Association analyses of predicted loss-of-function variants prioritized 15 genes as blood pressure regulators

Estelle Lecluze, PhD¹,² and Guillaume Lettre, PhD¹,²

¹Montreal Heart Institute, Montreal, Québec, H1T 1C8, Canada
²Faculté de Médecine, Université de Montréal, Montreal, Québec, H3T 1J4, Canada

Short title: Blood pressure-associated loss-of-function variants

Correspondence:
Guillaume Lettre
Montreal Heart Institute
5000 Belanger Street
Montreal, Quebec, Canada
H1T 1C8
514-376-3330 ext. 2657
guillaume.lettre@umontreal.ca

Total word count: 8653 words
ABSTRACT

Background: Hypertension, clinically defined by elevated blood pressure (BP), is an important cause of mortality and morbidity worldwide. Many risk factors for hypertension are known, including a positive family history, which suggests that genetics contribute to inter-individual BP variation. Genome-wide association studies (GWAS) have identified >1000 loci associated with BP, yet the identity of the genes responsible for these associations remains largely unknown.

Methods: To pinpoint genes that causally impact BP variation in humans, we analyzed predicted loss-of-function (pLoF) variants in the UK Biobank whole-exome sequencing dataset (n=454,709 participants, 6% non-European ancestry). We analyzed genetic associations between systolic or diastolic BP (SBP/DBP) and single pLoF variants (additive and recessive genetic models) as well as with the burden of very rare pLoF variants (minor allele frequency [MAF] <0.01%).

Results: Single pLoF variants in ten genes associated with BP (ANKDD1B, ENPEP, PNCK, BTN3A2, C1orf145 [OBSCN-AS1], CASP9, DBH, KIAA1161 [MYORG], OR4X1, and TMC3). We also found a burden of rare pLoF variants in five additional genes associated with BP (TTN, NOS3, FES, SMAD6, COL21A1). Except for PNCK, which is located on the X-chromosome, these genes map near variants previously associated with BP by GWAS, validating the study of pLoF variants to prioritize causal genes at GWAS loci.

Conclusions: Our study highlights 15 genes that likely modulate BP in humans, including five genes that harbor pLoF variants associated with lower BP.

KEYWORDS
Blood pressure, hypertension, loss-of-function, genetic associations, PNCK, FES

ABBREVIATIONS
BP: blood pressure
DBP: diastolic blood pressure
GWAS: genome-wide association study
LD: linkage disequilibrium
MAC: minor allele count
MAF: minor allele frequency
mmHg: millimeters of mercury
pLoF: predicted loss-of-function
PRS: polygenic risk score
SBP: systolic blood pressure
WES: whole-exome DNA sequencing
INTRODUCTION

Hypertension, defined as SBP >130 mmHg and/or DBP >80 mmHg, affects over 1.28 billion people worldwide and is directly or indirectly responsible for ~13% of all annual deaths. These numbers are alarming, especially because recent predictions suggest that they will continue to rise in the coming years despite simple disease modifying interventions (e.g. diet, physical activity) and many widely available BP-lowering drugs. Regulating BP is clinically challenging because different patients often respond differently to the same treatments. In part, this is explained by the fact that hypertension results from the complex interplay of the cardiovascular, renal, endocrine and neural systems together, but also in response to variable environmental stimuli.

Genetics also contribute to the risk of developing hypertension: a positive family history of hypertension is a predictive risk factor, and both SBP and DBP are heritable traits ($h^2$ estimates range from 15-40% and 15-30%, respectively). Over the last 15 years, increasingly large GWAS have identified >1,000 genetic loci associated with BP variation. These genetic associations, enriched for non-coding regulatory common variants, often fail to pinpoint the causal DNA sequence variants and genes. However, this limitation of the GWAS approach is not fatal to all downstream analyses. For instance, the development of BP polygenic risk scores (PRS), which in their simplest forms are the weighted sums of BP-associated variants, is agnostic of the underlying biology. PRS can stratify individuals more at risk to develop high BP, and also associate with several co-morbidities such as stroke and myocardial infarction.

For applications such as drug targeting or repurposing, knowing the identity of the candidate causal BP genes at GWAS loci is however essential. One strategy is to painstakingly characterize candidate genes in cellular and animal models using gain- or loss-of-function experimental designs. The emergence of genome editing technologies (e.g. CRISPR/Cas9) has accelerated the pace of these experiments, yet they remain time-consuming. One complementary approach is to identify and characterize DNA sequence variants that are predicted to completely abrogate the functions of a gene. Such predicted loss-of-function (pLoF) variants are instrumental because they yield biological insights and can also help guide the development of new drugs. Indeed, a pLoF variant should mimic the effect of a drug developed to block the same gene, providing an opportunity to study side effects (e.g. pleiotropy) and therapeutic impact.

In this study, we performed genetic association testing between BP variation and pLoF variants identified by whole-exome sequencing (WES) in the UK Biobank. A combination of additive, recessive, and gene-based statistical models implicated 15 genes in the regulation of BP in humans.
METHODS

Description of the UK Biobank WES sets
The UK Biobank cohort includes 502,412 individuals from White, Black, Asian, Mixed and other ancestral backgrounds (Supplementary Table 1). To perform our analyses, we defined three sets of WES data: (i) the initial 200k WES release (N=200,602) was used as a discovery cohort, (ii) WES data from 254,196 individuals newly available in the 500k release were designated as the 300k set and used as a replication cohort, and (iii) the whole 500k set was also used to perform similar analyses. We summarize our study design in Figure 1. This project was approved by the Montreal Heart Institute Ethics Committee (protocol #2017-2247). This work was conducted using the UK Biobank resource (Project number 62518).

Phenotype normalization
The phenotype was normalized using the whole UK Biobank, including participants that were not included in the WES data. The mean SBP and DBP were calculated using two automated BP measurements (N_{SBP} = 472,371, N_{DBP} = 472,376) or, if unavailable, manual measurements (N_{SBP} = 28,793, N_{DBP} = 28,790). For individuals taking BP medication, 15 mmHg and 10 mmHg were added to their SBP and DBP, respectively. Phenotype normalization was performed separately for each ancestral group. To analyze the SBP and DBP data, a linear regression model was used to calculate residuals of the measure, using sex, age, body mass index, and center of measurement as covariates. A rank-based inverse normal transformation was then applied to these residuals. These inverse normal residuals were scaled to match the standard deviation of the original trait distribution. Therefore, throughout this manuscript, we report genetic effect sizes in standardized mmHg.

Genotyping array and WES quality control
We then applied quality control measures to the data using plink2, and excluded all unplaced variants, selecting variants with a minor allele count (MAC) >5, filtering out variants and samples with a missing call rate >0.1, and filtering out all variants that had a Hardy-Weinberg equilibrium exact test p-value <1e-15. The snpEff tool was used to identify pLoF variants in the UK Biobank WES dataset. While we tested all pLoF with a MAC >5 in the single variant association tests, we focused our downstream analyses on single pLoF variants with a MAF >0.01%. This effectively excluded two very rare variants in PYGL (14:50920952:TCA:T, MAF=3.32e-06) and ZNF804B (7:89327362:C:T, MAF=4.42e-06)(Supplementary Table 2).

Genetic association (statistical) analyses
Association analyses were performed using the REGENIE (V2.2.4) software. The first step of REGENIE involves fitting a whole-genome regression model to the SBP and DBP data on the complete genotype data. Normalized values of BP were used, and age, sex, and the first 10 principal components were included as covariates. The genotype block was set to 1000 base pairs. The model was used to produce a set of genomic predictions using the additive or recessive models.
The second step of REGENIE involves using WES data to test pLoF variants for association with SBP and DBP, conditional upon the prediction computed in the previous step. The single variant analysis used both an additive and a recessive test, including only pLoF variants with a MAC ≥ 5. Sex, age, and the first 10 principal components were used as covariates, and the genotype data was divided into blocks of 200 base pairs.

For the gene-based analyses, we only included pLoF variants with a MAF ≤ 0.01% and a MAC ≥ 5. We used a burden test to evaluate the aggregated effect of these variants, using the same covariates and genotype block size as in the single variant analyses. pLoF variants within the MHC region were excluded from the final results (hg38: chr6:28,510,120-33,480,577).

**Databases curation**

Each variant or gene of interest were investigated through the curation of GWAS (GWAS catalog\(^\text{11}\)), pheWAS (UK Biobank, FINGENN), and aggregated databases (OpenTarget\(^\text{12}\), VannoPortal\(^\text{13}\)). We used TOPLD resource\(^\text{14}\) to retrieve linkage disequilibrium (LD) proxies (\(r^2\geq 0.8\)) of our variants of interest in the European ancestral subset of TOPMed. We also queried single-cell RNA-sequencing data from the Tabula Sapiens Consortium\(^\text{15}\).
RESULTS

Single pLoF variants associated with BP in the UK Biobank

We re-analyzed the WES data generated in the UK Biobank (Figure 1)\textsuperscript{16}. This dataset includes 515,198 pLoF variants found in 454,709 participants, among whom 26,328 participants are of non-European ancestry (Supplementary Table 1). Because the WES data was generated by phase, we initiated our discovery effort in the initial 200k set (N=200,602) and replicated the significant results in the 300k set (N=254,196, independent from the 200k set) (Figure 1 and Supplementary Table 1). Under an additive genetic model (Supplementary Figures 1A and 2A, Supplementary Table 2), this framework identified three replicated pLoF variants in ANKDD1B, ENPEP, and PNCK (Table 1). We also tested associations under a recessive genetic model (Supplementary Figures 1B and 2B, Supplementary Table 2), reasoning that haplosufficient genes would only impact BP when both copies are inactivated by pLoF variants. Since most pLoF variants are rare, we could identify 959 pLoF variants with homozygous carriers, and only the same ANKDD1B pLoF variant reached statistical significance in the 200k set and replicated in the 300k set (Table 1). To maximize statistical power, we also carried out association testing for the additive and recessive genetic models in the full 500k UK Biobank WES dataset (N=454,709) (Figure 1, Supplementary Figure 3, and Supplementary Table 2). While promising, these results still need to be replicated in external cohorts. The 500k set analyses identified pLoF variants in 10 genes, including seven genes not found in the staged 200k+300k analyses: BTN3A2, C1orf145 (OBSCN-AS1), CASP9, DBH, KIAA1161 (MYORG), OR4X1, and TMC3 (Table 1).

The association between reduced BP and the PNCK splice donor pLoF variant is novel. Although this variant is common (rs5987128, MAF=12%), it might have escaped GWAS detection because it is located on the X-chromosome (which was often excluded from large GWAS consortia efforts). The pLoF variants in DBH and KIAA1161 were not identified in previous GWAS of BP, but other variants located at the same loci have been described (Tables 1-2). The seven remaining pLoF variants were reported before (Tables 1-2). We confirmed that the BP associations that we identified are driven by the White British participants from the UK Biobank (which represents 94% of the 500k set, Supplementary Figure 4) and that their effect sizes are consistent between women and men (Supplementary Figure 5).

Half of the pLoF alleles identified in the single-variant analyses are associated with lower BP (Figure 2). The strongest effect was observed for carriers of a rare (MAF=0.2%) splice donor pLoF variant in DBH, a gene that encodes a dopamine beta-hydroxylase. Rare Mendelian mutations in DBH caused orthostatic hypotension 1, and the splice donor pLoF variant identified here (rs74853476, c.339+2T>C) has been characterized as pathogenic in ClinVar\textsuperscript{17,18}. Individuals who carry one copy of this pLoF variant have a DBP that is 1.2 mmHg below the average (Figure 2). The four other pLoF variants that decreased BP, as well as the five pLoF that increased BP, are more common (MAF \(\geq1\%\)) (Figure 2). The BP-increasing pLoF variant with the strongest effect – 1.4 mmHg per allele – is a stop gained variant in ENPEP (MAF=1%), a gene that encodes an aminopeptidase that can upregulate BP by cleaving angiotensin II.
We queried publicly available databases (Methods) to determine if the pLoF variants identified in our analyses fell within loci previously associated with other human phenotypes. Except for the C1orf145 (OBSCN-AS1) pLoF variant, which is only associated with SBP and hypertension, the remaining variants (or the corresponding genes) are pleiotropic and associate with multiple human traits and diseases, including blood-cell and lipid indices (Table 2).

**Rare pLoF variants associated with BP variation by gene-based analyses**

To complement our single-variant association tests and increase power to identify rare pLoF variants associated with BP variation in the 500k set from the UK Biobank, we aggregated pLoF variants from the same gene and performed gene-level association tests (Figure 1, Methods and Supplementary Figure 6). We restricted these analyses to pLoF variants with MAF <0.1% for two reasons: First, including more common variants tended to recover the same variants (and genes) as the single-variant analyses described above. And second, by focusing on rare variants, we limited the possibility to capture association signals arising from residual LD with more common (and likely non-LoF) variants. Using this strategy, we found associations between increased BP traits and the burden of pLoF variants in five genes: NOS3, FES, COL21A1, SMAD6, and TTN (Figure 3 and Supplementary Table 2-3). Except for TTN, for which a burden of pLoF slightly decreases BP, pLoF in the remaining four genes are associated with increase BP (Figure 3). Leave-one-variant-out (LOVO) analyses confirmed that these gene-based associations are not due to a single pLoF variant (Supplementary Figure 7). NOS3 represents a strong validation of our strategy as it encodes endothelial nitric oxide (NO) synthase, an enzyme that produces the vasodilator nitric oxide (NO) in the vascular system that helps maintain BP homeostasis. GWAS had previously identified common non-coding variants near these genes, and our gene-based results now nominate NOS3, FES, COL21A1, SMAD6 and TTN as causal regulators of BP in humans (Table 2).

**Gene expression profiles at single-cell resolution implicate multiple cell-types**

To better understand how the 15 genes with pLoF variants (from the single-variant and gene-based tests) may regulate BP, we queried single-cell RNA-sequencing data from the Tabula Sapiens Consortium (Table 3)\textsuperscript{15}. We found specific expression for TTN and PNCK in cardiomyocytes, and for NOS3 and SMAD6 in endothelial cells, as expected (Supplementary Figure 8). The expression of ENPEP was largely restricted to kidney epithelial cells, although it was also detectable in smooth muscle cells and pericytes. BTN3A2 was expressed in T-cells from multiple tissues. The expression patterns of FES and COL21A1 were more diffused: FES was expressed in immune (monocytes, macrophages, neutrophils) and endothelial cells, whereas COL21A1 was found in smooth muscle cells, pericytes and fibroblasts. The expression of the remaining seven genes (ANKDD1B, CASP9, OR4X1, DBH, TMC3, C1orf145 [OBSCN-AS1], and KIAA1161 [MYORG]) was too low in the Tabula Sapiens dataset to be unambiguously assigned to specific cell-types.
DISCUSSION

Hypertension is a central risk factor for stroke, ischemic heart disease, renal dysfunction, dementia and other cardiovascular diseases. Using a combination of different statistical models and the large WES dataset from the UK Biobank, we highlighted 15 genes that likely impact the regulation of BP in humans. Indeed, the list is enriched with genes that encode proteins with well-defined role(s) in BP homeostasis: NOS3, DBH, and ENPEP. These positive controls genes are further discussed in the Supplementary Notes, along with CASP9 (renal functions), genes that have known roles in the cardiovascular system (SMAD6, COL21A1, TTN), and genes with more speculative functions in BP regulation and hypertension (BTN3A2, ANKDD1B, OR4X1, TMC3, C1orf145 [OBSCN-AS1], KIAA1161 [MYORG]). Below, we discuss two genes of particular interest: PCNK and FES.

PNCK encodes the pregnancy-upregulated non-ubiquitous calcium-calmodulin-dependent kinase (also known as calmodulin kinase 1b [CaMKIb]), a member of the protein serine/threonine kinase family. This gene, a paralog of the canonical CAMKI gene, is not well-characterized functionally. Its strongest expression is in the developing brain, but it remains expressed in numerous tissues, including the heart (Supplementary Figure 8) and the mammary gland during pregnancy19. PNCK is also differentially expressed in the placenta of pregnant women with pre-eclampsia, a condition characterized by hypertension10. In contrast to PNCK and its CAMKI family members, genes that encode members of the CAMKII family have well-known roles in regulating BP. In vascular smooth muscle cells, CaMKII responds to angiotensin II to increase BP21, whereas in endothelial cells, it can phosphorylate eNOS and promote the production of the vasodilator NO22. While not specifically developed for PNCK, CaMKI inhibitors have been tested in diabetic and obese mouse models. It would be interesting to know whether the same molecules can interfere with PNCK activity, and whether this affects BP23.

FES is a proto-oncogene that codes for a protein tyrosine kinase. It is expressed in multiple cell types, including myeloid and endothelial cells. GWAS have identified SNPs at the FES locus that associate with BP and coronary artery disease9,24. FURIN is located just upstream of FES and encodes a subtilisin-like proprotein convertase with pro-atherogenic functions25. Therefore, it represents a more obvious candidate causal gene to explain the GWAS signal at the locus. However, several lines of evidence now suggest that FES might also contribute to the GWAS signal for hypertension and CAD risk: (1) statistical and epigenomic fine-mapping has prioritized a SNP (rs12906125) in the promoter of FES as the likeliest causal variant at the locus26, (2) this variant is a stronger expression quantitative trait locus (eQTL) for FES than FURIN27, (3) CRISPR activation at this variant strongly upregulates FES expression and induces endothelial dysfunction28, and (4) specific base editing at rs12906125 impacts the expression of FES, but not FURIN, in endothelial cells treated with TNFα28. Our new analyses now show that rare pLoF in FES, which are independent from the GWAS signal, are associated with BP, further suggesting that FES also contributes to BP regulation and CAD risk. Importantly, our data do not rule out FURIN as an equally important local contributor to the GWAS signal. Indeed, while they do not reach statistical
significance after multiple testing correction, the 42 pLoF variants in *FURIN* found in the UK Biobank 500k set are nominally associated with BP (burden P-value=0.002).

Many phenotypes, including BP, have already been analyzed in the WES UK Biobank dataset\textsuperscript{16,29}. While there are small differences (e.g. focus on rare pLoF variants, adjustment for BP-lowering drugs, marginal differences in terms of association P-values due to the statistical tests used, no MAF threshold for gene-based tests ...), their results are largely consistent with ours. However, because our study is specifically focused on the genetics of BP, we bring into the spotlight 15 candidate BP genes, genes which were not specifically discussed by these other studies since they analyzed 1000s of other phenotypes. We acknowledge several limitations of our study. Most importantly, because 94% of the UK Biobank is of European ancestry, we were not well-powered to identify BP-associated pLoF variants in other ancestral groups. We were also limited in power to test the role of very rare alleles (e.g. singletons and doubletons), and expect that future studies in larger (and more diverse) datasets will make exciting new BP discoveries. Another area for future development is the inclusion of additional classes of genetic variants in pLoF variant association studies. For simplicity, we restricted our analyses to nonsense, frameshift and essential splice site variants, but we recognize that additional variants – such as a subset of missense variants – can also have null function. Including these variants in association tests will broaden the list of genes that we can test for a role in human phenotypic variation.

Motivated by work in animal models and Mendelian genetics, many studies have now considered the impact of pLoF variants on human diseases or traits\textsuperscript{30–34}. We extended this framework to the large UK Biobank WES dataset and identified 15 genes likely implicated in BP homeostasis. These genes are expressed in multiple tissues and cell-types, consistent with the network complexity that controls BP. Strikingly, only one of these genes – *PNCK* – is located at a locus that was not previously associated with BP by GWAS (likely because it is on the X-chromosome). This suggests that the true value of pLoF-based studies will not be in identifying new loci, but rather in prioritizing strong candidate genes within GWAS regions, thus streamlining downstream functional experiments. For BP and hypertension more specifically, our data highlight that the putative inactivation of five genes – *ANKDD1B, PNCK, CASP9, OR4X1*, and *DBH* – reduces BP, prompting further analyses to determine if the encoded proteins could be safe and efficacious BP-lowering drug targets.
Acknowledgements
We thank the participants from the UK Biobank who generously contributed their data to enable this research. This research was enabled in part by support provided by Calcul Quebec (https://www.calculquebec.ca/en/) and Compute Canada (www.computecanada.ca).

Funding
This work was funded by the Canadian Institutes of Health Research (MOP #136979), the Canada Research Chair Program, the Foundation Joseph C. Edwards and the Montreal Heart Institute Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosures
The authors declare that they have no competing interests.

Supplemental Material
Supplemental Text
Supplementary tables

Declarations
Ethics approval and consent to participate
This project was approved by the Montreal Heart Institute Ethics Committee (protocol #2017-2247). This research has been conducted using the UK Biobank resource under application number 62518.

Availability of data and materials

Author contributions
Conceptualization: Estelle Lecluze, Guillaume Lettre
Data curation: Estelle Lecluze, Guillaume Lettre
Formal analysis: Estelle Lecluze
Funding acquisition: Guillaume Lettre
Investigation: Estelle Lecluze
Project administration: Guillaume Lettre
Resources: Estelle Lecluze
Supervision: Guillaume Lettre
Validation: Estelle Lecluze
Visualization: Estelle Lecluze
Writing: Estelle Lecluze, Guillaume Lettre
References


53. Kurki MI, Karjalainen J, Palta P, et al. FinnGen: Unique genetic insights from combining isolated population and national health register data. Published online March 6, 2022;2022.03.03.22271360. doi:10.1101/2022.03.03.22271360


Table 1. Association results of predicted loss-of-function (pLoF) variants with systolic and diastolic blood pressure (SBP and DBP) (minor allele frequency (MAF) >0.01%). Variants that are significantly associated with a blood pressure trait in the 200k discovery set and replicated in 300k replication set are noted in the “200k (300k)” column. We marked significant P-values after multiple testing correction with a star (additive model : p-value <3.27e-07; recessive model: p-value <8.40e-06), and genome-wide significant variants with two stars (p-value <5e-08). Effect size (BETA) and standard errors are expressed in standardized mmHg. Genomic coordinates (CHR:POS:Allele0:Allele1) are on build GRCh38. The direction of the effect (BETA) is for Allele 1. Linkage disequilibrium (LD) metrics are from European-ancestry individuals in TOPMed.

<table>
<thead>
<tr>
<th>GENE</th>
<th>A1FREQ (500K)</th>
<th>PHENO</th>
<th>Additive (200k (300k))</th>
<th>Recessive (200k (300k))</th>
<th>Blood pressure associated SNPs in 500kb window (LD if r² ≥0.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>LOG1P BETA SE</td>
<td>LOG1P BETA SE</td>
<td></td>
</tr>
<tr>
<td>ANKDD1B</td>
<td>0.65</td>
<td>DBP</td>
<td>X</td>
<td>15.32** -0.18 0.02</td>
<td>X 15.32** -0.24 0.03</td>
</tr>
<tr>
<td>5:75669297:G:A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs34358</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Known pLoF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Known LD proxies : rs2307111*(0.73), rs10078021*(0.513), rs258494*(0.435)</td>
</tr>
<tr>
<td>PNCK</td>
<td>0.12</td>
<td>DBP</td>
<td>X</td>
<td>12.38** -0.19 0.03</td>
<td>9.25** -0.38 0.06</td>
</tr>
<tr>
<td>X:153674273:C:T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs5987128</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Novel pLoF and locus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CASP9</td>
<td>0.22</td>
<td>SBP</td>
<td>X</td>
<td>10.46** -0.30 0.05</td>
<td>6.57* -0.63 0.12</td>
</tr>
<tr>
<td>1:15493812:A:AC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs2234723</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Known pLoF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Known LD proxies : rs3820068*(0.767), rs7516462*(0.536), rs7887325*(0.639)</td>
</tr>
<tr>
<td>OR4X1</td>
<td>0.75</td>
<td>SBP</td>
<td>X</td>
<td>8.28** -0.25 0.04</td>
<td>7.02** -0.28 0.05</td>
</tr>
<tr>
<td>11:48264679:T:A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs10838851</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Novel pLoF</td>
</tr>
<tr>
<td>DBH</td>
<td>0.002</td>
<td>DBP</td>
<td>X</td>
<td>6.79* -1.20 0.23</td>
<td>- - -</td>
</tr>
<tr>
<td>9:133636712:T:C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs74853476</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Known pLoF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Known variants : rs3025380*, rs6271*,* rs10993958*</td>
</tr>
<tr>
<td>ENPEP</td>
<td>0.01</td>
<td>DBP</td>
<td>X</td>
<td>16.25** 0.81 0.10</td>
<td>2.02 2.94 1.14</td>
</tr>
<tr>
<td>4:110510288:G:A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs33966350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Known pLoF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Known LD proxy : rs34297584*(0.332). Known variants : rs6533515*,* rs10033071*, rs6828511*, rs4358460*, rs33966350*</td>
</tr>
<tr>
<td>BTN3A2</td>
<td>0.04</td>
<td>DBP</td>
<td>X</td>
<td>12.05** 0.39 0.05</td>
<td>2.70 1.19 0.38</td>
</tr>
<tr>
<td>6.26370605:T:G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs58367598</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Known pLoF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMC3</td>
<td>0.02</td>
<td>DBP</td>
<td>X</td>
<td>8.26** 0.42 0.07</td>
<td>0.57 0.74 0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The copyright holder for this preprint is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.
<table>
<thead>
<tr>
<th>Chromosome:Position:SNP</th>
<th>Gene</th>
<th>Stop Codon</th>
<th>pLoF Score</th>
<th>SBP Value</th>
<th>p-value</th>
<th>DBP Value</th>
<th>p-value</th>
<th>Known pLoF variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:81332588:G&gt;T</td>
<td>C1orf145 (OBSCN-AS1)</td>
<td>Stop gained</td>
<td>0.24</td>
<td>6.88*</td>
<td>0.04</td>
<td>1.50</td>
<td>0.24</td>
<td>rs150843673</td>
</tr>
<tr>
<td>rs11800309</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Known pLoF variants : rs2760061, rs649418</td>
</tr>
<tr>
<td>1:228204107:G&gt;T</td>
<td>KIAA1161 (MYORG)</td>
<td>Stop gained</td>
<td>0.21</td>
<td>6.73*</td>
<td>0.03</td>
<td>2.13</td>
<td>0.19</td>
<td>rs4879782</td>
</tr>
<tr>
<td>rs4879782</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Novel pLoF variants: rs4553000, rs3808869</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Known LD proxy: 0.274</td>
</tr>
</tbody>
</table>

Note: CC-BY-NC-ND 4.0 International license

It is made available under a license to display the preprint in perpetuity.
Table 2. Pleiotropic associations of blood pressure (BP)-associated predicted loss-of-function (pLoF) variants and corresponding genes. This table includes the most significant pleiotropic associations as identified in OpenTargets but is not exhaustive. To populate this table, we queried the Open Targets databases for the pLoF variant or its associated gene and included the phenotypes that were associated (pvalues ≤ 10e-8) in the “Previously known pLoF variant associations” and “Previously known gene-level associations” columns, respectively. Positive and negative reported effect are represented by ▲ and ▼ respectively.

BMI: body mass index; DBP: diastolic blood pressure; HDL: high density lipoprotein; HT: hypertension; LDL: low density lipoprotein; MAP: mean arterial pressure; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; MCV: mean corpuscular volume; PP: pulse pressure; RDW: red blood cell distribution width; SBP: systolic blood pressure; T2D: type 2 diabetes; WBC: white blood cell count.

<table>
<thead>
<tr>
<th>Gene (variant)</th>
<th>Animations previously known LoF variant associations</th>
<th>Gene functions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ANKD1B</strong></td>
<td>rs34358 ▼</td>
<td></td>
</tr>
<tr>
<td>DBP ▲ ▼</td>
<td>LDL cholesterol levels[5,40,42], total cholesterol levels[5,41,42], apolipoprotein B levels[59], BMI[44-46], weight[7,11], platelet count[44,46]...</td>
<td>LDL cholesterol levels[5,40,42,47], apolipoprotein B levels[59,40,41], BMI[47,49,50], weight, total cholesterol level[59,51,52], platelet count[44,45,46]... Mediation of protein-protein interaction, regulation of cell adhesion and migration.</td>
</tr>
<tr>
<td><strong>PNCK</strong></td>
<td>rs5987128 ▼</td>
<td></td>
</tr>
<tr>
<td>DBP ▲ ▼</td>
<td>Cystatin C levels[79], Creatinine levels[79]</td>
<td>T2D ▲ ▼, soft tissue disorders, RDW ▲ ▼ Member in calmodulin kinase I complex.</td>
</tr>
<tr>
<td><strong>CASP9</strong></td>
<td>rs2234723 ▼</td>
<td></td>
</tr>
<tr>
<td>DBP ▲ ▼</td>
<td>Heel bone mineral density t score[48,49], sex hormone-binding globulin levels[49]</td>
<td>Intrinsic apoptosis, autophagy regulation.</td>
</tr>
<tr>
<td><strong>OR4X1</strong></td>
<td>rs10838851 ▼</td>
<td></td>
</tr>
<tr>
<td>DBP ▲ ▼</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>DBH</strong></td>
<td>rs74853476 ▼</td>
<td>-</td>
</tr>
<tr>
<td>DBP ▲ ▼</td>
<td>Factor VIII levels[73,76], vWF levels[73,75], serum alkaline phosphatase levels[39,49,51,52], E-selectin levels[73,74,76,78-81], height[79,82-86]...</td>
<td>Dopamine to norepinephrine conversion.</td>
</tr>
<tr>
<td><strong>ENREP</strong></td>
<td>rs33966350 ▲</td>
<td></td>
</tr>
<tr>
<td>DBP ▲ ▼</td>
<td>Atrial fibrillation[55,56], arrhythmia[55,67], platelet</td>
<td>Angiotensin II degradation.</td>
</tr>
<tr>
<td>Gene</td>
<td>Sample ID</td>
<td>pLoF burden</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>BTN3A2 rs58367598</td>
<td>HT15</td>
<td>Blood cell traits (MCHC, MCH, RDW, MCV, ...), ischemic stroke</td>
</tr>
<tr>
<td>TMCS rs150843673</td>
<td>DBP8,19</td>
<td>Pro-interleukin 16 levels, height, lymphocytes count, MCHC</td>
</tr>
<tr>
<td>C1orf145 [OBSCN-AS1] rs11800309</td>
<td>HT15, SBP58</td>
<td>Putative mechanosensor</td>
</tr>
<tr>
<td>KIAA1161 (MYORG) rs4879782</td>
<td>-</td>
<td>Immature reticulocyte fraction, MAP, PP</td>
</tr>
<tr>
<td>NOS3</td>
<td>-</td>
<td>Serum levels of protein AOC1, appendicular lean mass, height, QT interval</td>
</tr>
<tr>
<td>FES</td>
<td>-</td>
<td>Encodes tyrosine kinase, control cell growth, differentiation, and adhesion</td>
</tr>
<tr>
<td>COL21A1</td>
<td>-</td>
<td>Heel bone mineral density, serum levels of protein GFRAL, cystatin C levels, lung function (FEV1/FVC)</td>
</tr>
<tr>
<td>TTN</td>
<td>-</td>
<td>Component of the sarcomere, essential to muscle contraction, maintenance and development, heart function.</td>
</tr>
<tr>
<td>SMAD6</td>
<td>-</td>
<td>Negative regulation of BMP and TGF/β signalling pathway: regulation of cell proliferation and differentiation, regulation of inflammation, cancer development, maintenance of vascular homeostasis.</td>
</tr>
</tbody>
</table>
Table 3. Expression of genes that carry predicted loss-of-function (pLoF) variants associated with blood pressure (BP) in the Tabula Sapiens Consortium single-cell RNA-sequencing dataset. We analyzed the whole dataset (n=483,152 cells, 45 tissues) using cell-type annotations provided by the Consortium. Seven genes could not be unambiguously assigned to specific cell-types: ANKDD1B, CASP9, OR4X1, DBH, TMC3, C1orf145 (OBSCN-AS1), and KIAA1161 (MYORG). See also Supplementary Figure 8 for details.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Cell-type(s) (Tabula Sapiens)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTN</td>
<td>Cardiomyocytes</td>
</tr>
<tr>
<td>PNCK</td>
<td>Cardiomyocytes</td>
</tr>
<tr>
<td>NOS3</td>
<td>Endothelial cells</td>
</tr>
<tr>
<td>SMAD6</td>
<td>Endothelial cells (mostly capillary and arterial)</td>
</tr>
<tr>
<td>ENPEP</td>
<td>Mostly kidney epithelial cells. Weaker signal in smooth muscle cells and pericytes.</td>
</tr>
<tr>
<td>COL21A1</td>
<td>Smooth muscle cells, pericytes, fibroblasts.</td>
</tr>
<tr>
<td>FES</td>
<td>Immune cells (monocytes, macrophages, neutrophils), endothelial cells.</td>
</tr>
<tr>
<td>BTN3A2</td>
<td>T-cells</td>
</tr>
</tbody>
</table>
Figures

**Figure 1.** Design of the study to identify predicted loss-of-function (pLoF) variants associated with blood pressure (BP) variation in the UK Biobank. Initially, we performed a discovery analysis in the 200k set followed by a replication analysis in the 300k set. We also analyzed BP associations in the whole 500k set. For each analysis, we report the Bonferroni-corrected statistical threshold. We performed gene-based association tests only in the 500k set. n.d., not done.
Figure 2. Predicted loss-of-function (pLoF) variants associated with blood pressure (BP) variation in single-variant analyses (minor allele frequency (MAF)>0.01%). For each variant, we report the gene, the coordinates on build GRCh38 of the human genome and the corresponding two alleles. Effect sizes and standard errors (SE) are in standardized mmHg. Stars in the panel on the right indicate in which UK Biobank cohort the variants are significantly associated with BP.
**Figure 3.** Gene-based association results found five genes that carry rare (minor allele frequency (MAF)≤0.01%) predicted loss-of-function (pLoF) variants associated with blood pressure (BP) variation. On the x-axis, we report the effect size (in standardized mmHg) for the gene-based results as well as for each of the pLoF variant taken individually (additive model). The horizontal dashed line indicates the threshold for gene-based statistical significance (p-val≤5.21e-05).
Supplementary Figure 1. Quantile-quantile (Q-Q) plots for predicted loss-of-function (pLoF) variant association results with diastolic blood pressure (DBP) and systolic blood pressure (SBP) in the 200k (discovery) set (N_{sample}=199,558) using (A) an additive model (N_{variant}=59,418) and (B) a recessive model (N_{variant}=959). In each plot, the dashed line represents the statistical significance threshold (additive model p-val≤8.41e-07; recessive model p-val≤5.21e-05) and the grey area is the 95% confidence interval.
Supplementary Figure 2. Quantile-quantile (Q-Q) plots for predicted loss-of-function (pLoF) variant association results with diastolic blood pressure (DBP) and systolic blood pressure (SBP) in the 300k (replication) set ($N_{\text{sample}}=252,898$) using (A) an additive model ($N_{\text{variant}}=101,030$) and (B) a recessive model ($N_{\text{variant}}=3,944$). In each plot, the dashed line represents the statistical significance threshold (additive model p-val$\leq 4.95\text{e-07}$; recessive model p-val$\leq 1.27\text{e-05}$) and the grey area is the 95% confidence interval.
Supplementary Figure 3. Quantile-quantile (Q-Q) plots for predicted loss-of-function (pLoF) variant association results with diastolic blood pressure (DBP) and systolic blood pressure (SBP) in the complete 500k set ($N_{\text{samples}}=452,385$) using (A) an additive model ($N_{\text{variants}}=152,840$) and (B) a recessive model ($N_{\text{variants}}=5,955$). In each plot, the dashed line represents the statistical significance threshold (additive model p-val$\leq 3.27e-07$; recessive model p-val$\leq 8.40e-05$) and the grey area is the 95% confidence interval.
Supplementary Figure 4. Comparison of single-variant association P-values obtained in the multi-ancestry 500k set (x-axis) and in the White British participants subset of the 500k set \(N_{\text{sample}}=428,381\). (A) Additive and (B) recessive genetic models.
Supplementary Figure 5. Comparison of effect sizes (in standardized mmHg) between men (x-axis) and women (y-axis) in the UK Biobank for each predicted loss-of-function (pLoF) variants associated with blood pressure. The differences are non-significant.
Supplementary Figure 6. Quantile-quantile (Q-Q) plots for gene-based results with diastolic blood pressure (DBP) and systolic blood pressure (SBP) in the complete 500k set (N\textsubscript{sample}=452,385, N\textsubscript{gene}=18,436). We only tested predicted loss-of-function (pLoF) variants with minor allele frequencies ≤0.1%. The grey area is the 95% confidence interval.
Supplementary Figure 8. Uniform Manifold Approximation and Projection (UMAP) for 483,152 cells from 45 human tissues analyzed by the Tabula Sapiens Consortium (https://cellxgene.cziscience.com/). For each of the eight blood pressure (BP) genes with distinguishable expression profiles, we highlighted the main cell-type(s) based on annotations provided by Tabula Sapiens. TTN and PNCK, cardiomyocytes; ENPEP, kidney epithelial cells (circle) and smooth muscles/pericytes (rectangle); NOS3 and SMAD6, endothelial cells; COL21A1, smooth muscle cells, pericytes, and fibroblasts; FES, myeloid cells (oval) and endothelial cells (rectangle); BTN3A2, T-cells.