

ONLINE SUPPLEMENT

Transcriptome-wide analysis identifies novel associations with blood pressure

Running title: Blood pressure related gene expression

Zeller T^{1,2*}, Schurmann C^{2,3*#}, Schramm K^{4*}, Müller C^{1,2,5*}, Kwon S^{6*}, Wild PS^{2,7,8}, Teumer A^{2,9}, Herrington D¹⁰, Schillert A^{2,5}, Iacoviello L¹¹, Kratzer A^{2,12}, Jagodzinski A^{1,2}, Karakas M^{1,2}, Ding J¹⁰, Neumann JT^{1,2}, Kuulasmaa K¹³, Gieger C^{14,15}, Kacprowski T^{2,3}, Schnabel RB^{1,2}, Roden M^{16,17,18}, Wahl S^{14,15,18}, Rotter JI⁶, Ojeda F¹, Carstensen-Kirberg M^{16,18}, Tregouet DA¹⁹, Dörr M^{2,20}, Meitinger T^{2,4,21}, Lackner KJ^{2,22}, Wolf P^{5,21}, Felix SB^{2,20}, Landmesser U^{2,12}, Costanzo S¹¹, Ziegler A^{2,5}, Liu Y²³, Völker U^{2,3,24}, Palmas W²⁵, Prokisch H^{4,21}, Guo X^{6*}, Herder C^{16,18*}, Blankenberg S^{1,2*}, Homuth G^{*3,24}

1. Department of General and Interventional Cardiology, University Heart Center Hamburg, Germany
2. DZHK (German Centre for Cardiovascular Research), Germany
3. Interfaculty Institute for Genetics and Functional Genomics, University Greifswald, Germany
4. Institute of Human Genetics, Helmholtz Zentrum München, Germany
5. Institut für Medizinische Biometrie und Statistik, Universitätsklinikum Schleswig-Holstein, Germany
6. Institute for Translational Genomics and Population Sciences, UCLA Medical Center, USA.
7. Preventive Cardiology and Preventive Medicine, University Medical Center of the Johannes Gutenberg-University Mainz, Germany
8. Center for Thrombosis and Hemostasis, Mainz, Germany
9. Institute for Community Medicine, University Medicine Greifswald, Germany
10. Department of Internal Medicine, Wake Forest School of Medicine, USA
11. IRCCS Istituto Neurologico Mediterraneo Neuromed, Department of Epidemiology and Prevention, Italy
12. Department of Cardiology and Pneumology, Charité, Universitätsmedizin Berlin, Germany
13. National Institute for Health and Welfare, Helsinki, Finland
14. Molecular Epidemiology, Helmholtz Zentrum München, Germany
15. Institute of Epidemiology II, Helmholtz Zentrum München, Germany
16. Institute for Clinical Diabetology, German Diabetes Center, Germany
17. Department of Endocrinology and Diabetology, Medical Faculty Düsseldorf, Germany
18. German Center for Diabetes Research (DZD), Germany
19. Sorbonne Universités, UPMC, INSERM, UMR_S 1166, ICAN Institute for Cardiometabolism and Nutrition, Paris, France.
20. Department of Internal Medicine B, University Medicine Greifswald, Germany
21. Institute of Human Genetics, Technische Universität München, Germany

- 22. Institute of Clinical Chemistry and Laboratory Medicine, University Medical Center of the Johannes Gutenberg-University Mainz, Germany
- 23. Department of Epidemiology & Prevention, Wake Forest School of Medicine, North Carolina, USA
- 24. ZIK_FunGene, Universität Greifswald, Germany
- 25. Columbia University Medical Center, New York, USA

* equal contribution

Supplementary Material

1. Description of study cohorts
2. Definition and measurement of subclinical phenotypes and biomarkers
3. Preparation and quality control of RNA
4. Genotyping and imputation in the Gutenberg Health Study
5. Measurement of Human Cystein Rich Protein 1 (CRIP1) protein levels in the Moli-Sani Study
6. Statistical analyses

Supplementary Tables

Table S1: Study characteristics

Table S2: Gene-specific transcripts associated with blood pressure traits in the discovery analysis (Excel file).

Tables S3: Gene-specific transcripts associated with blood pressure traits in the validation analysis (Excel file).

Tables S4: Gene-specific transcripts associated with blood pressure traits in the discovery and validation analysis after exclusion of individuals receiving antihypertensive treatment (Excel file).

Tables S5: Gene-specific transcripts associated with blood pressure traits in the discovery analysis including only subjects of Caucasian ethnicity in MESA (Excel file)

Table S6: Explained variance (%) in BP levels attributable to the expression of genes in the participating studies.

Table S7: Association of monocyte candidate gene expression with blood pressure-related, subclinical cardiovascular phenotypes in the Gutenberg Health Study.

Tables S8: Single nucleotide polymorphisms from the GRASP database (A) and published BP related variants (B) used for analyses of expression quantitative trait loci (Excel file).

Table S9: Blood pressure-related eQTLs and SNPs related to blood pressure

Table S10: Baseline characteristics of subjects selected for CRIP1 measurements from the Moli-Sani Study.

Supplementary Figures

Figure S1: Study workflow

Figure S2: Candidate gene expression by study and cell types.

Figure S3: Associations between blood pressure changes and changes in mRNA levels of candidate genes after 5-years follow-up in the Gutenberg Health Study

Figure S4: Reduction of blood pressure between baseline and 6 month follow up in a clinical trial.

Figure S5:

A) Single nucleotide polymorphisms from the GRASP database used for analyses of expression quantitative trait loci. B). Single nucleotide polymorphisms from the GWAS catalogue used for analyses of expression quantitative trait loci.

Figure S6: Median candidate gene expression in various tissues measured by RNA-sequencing in the GTEx project

Supplementary Material

1. Description of study cohorts

Ethics statement

The study followed the recommendations of the Declaration of Helsinki. The study protocols of all studies were approved by the respective ethics committees. Written informed consent was obtained from all study participants.

GHS

The Gutenberg Health Study (GHS) is designed as a community-based, prospective, observational, single-center cohort study in the Rhine-Main area of Western Germany.¹ The sample was drawn randomly from the governmental local registry offices in the city of Mainz and the district of Mainz-Bingen. The sample was stratified 1:1 for sex and residence (urban and rural) and in equal strata for decades of age. Individuals between 35 and 74 years of age were enrolled. Exclusion criteria were insufficient knowledge of the German language and physical or psychological inability to participate in the examinations at the study center. Baseline examination of 15,010 study participants was performed between 2007 and 2012. Starting in April 2012, a second clinical follow-up visit was started after 5 years. At baseline and after 5 years, detailed medical, biochemical, and molecular information and biomaterial have been acquired.

MESA

The Multi-Ethnic Study of Atherosclerosis (MESA) was designed to investigate the characteristics of subclinical cardiovascular disease (disease detected non-invasively before it has produced clinical signs and symptoms) and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease. The cohort is a diverse, population-based sample of 6,814 asymptomatic men and women aged 45-84. Approximately 38 percent of the recruited participants are white, 28 percent African American, 22 percent Hispanic, and 12 percent Asian (predominantly of Chinese descent). Participants were recruited during 2000-2002 from 6 field centers across the U.S. (at Wake Forest University; Columbia University; Johns Hopkins University; the University of Minnesota; Northwestern University, and the University of California – Los Angeles). Since its inception in 2000, five clinic visits (exams) collected extensive clinical, socio-demographic, lifestyle and behavior, laboratory, nutrition, and medication data.² The present analysis is primarily based on analyses of purified monocyte samples from the April 2010-February 2012 examination (exam 5) of 1,264 randomly selected MESA participants (55-94 years old, Caucasian (47%), African American (21%) and Hispanic (32%), female 51%) from four MESA field centers (Baltimore, MD; Forsyth County, NC; New York, NY; and St. Paul, MN).

KORA

KORA (Cooperative Health Research in the Augsburg Region) exists since 1996 in the region of Augsburg in the southwest of Germany, and builds on the MONICA (Monitoring of trends and determinants in cardiovascular disease) project initiated in 1984³. KORA is a regional research platform for population-based surveys and a cohort of more than 18,000 subjects are actively followed up to date. The KORA F4 survey (2006-2008) was

the 7-year follow-up survey of KORA S4 (1999-2001), which included a population-based sample aged 25 to 74 years from the city of Augsburg and two adjacent counties.

SHIP-TREND

The Study of Health in Pomerania (SHIP) is a longitudinal population-based cohort in West Pomerania, a region in the northeast of Germany. The study assesses the prevalence and incidence of common population-relevant diseases and their risk factors. Baseline examinations for SHIP-TREND were carried out between 2008 and 2012, comprising 4,420 participants. Study design and sampling methods were previously described ⁴. The present project is based on a subset of 997 individuals aged 20 to 81 years of the SHIP-TREND study population.

Blood pressure lowering clinical trial

The influence of routinely used BP lowering medication over 6 month on gene expression was assessed in a clinical trial (EudraCT No.: 2009-017010-68). The study used the combination of telmisartan and amlodipine versus olmesartan and hydrochlorothiazide in hypertensive patients. Participants were selected at an age of 35 years or older, and being treated, but with uncontrolled hypertension (defined as 20/10 mmHg above target blood pressure of <140/90 mmHg [$<130/80$ mmHg for renal impaired and/ or diabetic patients]) or controlled hypertension and ≥ 3 cardiovascular risk factors and/or metabolic syndrome and/or diabetes mellitus and/or end organ damage. Exclusion criteria included pretreatment with amlodipine, Diuretics and AT1Blocker/ACEInhibitor within the last 3 months; pretreatment with telmisartan within the last 3 months, myocardial infarction within the last 6 months, previous stroke or hemodynamically relevant stenosis of carotid artery and cardiac or peripheral bypass surgery within the last 6 month. Laboratory and clinical phenotyping includes various markers and collection of biomaterial available in all 600 individuals at 2 time points (baseline and 6 month follow up).

The Moli Sani Study

The cohort of the Moli-Sani Study was recruited in the Molise region, Italy, from city hall registries by a multistage sampling. First, townships were sampled in major areas by cluster sampling; then, within each township, participants aged 35 years or over were selected by simple random sampling. Exclusion criteria were pregnancy at the time of recruitment, lack of understanding, current multiple trauma or coma, or refusal to sign the informed consent. A total of 24,325 men (47%) and women (53%) over the age of 35 were examined at baseline from 2005 to 2010. Participation was 70%. The cohort was followed-up for a median of 4.2 years (maximum 6.5 years) at December 2011 and will be followed-up every 5 years ⁵. (<http://www.moli-sani.org>). To determine CRIP1 serum levels by ELISA, incident cases of cardiovascular endpoints (stroke, heart failure, coronary heart disease) as well as a random sample of the cohort were selected. In total, 107 cases of incident coronary heart disease, 50 cases of incident stroke, 139 cases of incident heart failure and a random sample of 133 subjects were selected. The baseline characteristics of the selected subjects is given in Table S 10.

2. Definition and measurement of clinical phenotypes and biomarkers

Definitions of SBP and DBP were standardized between all studies. PP was calculated as the difference between the systolic and diastolic pressure readings. Hypertension was defined as SBP \geq 140 mmHg or diastolic BP \geq 90 mmHg at rest obtained as the mean of the second and third measurement, or by taking any antihypertensive drugs within the last 2 weeks. BP was measured with an Omron HEM-705CP device in GHS, SHIP-TREND and the clinical trial, using a Hawksley random-zero sphygmomanometer in KORA and a Dinamap PRO 100 automated oscillometric device in MESA.

Hyperlipidemia was defined as LDL/HDL ratio $>$ 3.5. Echocardiography was performed in every individual using a standardized protocol according to current American and European guidelines and as described in ¹. An iE33 echocardiography system with an S5-1 sector array transducer was used (Philips Electronics, Amsterdam, The Netherlands). Trained and certified medical technical assistants performed the examination. The following linear echocardiographic variables were studied in the present investigation: interventricular end-diastolic septum diameter (IVSD), LV internal end-diastolic diameter (LVIDD), LV end-diastolic posterior wall diameter (LVPWD), and LV end-systolic diameter (LVESD), LV hypertrophy (LVH). Derived from these variables, RWT as (IVSD_LVPWD)/LVEDD, and LV mass (LVM) according to the American Society of Echocardiography (ASE). The estimated glomerular filtration rate (eGFR) was calculated by the CKD-EPI formula. ⁶.

NT-proBNP levels were measured using a commercially available assay on the ELECSYS 2010 using the Elecsys proBNP II assay (ECLIA, Roche Diagnostics, Mannheim, Germany) with a LoD of 5 ng/L. The assay range of the assay was 5–35,000 ng/L. The inter- and intra-coefficient of variation were 5.4% and 2.3%, respectively. hsCRP levels were measured with the routine laboratory using an Abbott Architect c8000 system and the CRP Vario immunoassay. The inter- and intra-coefficient of variation were 5.6% and 3.7%, respectively. hsTnI levels were measured using a high-sensitivity cardiac troponin assay (ARCHITECT STAT highly sensitive Troponin I immunoassay, Abbott Diagnostics, USA, ARCHITECT i2000SR). The limit of detection for the assay was 1.9 ng/L (range 0-50000 ng/L). The assay had a ten percent coefficient of variation at a concentration of 5.2 ng/L. The inter- and intra-coefficient of variation were 4.5% and 5.8%, respectively.

3. Preparation and quality control of RNA

Preparation of RNA samples in GHS, KORA F4, and SHIP-TREND has been described previously. ⁷. Briefly, total RNA was isolated from whole blood in KORA and SHIP-TREND, and from monocytes in GHS. For all three studies, purity and concentration of RNA were determined by NanoDrop measurement and RNA quality was measured on a 2100 Bioanalyzer and the RNA 6000 Nano Lab Chips (Agilent Technology, Inc., Santa Clara, CA) by the RNA integrity number (RIN). Only samples with a RIN above 7 were used for gene expression profiling. In MESA, monocytic RNA was isolated using the AllPrep DNA/RNA Mini Kit (Qiagen, Inc., Hilden, Germany). RNA quality metrics included optical density measurements, using a NanoDrop spectrophotometer and evaluation of the RIN using the Agilent 2100 Bioanalyzer with RNA 6000 Nano Lab chips (Agilent Technology, Inc., Santa Clara, CA) following manufacturer's instructions. RNA with a RIN above 9 was used for gene expression profiling.

In the blood pressure lowering clinical trial, total RNA was isolated from peripheral blood mononuclear cells (PBMCs) using Trizol/Chloroform extraction.

Separation of PBMCs was conducted within 20 min after blood collection. Briefly, 8 mL blood was collected using the Vacutainer CPT Cell Preparation Tube System (BD, Heidelberg, Germany) and centrifuged. After separation, cells were washed and were resuspended in 1.5 mL Trizol Reagent (Invitrogen, Karlsruhe, Germany). 300 µL chloroform was added and phases were separated by centrifugation. For precipitation of RNA isopropanol was added. After a final washing step, total RNA was eluted in 40 µL RNase-free water. RNA quality metrics included optical density measurements, using a NanoDrop spectrophotometer. Integrity of total RNA was evaluated on an Agilent Bioanalyzer 2100 (Agilent Technologies, Boeblingen, Germany). Only samples with a RIN above 7 were used for subsequent gene expression analysis by semi-quantitative real-time PCR.

Semi-quantitative Polymerase Chain Reaction (PCR)

For semi-quantitative PCR, total RNA (100ng) was reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit including RNase Inhibitor (Life Technologies, Darmstadt). The resulting cDNA was diluted 1:20 with RNase-/DNase-free water and 2.5 µL were used for gene expression analysis by real-time PCR amplification. The amplification was carried out on a 7900HT Fast Real-Time PCR System (Life Technologies, Darmstadt) using the TaqMan Gene Expression Master Mix (Life Technologies, Darmstadt) and TaqMan Gene Expression assays for the following genes: CEBPA (Hs00269972_s1), CRIP1 (Hs00906229_g1), F12 (Hs01557543_g1), LMNA (Hs00153462_m1), MYADM (Hs01880197_s1), TIPARP (Hs00296054_m1), TPPP3 (Hs00372228_g1), TSC22D3 (Hs00929365_m1). For normalization, GAPDH endogenous control (Life Technologies, Darmstadt) were used.

4. Genotyping and imputation in the Gutenberg Health Study

Genomic DNA was extracted from buffy coats prepared from EDTA blood samples. Genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 as described by the Affymetrix user manual. Genotypes were called using the Affymetrix Birdseed v2 calling algorithm, and quality control was performed using GenABEL (<http://mga.bionet.nsc.ru/nlru/GenABEL>). Individuals with a call rate <97% or an autosomal heterozygosity >3 SDs around the mean were excluded. Standard quality criteria were applied to exclude SNPs (MAF<1%, genotype call rate <98%, and P value of deviation from HWE <10⁻⁴). Imputation based on 1,000 genomes Phase I Integrated Release Version 2 (NCBI Build 37) was performed using MACH version 1.0.18.c. In total, 1,133 individuals with genotype and monocytic gene expression data were available for expression quantitative trait loci (eQTL) analysis.

5. Measurement of Human Cystein Rich Protein 1 (CRIP1) protein levels in the Moli-Sani Study

To investigate whether the protein encoded by CRIP1 may be utilized as a circulating protein biomarker for clinically relevant endpoints, serum CRIP1 levels were determined in a population setting of 379 individuals from the cohort of the Moli-Sani Study⁸. This subsample was selected from the overall Moli-Sani cohort using a case cohort sampling design including incident cardiovascular cases of stroke (n=50), heart failure (n= 139) and coronary heart disease (CHD, a composite endpoint of myocardial infarction and coronary death, n=107) and an age-weighted random samples of the cohort (n=133).

Subjects with multiple events were included in each of the respective groups, resulting in a total overlap between groups of 50 subjects. CRIP1 protein levels were measured using an enzyme-linked immunosorbent assay (EKU03572, Biomatik, Wilmington, Delaware, USA). The lower limit of detection was 0.057 ng/mL. The inter-assay coefficient of variation (CV) was 15.5%; the intra-assay CV was 11.9%.

6. Statistical Methods

Proportion of BP variance attributable to candidate transcript expression

The proportion of SBP, DBP and PP explained by mRNA expression was calculated separately for each candidate transcript. Therefore, a generalized R^2 was computed based on the likelihoods from linear regressions, i) containing all covariates plus the candidate transcript expression and ii) with the covariates only as independent variables, as proposed by Magee⁹. Confidence intervals were estimated in 5,000 bootstrap iterations. The aggregated variance was calculated analogously, but with all candidate transcripts as independent variables.

Analysis of differential gene expression in the clinical trial

Expression values of candidate transcripts measured by qPCR were normalized for GAPDH Ct values prior to association analysis and are represented as deltaCt values ($\text{deltaCt}_{\text{mRNA}} = \text{Ct}_{\text{mRNA}} - \text{Ct}_{\text{GAPDH}}$). Differential gene expression in the clinical trial was assessed comparing gene expression at baseline and 6-month follow-up. Associations with Benjamini-Hochberg¹⁰ based FDR ≤ 0.05 were considered significant.

Long term changes in gene expression analysis in the Gutenberg Health Study

Longitudinal gene expression data (Baseline and 5-year follow up) of 1,092 GHS individuals was pre-processed, \log_2 -transformed and batch effects were removed by quantile normalization followed by ComBat.¹¹ The difference between SBP and DBP and expression of each candidate transcript at baseline and 5-year follow-up was calculated for each individual. Association analyses between changes of each BP trait and candidate transcript expression were performed using linear regression and were adjusted for sex, age at baseline and BMI change between baseline and follow-up. Adjustments for multiple testing were performed using the Benjamini-Hochberg method¹⁰, and the significance level was set to 0.05. Results are given as percent mRNA change (%mRNA change) after 6 months \pm standard error.

Expression quantitative trait loci (eQTL) analysis in the GHS

eQTL analyses were performed in 1,333 individuals from the GHS with available gene expression and available genome-wide genotyping data based on Affymetrix SNP 6.0 microarrays with imputation based on the 1,000 genome reference¹². Two approaches were used to identify eQTLs related to BP or cardiovascular disease:

A.) For each candidate transcript, cis-eQTLs were calculated for SNPs within $\pm 250\text{kb}$ around the transcription start site and a minor allele frequency (MAF) $\geq 1\%$. The Genome-wide Repository of Associations between SNPs and Phenotypes (GRASP) database version 2.0.0.0¹³ was studied for significant cis- eQTLs that were at least moderately associated to a BP or a cardiovascular disease related trait ($p\text{-value} \leq 10^{-4}$).

B.) Published GWAS results of BP-related traits were retrieved from the GWAS catalogue (2017-03-20)¹⁴. Latest identified novel BP variants not present in the GWAS catalogue were retrieved from the relevant publications¹⁵⁻¹⁸.

In total, 270 unique SNPs with genome-wide association to at least one BP trait were found. Of those, 191 SNPs with a MAF $\geq 1\%$ present in GHS were tested, leading to a significance level of 2.6×10^{-4} .

All eQTL analyses were performed using linear models assuming an additive genetic model, and were adjusted for age, sex and technical covariates ⁷.

Association of candidate transcript expression and sub-clinical phenotypes in the GHS

In 1,285 GHS subjects, associations between sub-clinical phenotypes and candidate transcript expression from microarrays were calculated using linear regression models adjusted for age, sex and technical covariates as described above. For each sub-clinical phenotype independently, associations with a Benjamini-Hochberg based FDR ≤ 0.05 were considered significant. Results are given as mRNA change per 10mmHg increase in blood pressure trait (delta mRNA) \pm standard error.

Association of CRIP1 serum levels with clinically relevant endpoints

In the Moli-Sani Study, CRIP1 serum levels were tested for association with cardiovascular endpoints. For this, multivariate associations of stroke, heart failure, and coronary heart disease with CRIP1 serum levels were calculated using Cox regression adjusted for age, sex, systolic BP and plate layout. Since subjects were selected to the case-cohort sample with unequal probabilities, inverse of the sampling probabilities were used as weights in the Cox regression as proposed by Kulathinal et al. ¹⁹.

References

1. Wild PS, Sinning CR, Roth A, Wilde S, Schnabel RB, Lubos E, Zeller T, Keller T, Lackner KJ, Blettner M, Vasan RS, Munzel T, Blankenberg S. Distribution and categorization of left ventricular measurements in the general population: Results from the population-based Gutenberg heart study. *Circulation. Cardiovascular imaging*. 2010;3:604-613
2. Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: Objectives and design. *American journal of epidemiology*. 2002;156:871-881
3. Holle R, Happich M, Lowel H, Wichmann HE. Kora--a research platform for population based health research. *Gesundheitswesen*. 2005;67 Suppl 1:S19-25
4. Volzke H, Alte D, Schmidt CO, et al. Cohort profile: The study of health in pomerania. *International journal of epidemiology*. 2011;40:294-307
5. Di Castelnuovo A, de Curtis A, Costanzo S, Persichillo M, Olivieri M, Zito F, Donati MB, de Gaetano G, Iacoviello L, Investigators M-SP. Association of d-dimer levels with all-cause mortality in a healthy adult population: Findings from the moli-sani study. *Haematologica*. 2013;98:1476-1480
6. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J. A new equation to estimate glomerular filtration rate. *Annals of internal medicine*. 2009;150:604-612
7. Schurmann C, Heim K, Schillert A, et al. Analyzing illumina gene expression microarray data from different tissues: Methodological aspects of data analysis in the metaxpress consortium. *PloS one*. 2012;7:e50938
8. Bonaccio M, Di Castelnuovo A, Rago L, de Curtis A, Assanelli D, Badilini F, Vaglio M, Costanzo S, Persichillo M, Cerletti C, Donati MB, de Gaetano G, Iacoviello L. T-wave axis deviation is associated with biomarkers of low-grade inflammation. Findings from the moli-sani study. *Thrombosis and haemostasis*. 2015;114:1199-1206
9. Magee L. R 2 measures based on wald and likelihood ratio joint significance tests. *The American Statistician*. 1990;44:250-253
10. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing *Journal of the Royal Statistical Society* 1995;57:289-300
11. Muller C, Schillert A, Rothemeier C, Tregouet DA, Proust C, Binder H, Pfeiffer N, Beutel M, Lackner KJ, Schnabel RB, Tiret L, Wild PS, Blankenberg S, Zeller T, Ziegler A. Removing batch effects from longitudinal gene expression - quantile normalization plus combat as best approach for microarray transcriptome data. *PloS one*. 2016;11:e0156594
12. The 1000 Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature*. 2015;526:68-74
13. Leslie R, O'Donnell CJ, Johnson AD. Grasp: Analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics*. 2014;30:i185-194
14. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. The nhgri gwas catalog, a curated resource of snp-trait associations. *Nucleic acids research*. 2014;42:D1001-1006
15. Liu C, Kraja AT, Smith JA, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nature genetics*. 2016;48:1162-1170
16. Ehret GB, Ferreira T, Chasman DI, et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nature genetics*. 2016;48:1171-1184

17. Surendran P, Drenos F, Young R, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nature genetics*. 2016;48:1151-1161
18. Warren HR, Evangelou E, Cabrera CP, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nature genetics*. 2017;49:403-415
19. Kulathinal S, Karvanen J, Saarela O, Kuulasmaa K. Case-cohort design in practice - experiences from the morgam project. *Epidemiologic perspectives & innovations : EP+I*. 2007;4:15
20. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eqtls as putative drivers of known disease associations. *Nature genetics*. 2013;45:1238-1243
21. Newton-Cheh C, Johnson T, Gateva V, et al. Genome-wide association study identifies eight loci associated with blood pressure. *Nature genetics*. 2009;41:666-676
22. Wain LV, Verwoert GC, O'Reilly PF, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nature genetics*. 2011;43:1005-1011
23. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nature genetics*. 2010;42:295-302
24. Zhernakova A, Stahl EA, Trynka G, et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-hla shared loci. *PLoS genetics*. 2011;7:e1002004
25. Kottgen A, Pattaro C, Boger CA, et al. New loci associated with kidney function and chronic kidney disease. *Nature genetics*. 2010;42:376-384
26. Kottgen A, Albrecht E, Teumer A, et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nature genetics*. 2013;45:145-154
27. Levy D, Ehret GB, Rice K, et al. Genome-wide association study of blood pressure and hypertension. *Nature genetics*. 2009;41:677-687
28. Gieger C, Radhakrishnan A, Cvejic A, et al. New gene functions in megakaryopoiesis and platelet formation. *Nature*. 2011;480:201-208
29. Shameer K, Denny JC, Ding K, et al. A genome- and phenome-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. *Human genetics*. 2014;133:95-109
30. Ehret GB, Munroe PB, Rice KM, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011;478:103-109
31. Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. *Nature genetics*. 2014;46:543-550
32. Gudbjartsson DF, Bjornsdottir US, Halapi E, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nature genetics*. 2009;41:342-347
33. Eriksson N, Tung JY, Kiefer AK, Hinds DA, Francke U, Mountain JL, Do CB. Novel associations for hypothyroidism include known autoimmune risk loci. *PloS one*. 2012;7:e34442

Supplementary Tables

Table S1. Study characteristics. Data presented as mean \pm standard deviation or number and percent. *Prevalent hypertension defined as SBP \geq 140 mmHg or DBP \geq 90 mmHg or antihypertensive drug treatment during the last two weeks. **P-values were calculated by Student's t-test for differences of continuous traits and by chi-squared test for differences of dichotomous traits.

Study Characteristics	Discovery			Replication		
	GHS n = 1,285	MESA n = 1,264	P-value**	SHIP-TREND n = 993	KORA F4 n = 997	P-value**
Mean age [years]	54.7 \pm 11.0	60.2 \pm 9.5	<0.001	50.1 \pm 13.7	70.38 \pm 5.37	<0.001
Females (%)	663 (51.6)	650 (51.4)	0.963	554 (56.0)	493 (49.6)	0.005
Current smoker, n (%)	239 (18.6)	170 (13.5)	<0.001	213 (21.5)	66 (6.7)	<0.001
Body mass index (BMI) [kg/m ²]	27.0 \pm 4.6	29.3 \pm 5.2	<0.001	27.3 \pm 4.6	28.9 \pm 4.5	<0.001
Hyperlipidemia, n (%)	351 (27.3)	206 (16.3)	<0.001	141 (14.2)	155 (15.7)	0.435
Diabetes Mellitus, n (%)	79 (6.2)	104 (8.2)	0.050	28 (2.8)	211 (21.2)	<0.001
CAD, n (%)	58 (4.6)	55(4.4)	0.918	72 (7.3)	117 (11.8)	<0.001
Systolic blood pressure (SBP) [mmHg]	132.2 \pm 17.8	122.9 \pm 18.9	<0.001	124.4 \pm 16.9	128.7 \pm 20.0	<0.001
Diastolic blood pressure (DBP) [mmHg]	83.5 \pm 9.7	71.4 \pm 9.6	<0.001	76.6 \pm 9.8	74.0 \pm 10.1	<0.001
Pulse pressure (PP) [mmHg]	48.7 \pm 13.0	51.5 \pm 15.0	<0.001	47.8 \pm 12.2	54.7 \pm 14.7	<0.001
Heart rate, bpm	68.7 \pm 10.9	62.9 \pm 9.3	<0.001	71.0 \pm 10.7	71.6 \pm 11.3	0.169
Prevalent Hypertension, n (%)*	618 (48.1)	557 (44.1)	0.046	460 (46.5)	682 (68.7)	<0.001
Antihypertensive drug treatment	344 (26.8)	449 (35.5)	<0.001	388 (39.2)	421 (42.4)	0.166
Caucasian, n (%)	1,285	590 (46.7)	<0.001	993	997	1

Tables S2: Gene-specific transcripts associated with blood pressure traits in the discovery analysis (see Excel file Supplementary Table S2).

Tables S3: Gene-specific transcripts associated with blood pressure traits in the validation analysis (see Excel file Supplementary Table S3).

Tables S4: Gene-specific transcripts associated with blood pressure traits in the discovery and validation analysis after exclusion of individuals receiving antihypertensive treatment (see Excel file Supplementary Table S4).

Tables S5: Gene-specific transcripts associated with blood pressure traits in the discovery analysis including only subjects of Caucasian ethnicity in MESA (see Excel file Supplementary Table S5)

Table S6: Variance (%) in BP levels attributable to transcript expression. Values refer to the generalized R^2 which was computed based on the likelihoods from linear regressions for each BP trait / candidate gene combination. The total attributable variance refers to the aggregated proportion of variance by the transcripts, which were significantly associated with the given trait across all studies (bold).

Gene	Monocytes						Whole blood					
	GHS			MESA			KORA			SHIP		
	SBP	DBP	PP	SBP	DBP	PP	SBP	DBP	PP	SBP	DBP	PP
	Attributable variance (AV) % (95% confidence interval)											
CEBPA	1.63 (0.19, 2.95)	1.42 (0.03, 2.67)	0.66 (0.33, 1.49)	0.12 (0.00, 0.46)	0.39 (0.00, 0.66)	0.00 (0.00, 0.21)	0.35 (0.00, 1.14)	0.47 (0.00, 1.27)	0.14 (0.00, 0.66)	1.02 (0.00, 2.20)	1.04 (0.00, 2.15)	0.26 (0.00, 0.84)
CRIP1	5.52 (2.93, 8.03)	3.68 (1.47, 5.72)	2.96 (0.94, 4.81)	2.98 (0.61, 4.12)	1.86 (0.23, 3.11)	1.94 (0.00, 2.69)	0.36 (0.00, 1.09)	0.32 (0.00, 1.02)	0.22 (0.00, 0.79)	1.18 (0.00, 2.41)	0.61 (0.00, 1.51)	0.68 (0.00, 1.66)
F12	1.46 (0.02, 2.74)	0.42 (0.40, 1.09)	1.36 (0.12, 2.65)	0.20 (0.00, 0.67)	0.02 (0.00, 0.35)	0.27 (0.00, 0.65)	0.58 (0.00, 1.56)	0.17 (0.00, 0.67)	0.68 (0.00, 1.73)	0.02 (0.00, 0.25)	0.46 (0.00, 1.26)	0.59 (0.00, 1.45)
LMNA	1.52 (0.00, 2.90)	1.25 (0.07, 2.43)	0.66 (0.41, 1.56)	0.74 (0.00, 1.07)	0.33 (0.00, 1.05)	0.59 (0.00, 0.63)	0.17 (0.00, 0.69)	0.65 (0.00, 1.62)	0.00 (0.00, 0.21)	0.20 (0.00, 0.76)	0.58 (0.00, 1.53)	0.00 (0.00, 0.19)
MYADM	3.96 (1.65, 6.15)	2.17 (0.51, 3.70)	2.50 (0.54, 4.30)	0.77 (0.00, 1.26)	0.03 (0.00, 0.20)	1.13 (0.00, 1.67)	1.65% (0.00, 3.26)	1.89 (0.10, 3.53)	0.77 (0.00, 1.83)	1.02 (0.00, 2.14)	0.61 (0.00, 1.53)	0.52 (0.00, 1.31)
TIPARP	3.64 (1.60, 5.51)	1.56 (0.12, 2.86)	2.72 (0.88, 4.43)	0.55 (0.00, 0.91)	0.49 (0.00, 1.18)	0.27 (0.00, 0.39)	0.63 (0.00, 1.65)	0.23 (0.00, 0.84)	0.68 (0.00, 1.70)	1.16 (0.00, 2.44)	0.36 (0.00, 1.07)	0.93 (0.00, 2.09)
TPPP3	5.12 (2.54, 7.54)	4.71 (2.34, 6.89)	1.94 (0.25, 3.44)	2.12 (0.05, 3.27)	0.95 (0.00, 1.82)	1.69 (0.00, 2.61)	1.05 (0.00, 2.41)	0.83 (0.00, 1.99)	0.70 (0.00, 1.87)	0.10 (0.00, 0.49)	0.42 (0.00, 1.19)	0.01 (0.00, 0.25)
TSC22D3	4.05 (1.75, 6.22)	1.03 (0.21, 2.13)	3.96 (1.56, 6.18)	0.89 (0.00, 1.48)	0.35 (0.00, 0.75)	0.75 (0.00, 1.30)	0.79 (0.00, 1.77)	1.25 (0.00, 2.57)	0.23% (0.00, 0.77)	1.24 (0.00, 2.49)	0.37 (0.00, 1.04)	1.01 (0.00, 2.17)
Total AV by candidate genes	11.33 (7.57, 14.48)	8.31 (4.95, 10.79)	4.74 (1.92, 7.10)	3.43 (0.44, 4.39)	2.11 (0.00, 2.97)	2.67 (0.16, 3.83)	2.82 (0.09, 4.44)	4.05 (0.92, 6.00)	1.36 (0.00, 2.73)	4.21 (1.25, 6.17)	3.03 (0.15, 4.70)	1.37 (0.00, 2.60)

Table S7: Association of candidate transcript expression with blood pressure-related, subclinical cardiovascular phenotypes in the Gutenberg Health Study. Each trait was tested for association to mRNA by a linear regression adjusted for age, sex and technical covariates, and significant associations (FDR < 0.05) are shown in bold. IVSD = Interventricular septal thickness end-diastolic (mm) LVIDD = left ventricular internal diameter end-diastolic (cm); LVPWD = left ventricular posterior wall thickness end-diastolic (cm); LVM= left ventricular mass (g); RWT = relative wall thickness (cm); LVH = left ventricular hypertrophy (0: LVH; 1: no LVH), NTproBNP = N-terminal pro B-type natriuretic peptide (pg/mL).

Gene	IVSD	LVIDD	LVPWD	LVM	RWT	LVH	Stiffness index	NT-pro BNP
P-value (log ₂ -fold mRNA change per unit trait)								
CRIP1	0.0019 (0.0198)	0.3877 (0.0021)	0.0001 (0.03)	0.0005 (0.0006)	0.0088 (0.3697)	0.0052 (0.1276)	0.651 (-0.0017)	0.4461 (0.0069)
MYADM	0.4948 (-0.0034)	0.4896 (0.0013)	0.4607 (0.0044)	0.7331 (0)	0.6963 (-0.0425)	0.1333 (-0.0515)	0.4338 (-0.0023)	0.259 (0.0079)
TIPARP	0.2419 (-0.0054)	0.2605 (-0.0019)	0.4515 (-0.0042)	0.1338 (-0.0002)	0.8738 (-0.0162)	0.1151 (-0.0507)	0.1534 (-0.004)	0.0468 (0.013)
TSC22D3	0.0048 (0.0164)	0.7672 (0.0006)	0.1148 (0.0111)	0.0189 (0.0004)	0.0502 (0.2517)	0.0191 (0.0963)	0.0145 (-0.0085)	0.796 (-0.0021)
CEBPA	0.1006 (0.007)	0.588 (0.0009)	0.0955 (0.0086)	0.0706 (0.0002)	0.2875 (0.1002)	0.1369 (0.0419)	0.8103 (0.0006)	0.4772 (0.0043)
F12	0.3003 (0.0044)	0.7302 (-0.0005)	0.8679 (0.0009)	0.6823 (0)	0.4496 (0.0708)	1 (0)	0.645 (0.0012)	0.4218 (0.0048)
LMNA	0.0002 (0.031)	0.7639 (0.0009)	0.0119 (0.0252)	0.0026 (0.0007)	0.0041 (0.5258)	0.0015 (0.1788)	0.595 (0.0026)	0.1599 (-0.0166)
TPPP3	0.1193 (0.0112)	0.7369 (-0.0009)	0.0037 (0.0251)	0.1082 (0.0003)	0.0267 (0.3513)	0.4471 (0.0387)	0.4822 (0.003)	0.8027 (0.0025)

Tables S8: Single nucleotide polymorphisms from the GRASP database (A) and published BP related variants (B) used for analyses of expression quantitative trait loci (see Excel file Supplementary Table S8).

Table S9: Blood pressure-related eQTLs and SNPs related to blood pressure. Genome-wide significant associations for blood pressure and other disease traits were taken from the GWAS Catalog (www.ebi.ac.uk/gwas/) and novel BP GWAS publications which were not present in the GWAS Catalog in March 2017 ^{15, 17, 16, 18}. Associations between the SNP and gene expression were assessed in 1,133 samples from the GHS (with available gene expression and genotyping data) by linear regression under the assumption of an additive genetic model and in publically available data of 5,311 whole blood samples from Westra et al ²⁰; (genenetwork.nl/blooddeqtlbrowser/). Associations with a p-value < 2.6×10^{-4} (=0.05/191 SNPs tested) were considered significant. The linkage disequilibrium between rs3184504 and rs653178 is r^2 0.99. n.s: not significant.

SNP / allele	SNP annotation			SNP-BP associations		SNP-trait associations	SNP-gene expression associations		
	Chr	Gene	MAF	SBP	DBP	Other traits	Gene	Monocytes	Whole blood ²⁰
rs653178 / T	12	ATXN2 (intron)	48.6%		3×10^{-18} ²¹	Mean arterial pressure ²²	CRIP1	3.59×10^{-7}	1.02×10^{-6}
						Celiac disease ^{23, 24}	TPPP3	1.35×10^{-4}	n.s.
						Chronic kidney disease ²⁵	MYADM	7.86×10^{-5}	5.12×10^{-7}
						Urate levels ²⁶	TIPARP	2.6×10^{-7}	n.s.
rs3184504 / C	12	SH2B3 (exon, missense)	48.9%	5×10^{-9} ²⁷	3×10^{-14} ²⁷	Platelet count ^{28, 29}	CRIP1	6.53×10^{-7}	1.32×10^{-6}
						Kynurenine levels ³¹	TPPP3	1.64×10^{-4}	n.s.
						Eosinophil counts ³²	MYADM	6.47×10^{-5}	8.06×10^{-7}
						Hypothyroidism ³³	TIPARP	3.83×10^{-7}	n.s.

Table S10: Baseline characteristics of subjects selected for CRIP1 measurements from the Moli-Sani Study.

Study Characteristics	Moli-Sani Study n = 379
Mean age [years]	54.5 ± 9.4
Females, %	47.2
Current smoker, %	17.5
Body mass index (BMI) [kg/m ²]	28.5 ± 4.6
Diabetes Mellitus, %	5.3
CAD, %	1.3
Systolic blood pressure (SBP) [mmHg]	140.7 ± 20.4
Diastolic blood pressure (DBP) [mmHg]	83.2 ± 9.6
Pulse pressure (PP) [mmHg]	57.5 ± 16.5
Heart rate, bpm	68.0 ± 11.5
Prevalent Hypertension, %	55.0
Antihypertensive drug treatment, %	30.5

Supplementary Figures

Figure S1. Study workflow.

I) A meta-analysis of transcriptomics data and BP traits was performed for discovery in GHS and MESA. A FDR < 0.01 was used to indicate statistical significance.

II) External validation in a whole blood data set of independent studies (SHIP-Trend and KORA F4). A statistical significance of $p < 0.05$ in each study for at least one BP trait and consistent direction of effects were used to selected candidate genes. SBP: systolic blood pressure; DBP: diastolic blood pressure; PP: pulse pressure; CVD: cardiovascular disease; RCT randomized clinical trial.

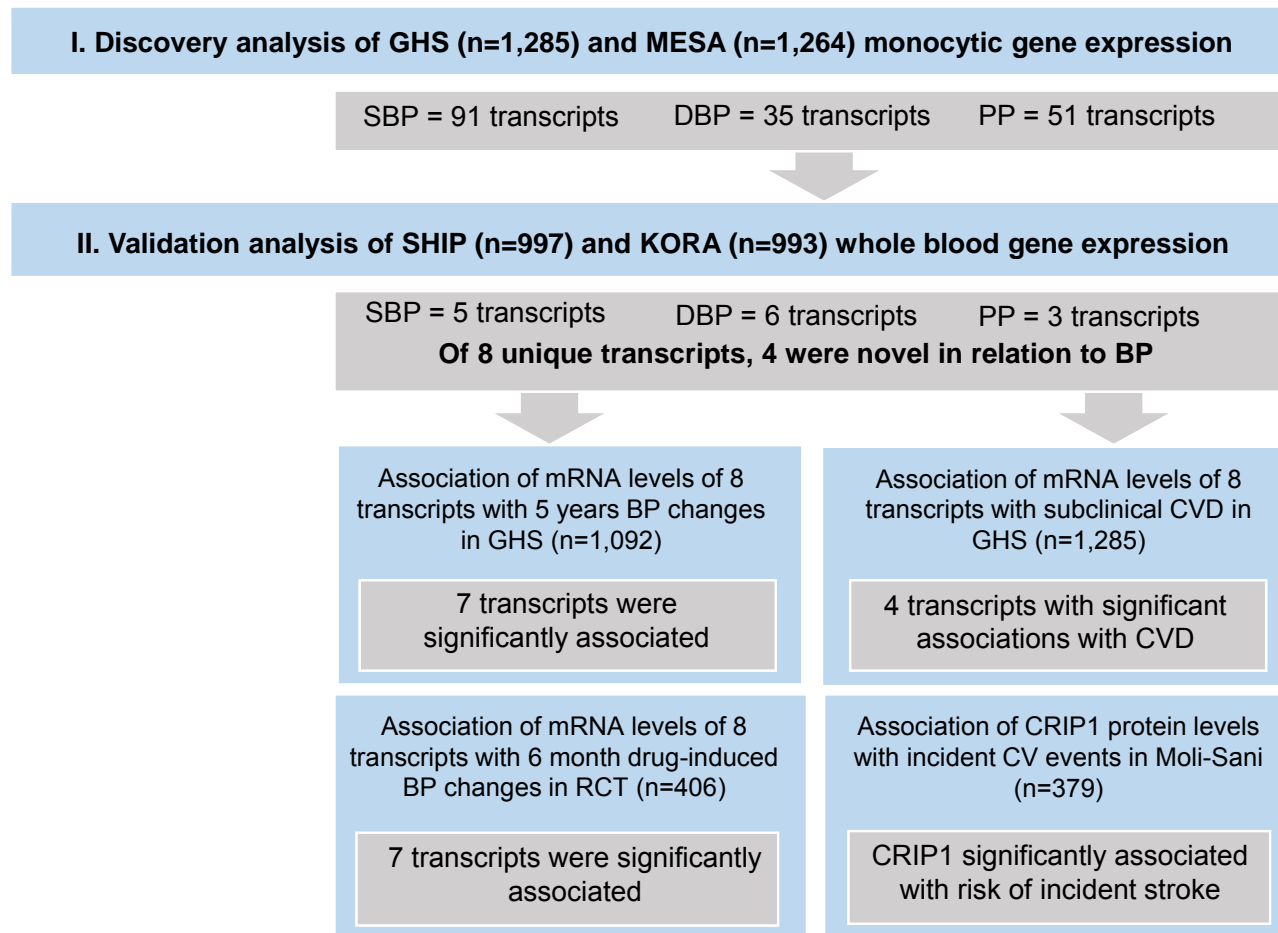


Figure S2: Candidate gene expression by study and cell types.

To consider study specific technical differences between studies, expression levels were normalized to the overall mean of all genes on the microarrays.

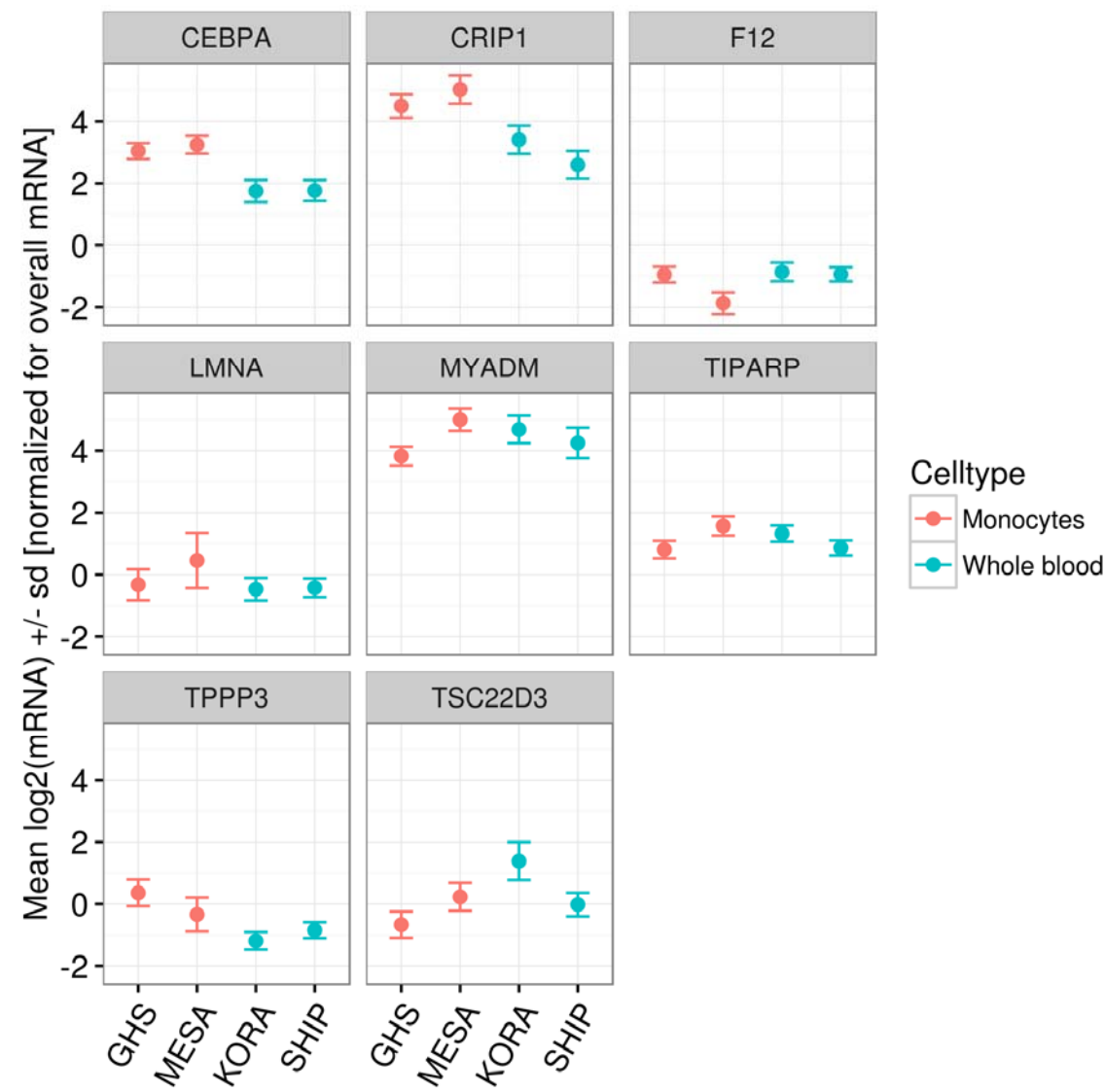


Figure S3a: Association between BP changes and *CRIP1* mRNA expression changes after 5-years follow-up in the GHS. A) Association to systolic BP; B) Association to diastolic BP. Numbers indicate the total number of individuals per group. BP: blood pressure

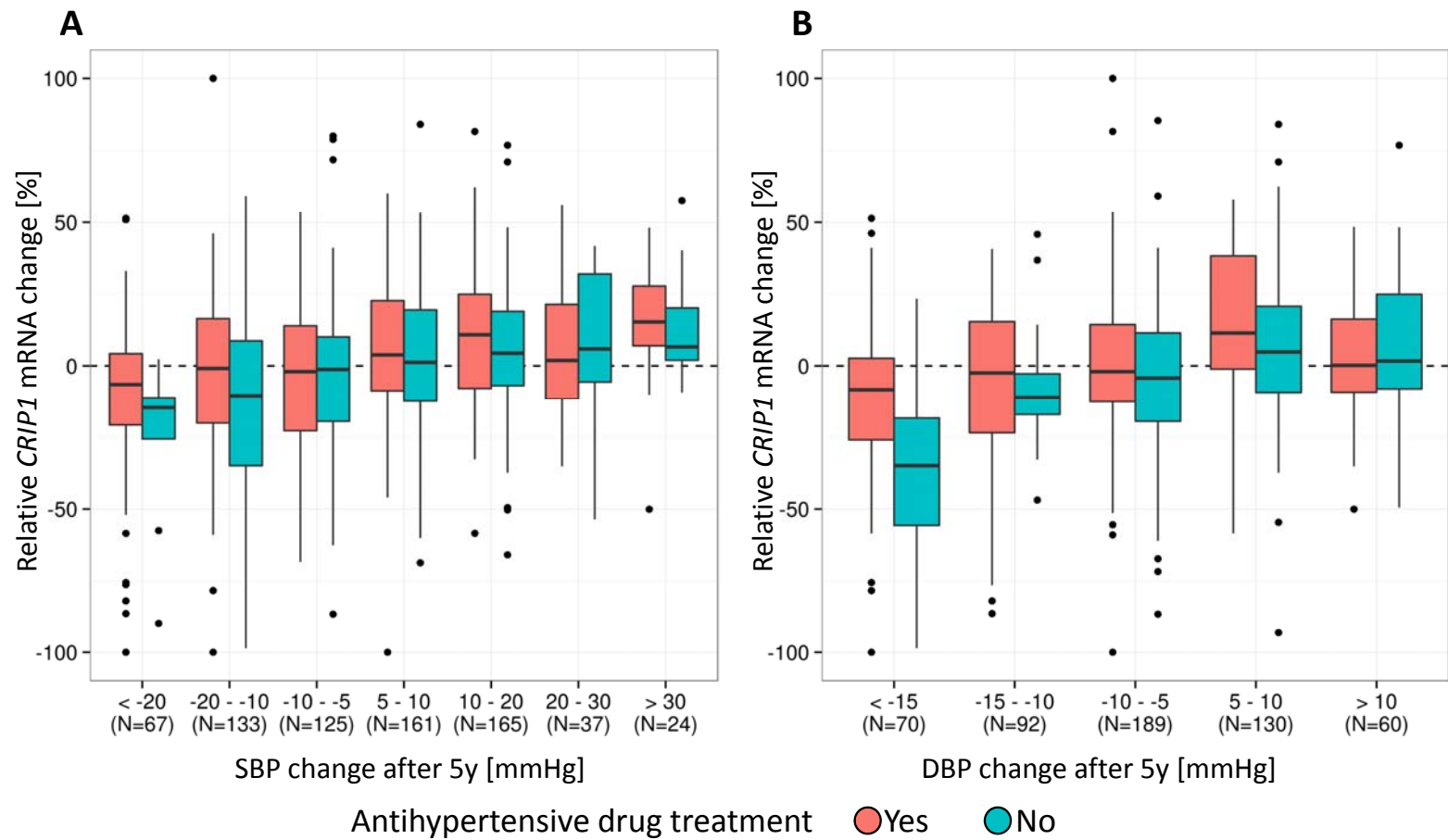


Figure S3b: Association between BP changes and *MYADM* mRNA expression changes after 5-years follow-up in the GHS. A) Association to systolic BP; B) Association to diastolic BP. Numbers indicate the total number of individuals per group. BP: blood pressure

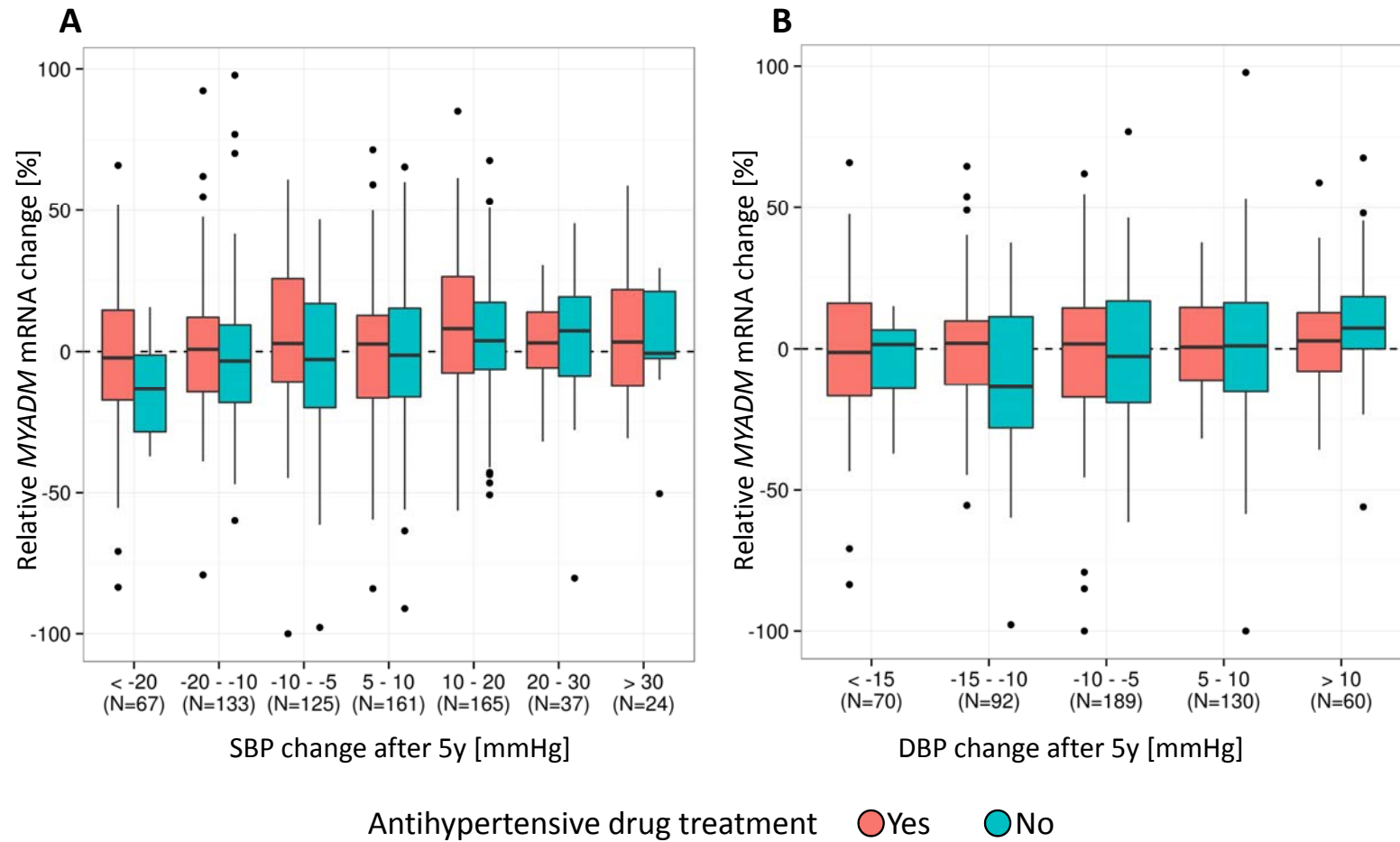


Figure S3c: Association between BP changes and *TIPARP* mRNA expression changes after 5-years follow-up in the GHS. A) Association to systolic BP; B) Association to diastolic BP. Numbers indicate the total number of individuals per group. BP: blood pressure

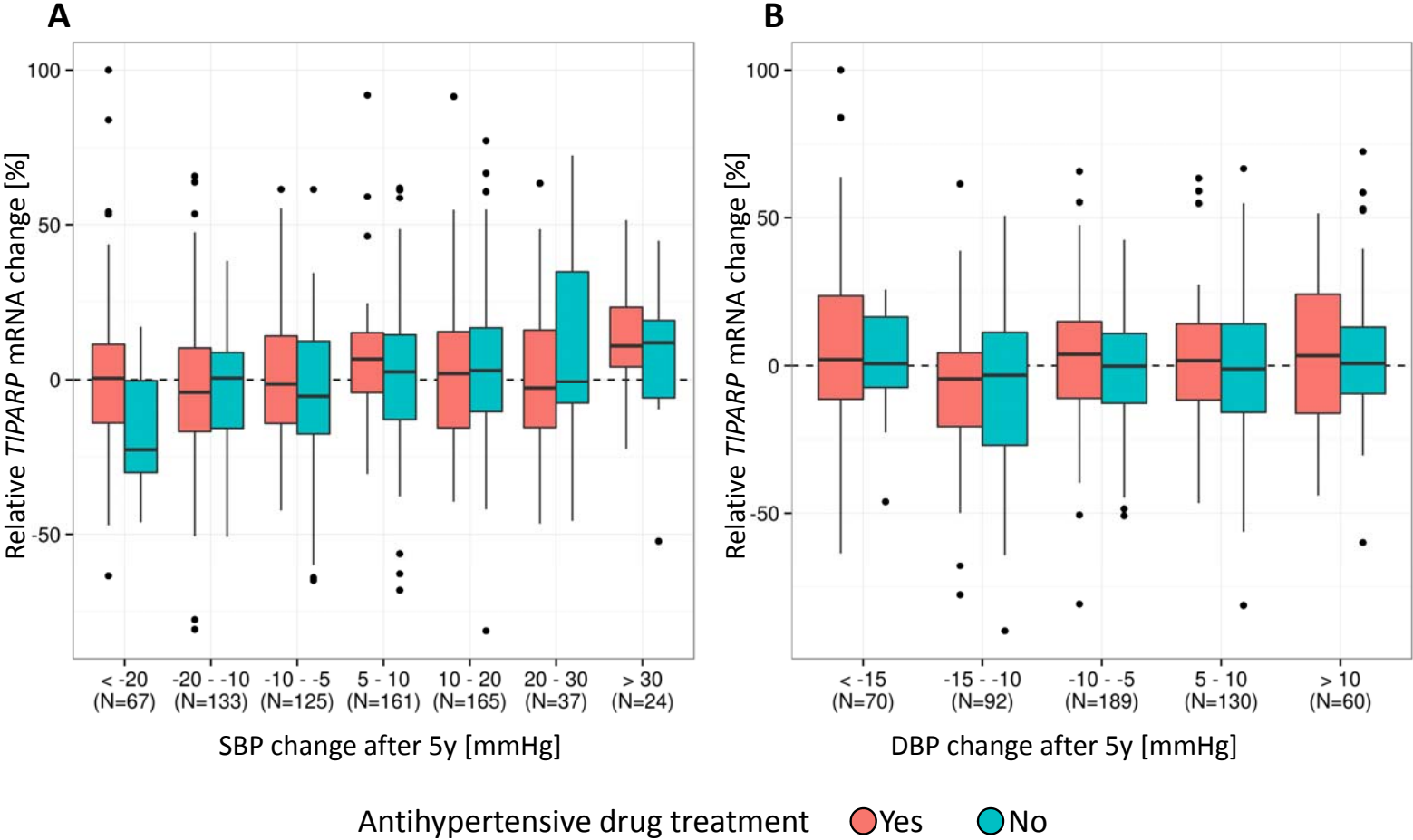


Figure S3d: Association between BP changes and *TSC22D3* mRNA expression changes after 5-years follow-up in the GHS. A) Association to systolic BP; B) Association to diastolic BP. Numbers indicate the total number of individuals per group. BP: blood pressure

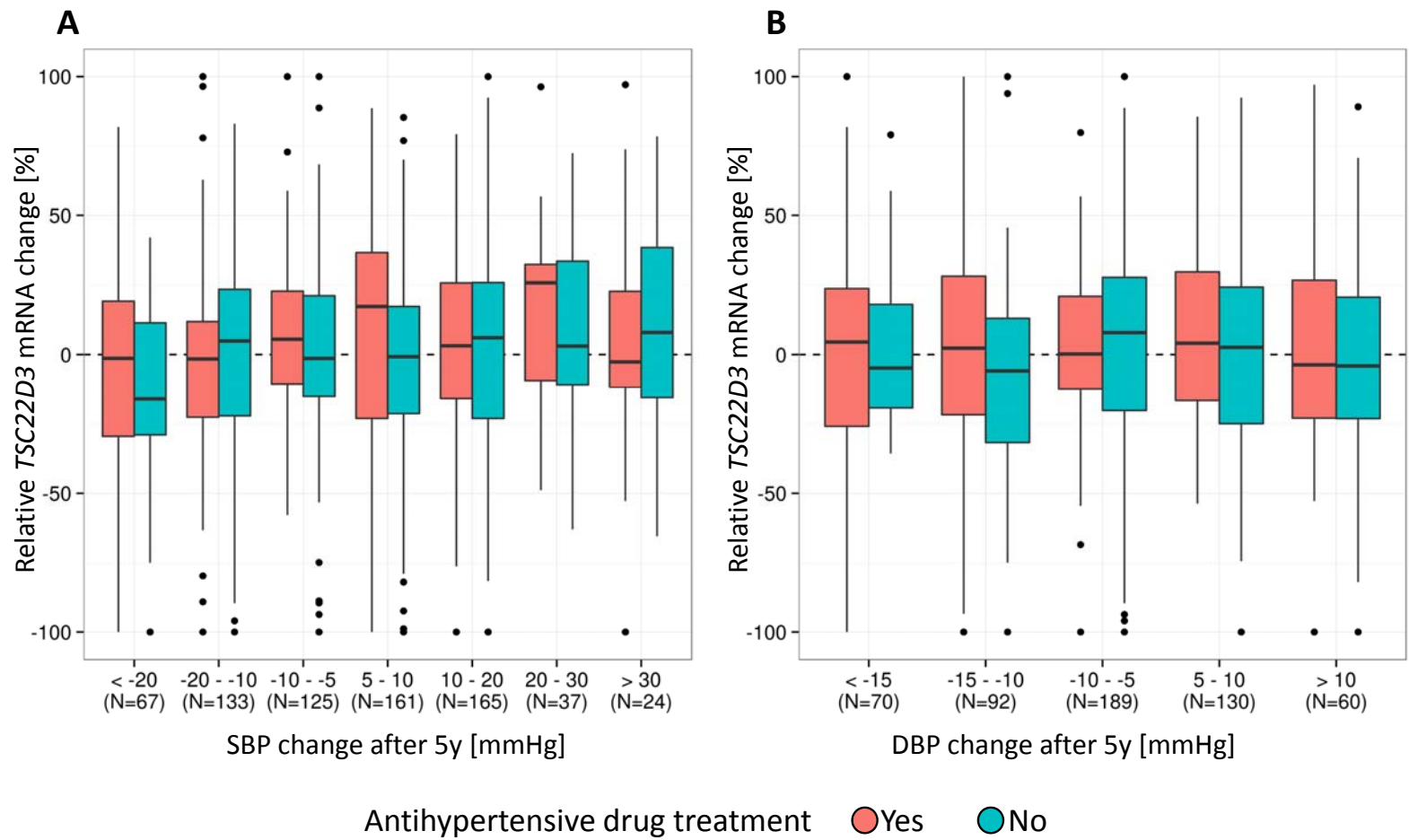


Figure S3e: Association between BP changes and *CEBPA* mRNA expression changes after 5-years follow-up in the GHS. A) Association to systolic BP; B) Association to diastolic BP. Numbers indicate the total number of individuals per group. BP: blood pressure

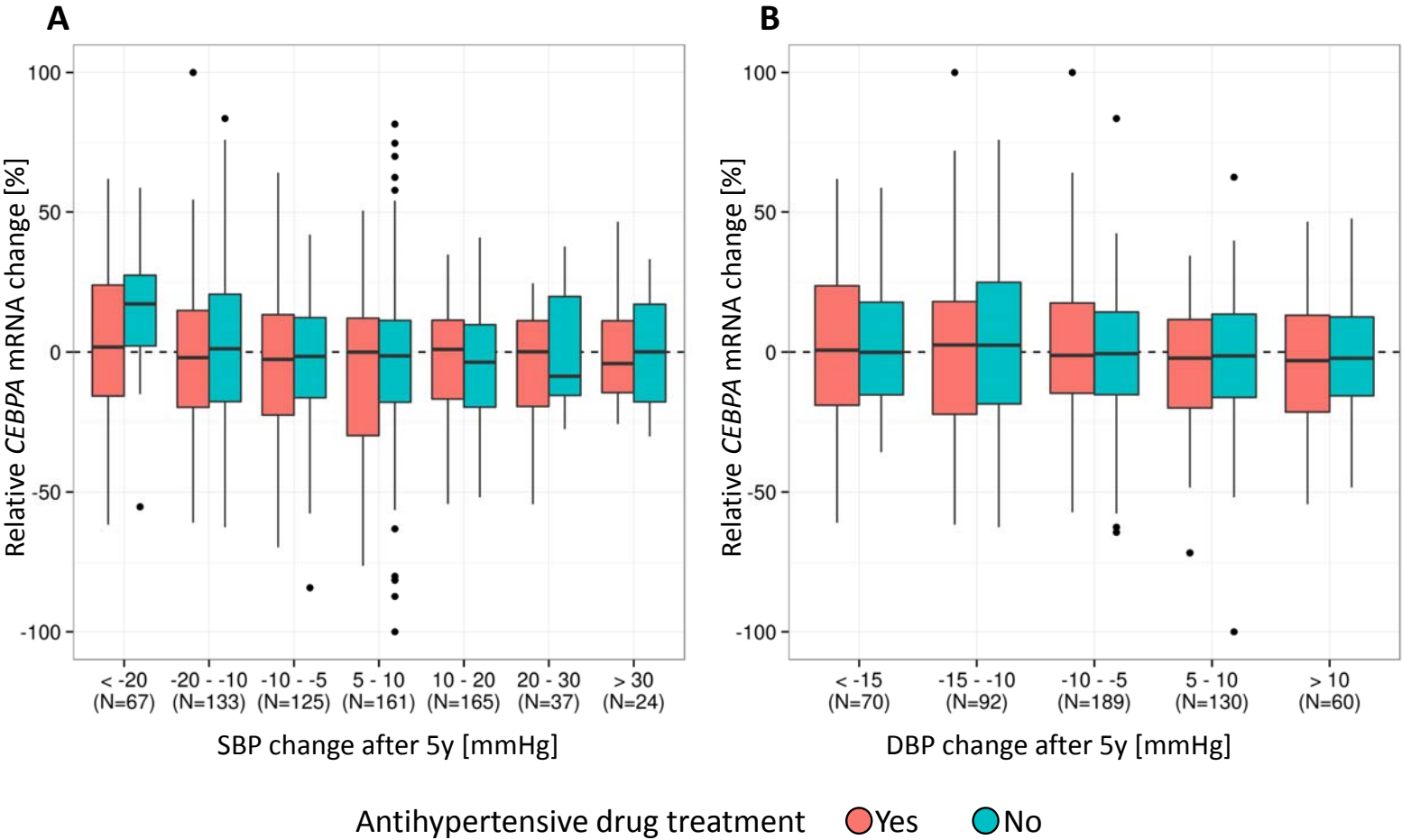


Figure S3f: Association between BP changes and *LMNA* mRNA expression changes after 5-years follow-up in the GHS. A) Association to systolic BP; B) Association to diastolic BP. Numbers indicate the total number of individuals per group. BP: blood pressure

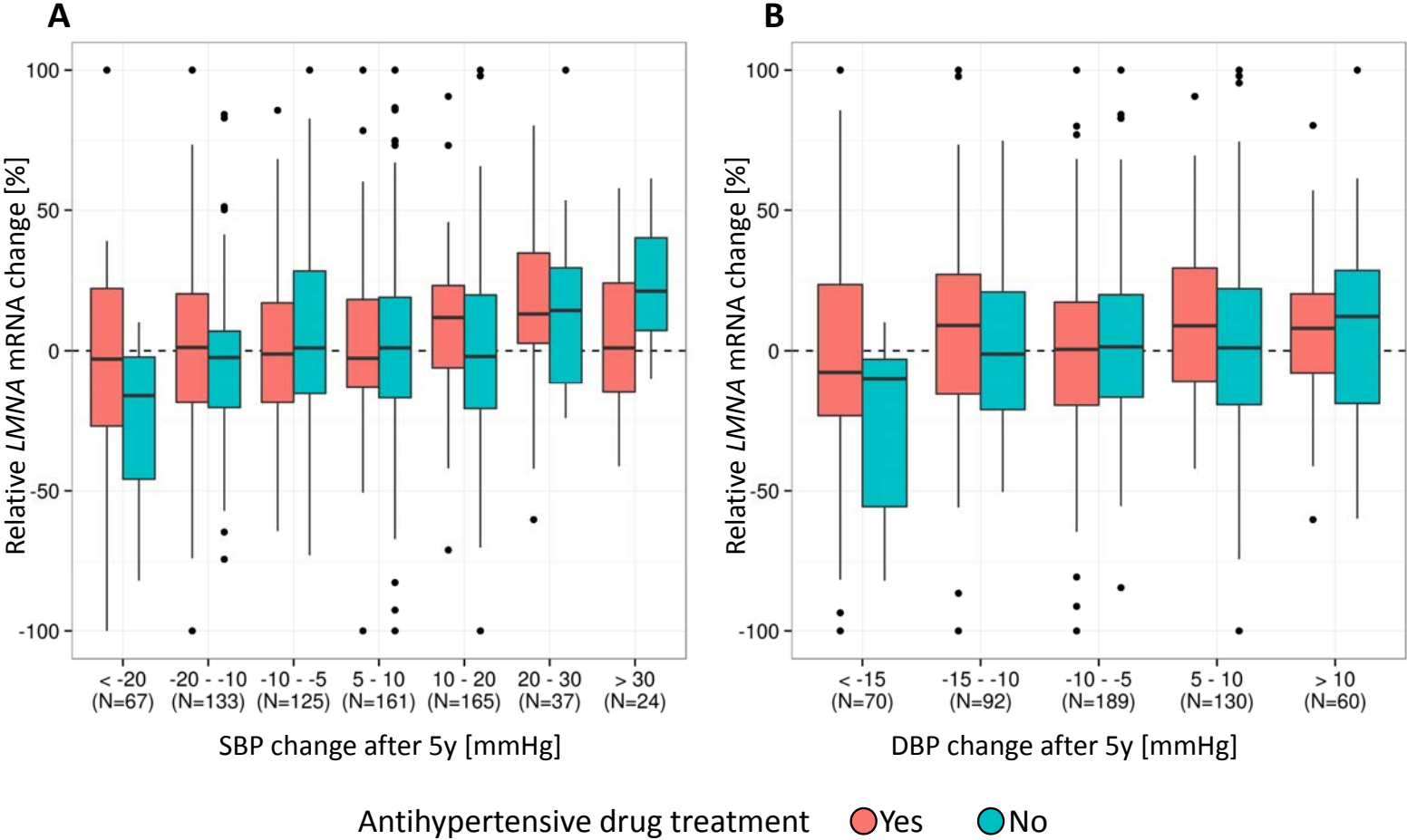


Figure S3g: Association between BP changes and *TPPP3* mRNA expression changes after 5-years follow-up in the GHS. A) Association to systolic BP; B) Association to diastolic BP. Numbers indicate the total number of individuals per group. BP: blood pressure

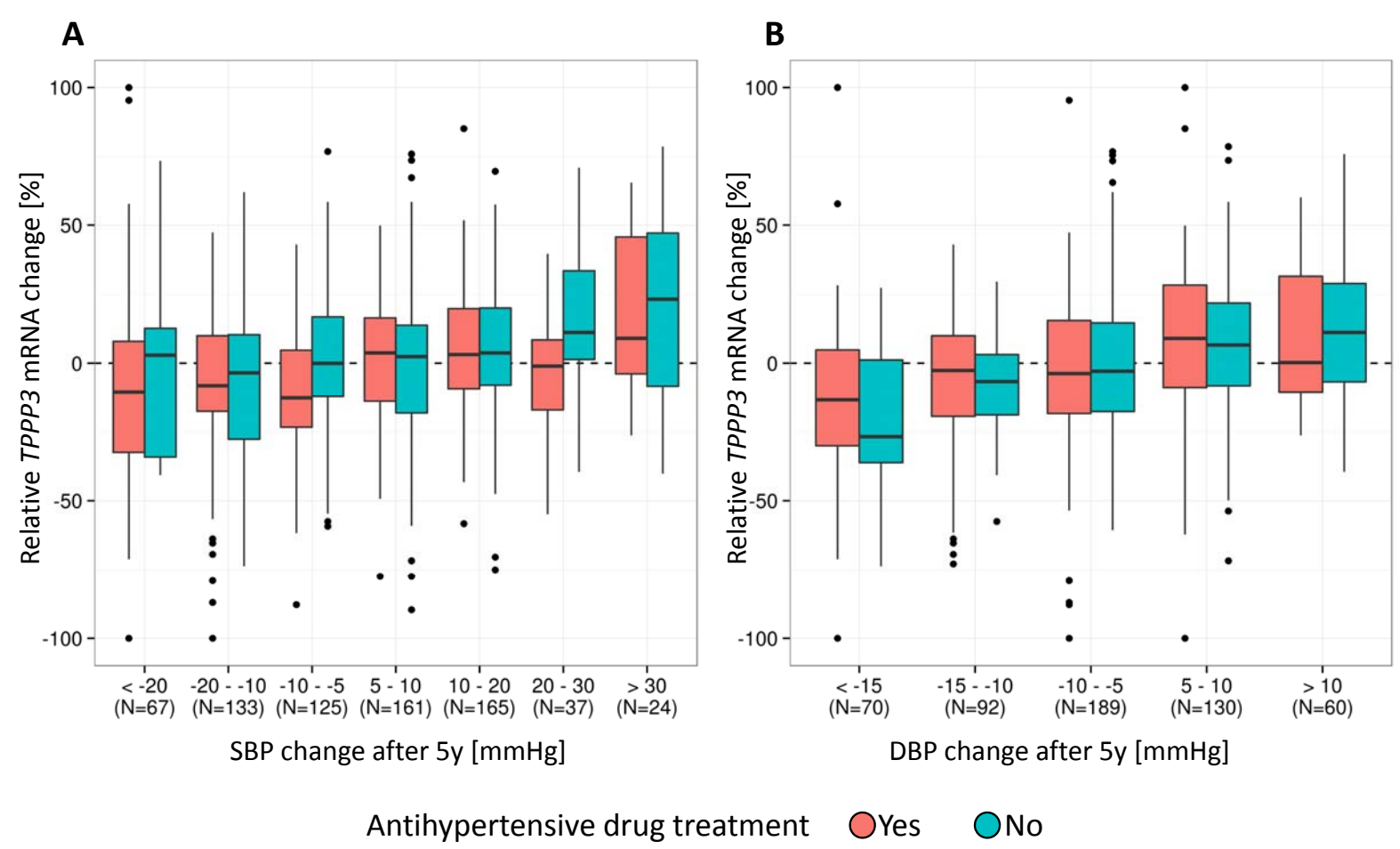


Figure S4: Reduction of blood pressure between baseline and 6 month follow up in a clinical trial.

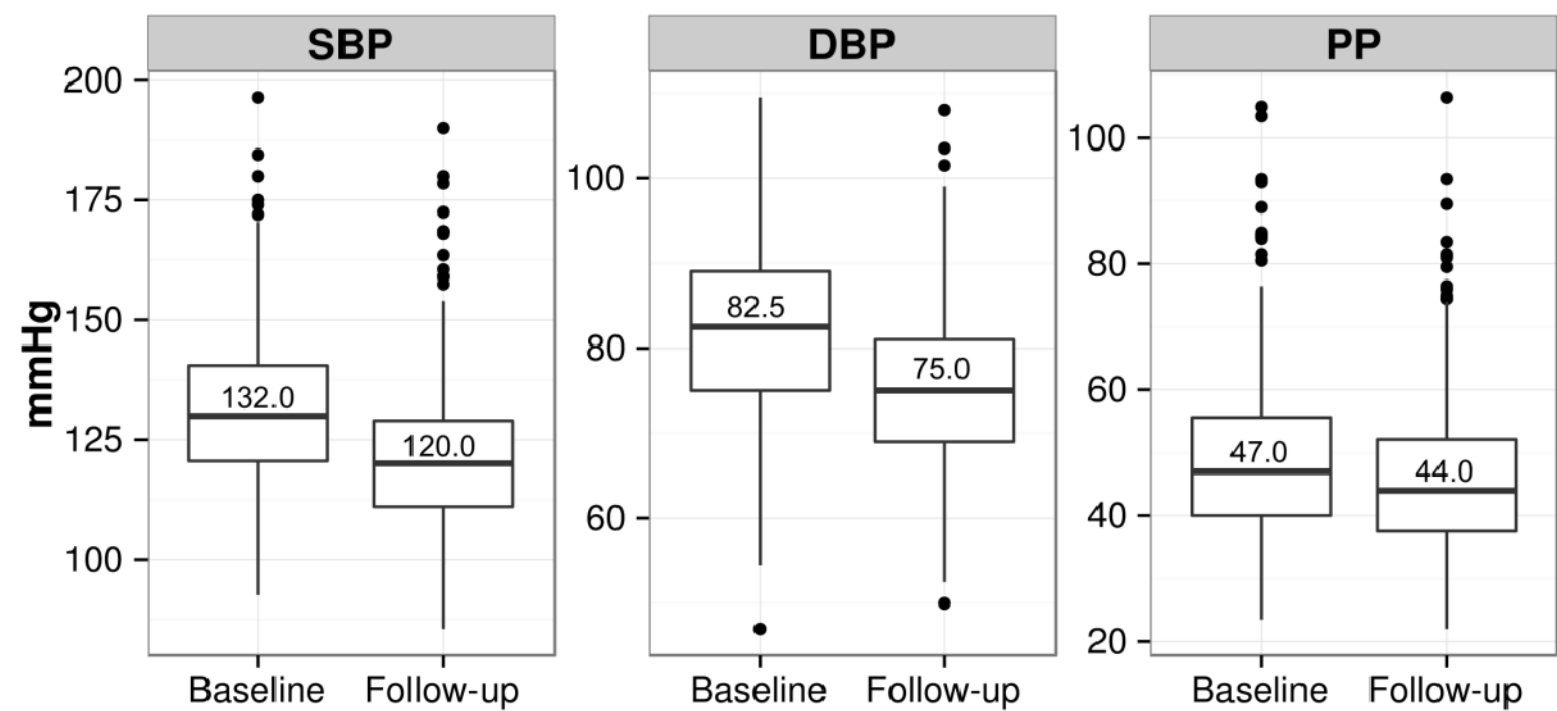


Figure S5A: Single nucleotide polymorphisms from the GRASP database used for analyses of expression quantitative trait loci. For each candidate transcript, cis-eQTLs were calculated for SNPs within $\pm 250\text{kb}$ around the transcription start site and a minor allele frequency $\geq 1\%$. The Genome-wide Repository of Associations between SNPs and Phenotypes (GRASP) database version 2.0.0.0 (<https://grasp.nhlbi.nih.gov>) was studied for significant cis-eQTLs which were at least moderately associated to a BP or a cardiovascular disease related trait ($p\text{-value} \leq 10^{-4}$).

Gene	CRIP1	MYADM	TIPARP	TSC22D3	CEBPA	F12	LMNA	TPPP3
	1,000G SNPs in candidate gene ($\pm 250\text{kb}$)							
# SNPs in gene	224	684	1209	0	1263	521	725	285
	Significant cis-eQTL after Bonferroni correction ($p\text{-value} \leq 0.05$)							
# significant eQTLs	25	0	0	0	0	0	191	30
	SNPs associated to CVD traits in the GRASP database ($p < 10^{-4}$) ?							
# significant eQTLs related to CVD traits	2	0	0	0	0	0	5	0
	<u>Mitral annular calcium (aortic valve stenosis)</u> rs10151805 ($p = 4.55 \times 10^{-5}$) rs8003942 ($p = 4.91 \times 10^{-5}$) Thanassoulis et al. NEJM 2013, PMID: 23388002				<u>Body mass index (BMI)</u> rs1475766 ($p = 6.36 \times 10^{-5}$) rs11577179 ($p = 7.48 \times 10^{-5}$) rs10908486 ($p = 7.88 \times 10^{-5}$) rs12032631 ($p = 9.68 \times 10^{-5}$) rs2364400 ($p = 9.71 \times 10^{-5}$) Speliotes et al. Nat Genet 2010, PMID: 20935630			

Figure S5B: Single nucleotide polymorphisms from the GWAS catalogue used for analyses of expression quantitative trait loci. Published GWAS results of blood pressure (BP) related traits were retrieved from the GWAS catalogue (<https://www.ebi.ac.uk/gwas/>). In total, 191 unique SNPs with genome-wide association to at least one BP trait and a minor allele frequency (MAF) $\geq 1\%$ in the Gutenberg Health Study (GHS) were found, leading to a significance level of 2.6×10^{-4} . Two SNPs within the SH2B adaptor protein 3 (SH2B3) genomic locus were identified as putative trans-regulators of *CRIP1*, *MYADM*, *TIPARP* and *TPPP3* mRNA expression in monocytes.

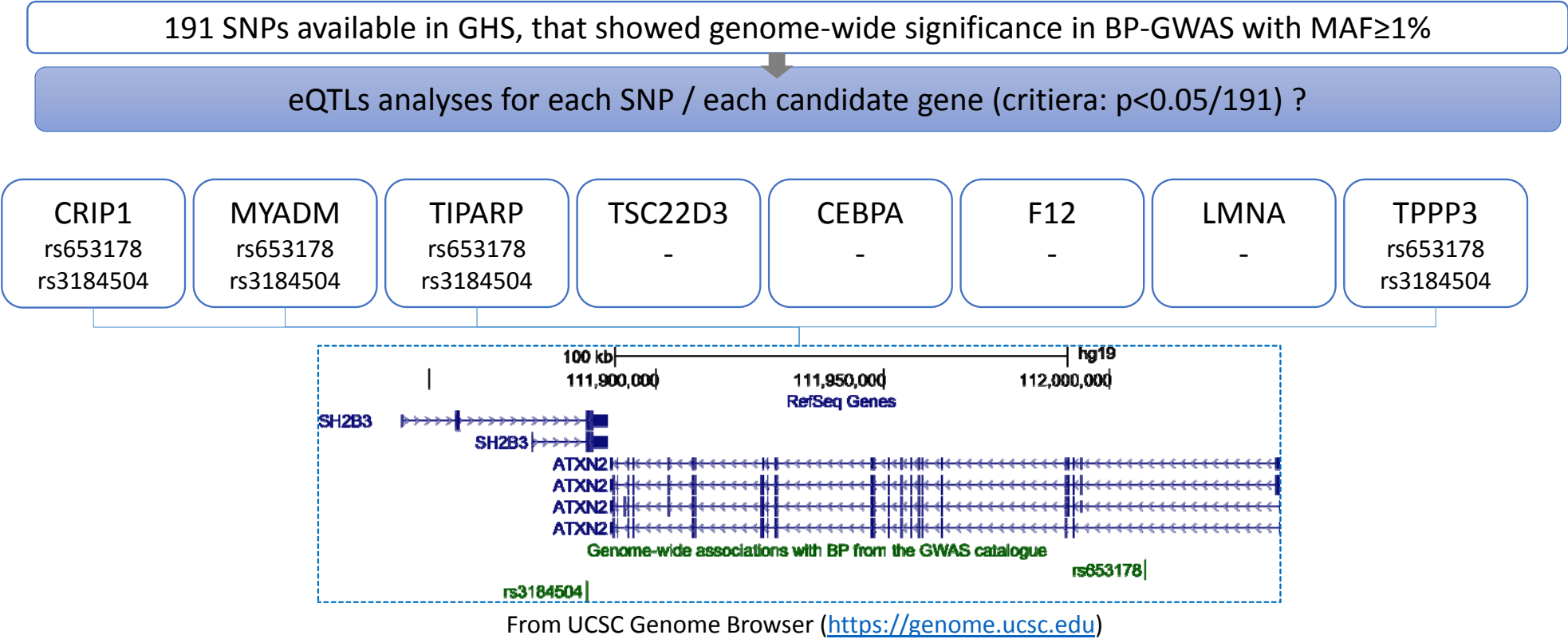


Figure S6: Median candidate gene expression in various tissues measured by RNA-sequencing in the GTEx project. RPKM: Reads Per Kilobase of transcript per Million mapped reads

