-0.174	0.03	$2.5x10^{-10}$	SBP	PPL[aefg]
FBXL19	16	30,936,743	rs72799341	A	0.24	0.235	0.05	$3.0x10^{-7}$	0.139	0.04	$1.6x10^{-3}$	324,502	0.185	0.03	$5.8x10^{-9}$		CTF1[cdf], FBXL19[efg]
CMIP	16	81,574,197	rs8059962	T	0.42	-0.241	0.04	$2.0x10^{-9}$	-0.103	0.04	$8.5x10^{-3}$	319,839	-0.170	0.03	$1.3x10^{-9}$		CMIP[g]
ACE	17	61,559,625	rs4308	A	0.37	0.242	0.04	$3.2x10^{-9}$	0.186	0.04	$2.7x10^{-6}$	319,394	0.213	0.03	$6.8x10^{-14}$	SBP	ACE[cdefg]
MAPK4	18	48,142,854	rs745821	T	0.76	0.236	0.05	$3.2x10^{-7}$	0.150	0.04	$4.2x10^{-4}$	330,954	0.189	0.03	$1.4x10^{-9}$		MAPK4[defg]
CCNE1	19	30,294,991	rs62104477	T	0.33	0.209	0.04	$7.1x10^{-7}$	0.148	0.04	$2.4x10^{-4}$	320,347	0.177	0.03	$1.2x10^{-9}$		CCNE1[efg]
PLCB1	20	8,626,271	rs6108168	A	0.25	-0.305	0.05	$1.5x10^{-11}$	-0.127	0.04	$2.9x10^{-3}$	327,368	-0.211	0.03	$1.1x10^{-11}$	SBP	PLCB1[defg]
	Pulse pressure																
chr1mb9	1	9,441,949	rs9662255	A	0.43	-0.303	0.05	$4.7x10^{-10}$	-0.130	0.04	$3.0x10^{-3}$	310,618	-0.207	0.03	$1.9x10^{-10}$		SPSB1[efg]
SF3A3	1	38,455,891	rs4360494	C	0.55	0.332	0.05	$5.7x10^{-12}$	0.224	0.05	$3.6x10^{-6}$	282,851	0.278	0.03	$3.7x10^{-16}$		SF3A3[bfg], FHL3[bef]
RP4-710M16.1-PPAP2B	1	56,576,924	rs112557609	A	0.35	0.280	0.05	$3.2x10^{-8}$	0.187	0.04	$1.8x10^{-5}$	325,952	0.227	0.03	$6.8x10^{-12}$	SBP	PLPP3[cefg]
FGGY	1	59,653,742	rs3889199	A	0.71	0.462	0.05	$3.3x10^{-18}$	0.271	0.05	$1.9x10^{-9}$	329,486	0.351	0.03	$1.8x10^{-24}$	SBP	FGGY[dfg], HSD52[ef]
C2orf43	2	20,881,840	rs2289081	C	0.36	-0.251	0.05	$5.3x10^{-7}$	-0.203	0.04	$1.7x10^{-6}$	329,140	-0.223	0.03	$5.5x10^{-12}$		GDF7[efg]
PRKCE	2	46,363,336	rs11690961	A	0.88	0.437	0.07	$4.2x10^{-9}$	0.266	0.07	$4.6x10^{-5}$	327,847	0.340	0.05	$3.9x10^{-12}$		PRKCE[dfg]
CEP68	2	65,283,972	rs74181299	T	0.62	0.296	0.05	$2.1x10^{-9}$	0.181	0.04	$2.0x10^{-5}$	324,224	0.230	0.03	$9.6x10^{-13}$	SBP	CEP68[efg]
TCF7L1	2	85,491,365	rs11689667	T	0.54	0.256	0.05	$1.1x10^{-7}$	0.118	0.04	$3.8x10^{-3}$	330,634	0.176	0.03	$1.7x10^{-8}$		TCF7L1[cefg]
FN1	2	216,300,482	rs1250259	A	0.74	-0.457	0.05	$5.5x10^{-17}$	-0.210	0.05	$7.7x10^{-6}$	325,485	-0.314	0.04	$8.7x10^{-19}$	SBP	FN1[cedfg]
GATA2	3	128,201,889	rs62270945	T	0.03	0.861	0.14	$2.6x10^{-9}$	0.366	0.14	$9.5x10^{-3}$	279,925	0.607	0.10	$1.8x10^{-9}$		GATA2[cefg]
PALLD	4	169,717,148	rs1566497	A	0.42	0.320	0.05	$6.6x10^{-11}$	0.173	0.04	$4.8x10^{-5}$	320,948	0.236	0.03	$1.9x10^{-13}$		PALLD[cdfg]
chr4mb174	4	174,584,663	rs17059668	C	0.92	-0.442	0.09	$9.0x10^{-7}$	-0.245	0.08	$2.2x10^{-3}$	313,277	-0.332	0.06	$2.8x10^{-8}$		HAND2-AS1[g]
LHFPL2	5	77,837,789	rs10057188	A	0.46	-0.280	0.05	$5.5x10^{-9}$	-0.149	0.04	$3.3x10^{-4}$	325,985	-0.205	0.03	$6.7x10^{-11}$	SBP	LHFPL2[efg]
GJA1	6	121,781,390	rs11154027	T	0.47	0.311	0.05	$1.1x10^{-10}$	0.125	0.04	$3.7x10^{-3}$	316,708	0.207	0.03	$1.1x10^{-10}$		GJA1[cdfg]
ESR1	6	152,397,912	rs36083386	I	0.11	0.651	0.08	$4.6x10^{-17}$	0.289	0.07	$1.0x10^{-5}$	323,303	0.439	0.05	$1.5x10^{-18}$		ESR1[ecdfg]
FNDC1	6	159,699,125	rs449789	C	0.14	0.480	0.07	$2.2x10^{-12}$	0.264	0.06	$1.3x10^{-5}$	325,584	0.359	0.05	$2.4x10^{-15}$		FNDC1[defg]

THBS2	6	169,587,103	rs1322639	A	0.78	0.433	0.06	$7.7x10^{-14}$	0.230	0.05	$3.4x10^{-6}$	319,866	0.316	0.04	$4.8x10^{-17}$		THBS2[cdefg]
SUGCT	7	40,447,971	rs76206723	A	0.10	-0.405	0.08	$2.6x10^{-7}$	-0.305	0.07	$3.8x10^{-6}$	328,162	-0.346	0.05	$7.4x10^{-12}$		SUGCT[g]
SLC20A2	8	42,324,765	rs2978456	T	0.55	-0.253	0.05	$1.3x10^{-7}$	-0.130	0.05	$4.4x10^{-3}$	304,964	-0.188	0.03	$1.2x10^{-8}$		SLC20A2[defg]
TRAPPC9	8	141,060,027	rs4454254	A	0.63	-0.320	0.05	$9.4x10^{-11}$	-0.217	0.04	$2.9x10^{-7}$	330,022	-0.261	0.03	$5.1x10^{-16}$		TRAPPC9[efg]
SCAI	9	127,900,996	rs72765298	T	0.87	-0.392	0.07	$5.9x10^{-8}$	-0.358	0.07	$8.6x10^{-8}$	316,271	-0.374	0.05	$2.7x10^{-14}$	SBP	RABEPK[af], SCAI[efg]
KIAA1462	10	30,317,073	rs9337951	A	0.34	0.301	0.05	$7.6x10^{-9}$	0.262	0.05	$5.5x10^{-8}$	299,646	0.280	0.04	$2.5x10^{-15}$		KIAA1462[defg]
ARHGAP12	10	32,082,658	rs10826995	T	0.71	-0.317	0.05	$2.2x10^{-9}$	-0.133	0.05	$3.9x10^{-3}$	327,373	-0.212	0.03	$1.1x10^{-9}$		ZEB1[cf], ARHGAP12[efg]
PRDM11	11	45,208,141	rs11442819	I	0.11	-0.412	0.07	$3.8x10^{-8}$	-0.185	0.06	$3.3x10^{-3}$	326,483	-0.279	0.05	$7.1x10^{-9}$		PRDM11[efg]
NOX4	11	89,224,453	rs2289125	A	0.21	-0.481	0.06	$3.1x10^{-16}$	-0.293	0.05	$2.9x10^{-8}$	307,682	-0.377	0.04	$9.1x10^{-22}$		NOX4[acdefg]
CEP164	11	117,283,676	rs8258	T	0.38	0.341	0.05	$5.3x10^{-12}$	0.157	0.04	$2.4x10^{-4}$	327,038	0.236	0.03	$2.9x10^{-13}$		CEP164[defg]
CCDC41	12	94,880,742	rs139236208	A	0.10	-0.442	0.08	$5.7x10^{-8}$	-0.288	0.08	$2.8x10^{-4}$	291,244	-0.363	0.06	$1.6x10^{-10}$		CEP83-AS1[ef], CEP83[g]
RP11-61O1.1	14	98,587,630	rs9323988	T	0.63	-0.291	0.05	$5.6x10^{-9}$	-0.156	0.04	$2.0x10^{-4}$	327,551	-0.212	0.03	$4.1x10^{-11}$		C14orf177[g]
VAC14	16	70,755,610	rs117006983	A	0.01	1.448	0.30	$9.4x10^{-7}$	0.847	0.16	$1.8x10^{-7}$	250,766	0.986	0.14	$4.1x10^{-12}$		VAC14[efg]
CDH13	16	83,045,790	rs7500448	A	0.75	0.386	0.06	$4.2x10^{-12}$	0.288	0.05	$1.8x10^{-9}$	321,958	0.329	0.04	$1.1x10^{-19}$		CDH13[bdefg]
KIAA0753	17	6,473,828	rs7226020	T	0.56	-0.348	0.05	$1.3x10^{-12}$	-0.175	0.05	$1.4x10^{-4}$	303,389	-0.256	0.03	$2.3x10^{-14}$		KIAA0753[bfg], PITPNM3[ef]
TP53-SLC2A4	17	7,571,752	rs78378222	T	0.99	1.530	0.22	$8.9x10^{-12}$	0.487	0.18	$7.9x10^{-3}$	294,053	0.904	0.14	$1.8x10^{-10}$	DBP	TP53[cdefg]
KCNH4-HSD17B1	17	40,317,241	rs79089478	T	0.97	0.842	0.15	$1.2x10^{-8}$	0.377	0.13	$4.4x10^{-3}$	318,326	0.584	0.10	$3.1x10^{-9}$		KCNH4[g]
PYY	17	42,060,631	rs62080325	A	0.66	-0.260	0.05	$3.6x10^{-7}$	-0.128	0.05	$4.8x10^{-3}$	315,689	-0.186	0.03	$4.0x10^{-8}$		PYY[cefg]
MRC2	17	60,767,151	rs740698	T	0.56	-0.307	0.05	$2.1x10^{-10}$	-0.161	0.04	$2.8x10^{-4}$	311,450	-0.228	0.03	$3.1x10^{-12}$		MRC2[efg]
SLC14A2	18	43,097,750	rs7236548	A	0.18	0.462	0.06	$1.1x10^{-13}$	0.273	0.05	$2.2x10^{-7}$	330,075	0.352	0.04	$2.0x10^{-18}$		SLC14A2[cdefg]
SLC24A3	20	19,465,907	rs6081613	A	0.28	0.326	0.05	$1.2x10^{-9}$	0.213	0.05	$8.1x10^{-6}$	315,546	0.263	0.04	$1.6x10^{-13}$		SLC24A3[efg]
ARVCF	22	19,967,980	rs12628032	T	0.30	0.269	0.05	$2.4x10^{-7}$	0.216	0.05	$3.8x10^{-6}$	310,292	0.240	0.03	$5.5x10^{-12}$	SBP	ARVCF[efg]
XRCC6	22	42,038,786	rs73161324	T	0.05	0.611	0.11	$6.5x10^{-9}$	0.380	0.11	$3.1x10^{-4}$	267,722	0.496	0.07	$2.8x10^{-11}$		XRCC6[g]
(b) UK Biobank exome																	
Systolic blood pressure																	
SSPN	12	26,438,189	rs6487543	A	0.77	0.345	0.09	$5.9x10^{-5}$	0.279	0.06	$2.1x10^{-6}$	244,842	0.300	0.05	$6.3x10^{-10}$	DBP	SSPN[dfg]
Diastolic blood pressure																	
MRAS	3	138,119,952	rs2306374	T	0.84	-0.237	0.05	$9.3x10^{-6}$	-0.155	0.04	$9.3x10^{-5}$	281,715	-0.184	0.03	$7.4x10^{-9}$	SBP	MRAS[defg]
Pulse pressure																	

CD34	1	208,024,820	rs12731740	T	0.10	-0.360	0.08	$5.8x10^{-6}$	-0.202	0.05	$1.1x10^{-4}$	279,078	-0.249	0.04	$1.1x10^{-8}$		MIR29B2[df], CD34[dfg], LOC148696[ef]
ZNF638	2	71,627,539	rs3771371	T	0.57	-0.223	0.05	$4.1x10^{-6}$	-0.130	0.03	$9.6x10^{-5}$	280,285	-0.160	0.03	$5.8x10^{-9}$		DYSF[df], ZNF638[efg]
CRACR2B	11	828,916	rs7126805	A	0.73	0.262	0.05	$1.1x10^{-6}$	0.184	0.05	$4.6x10^{-4}$	145,162	0.222	0.04	$3.3x10^{-9}$		CD151[cdfg], CRACR2B[ef]

Locus: named according to the nearest annotated gene(s); Pos: build 37; EA: effect allele; 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; Traits: the other BP traits which reached genome-wide significance in the combined meta-analysis. Note: within the UK Biobank discovery analysis sample size was N=140,882/140,886 for systolic and pulse pressure / diastolic pressure, imputation quality score from SNPTEST ≥ 0.93 for all loci.

Candidate genes have been identified by one or multiple strategies:

[a]coding, nonsynonymous variant;

[b]GTEX eQTL

[c]CV KO Phenotype

[d]supporting biology

[e]Hi-C support

[f]vascular expression

[g]nearest to lead SNP

Table 2: Association results for new independent secondary variants identified at (a) validated 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) Validated 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	$3.1x10^{-8}$	-0.243	0.05	$5.6x10^{-6}$	327,614	-0.273	0.04	$1.1x10^{-12}$
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.) The variants with * denotes secondary signals which are in LD ($r^2 \geq 0.2$) with secondary signals which have been published since the time of our study[10,11]

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. UKBBvsUKBL: a binary indicator variable for UK Biobank vs UK BiLEVE to adjust for the different genotyping chips

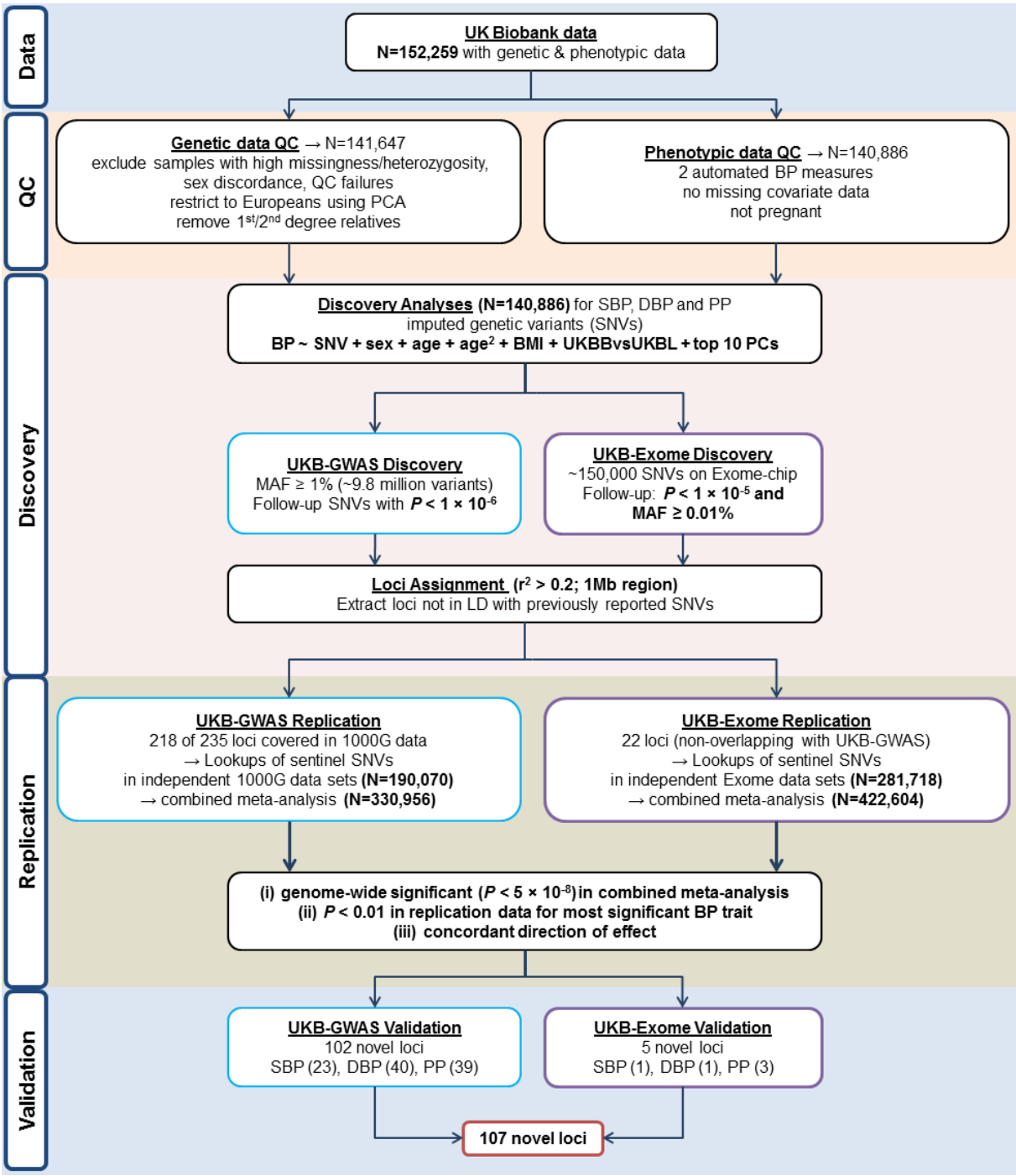


Figure 2: UK Biobank GWAS discovery Venn diagram of 107 validated novel loci showing concordance of significant associations across the three blood pressure phenotypes for the 107 novel sentinel variants (Table 1) from both the GWAS and exome analyses, according to genome-wide significance in the combined meta-analysis. The locus names labelled within the Venn Diagram correspond to Table 1, and relate to the nearest annotated gene.

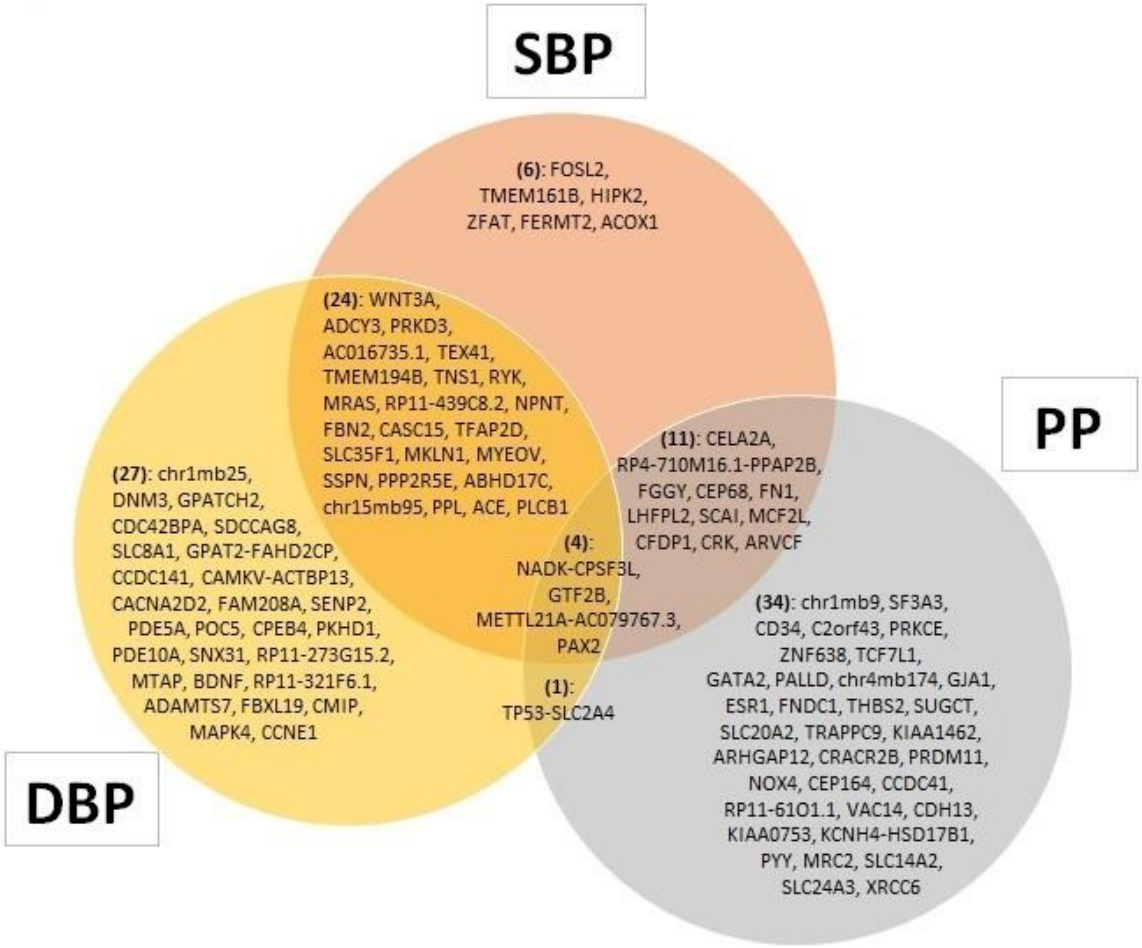


Figure 3: Association of blood pressure loci with other traits. Plot (A) shows results for associations with other traits which were extracted from the PhenoScanner database for the sentinel novel variants from Table 1, including proxies in Linkage Disequilibrium ($r^2 \geq 0.8$), with genome-wide significant 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 validated novel and previously reported variants. Enrichment was performed using the GREAT tool (http://bejerano.stanford.edu/great) with the sentinel SNVs as query.

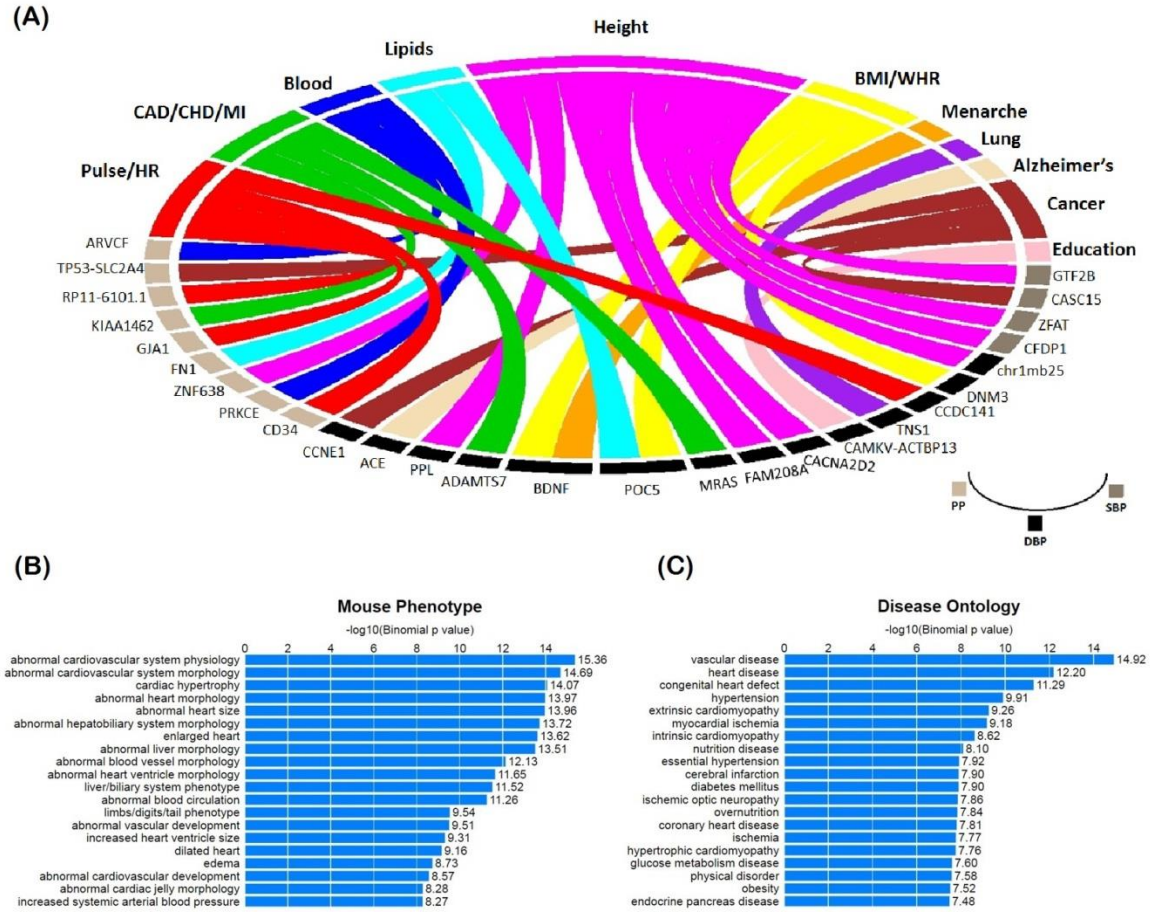
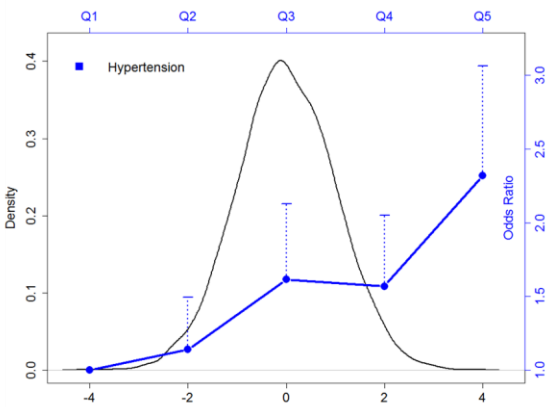
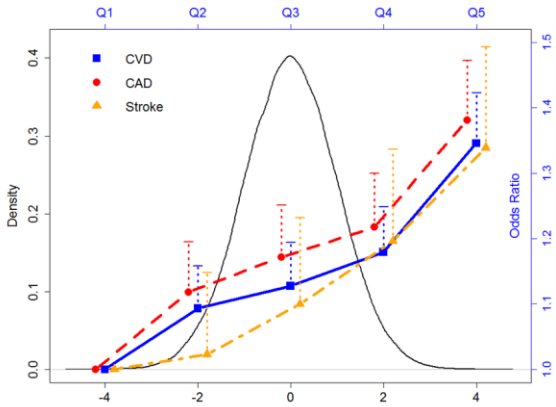


Figure 4: Distribution of Genetic Risk Score (GRS) based on previously reported and validated novel blood pressure variants and its relationship with blood pressure values, hypertension and CVD outcomes. A, Distribution of GRS in Airwave and sex-adjusted odds ratio of hypertension in age 50+ comparing each of the upper four GRS quintiles with the lowest quintile; dotted lines represent the upper 95% confidence intervals. B, Mean blood pressures and standard deviation in bracket in Airwave age 50+ across GRS quintiles. C, Distribution of GRS in UKB and sex-adjusted odds ratio of CVD, CAD and stroke comparing each of the upper four GRS quintiles with the lowest quintile; dotted lines represent the upper 95% confidence intervals. D, Count of CVD, CAD and stroke (events and deaths) across GRS quintiles in UKB participants

(A)



(C)



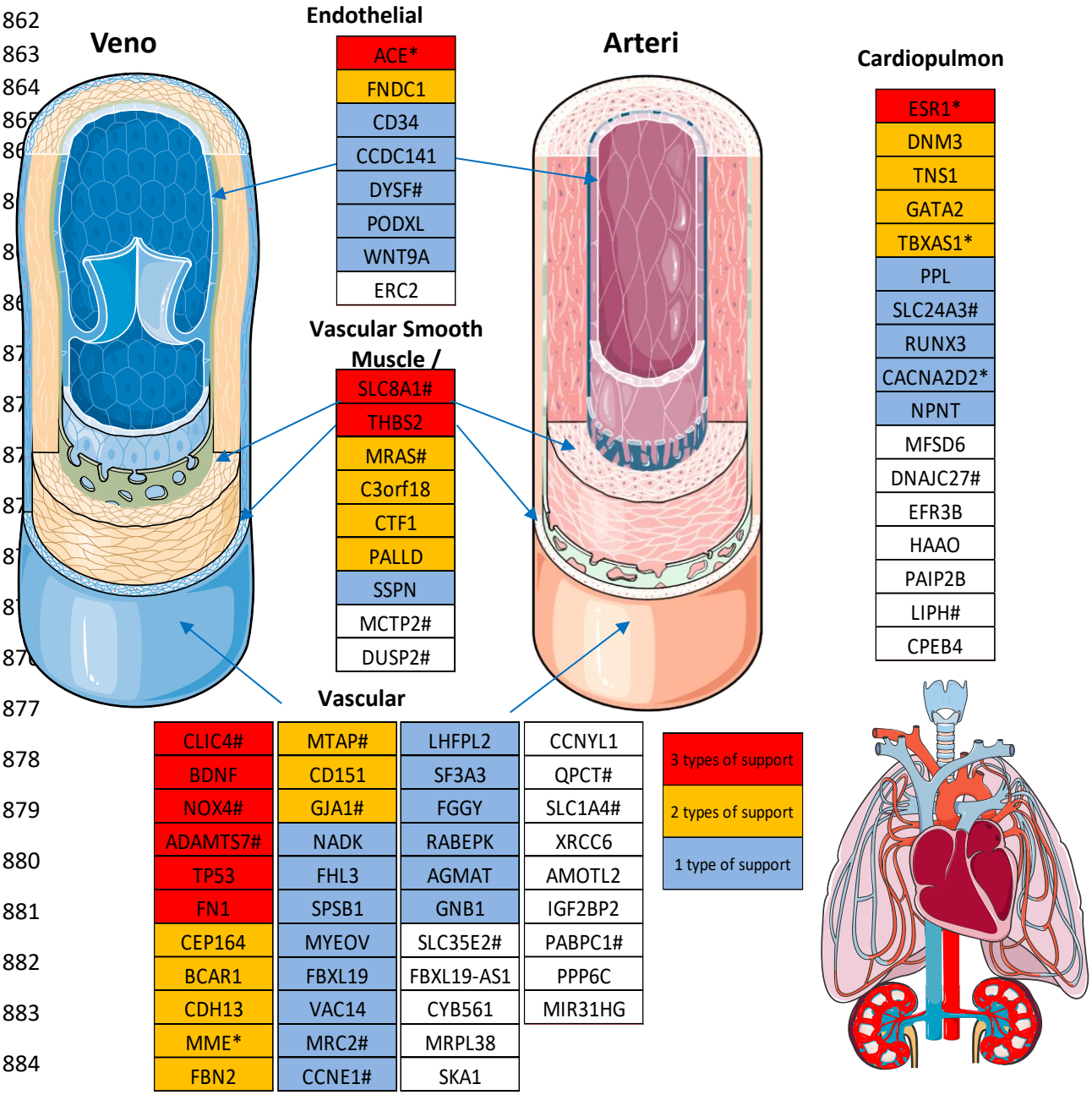
(B)

(mmHg)	Q1	Q2	Q3	Q4	Q5
SBP	136.4 (17.2)	138.7 (16.5)	142.5 (18.8)	142.5 (19.4)	145.7 (19.7)
DBP	82.9 (9.9)	84.3 (10.3)	85.9 (11.1)	86.0 (10.9)	87.6 (11.9)
PP	53.4 (11.6)	54.4 (11.3)	56.7 (12.3)	56.5 (12.4)	58.2 (12.4)

(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. We note that some druggable genes may carry a safety liability, such as GJA1, which has known association with QT interval[23]



887

888 **Online Methods**

889 **UK Biobank data**

890 Our Genome Wide Association Study (GWAS) analysis is performed using data from the
891 interim release of the first ~150k UK Biobank participants (Supplementary Methods)[17]. These
892 consist of ~100k individuals from UK Biobank genotyped at ~800,000 single nucleotide
893 variants (SNVs) with a custom Affymetrix UK Biobank Axiom Array chip[68] and ~50k individuals
894 genotyped with a custom Affymetrix UK BiLEVE Axiom Array chip from the UK BiLEVE study[69],
895 which is a subset of UK Biobank. SNVs were imputed centrally by UK Biobank using a merged
896 UK10K sequencing + 1000 Genomes imputation reference panel.

897 **Quality control**

898 Following quality control (QC) procedures already carried out centrally by UK Biobank, we
899 exclude discordant SNVs and samples with QC failures, gender discordance and high
900 heterozygosity/missingness. We further restrict our data to a subset of individuals of
901 European ancestry. By applying *kmeans* clustering to the Principal Component Analysis (PCA)
902 data a total of N=145,315 Europeans remain. Then we use the kinship data to exclude 1st and
903 2nd degree relatives, with N=141,647 unrelated individuals remaining. Finally we restrict our
904 data to non-pregnant individuals with two automated BP measurements available, resulting
905 in a maximum of N=140,886 individuals for analysis (Supplementary Methods).

906 **Phenotypic data**

907 After calculating the mean systolic and diastolic pressure values from the two blood pressure
908 measurements, we adjust for medication use by adding 15 and 10 mmHg to systolic and
909 diastolic pressure, respectively, for individuals reported to be taking blood pressure-lowering
910 medication (21.4% of individuals)[70]. Pulse Pressure is calculated as systolic minus diastolic
911 pressure, according to the medication-adjusted traits. Hypertension, used in secondary
912 analyses, is defined as: (i) systolic pressure ≥ 140 mmHg, or (ii) diastolic pressure ≥ 90 mmHg,
913 (iii) or taking blood pressure-lowering medication; otherwise individuals are classified as non-
914 hypertensive. Descriptive summary statistics are provided for all individuals, and stratified by
915 UK Biobank vs UK BiLEVE participants (**Supplementary Table 1**).

916 **Analysis models**

917 For the GWAS, we perform linear regression analyses of the three (untransformed)
918 continuous, medication-adjusted BP traits (systolic, diastolic and pulse pressure) for all
919 measured and imputed genetic variants in dosage format using SNPTEST software[71] under an
920 additive genetic model. We carry out a similar analysis for the exome content. Each analysis
921 includes the following covariates: sex, age, age^2, body mass index, top ten PCs and a binary
922 indicator variable for UK Biobank vs UK BiLEVE to adjust for the different genotyping chips.
923 We also run an association analysis within UK Biobank for validated novel blood pressure SNVs
924 and hypertension using logistic regression under an additive model with adjustments as
925 above. There are 76,554 hypertensive cases and the 64,384 remaining participants are

926 treated as non-hypertensive controls. This sample size is slightly larger than the N=140,866
927 used in the main analyses, since participants with only one blood pressure measurement, but
928 with reported blood pressure-lowering medication, could be included as hypertensive.

929 **Previously reported variants**

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

939 **Replication strategy**

940 We use three independent external data sets for replication (Supplementary Methods). First,
941 for the GWAS analysis based on advanced 1000 Genomes imputation enhanced by UK10K
942 data we consider SNVs with MAF ≥ 1% and perform a reciprocal replication exchange with the
943 International Consortium of Blood Pressure (ICBP) 1000 Genomes meta-analysis (max N =
944 150,134). The imputation strategy for ICBP 1000 Genomes meta-analysis is based on an
945 earlier imputation grid for the 1000 Genomes project. In addition, we recruit further cohorts
946 with 1000 Genomes data which had not contributed to the ICBP-1000 Genomes discovery
947 meta-analysis: ASCOT-UK (N = 3,803), ASCOT-SC (N = 2,462), BRIGHT (N = 1,791), Generation
948 Scotland (GS) (N = 9,749), EGCUT (N = 5,468), Lifelines (N = 13,292) and PREVEND (N = 3,619).
949 This gives a total of N = 190,318 independent replication samples for the GWAS discovery.

950 Second, because the UK Biobank and UK BiLEVE genotyping chips contain exome content, we
951 sought replication from two blood pressure exome consortia (European exome consortium
952 and the Cohorts for Heart and Ageing research in Genome Epidemiology – CHARGE BP exome
953 consortium), to allow validation of coding variants and variants with lower frequency. The
954 European exome consortium (N = 161,926) and CHARGE consortium (N = 119,792) give a total
955 of N = 281,718 independent replication samples for the UK Biobank exome discovery.
956
957 Note that the lookups for GWAS and exome discovery are distinct sets of SNVs. Loci are
958 assigned sequentially, prioritising the primary GWAS discovery first, then considering any
959 remaining loci with non-overlapping exome content for replication in the independent exome
960 replication resources.
961
962 **Statistical criteria for replication**

963 For the GWAS discovery, there are ~9.8 million SNVs with MAF ≥ 1% and INFO > 0.1. We
964 consider for follow-up any SNVs with $P < 1\text{x}10^{-6}$ for any of the three blood pressure traits. For
965 the exome discovery, there are 149,026 exome SNVs (Supplementary Methods) which were
966 polymorphic with INFO > 0.1; for follow-up we consider all SNVs with MAF ≥ 0.01% and $P <$
967 $1\text{x}10^{-5}$. All such SNVs are annotated to loci according to both an LD threshold of $r^2 \geq 0.2$ and a

968 1Mb interval region (see Supplementary Methods), and signals are classified either as
969 belonging to novel loci, or being potential secondary signals at previously reported loci.

970 **Selection of variants for follow-up**

971 The sentinel (most significant) SNV from each association signal is selected for follow-up, all
972 of which are pairwise-independent by LD ($r^2 < 0.2$). For the GWAS discovery, we check that
973 potential lookup SNVs are covered within the ICBP-1000G replication data (Supplementary
974 Methods). Of the 235 novel loci containing previously unreported SNVs with MAF ≥ 1%, INFO
975 > 0.1 and $P<1x10^{-6}$, 218 are covered, and similarly 100 of the 123 potential secondary SNVs at
976 51 of the 54 previously reported BP loci are available for follow-up. For the exome discovery,
977 by following up SNVs with MAF ≥ 0.01%, INFO > 0.1 and $P < 1x10^{-5}$ across the three blood
978 pressure traits, we carry forward for replication sentinel SNVs at 22 novel loci, and potential
979 secondary SNVs at three previously reported loci. We produce locus zoom plots for each of
980 the lookup variants.

981 **Replication meta-analyses**

982 The replication and combined meta-analyses are performed within METAL software[72] using
983 fixed effects inverse variance weighted meta-analysis (Supplementary Methods). The
984 combined meta-analysis of both the UK Biobank discovery (N = 140,886) and GWAS
985 replication meta-analysis (max N = 190,070) include a total maximum sample size of N =
986 330,956. For the exome combined meta-analysis, we synthesize data from the UK Biobank
987 discovery exome content (max N=140,866), with the replication dataset from both exome
988 consortia (total max N=281,718), giving a maximum sample size of N=422,604.

989 **Validation Criteria**

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

991 (i) the sentinel SNV is genome-wide significant ($P < 5\times10^{-8}$) in the combined meta-
992 analysis for any of the three blood pressure traits;
993 (ii) the sentinel SNV shows evidence of support ($P < 0.01$) in the replication meta-
994 analysis alone for association with the most significantly associated blood pressure
995 trait from the combined meta-analysis;
996 (iii) the sentinel SNV has concordant direction of effect between the UK Biobank
997 discovery and the replication meta-analysis for the most significantly associated
998 blood pressure trait from the combined meta-analysis.

999 **Secondary signals**

1000 By conditional analysis within UK Biobank data we assess all validated secondary signals from
1001 novel and previously reported loci for independence from the sentinel or previously reported
1002 SNV, respectively (Supplementary Methods). We declare a secondary signal to be
1003 independent of the previously reported SNV if there is less than a 1.5 fold difference between
1004 the main association and conditional association *P*-values on a –log10 scale, i.e. if –log10(*P*) /
1005 -log10(*P*_cond) < 1.5. Note that the lookup criteria already ensure that the secondary variant

1006 is not in LD ($r^2 < 0.2$) with the previously reported SNV. If more than one SNV in a region is
1007 found to be independent we undertake further rounds of iterative conditional analysis.

1008 **Lookups in non-European ancestries**

1009 As a secondary analysis, we look up 102 and 5 validated novel SNVs from the UK Biobank-
1010 GWAS and exome analyses, respectively, in non-European ancestry samples. These comprise
1011 analysis of East Asian (N = 31,513) and South Asian (N = 33,115) ancestry data from the iGEN-
1012 BP consortium[13] for the GWAS lookups, and South Asian (N = 25,937), African American (N =
1013 21,488) and Hispanic (N = 4,581) ancestry data from the CHARGE BP exome consortium[12] and
1014 CHD+ Exome consortium[11], for the exome content lookups (Supplementary Methods). We
1015 carry out a binomial (sign) test based on the number of SNVs with consistent directions of
1016 effect between UK Biobank and each of the non-European ancestry samples.

1017 **Monogenic blood pressure gene lookups**

1018 The UK Biobank and UK BiLEVE arrays include some rare coding variants for monogenic
1019 disorders. We collate a list of all specific mutation variants within genes known to be
1020 associated with monogenic blood pressure disorders[22]. Results from the UKB discovery
1021 association analyses for all three blood pressure traits are extracted for any of these SNVs
1022 directly covered within the UK Biobank dataset (**Supplementary Table 13**). Note that a search
1023 of proxies did not augment the list of available variants, so results are reported for the specific
1024 variants only.

1025 **Functional analyses**

1026 In order to prioritise associated SNVs, we use an integrative bioinformatics approach to
1027 collate functional annotation at both the variant and gene level for each SNV within the blood
1028 pressure loci (all SNVs in LD $r^2 \geq 0.8$ with the blood pressure-associated SNVs). At the variant
1029 level we use ANNOVAR[73] to obtain comprehensive functional characterisation of variants,
1030 including gene location, conservation and amino acid substitution impact based on a range of
1031 prediction tools including SIFT and polyphen2. All nonsynonymous variants were predicted
1032 damaging by two or more methods.

1033 We use the University of California Santa Cruz (UCSC) genome browser to review sequence
1034 specific context of SNVs in relation to function, particularly in the Encyclopedia of DNA
1035 Elements (ENCODE) dataset[74]. We use the UCSC table browser to annotate SNVs in ENCODE
1036 regulatory regions. We evaluate SNVs for impact on putative micro RNA target sites in the 3'
1037 un-translated regions (3'UTR) of transcripts by a query of the miRNASNP database[75]. We
1038 evaluate all SNVs in LD ($r^2 \geq 0.8$) with our novel sentinel SNVs for evidence of mediation of
1039 expression quantitative trait loci (eQTL) in all 44 tissues using the Genotype-Tissue Expression
1040 (GTEx) database (www.gtexportal.org), in order to identify novel loci which are highly
1041 expressed, and to highlight specific tissue types which show eQTLs for a large proportion of
1042 novel loci. We further seek to identify novel loci with the strongest evidence of eQTL
1043 associations in arterial tissue, in particular.

1044 At the gene level, we use Ingenuity Pathway Analysis (IPA) software (IPA®,QIAGEN Redwood
1045 City,www.qiagen.com/ingenuity) to review genes with prior links to blood pressure, based on

1046 annotation with the "Blood Pressure" Medline Subject Heading (MESH) term which is
1047 annotated to 684 genes. We also use IPA to identify genes which interact with blood pressure
1048 MESH annotated genes, and evaluate genes for evidence of small molecule druggability based
1049 on queries of Chembl (www.ebi.ac.uk/chembl/) and Drug Gene Interaction database
1050 (dgidb.genome.wustl.edu).

1051 We then perform overall enrichment testing across all loci. Firstly, we use DEPICT[76] (Data-
1052 driven Expression Prioritized Integration for Complex Traits) to identify highly expressed
1053 tissues and cells within the blood pressure loci. DEPICT uses a large number of microarrays
1054 (~37k) to identify cells and tissues where the genes are highly expressed and uses
1055 precomputed GWAS phenotypes to adjust for co-founding sources. DEPICT provides a *P*-value
1056 of enrichment and false discovery rates adjusted *P*-values for each tissue/cells tested.

1057 Furthermore, to investigate regulatory regions, we employ a two tiered approach to
1058 investigate cell type specific enrichment within DNase I sites using FORGE, which tests for
1059 enrichment of SNVs within DNase I sites in 123 cell types from the Epigenomics Roadmap
1060 Project and ENCODE[77] (Supplementary Methods). Validated novel sentinel SNVs discovered
1061 in our study are analysed along with previously reported SNVs and secondary signals (with *P*-
1062 value $< 1\times10^{-4}$) to evaluate the overall tissue specific enrichment of blood pressure associated
1063 variants. In a second analysis we use FORGE (with no LD filter) to investigate directly our
1064 curated candidate regulatory SNVs for overlap with cell-specific DNase I signals.

1065 GenomeRunner[78] is used to search for enrichment of validated novel and previously reported
1066 sentinel SNVs with histone modification mark genomic features (Supplementary Methods).
1067 Relevant cardiovascular tissue expression is investigated using Fantom5 reference transcript
1068 expression data (fantom.gsc.riken.jp/5) (Supplementary Methods).

1069 We use IPA (IPA®,QIAGEN Redwood City,www.qiagen.com/ingenuity) to identify biological
1070 pathways and transcriptional upstream regulators enriched for genes within the blood
1071 pressure loci. The transcriptional upstream regulator analysis aims to identify transcription
1072 factors, compounds, drugs, kinases and other molecules, for which the target is one of the
1073 blood pressure genes under investigation.

1074 We query SNVs against PhenoScanner[19] to investigate trait pleiotropy, extracting all
1075 association results with nominal significance at $P < 0.05$ for full reporting (**Supplementary**
1076 **Table 15**), and then extract genome-wide significant results to highlight the validated novel
1077 loci with strongest evidence of association with other traits (**Fig. 3a**). We also use the Genomic
1078 Regions Enrichment of Annotations Tool (GREAT) to study gene set enrichment of mouse
1079 phenotype and disease ontology terms within our validated novel and previously reported
1080 loci, using default SNV to gene mapping settings[79].

1081 We carry out metabolomics analysis using two sets of data. First we use ^{1}H NMR lipidomics
1082 data on plasma from a subset of 2,000 participants of the Airwave Health Monitoring
1083 Study[80,81] (Supplementary Methods). For each replicated blood pressure-associated SNV we
1084 ran association tests with the lipidomics data using linear regression analyses, adjusted for
1085 age and sex. We computed significance thresholds using a permutation derived family wise
1086 error rate (5%) to account for the high correlation structure of these data (ENT=35)[82]. We also

1087 test each replicated SNV against published genome-wide vs metabolome-wide associations
1088 in plasma and urine using publicly available data from the "Metabolomics GWAS Server" to
1089 identify metabolites that have been associated with variants of interest at $P < 3.0 \times 10^{-4}$
1090 (Bonferroni corrected *P* for validated signals)[25,26].

1091 **Experimental methods**

1092 We prioritise novel genes for laboratory testing on the basis of evidence for SNV function
1093 (including coding variants, eQTLs and Hi-C interactions), biological support for relevance to
1094 blood pressure (from literature review) and transgenic phenotype. We perform genotyping
1095 and Quantitative Reverse-Transcription Polymerase Chain Reaction (q RT-PCR) for the
1096 selected sentinel variants of interest using human vascular smooth muscle cells and
1097 endothelial cells and test for expression levels (Supplementary Methods). All three SNVs were
1098 tested using an additive model.

1099 **Genetic risk scores**

1100 First, by calculating genetic risk scores (GRS), we use the Airwave study[80] data to assess the
1101 effect in an independent cohort of the blood pressure-associated variants on blood pressure
1102 and risk of hypertension (Supplementary Methods). This provides an estimate of the
1103 combined effect of the blood pressure raising variants avoiding bias by "winners curse". We
1104 create weighted GRSs for all pairwise-independent, LD-filtered ($r^2 < 0.2$) previously reported
1105 variants and validated novel variants (sentinel and secondary SNVs) combined, using SNVs
1106 available in Airwave (**Supplementary Table 21**). For the previously reported variants, we
1107 weight blood pressure increasing alleles by the beta coefficients from the UK Biobank
1108 discovery GWAS. For the novel variants, beta coefficients of the replication meta-analysis are
1109 used as independent, unbiased weights.

1110 For the analyses of trait variance explained, we use three trait-specific GRSs (i.e. systolic,
1111 diastolic and pulse pressure). Each GRS includes all variants, but weights are trait-specific,
1112 using the beta coefficients from the analysis of each of the three different blood pressure
1113 traits, e.g. the systolic GRS is weighted by the beta coefficients from the systolic GRS. To
1114 calculate the percent of variance for each blood pressure trait explained by its corresponding
1115 trait-specific GRS, not accounted for by known factors, we generate the residuals from the
1116 regression model of each trait against covariates of age, age^2, sex and body mass index. We
1117 then fit a second linear model for the trait residuals with all the variants in the GRS plus the
1118 top 10 principal components. We calculate these percentage variance explained results
1119 within an independent population (Airwave).

1120 For risk score analyses we calculate a single blood pressure GRS, as the average of the systolic
1121 and diastolic pressure GRSs. We standardize the average GRS to have mean of zero and
1122 standard deviation of one. We assess the association of the continuous average GRS variable
1123 with each blood pressure trait by simple linear regression. We also run a logistic regression to
1124 examine the association of the average GRS with risk of hypertension. We perform each
1125 analysis both with and without adjustment for sex. We test for interaction between age
1126 (below age 50, and 50 years and above) and the effect of the GRS on blood pressure. We then
1127 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.

We also assess the association of the average blood pressure GRS 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

68. The UK Biobank Array Design Group. UK Biobank Axiom Array Content Summary. (2014).
69. 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).
70. 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).
71. 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).
72. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
73. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).

1167 74. Barnes, M.R. Exploring the landscape of the genome. *Methods Mol Biol* **628**, 21-38
1168 (2010).
1169 75. Gong, J. *et al.* Genome-wide identification of SNPs in microRNA genes and the SNP
1170 effects on microRNA target binding and biogenesis. *Hum Mutat* **33**, 254-63 (2012).
1171 76. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using
1172 predicted gene functions. *Nat Commun* **6**(2015).
1173 77. Dunham, I., Kulesha, E., Iotchkova, V., Morganella, S. & Birney, E. *FORGE: A tool to*
1174 *discover cell specific enrichments of GWAS associated SNPs in regulatory regions*
1175 *[version 1; referees: 2 approved with reservations]*, (2015).
1176 78. Dozmorov, M.G., Cara, L.R., Giles, C.B. & Wren, J.D. GenomeRunner: automating
1177 genome exploration. *Bioinformatics* **28**, 419-20 (2012).
1178 79. McLean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory
1179 regions. *Nat Biotechnol* **28**, 495-501 (2010).
1180 80. Elliott, P. *et al.* The Airwave Health Monitoring Study of police officers and staff in
1181 Great Britain: rationale, design and methods. *Environ Res* **134**, 280-5 (2014).
1182 81. Petersen, M. *et al.* Quantification of lipoprotein subclasses by proton nuclear
1183 magnetic resonance-based partial least-squares regression models. *Clin Chem* **51**,
1184 1457-61 (2005).
1185 82. Chadeau-Hyam, M. *et al.* Metabolic profiling and the metabolome-wide association
1186 study: significance level for biomarker identification. *J Proteome Res* **9**, 4620-7
1187 (2010).

1188

1189

1190