ONLINE METHODS

Cohorts contributing to systolic (SBP) and diastolic blood pressure (DBP) analyses

Studies contributing to BP association discovery including community- and population-based collections as well as studies of non-BP traits, analyzed as case and control samples separately. Details on each of the studies including study design and BP measurement are provided in **Supplementary Table 1**, genotyping information in **Supplementary Table 2**, and participant characteristics in **Supplementary Table 3.** All participants provided written informed consent and the studies were approved by local Research Ethics Committees and/or Institutional Review Boards.

European ancestry meta-analysis

A meta-analysis of 340,934 individuals of European descent was undertaken in four stages with subsequent validation in an independent cohort. Because stage 1 Cardio-MetaboChip samples included many SNPs selected on the basis of association with BP in earlier GWAS, we performed genomic control using a set of putative null SNPs based on $P > 0.10$ in earlier GWAS of SBP and DBP or both. Stage 2 samples with genome-wide genotyping used the entire genome-wide set of SNPs for genomic control given the lack of ascertainment. The study design is summarized in **Supplementary Figure 1**, and further details are provided in **Supplementary Tables 2-5** and the **Supplementary Note**.

Systematic PubMed search +/- 100kb of each newly discovered index SNP

All genes with any overlap with a 200kb region centered around each of the 20 newly discovered lead SNPs were identified using the UCSC Genome Browser. A search term was constructed for each gene including the short and long gene name and the terms "blood pressure" and "hypertension" (e.g. for *NPPA* on chr 1: "NPPA OR natriuretic peptide A AND (blood pressure OR hypertension)") and the search results of each search term from PubMed were individually reviewed.

Trait variance explained

The trait variance explained by 66 lead SNPs at novel and known loci was evaluated in one study that contributed to the discovery effort: the Atherosclerosis Risk in Communities (ARIC) study. We constructed a linear regression model with all 66 or the subset of 49 known SNPs as a set of predictors of the BP residual after adjustment for covariates of the adjusted treatment-corrected BP phenotype (SBP or DBP). The r^2 from the regression model was used as the estimate of trait variance explained.

European ancestry GCTA-COJO analysis

To identify multiple distinct association signals at any given BP locus, we undertook approximate conditional analyses using a model selection procedure implemented in the GCTA-COJO software package[44,45]. To evaluate the robustness of the GCTA-COJO results to the choice of reference data set, model selection was performed using the LD between variants in separate analyses from two datasets of European descent, both with individuals from the UK with Cardio-MetaboChip genotype data: GoDARTS with 7,006 individuals and WTCCC1-T2D/58BC with 2,947 individuals. Assuming that the LD between SNPs more than 10 Mb away or on different chromosomes is zero, we undertook the GCTA-COJO step-wise model selection to select SNPs that were conditionally-independently associated with SBP and DBP, in turn, at a genome-wide significance, given by $P < 5\times10^{-8}$ (**Supplementary Tables 6-8**) using the stage 4 combined European GWAS+ Cardio-MetaboChip meta-analysis.

Conditional analyses in the Women's Genome Health Study (WGHS)

Multivariable regression modeling was performed for each possible combination of putative independent SNPs from a) model selection implemented in GCTA-COJO and b) a comprehensive manual

review of the literature (**Supplementary Table 9**). Any SNP with $P < 5x10^{-8}$ in a previous reported BP GWAS was considered. A total of 46 SNPs were examined (**Supplementary Table 10**). Genome-wide genotyping data imputed to 1000 Genomes in the WGHS (N = 23,047) were used. Regression modeling was performed in the R statistical language (**Supplementary Table 10**).

Fine mapping and determination of credible sets of causal SNPs

The GCTA-COJO and WGHS conditional analyses identified multiple distinct signals of association at multiple loci (**Supplementary Tables 6 and 10**). Of the 24 loci considered in fine-mapping analyses, 16 had no evidence for the existence of multiple distinct association signals, so it is reasonable to assume that there is a single causal SNP and therefore the credible sets of variants could be constructed using the association summary statistics from the unconditional meta-analyses. However, in the remaining eight loci, where evidence of secondary signals was observed from GCTA-COJO, we performed approximate conditional analyses across the region by conditioning on each index SNP (**Supplementary Table 11**). By adjusting for the other index SNPs at the locus, we can therefore assume a single variant is driving each "conditionally-independent" association signal, and we can construct the 99% credible set of variants on the basis of the approximate conditional analysis from GCTA-COJO (**Supplementary Tables 12-13**). At five of the eight loci with multiple distinct signals of association, one index SNP mapped outside of the fine-mapping region, so a credible set could not be constructed.

eQTL analysis: Whole Blood

NESDA/NTR: Whole blood eQTL analyses were performed in samples from the Netherlands Study of Depression and Anxiety (NESDA)[46] and the Netherlands Twin Registry (NTR)[47] studies. RNA expression analysis was performed in the statistical software R. The residuals resulting from the linear regression analysis of the probe set intensity values onto the covariates sex, age, body mass index (kg/m^2), smoking status coded as a categorical covariate, several technical covariates, and three principal components were used. The eQTL effects were detected using a linear mixed model approach, including for each probe set the expression level (normalized, residualized and without the first 50 expression PCs) as dependent variable; the SNP genotype values as fixed effects; and family identifier and zygosity (in the case of twins) as random effects to account for family and twin relations[48].

The eQTL effects were defined as *cis* when probe set–SNP pairs were at distance < 1M base pairs. At a FDR of 0.01 used genome-wide, therefore not only considering the candidate SNPs, for cis-eQTL analysis the *P* value threshold was $1x10^{-4}$. For each probe set that displayed a statistically significant association with at least one SNP located within its *cis* region, we identified the most significantly associated SNP and denoted this as the top *cis*-eQTL SNP. See **Supplementary Note** for details.

eQTL analysis: Selected published eQTL datasets

Lead BP SNP and proxies ($r^2 > 0.8$) were searched against a collected database of expression SNP (eSNP) results. The reported eSNP results met criteria for statistical thresholds for association with gene transcript levels as described in the original papers. The non-blood cell tissue eQTLs searched included aortic endothelial cells[49], left ventricle of the heart [50], cd14+ monocytes [51] and the brain [52]. The results are presented in **Supplementary Tables 14-15**.

Enrichment analyses: Analysis of cell-specific DNase hypersensitivity sites (DHSs) using an OR method

The overlap of Cardio-MetaboChip SNPs with DHSs was examined using publicly available data from the Epigenomics Roadmap Project and ENCODE, choosing different cutoffs of Cardio-MetaboChip *P* values. The DHS mappings were available for 123 mostly adult cells and tissues [53] (downloaded from http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDnase/). The DHS

mappings were specified as both "narrow" and "broad" peaks, referring to reduction of the experimental data to peak calls at 0.1% and 1.0% FDR thresholds, respectively. Thus, the "narrow" peaks are largely nested within the "broad" peaks. Experimental replicates of the DHS mappings (typically duplicates) were also available for the majority of cells and tissues.

SNPs from the Cardio-MetaboChip genome-wide scan were first clumped in PLINK in windows of 100kb and maximum $r^2 = 0.1$ among LD relationships from the 1000 Genomes European data. Then, the resulting index SNPs at each *P* value threshold were tagged with $r^2 = 0.8$ in windows of 100kb, again using LD relationships in the 1000 Genomes, restricted to SNPs with MAF > 1% and also present in the HapMap2 CEU population. A reference set of SNPs was constructed using the same clumping and tagging procedures applied to GWAS catalog SNPs (available at http://www.genome.gov/gwastudies/, accessed 3/13/2013)[54] with discovery $P < 5x10^{-8}$ in European populations. A small number of reference SNPs or their proxies overlapping the BP SNPs or their proxies were excluded. After LD pruning and exclusions, there were a total of 1,196 reference SNPs. For each cell type and *P* value threshold, the enrichment of SBP or DBP SNPs (or their LD proxies) mapping to DHSs was expressed as an odds ratio (OR) relative to the GWAS catalog reference SNPs (or their LD proxies), using logistic mixed effect models treating the replicate peak determinations as random effects (glmer package in R). The significance of the enrichment ORs was derived from the significance of beta coefficients for the main effects in the mixed models (**Figure 3, Supplementary Table 16**).

Enrichment analyses: Analysis of tissue-specific enrichment of BP variants and H3K4me3 sites

An analysis to test for significant cell-specific enrichment in the overlap of BP SNPs (or their proxies) with H3K4me3 sites was performed as described in Trynka et al, 2013[55]. The measure of overlap is a "score" that is constructed by dividing the height of an H3K4me3 ChIP signal in a particular cell by the distance between the nearest test SNP. The significance of the scores (i.e. *P* value) for all SNPs was determined by a permutation approach that compares the observed scores to scores of SNPs with similar properties to the test SNPs, essentially in terms of LD and proximity to genes (**Supplementary Note**). The number of significant digits in the *P* values is determined by the number of permutations and we conducted 10,000 iterations. Results are shown in **Supplementary Table 19**.

Enrichment analyses: Analysis of tissue-specific DHSs and chromatin states using GREGOR

The DNase-seq ENCODE data for all available cell types were downloaded in the processed "narrowPeak" format. The local maxima of the tag density in broad, variable-sized "hotspot" regions of chromatin accessibility were thresholded at FDR 1% with peaks set to a fixed width of 150bp. Individual cell types were further grouped into 41 broad tissue categories (http://genome.ucsc.edu/ENCODE/cellTypes.html) by taking the union of DHSs for all related cell types and replicates. For each GWAS locus, a set of matched control SNPs was selected based on three criteria: 1) number of variants in LD ($r^2 > 0.7$; ± 8 variants), 2) MAF (± 1%), and 3) distance to nearest gene (± 11,655 bp). To calculate the distance to the nearest gene, the distance to the 5' flanking gene (start and end position) and to the 3' flanking gene was calculated and the minimum of these 4 values was used. If the SNP fell within the transcribed region of a gene, the distance was 0. The probability that a set of GWAS loci overlap with a regulatory feature more often than we expect by chance was estimated.

Enrichment analyses: FAIRE analysis of BP variants in fine-mapping regions in lymphoblastoid cell lines

FAIRE analysis was performed on a sample of 20 lymphoblastoid cell lines of European origin. All samples were genotyped using the Cardio-MetaboChip genotyping array, and BP SNPs and LD proxies ($r^2 > 0.8$) at the fine mapping loci (N = 24, see **Supplementary Table 23**) were assessed to identify heterozygous imbalance between non-treated and FAIRE-treated chromatin. A paired t-test was used to

compare the B allele frequency (BAF) arising from formaldehyde-fixed chromatin sheared by sonication and DNA purified to the BAF when the same chromatin sample underwent FAIRE to enrich for open chromatin. Three hundred and fifty-seven Cardio-MetaboChip BP SNPs were directly genotyped across the fine mapping regions. The Bonferroni-corrected threshold of significance is $P < 0.0001$ (0.05/357). The results for SNPs with $P < 0.05$ are reported in (**Supplementary Table 23**). FAIRE results were not available for some SNPs with missing data due to genotype failure or not having >3 heterozygous individuals for statistical analysis. Therefore there are no results for three lower frequency BP loci (*SLC39A8*, *CYP17A1-NT5C2* and *GNAS-EDN3*) and for the second signal at the following loci: *MTHFR-NPPB* (rs2272803), *MECOM* (rs2242338) and *HFE* rs1800562).

Pathway analyses: MAGENTA

MAGENTA tests for enrichment of gene sets from a precompiled library derived from GO, KEGG, PATHTER, REACTOME, INGENUITY, and BIOCARTA was performed as described by Segré et al, 2010[56]. Enrichment of significant gene-wide *P* values in gene sets is assessed by 1) using LD and distance criteria to define the span of each gene, 2) selecting the smallest *P* value among SNPs mapping to the gene span, and 3) adjusting this *P* value using a regression method that accounts for the number of SNPs, the LD, etc. In the second step, MAGENTA examines the distribution of these adjusted *P* values and defines thresholds for the 75%-ile and the 95%-ile. In the third step, MAGENTA calculates an enrichment for each gene set by comparing the number of genes in the gene set with *P* value less than either the 75th or 95th %ile to the number of genes in the gene set with *P* value greater than either the 75th or 95th %ile, and then comparing this quotient to the same quotient among genes not in the gene set. This gene-set quotient is assigned a *P* value based on reference to a hypergeometric distribution. The results based on our analyses are indicated in **Supplementary Table 21**.

Pathway analyses: DEPICT

We applied the DEPICT [57] analysis separately on genome-wide significant loci from the overall blood pressure (BP) Cardio-MetaboChip analysis including published blood pressure loci (see **Supplementary Table 22**). SNPs at the *HFE* and *BAT2-BAT5* loci (rs1799945, rs1800562, rs2187668, rs805303, rs9268977) could not be mapped. As a secondary analysis, we additionally included associated loci ($P < 1x10^{-5}$) from the Cardio-MetaboChip stage 4 combined meta-analyses of SBP and the DBP. DEPICT assigned genes to associated regions if they overlapped or resided within associated LD blocks with $r^2 > 0.5$ to a given associated SNP.

Literature review for genes at the newly discovered loci

Recognizing that the most significantly associated SNP at a locus may not be located in the causal gene and that the functional consequences of a SNP often extends beyond 100kb, we conducted a literature review of genes in extended regions around newly discovered BP index SNPs. The genes for this extensive review were identified by DEPICT (**Supplementary Table 22**).

Non-European meta-analysis

To assess the association of the 66 significant loci from the European ancestry meta-analysis in non-European ethnicities, we obtained lookup results for the 66 index SNPs for participants of South-Asian ancestry (8 datasets, total N = 20,875), East-Asian ancestry (5 datasets, total N = 9,637), and African- and African-American ancestry (6 datasets, total N = 33,909). The association analyses were all conducted with the same covariates (age, age^2, sex, BMI) and treatment correction (+15/10 mm Hg in the presence of any hypertensive medication) as the association analyses for the discovery effort in Europeans. Tests for heterogeneity across effect estimates in European, South Asian, East Asian and African derived samples were performed using GWAMA[58].

Genetic risk score and cardiovascular outcomes

The gtx package for the R statistical programming language was used to estimate the effect of the SNP-risk score on the response variable in a regression model[45].

REFERENCES

44. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
45. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
46. Penninx, B.W. *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int J Methods Psychiatr Res* **17**, 121-40 (2008).
47. Boomsma, D.I. *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet* **9**, 849-57 (2006).
48. Visscher, P.M., Benyamin, B. & White, I. The use of linear mixed models to estimate variance components from data on twin pairs by maximum likelihood. *Twin Res* **7**, 670-4 (2004).
49. Romanoski, C.E. *et al.* Network for activation of human endothelial cells by oxidized phospholipids: a critical role of heme oxygenase 1. *Circ Res* **109**, e27-41 (2011).
50. Koopmann, T.T. *et al.* Genome-wide identification of expression quantitative trait loci (eQTLs) in human heart. *PLoS One* **9**, e97380 (2014).
51. Fairfax, B.P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).
52. Ramasamy, A. *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci* **17**, 1418-28 (2014).
53. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-5 (2012).
54. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001-6 (2014).
55. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* **45**, 124-30 (2013).
56. Segre, A.V. *et al.* Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* **6**(2010).
57. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
58. Magi, R. & Morris, A.P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).