|  |
| --- |
|  |
| **Supplementary Figure 1** |
| GWAS discovery Manhattan plots. |
| Manhattan plots (a), (b) and (c) for systolic blood pressure (SBP), diastolic blood pressure (DBP) and pulse pressure (PP) respectively. P-value results from the GWAS discovery meta-analysis (N=757,601), were derived using inverse variance fixed effects meta-analysis and they are plotted on a – log10 scale for all SNPs with Minor Allele Frequency (MAF) ≥ 1%. SNPs within the 274 known loci (±500kb; Linkage Disequilibrium r2 ≥ 0.1) are highlighted in green. |
|  |
| **Supplementary Figure 2** |
| Effect Sizes of all Blood Pressure associated loci. |
| Plot (A) shows strong correlation between the published effect size estimates (x-axis) from literature vs the effect sizes from our discovery meta-analysis (y-axis), for known SNPs, colour-coded according to the published primary trait from the first published report. From the 357 validated SNPs listed in Supplementary Table 4 from the 274 published loci, 327 are available within the MAF ≥ 1% HRC-imputed data. For comparison of effect sizes, we only consider 299 such SNPs which have been identified from main-effect genetic association studies within Europeans (i.e. excluding any SNPs from interaction/stratified/multi-phenotype analysis, or from studies of other ancestries). For reliable comparison of effect sizes, we further restrict to the 284 known SNPs which reach genome-wide significance within the discovery meta-analysis for at least one BP trait. The r2 value is presented to show the correlation between published and observed effect sizes. Plots (B), (C) and (D) are trait-specific plots for SBP, DBP and PP, respectively (SBP: systolic blood pressure; DBP: diastolic blood pressure; PP: pulse pressure). Across all plots, the 284 “known” SNPs (black squared) from plot (A) are compared against the 325 novel sentinel SNPs from the 2-stage analysis (red circles), the 210 novel sentinel SNPs from the 1-stage analysis (green triangles), and the 92 SNPs (blue diamonds) replicated for the first time from Hoffman et al9. Each SNP is only plotted in one of the trait-specific plots, according to the published primary trait for the known SNPs, or the primary trait for the novel / replicated SNPs. For all SNPs we show the relationship between Minor Allele Frequency (MAF) on the x-axis and the effect size (mmHg) on the y-axis, where results are taken from the UKB+ICBP discovery meta-analysis. All meta-analysis results were computed using inverse variance fixed effects models. The different symbols and colours distinguish the “known” vs “novel-2stage” vs “novel-1stage” vs “replicated-Hoffman” SNPs, and show that in general, the novel SNPs have smaller effect sizes than known SNPs, and that there is no significant difference (P=0.447) between the effect sizes of the 1-stage (N=757,601) and 2-stage (N=1,006,863) novel SNPs. (UKB: UK Biobank; ICBP: International Consortium of Blood Pressure). |
|  |
| **Supplementary Figure 3** |
| Overview of functional annotation and prioritisation of genome-wide associated variants and genes. |
| SNPs: single nucleotide polymorphisms; LD: Linkage Disequilibrium; eQTL: expression Quantitative Trait Loci; UCSC: University of California Santa Cruz (UCSC) genome browser; IPA: Ingenuity Pathway Analysis (IPA) software (IPA®,QIAGEN Redwood City,www.qiagen.com/ingenuity); DEPICT: Data-driven Expression Prioritized Integration for Complex Traits; GREAT: Genomic Regions Enrichment of Annotations Tool. |
|  |
| **Supplementary Figure 4** |
| Enrichments of eQTLs. |
| 535 novel blood pressure associated SNPs and the SNPs in LD r2>0.8 were annotated for their effect on gene expression using the GTEx portal. The number of eGenes associated with BP SNPs in a given tissue/ cell type was normalised with the total number of eGenes in that tissue and z-score was calculated using the trimmed mean and standard deviation of the normalised scores.  Tissues of the same tissue group were coloured the same. |
|  |
| **Supplementary Figure 5** |
| DEPICT enrichment analysis. |
| DEPICT software was used to investigate enrichment of a range of biological properties, in each case we compared known sentinel SNPS (N=357) to all known and novel SNPs with P-value <1x10-12 (N=227). The gene set enrichment analysis algorithm is described in Pers et al66. Enrichment –log p value is reported for both groups, we also present delta –log p value as a measure of novelty introduced by novel associations reported. Enrichment categories are as follows. a) Enrichment of tissues and cell types. b) GO annotation. c) Protein-protein interaction subnetwork annotation. d) Mammalian phenotype annotation. |
|  |
| **Supplementary Figure 6** |
| FORGE Dnase I sensitive region enrichment in known sentinel SNPs, compared to known and novel sentinel SNP associations for blood pressure. |
| Sentinel SNPs were investigated for enrichment in ENCODE DNase I regulatory regions using FORGE. The background probability of overlap is determined from the 1000 background set overlap counts and the probability of the observed test result under a binomial distribution is calculated. The *P*-value thresholds of 0.05 and 0.01 are corrected for multiple testing by division by the number of tissue groupings tested, and the corrected threshold is used. Strongest enrichment in known SNPs was seen in vasculature (Human Aortic artery fibroblast (AoAF) and also Human Villous Mesenchymal Fibroblasts (HVMF) found in placenta). Enrichment in all known and novel SNPS was increased across vasculature (AoAF; HMVEC, Human microvascular endothelial cells) and highly vascularised tissues. Tissues in red are significant after correction for false discovery. |
|  |
| **Supplementary Figure 7** |
| Ingenuity pathway analysis of BP genes. |
| For genes mapped to 357 sentinel SNPs at 274 known loci and genes mapped to all 901 loci. Sentinel gene mapping is compared to genes identified by extended LD (r2>0.8). Pathway enrichment is represented as – log p value. A) Canonical pathway enrichment. B) Upstream regulator enrichment. C) Disease and Biofunction enrichment. |
|  |
| **Supplementary Figure 8** |
| Exploring known and novel drug mechanisms in blood pressure. |
| The figure summarises known and novel target opportunities highlighted by blood pressure genetics. Ingenuity pathway analysis was used to create a network of 6,562 genes showing direct interaction with 147 known blood pressure target genes. This network was compared with 1,738 genes that are either directly associated with BP or linked by LD (r2>0.8). Overlap between genetic associated genes and the BP drug interactome demonstrates genetic support for known drug mechanisms. Drugged or druggable genes showing genetic association with BP, but no interaction with the known BP drug interactome, represent potentially new mechanisms in blood pressure drug development and repositioning. Number of known and novel drugged/druggable gene associations are shown in parentheses. |

|  |
| --- |
|  |
| **Supplementary Figure 9** |
| Comparison of beta effect sizes between individuals of European (N=757,601), African (N=7,782) and South Asian (N=10,323) ancestry. |
| Scatterplots showing the direction of the standardized regression coefficient (beta) of novel (red) and known (grey) BP variants between Europeans and Africans (a,b,c) and South Asians (d,e,f), on the three studied BP phenotypes. |
|  |
| **Supplementary Figure 10** |
| Correlation and distribution of minor allele frequencies (MAF) of BP variants in individuals of European (N=757,601), African (N=7,782) and South Asian (N=10,323) ancestry. |
| Scatterplot showing the correlation and the distribution of MAF of novel (red) and known (grey) BP variants between a) Europeans and Africans and b) Europeans and South Asians. ρ is the Pearson correlation coefficient. |
|  |
| **Supplementary Figure 11** |
| Ethnicity clustering performed using PCA. |
| PC1 is plotted against PC2 for all N=486,683 UK Biobank participants post-QC, colour-coded according to the five ethnic clusters created from our K-means PCA clustering, from which only “White” Caucasians are selected for analysis of individuals of European ancestry. Plot (A) shows the clustering for all subjects, whereas plot (B) only shows the subsets of individuals selected for race-stratified analysis, after combining information together from both the PCA clustering and the self-reported ethnicity. PCA: Principal Component Analysis; QC: Quality Control; PCs: Principal Components. |
|  |
| **Supplementary Figure 12** |
| Quantile-Quantile plots. |
| QQ plots of results for (A) systolic blood pressure (SBP), (B) diastolic blood pressure (DBP), (C) pulse pressure (PP) from GWAS discovery (N=757,601). The black curves are based on all SNPs in the corresponding analysis, with Minor Allele Frequency ≥ 1%. The green curves are results after excluding SNPs within the 274 known loci (±500kb; Linkage Disequilibrium r2 ≥ 0.1). The *P*-values have been derived from inverse variance fixed effects meta-analysis. |