

# DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation

Melissa A. Richard,<sup>1,5,2,\*</sup> Tianxiao Huan,<sup>2,3,5,2</sup> Symen Ligthart,<sup>4,5,2</sup> Rahul Gondalia,<sup>5</sup> Min A. Jhun,<sup>6</sup> Jennifer A. Brody,<sup>7</sup> Marguerite R. Irvin,<sup>8</sup> Riccardo Marioni,<sup>9,10,11</sup> Jincheng Shen,<sup>12</sup> Pei-Chien Tsai,<sup>13</sup> May E. Montasser,<sup>14</sup> Yucheng Jia,<sup>15</sup> Catriona Syme,<sup>16</sup> Elias L. Salfati,<sup>17</sup> Eric Boerwinkle,<sup>18,19</sup> Weihua Guan,<sup>20</sup> Thomas H. Mosley, Jr.,<sup>21</sup> Jan Bressler,<sup>18</sup> Alanna C. Morrison,<sup>18</sup> Chunyu Liu,<sup>2,3,22</sup> Michael M. Mendelson,<sup>2,3,23</sup> André G. Uitterlinden,<sup>24</sup> Joyce B. van Meurs,<sup>24</sup> BIOS Consortium, Oscar H. Franco,<sup>4</sup> Guosheng Zhang,<sup>25,26,27</sup> Yun Li,<sup>25,28</sup> James D. Stewart,<sup>5,29</sup> Joshua C. Bis,<sup>7</sup> Bruce M. Psaty,<sup>30</sup> Yii-Der Ida Chen,<sup>15</sup> Sharon L.R. Kardia,<sup>6</sup> Wei Zhao,<sup>6</sup> Stephen T. Turner,<sup>31</sup>

(Author list continued on next page)

Genome-wide association studies have identified hundreds of genetic variants associated with blood pressure (BP), but sequence variation accounts for a small fraction of the phenotypic variance. Epigenetic changes may alter the expression of genes involved in BP regulation and explain part of the missing heritability. We therefore conducted a two-stage meta-analysis of the cross-sectional associations of systolic and diastolic BP with blood-derived genome-wide DNA methylation measured on the Infinium HumanMethylation450 BeadChip in 17,010 individuals of European, African American, and Hispanic ancestry. Of 31 discovery-stage cytosine-phosphate-guanine (CpG) dinucleotides, 13 replicated after Bonferroni correction (discovery:  $N = 9,828$ ,  $p < 1.0 \times 10^{-7}$ ; replication:  $N = 7,182$ ,  $p < 1.6 \times 10^{-3}$ ). The replicated methylation sites are heritable ( $h^2 > 30\%$ ) and independent of known BP genetic variants, explaining an additional 1.4% and 2.0% of the interindividual variation in systolic and diastolic BP, respectively. Bidirectional Mendelian randomization among up to 4,513 individuals of European ancestry from 4 cohorts suggested that methylation at cg08035323 (*TAF1B-YWHAQ*) influences BP, while BP influences methylation at cg00533891 (*ZMIZ1*), cg00574958 (*CPT1A*), and cg02711608 (*SLC1A5*). Gene expression analyses further identified six genes (*TSPAN2*, *SLC7A11*, *UNC93B1*, *CPT1A*, *PTMS*, and *LPCAT3*) with evidence of triangular associations between methylation, gene expression, and BP. Additional integrative Mendelian randomization analyses of gene expression and DNA methylation suggested that the expression of *TSPAN2* is a putative mediator of association between DNA methylation at cg23999170 and BP. These findings suggest that heritable DNA methylation plays a role in regulating BP independently of previously known genetic variants.

## Introduction

Elevated blood pressure (BP) confers a higher risk of heart disease, stroke, diabetes, dementia, renal failure, and pregnancy-related complications and is a leading risk factor for

death worldwide.<sup>1</sup> BP is a highly heritable trait<sup>2</sup> and recent genetic studies have revealed part of its complex genetic architecture,<sup>3–11</sup> yet the genetic variants identified to date account for only a small fraction of its phenotypic variance.<sup>3,6,8,12</sup> Complex phenotypes, such as BP, often

<sup>1</sup>Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center, Houston, TX 77030, USA; <sup>2</sup>Population Sciences Branch, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892, USA; <sup>3</sup>Framingham Heart Study, Framingham, MA 01702, USA; <sup>4</sup>Department of Epidemiology, Erasmus University Medical Center, Rotterdam 3000, the Netherlands; <sup>5</sup>Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27599, USA; <sup>6</sup>Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48108, USA; <sup>7</sup>Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98195, USA; <sup>8</sup>Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA; <sup>9</sup>Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK; <sup>10</sup>Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK; <sup>11</sup>Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia; <sup>12</sup>Department of Population Health Sciences, University of Utah School of Medicine, Salt Lake City, UT 84108, USA; <sup>13</sup>Department of Twin Research and Genetic Epidemiology, Kings College London, SE17EH London, UK; <sup>14</sup>Division of Endocrinology, Diabetes, and Nutrition, Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA; <sup>15</sup>Institute for Translational Genomics and Population Sciences, Departments of Pediatrics and Medicine, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA; <sup>16</sup>Hospital for Sick Children, University of Toronto, Toronto, ON M5G 0A4, Canada; <sup>17</sup>Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA; <sup>18</sup>Human Genetics Center, School of Public Health, University of Texas Health Science Center, Houston, TX 77030, USA; <sup>19</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; <sup>20</sup>Department of Biostatistics, University of Minnesota, Minneapolis, MN 55454, USA; <sup>21</sup>Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA; <sup>22</sup>Department of Biostatistics, Boston University, Boston, MA 02118, USA; <sup>23</sup>Department of Cardiology, Boston Children's Hospital, Boston, MA 02115, USA; <sup>24</sup>Department of Internal Medicine, Erasmus University Medical Center, Rotterdam 3000, the Netherlands; <sup>25</sup>Department of Genetics, University of North Carolina, Chapel Hill, NC 27514, USA; <sup>26</sup>Curriculum in Bioinformatics and Computational Biology, University of North Carolina, Chapel Hill, NC 27514, USA; <sup>27</sup>Department of Statistics, University of North Carolina, Chapel Hill, NC 27514, USA; <sup>28</sup>Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA; <sup>29</sup>Carolina Population Center, University of North Carolina, Chapel Hill, NC 27514, USA; <sup>30</sup>Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA 98101, USA; Kaiser Permanente Washington Health Research Unit, Seattle, WA 98101, USA; <sup>31</sup>Division of Nephrology and Hypertension, Mayo Clinic, Rochester, MN 55905, USA; <sup>32</sup>HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA; <sup>33</sup>Alzheimer Scotland

(Affiliations continued on next page)

Devin Absher,<sup>32</sup> Stella Aslibekyan,<sup>8</sup> John M. Starr,<sup>9,33</sup> Allan F. McRae,<sup>11,34</sup> Lifang Hou,<sup>35</sup> Allan C. Just,<sup>36</sup> Joel D. Schwartz,<sup>37</sup> Pantel S. Vokonas,<sup>38</sup> Cristina Menni,<sup>13</sup> Tim D. Spector,<sup>13</sup> Alan Shuldiner,<sup>14,39</sup> Coleen M. Damcott,<sup>14</sup> Jerome I. Rotter,<sup>15</sup> Walter Palmas,<sup>40</sup> Yongmei Liu,<sup>41</sup> Tomáš Paus,<sup>42,43,44</sup> Steve Horvath,<sup>45</sup> Jeffrey R. O'Connell,<sup>14</sup> Xiuqing Guo,<sup>15</sup> Zdenka Pausova,<sup>16,46</sup> Themistocles L. Assimes,<sup>17</sup> Nona Sotoodehnia,<sup>47</sup> Jennifer A. Smith,<sup>6</sup> Donna K. Arnett,<sup>48</sup> Ian J. Deary,<sup>9,49</sup> Andrea A. Baccarelli,<sup>37</sup> Jordana T. Bell,<sup>13</sup> Eric Whitsel,<sup>5,50</sup> Abbas Dehghan,<sup>4,51</sup> Daniel Levy,<sup>2,3,53</sup> and Myriam Fornage<sup>1,18,53,\*</sup>

result from the interplay between genetic and environmental influences. DNA methylation, the covalent binding of a methyl group to the 5' carbon of cytosine-phosphate-guanine (CpG) dinucleotide sequences in the genome, plays a critical role in the regulation of gene expression and may reflect a link between genes, environment, and complex phenotypes such as BP. Evidence is beginning to emerge that epigenetic modifications in genes relevant to BP may account for part of its regulation.<sup>13</sup> Variation in DNA methylation may thus explain additional phenotypic variation in BP and provide new clues to the biological processes influencing its regulation.

We conducted genome-wide DNA methylation meta-analyses for systolic and diastolic BP with a discovery phase and independent replication among 17,010 individuals of European (EA), African American (AA), and Hispanic ancestries in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. DNA methylation was measured in peripheral blood samples. We further sought to identify transcriptional changes for the replicated CpG sites and used Mendelian randomization techniques to explore the causal relationship between DNA methylation and BP. We report that the effect of DNA methylation on BP is likely independent of previously known genetic variants, representing new insights into the biological mechanisms underlying BP regulation.

## Material and Methods

### Study Populations

The discovery and replication studies were conducted in the framework of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, which comprises multiple population-based cohort studies.<sup>14</sup> Cohorts participating at the discovery stage included 9,828 individuals of EA and AA

ancestries in the Atherosclerosis Risk in Communities (ARIC) study, Cardiovascular Health Study (CHS), Framingham Heart Study (FHS), Genetic Epidemiology Network of Arteriopathy (GENOA) study, Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study, Lothian Birth Cohort 1936 (LBC1936), Normative Aging Study (NAS), Rotterdam Study (RS), and TwinsUK registry. Cohorts participating at the replication stage consisted of 7,182 additional individuals of EA, AA, and Hispanic ancestries in the Amish Complex Disease Research Studies (Amish), ARIC, the Multi-Ethnic Study of Atherosclerosis (MESA), RS, adults in the Saguenay Youth Study (SYS), and the Women's Health Initiative (WHI). Details for each cohort are provided in the [Supplemental Data](#). All studies obtained written informed consent from participants and were approved by local institutional review boards and ethics committees.

### Blood Pressure Measurements

Epigenome-wide association studies (EWASs) were conducted for systolic and diastolic BP, in mmHg. In each cohort, BP was measured in a sitting position after a period of rest and an average of sequential readings was used as the phenotype for each analysis. For most cohorts, BP was measured concurrently at the time of tissue collection for DNA methylation profiling, or in as close proximity as available for TwinsUK (0.8 years) and SYS adults (3.1 years). To adjust for the use of antihypertensive medication, we used the standard adjustment of adding 15 mmHg and 10 mmHg to measured systolic and diastolic BPs, respectively, when the use of any antihypertensive medications were self-reported.

### DNA Methylation Profiling

DNA methylation was measured on the Infinium HumanMethylation450 (450k) BeadChip (Illumina) in all cohorts using whole-blood samples, excepting that GOLDN measured DNA methylation in CD4<sup>+</sup> T cells. To correct the beta value distributions of the two types of probes on the 450k array, each cohort normalized methylation beta values using BMIQ,<sup>15</sup> DASEN,<sup>16</sup> ComBat,<sup>17</sup>

Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, UK; <sup>34</sup>Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia; <sup>35</sup>Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA; <sup>36</sup>Department of Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; <sup>37</sup>Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA; <sup>38</sup>Veterans Affairs Normative Aging Study, Veterans Affairs Boston Healthcare System, Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA; <sup>39</sup>The Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY 10591, USA; <sup>40</sup>Department of Medicine, Columbia University, New York, NY 10032, USA; <sup>41</sup>Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; <sup>42</sup>Departments of Psychology and Psychiatry, University of Toronto, Toronto, ON M5S 3G3, Canada; <sup>43</sup>Rotman Research Institute, Baycrest, Toronto, ON M6A 2E1, Canada; <sup>44</sup>Child Mind Institute, New York, NY 10022, USA; <sup>45</sup>Department of Human Genetics, Gonda Research Center, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA; <sup>46</sup>Departments of Physiology and Nutritional Sciences, University of Toronto, Toronto, ON M5S 1A8, Canada; <sup>47</sup>Cardiovascular Health Research Unit, Division of Cardiology, University of Washington, Seattle, WA 98195, USA; <sup>48</sup>University of Kentucky, College of Public Health, Lexington, KY 40563, USA; <sup>49</sup>Department of Psychology, University of Edinburgh, Edinburgh EH9 8JZ, UK; <sup>50</sup>Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA; <sup>51</sup>Department of Biostatistics and Epidemiology, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, UK

<sup>52</sup>These authors contributed equally to this work

<sup>53</sup>These authors contributed equally to this work

\*Correspondence: [melissa.a.lee@uth.tmc.edu](mailto:melissa.a.lee@uth.tmc.edu) (M.A.R.), [myriam.fornage@uth.tmc.edu](mailto:myriam.fornage@uth.tmc.edu) (M.F.)

<https://doi.org/10.1016/j.ajhg.2017.09.028>

SWAN,<sup>18</sup> or quantile normalization. LBC1936 did not normalize methylation beta values prior to analyses.

### Cohort-Level Association Analyses

In each cohort, race-stratified linear mixed effect models were used to estimate associations adjusting for age, sex (in samples including men and women), blood cell counts, body mass index, smoking (current/former/never), and ancestry, as well as fixed and/or random effects for technical covariates to control for batch effects. Surrogate variables were calculated and adjusted in the modeling for ARIC AAs and FHS due to batch effects not controlled by other modeling techniques. The Amish, FHS, GOLDN, and TwinsUK accounted for sample relatedness in all analyses. Study-specific modeling details can be found in the [Supplemental Data](#).

### Epigenome-wide Meta-Analyses

Effect estimates from all cohorts were combined using inverse variance fixed effects meta-analysis using GWAMA.<sup>19</sup> We assessed heterogeneity of effect estimates between strata of races, sexes, and methylation tissue source among discovery cohorts using a 1 degree of freedom chi-square test for effect differences between strata; no heterogeneous effects were observed so all cohorts were included in a single meta-analysis. Meta-analyses were conducted separately for the discovery and replication cohorts to identify probes associated with BP. Statistical significance was Bonferroni corrected for the epigenome-wide discovery meta-analysis ( $p < 1.0 \times 10^{-7}$ ) and the number of discovery CpG sites sought for replication in the second meta-analysis. An overall meta-analysis was additionally performed to combine effect estimates across all cohorts. Significant CpG sites were annotated using information provided by Illumina, including chromosome, position (GRCh37/hg19), UCSC gene names, relationship to CpG islands, location in gene enhancer regions, and DNase I hypersensitivity sites (DHS). To assess the impact of antihypertensive medication use on our top findings, we additionally performed an overall meta-analysis among all individuals reporting no use of antihypertensive medications. For the top findings in the discovery meta-analysis, we compared effect and standard error estimates to those estimated in the non-medicated meta-analysis.

### Percent Variance Explained

Percent variance explained was calculated in the ARIC AA and EA samples included in discovery and replication meta-analyses, as well as validated in a sample from the FHS Third Generation not included in the meta-analysis ( $N = 1,516$ ). Methylation profile scores for BP were calculated as the weighted sum of CpG sites significant for either BP trait in the replication and overall meta-analyses, with weights coming from the magnitude and direction of effects in the overall meta-analysis. Selection of CpGs from meta-analyses including the prediction samples could overestimate percent variance explained, so additional meta-analyses were conducted excluding the ARIC samples to identify CpGs for their respective methylation profile scores. The probe sets based on exclusion of the ARIC samples and the probes identified in the primary replication and overall meta-analyses were used to generate methylation profile scores in the FHS sample. Race- and cohort-stratified linear regression models were used to estimate the percent of age-, sex-, and BMI-adjusted systolic and diastolic BP variances explained by each methylation profile score; ARIC models were additionally adjusted for visit and study site, and ARIC AA and FHS models included surrogate variables. Percent

variance explained by the methylation profile scores is reported as the adjusted  $R^2$  from each model and compared to models without methylation profile scores (covariate-only models). We additionally assessed genetic risk scores derived using effect estimates from the UK Biobank for 146 previously reported independent variants ( $r^2 < 0.2$ ) and 115 validated novel variants<sup>11</sup> among the FHS Third Generation sample with available genetic data ( $N = 1,421$ ).

### Heritability

The narrow-sense heritability estimate of a DNA methylation trait ( $\beta$  score) (denoted as  $h^2_{CpG\_methy}$ ) was the proportion of the additive polygenic genetic variance of the total phenotypic variance of a DNA methylation trait:  $h^2_{CpG\_methy} = \sigma^2_A / \sigma^2_{CpG\_methy}$ , where  $\sigma^2_A$  denotes the additive polygenic genetic variance and  $\sigma^2_{CpG\_methy}$  denotes the total phenotypic variance of a DNA methylation trait. Heritability estimation for all DNA methylation traits was performed using the FHS-Offspring participants ( $N = 2,377$ ).

### Functional Tissue and Gene Set Enrichment Analyses

Functional DNA elements regulated by methylation may be tissue specific, so the set of replicated CpGs was used to identify tissue- and cell type-specific signals using experimentally derived Functional element Overlap analysis of ReGions from EWAS (eFORGE).<sup>20</sup> After pruning results for CpG sites within 1 kb (2 probes removed), we matched the top 11 EWAS signals for overlap with DNase I hypersensitive sites using data from ENCODE and Roadmap Epigenomics. 1,000 matched sets were used with the 450k array as the background set. FDR correction was applied to the results.

Gene Set Enrichment Analysis (GSEA)<sup>21</sup> was conducted on the results of the overall meta-analyses for systolic and diastolic BP. For each gene annotated to DNA methylation measured on the 450k array, a composite ranking for BP was generated based on the CpG site with the minimum p value for either trait. All gene ontology biological process categories (c5.bp.v5.1) were assessed for enrichment at  $FDR Q < 0.05$ .

### Methylation Quantitative Trait Loci

To determine methylation levels at CpG sites that may be influenced by nearby DNA sequence, methylation quantitative trait loci (meQTL) analyses were performed for the 13 replicated BP CpGs in EA individuals from ARIC ( $N = 948$ ), FHS ( $N = 2,357$ ), and RS ( $N = 731$ ) and AA individuals from ARIC ( $N = 2,173$ ) and GENOA ( $N = 422$ ). Residuals were obtained from regressing inverse-normal transformed methylation beta values on the first ten methylation principal components (PCs) and up to the first ten genetic PCs. The residuals were then regressed on 1000 Genomes Phase I imputed SNPs within 50 kb of the probe (CpG position  $\pm 25$  kb, GRCh37/hg19). SNPs with low imputation quality ( $r^2 < 0.3$ ), low frequency variants ( $MAF < 0.05$ ), and SNPs present in only one cohort were removed from analyses. Results for each probe were combined using race-stratified p value-based meta-analysis weighted by sample size and direction of effects using METAL.<sup>22</sup> Significant meQTLs were determined using a Bonferroni correction for all meQTLs tested in each race (EA:  $0.05/1,447 = 3.5 \times 10^{-5}$ ; AA:  $0.05/1,952 = 2.6 \times 10^{-5}$ ). To maximize statistical power for identifying meQTLs associated with BP, we then searched the largest genome-wide association studies (GWASs) for BP in each race for suggestive association of meQTL regions with BP.

To assess the association of SNPs reported by Kato et al.<sup>23</sup> whose association may be mediated by DNA methylation, we additionally performed meQTL analyses for 35 sentinel SNPs and additional GWAS loci in high linkage disequilibrium (LD) with these regions.<sup>3–5,23–30</sup> We assessed the association of DNA methylation within 1 Mb (CpG position  $\pm$  500 kb) of GWAS SNPs among ARIC EAs (N = 790) using the previously described methodology. SNPs associated with methylation after Bonferroni correction for the 28 meQTLs reported by Kato et al.<sup>23</sup> ( $p < 0.0018$ ) were then assessed for association with BP before and after adjustment for methylation at the CpG site. We additionally assessed the association of these CpG sites with BP in our overall meta-analysis.

### Bidirectional Mendelian Randomization

To assess the directional association of DNA methylation and BP, we conducted bidirectional Mendelian randomization (MR) using 1000 Genomes imputed SNPs among EA individuals in ARIC, FHS, RS, and WHI-EMPC (N = 4,513). Forward MR was used to identify replicated CpG sites which may have an effect on BP. Instrumental variables (IVs) for DNA methylation were drawn from the meQTLs estimated among EAs and pruned for independence ( $r^2 < 0.2$ ). Forward MR was conducted for the six sentinel CpG sites with at least three independent meQTLs, which is the minimum number of IVs needed to perform multi-instrument MR. Reverse MR was used to identify DNA methylation at the 11 sentinel CpG sites that may be caused by BP. The 29 independent loci reported as associated with BP by the International Consortium for Blood Pressure (ICBP) were selected as IVs. The SNP rs805303 was not imputed in 1000 Genomes and rs805301 was used as a proxy when available ( $r^2 = 1.0$  in HapMap).

Each cohort estimated the associations of IVs with systolic BP, diastolic BP, and DNA methylation at the respective CpG sites. Cohort-level effect estimates for each IV were combined using inverse variance-weighted meta-analyses in METAL.<sup>22</sup> For each CpG in forward and reverse MR, causation was formally tested based on the inverse variance-weighted effects across all IV-BP and IV-CpG estimates using the R package *MendelianRandomization*.<sup>31</sup> Tests for causation with  $p$  value  $< 0.05$  were considered significant. To ensure the validity of the inverse-variance weighted approach, the IVs were assessed for pleiotropy using the MR-Egger test. Inverse-variance weighted MR is invalid in the presence of pleiotropic effects of IVs, so Egger regression estimates of causality were assessed only when pleiotropy was indicated at a particular CpG site.

### Associations of DNA Methylation and Gene Expression

Association tests of BP-associated CpGs with transcripts that were located within  $\pm 1$  Mb distance of the corresponding CpGs were performed in 2,216 FHS-Offspring samples and 730 RS samples whose DNA methylation and gene expression data were both available. In FHS, linear mixed effect regression models were used with DNA methylation  $\beta$  scores as the dependent variable, gene expression as independent variables, age, sex, and technical covariates as fixed effects, and family structure as a random effect. In RS, we first created residuals for both DNA methylation and mRNA expression after regressing out age, sex, blood cell counts (fixed effect), and technical covariates (random effect). We then examined the association between the residuals of DNA methylation (independent variable) and mRNA expression (dependent variable) using a linear regression model. Estimates of the gene expression-methylation associations in RS and FHS were com-

puted using sample size weighted fixed effects meta-analysis based on  $p$  values and direction of effects using GWAMA.<sup>19</sup>

### Associations of Gene Expression and BP

Differential gene expression analysis of the transcripts assessed for association with DNA methylation were performed for systolic BP, diastolic BP, and hypertension in 3,679 FHS Offspring and 3<sup>rd</sup>-Gen participants who were not receiving anti-hypertensive treatments. Hypertension was defined as systolic BP  $\geq 140$  mmHg or diastolic BP  $\geq 90$  mmHg. See details in Huan et al.<sup>32</sup>

### Two-Step Mendelian Randomization for Relationship of DNA Methylation, Gene Expression, and BP

To identify gene transcription that functionally mediates the relationship of DNA methylation and BP, we performed a two-step MR technique for genes with expression associated with both DNA methylation and BP (FDR  $Q < 0.05$ ). The first step was to establish a directional relationship between DNA methylation and gene transcription. IVs for DNA methylation were drawn from estimated meQTLs pruned to be independent ( $r^2 < 0.2$ ). Using whole-blood eQTLs estimated in the Genotype-Tissue Expression (GTEx) project, we verified the association of each IV with the implicated gene expression. In the second step, IVs for each implicated gene were selected from the GTEx whole-blood dataset in order to establish a directional relationship between gene expression and BP. The top eQTL also present in the ICBP results was selected as the IV for each gene and assessed for association with systolic and diastolic BP in ICBP published GWAS results. Genes with  $p < 0.05$  at both steps were considered to mediate a directional relationship of the respective CpG and BP; correction for multiple testing is not used because strong associations of IVs with an outcome would violate the assumptions of Mendelian randomization.

## Results

### Cohort Characteristics

Characteristics of the 14 studies participating in discovery and replication meta-analyses are presented in Table 1. Each cohort included middle-aged and older adults with a wide range of BP values. Mean systolic BP ranged from 116 mmHg in GOLDN to 152 mmHg among CHS AAs. Mean diastolic BP ranged from 68 mmHg in GOLDN to 89 mmHg in the RS replication sample. Prevalence of anti-hypertensive medication use varied with cohort age and health, with no use among the Amish to more than 62% among the CHS AA sample.

### Identification of Epigenome-wide CpG Sites Associated with Blood Pressure

In the discovery stage, we conducted genome-wide associations of DNA methylation with systolic and diastolic BP in nine cohort studies (N = 9,828). Multiethnic meta-analyses identified methylation at 31 CpG sites associated with BP after Bonferroni correction for the number of DNA methylation CpG sites measured on the Illumina 450K array ( $p < 1.0 \times 10^{-7}$ ; Table S1, Figures S1 and S2). Replication of the 31 discovery CpG sites was sought in

**Table 1. Characteristics of the Discovery and Replication Cohorts**

Cohort	Race	n	Cohort Type	Tissue	Normalization	Age, years		SBP, mmHg		DBP, mmHg		HTN	AHT
						Mean	SD	Mean	SD	Mean	SD	%	%
<b>Discovery (N = 9,828)</b>													
ARIC	AA	2,743	unrelated	blood	BMIQ	56.6	5.9	135.0	23.4	80.2	12.4	65.5	48.9
CHS	AA	196	unrelated	blood	SWAN	73.0	5.4	151.5	23.9	83.2	12.5	78.6	62.2
CHS	EA	189	unrelated	blood	SWAN	76.0	5.1	142.8	23.9	76.7	10.9	65.6	49.2
FHS	EA	2,645	family	blood	DASEN	66.4	8.9	128.6	17.2	73.4	10.0	59.0	49.0
GENOA	AA	239	unrelated	blood	SWAN	60.1	8.4	146.1	25.6	82.5	12.4	72.0	58.2
GOLDN	EA	822	family	CD4 <sup>+</sup> T cells	ComBat	48.8	15.9	115.7	16.3	68.4	9.4	25.7	21.0
LBC1936	EA	903	unrelated	blood	–	69.5	0.8	149.4	19.0	81.3	10.1	40.7	43.0
NAS	EA	674	unrelated	blood	BMIQ	72.5	6.8	139.5	18.9	81.9	10.3	71.0	58.2
RS-III	EA	727	unrelated	blood	DASEN	59.7	8.2	138.9	22.0	85.8	12.5	53.2	30.1
TwinsUK	EA	690	twins	blood	BMIQ	58.4	9.3	126.0	16.6	77.2	9.8	25.7	21.5
<b>Replication (N = 7,182)</b>													
Amish	EA	192	family	blood	quantile	46.3	13.6	117.8	12.7	72.4	8.1	2.0	0.0
ARIC	EA	1,058	unrelated	blood	BMIQ	59.8	5.4	121.2	20.5	70.1	11.1	29.7	17.6
MESA	AA	236	unrelated	blood	quantile	60.6	9.2	127.5	19.6	73.3	9.5	55.2	48.0
MESA	EA	566	unrelated	blood	quantile	60.8	9.6	121.0	18.5	70.1	9.6	36.8	31.7
MESA	HL	381	unrelated	blood	quantile	59.0	9.5	122.6	18.4	72.0	9.3	37.8	31.3
RS-III	EA	711	unrelated	blood	DASEN	67.5	6.0	151.3	24.0	88.7	13.0	71.5	43.3
SYS adults	EA	111	unrelated	blood	SWAN	47.2	4.9	131.5	15.3	79.5	8.4	29.7	8.1
WHI-BAA23	AA	666	unrelated	blood	ComBat	62.8	6.7	140.9	21.1	83.3	10.9	65.0	54.7
WHI-BAA23	EA	965	unrelated	blood	ComBat	68.4	6.2	136.5	21.1	78.2	11.1	48.5	34.6
WHI-BAA23	HL	333	unrelated	blood	ComBat	62.3	6.8	133.3	20.7	78.6	10.8	47.3	35.2
WHI-EMPC	AA	556	unrelated	blood	BMIQ	62.8	7.0	131.5	18.1	77.4	9.6	60.4	55.2
WHI-EMPC	EA	1,092	unrelated	blood	BMIQ	64.7	7.1	127.5	17.7	74.5	9.4	42.9	30.5
WHI-EMPC	HL	315	unrelated	blood	BMIQ	61.6	6.2	127.2	18.2	74.8	9.5	41.9	29.5

Hypertension is defined as systolic BP  $\geq$  140 mmHg or diastolic BP  $\geq$  90 mmHg or the use of antihypertensive treatment. Antihypertensive treatment is defined as the self-reported use of any antihypertensive medication. WHI-EMPC normalized DNA methylation data using BMIQ and plate-adjusted using ComBat. The discovery and replication samples from RS-III do not include overlapping or related individuals. Abbreviations: AA, African American; AHT, antihypertensive treatment; BMIQ, Beta Mixture Quantile dilation; ComBat, combatting batch effects when COMbining BATches of microarray data; DASEN, background-adjusted (D) between-array (S) without dye bias correction (N); DBP, diastolic blood pressure; EA, European ancestry; HL, Hispanic/Latino; HTN, hypertension; SBP, systolic blood pressure; SD, standard deviation; SWAN, Subset-quantile Within Array Normalization.

multiethnic meta-analyses of an additional six cohort studies (N = 7,182). Methylation at 13 of the 31 discovery CpG sites was associated with BP at  $p < 0.0016$  in the replication meta-analysis (0.05/31; Table 2). A schematic of the overall study design, including subsequent integrative analyses, is found in Figure S3.

The top two CpG sites for both systolic and diastolic BP were at the *PHGDH* locus, cg14476101 (systolic BP: coefficient = 0.03% decrease in DNA methylation per 1 mmHg increase in BP,  $p = 2.7 \times 10^{-34}$ ; diastolic BP: coefficient = 0.04% decrease in DNA methylation per 1 mmHg increase in BP,  $p = 2.1 \times 10^{-21}$ ), and the *SLC7A11* locus, cg06690548 (systolic BP: coefficient = 0.02% decrease in DNA methylation per 1 mmHg increase in BP,  $p = 1.6 \times$

$10^{-32}$ ; diastolic BP: coefficient = 0.03% decrease in DNA methylation per 1 mmHg increase in BP,  $p = 7.9 \times 10^{-26}$ ). cg14476101 is located on chromosome 1p12 in the first intron of *PHGDH*, which encodes a phosphoglycerate dehydrogenase that catalyzes the rate-limiting step of serine biosynthesis. Located on chromosome 4q28.3, cg06690548 is in the first intron of *SLC7A11*, which encodes a sodium-independent cysteine/glutamate antiporter. All replicated CpG sites demonstrated associations of decreased DNA methylation with increases in BP (Table S1 and Figure S4). None of the replicated CpG sites cross-hybridized with sequence variation on the sex chromosomes, and one CpG, *SLC1A5* cg02711608, is polymorphic.<sup>33</sup> An additional CpG site in *SLC1A5*, cg22304262,

**Table 2. Results of Discovery, Replication, and Overall Meta-analyses for CpG Sites Replicated for Association with BP**

CpG site	Chr	Position	UCSC Gene	Systolic BP						Diastolic BP					
				Discovery		Replication		Overall		Discovery		Replication		Overall	
				Coeff	p Value	Coeff	p Value	Coeff	p Value	Coeff	p Value	Coeff	p Value	Coeff	p Value
cg23999170	1	115628111	TSPAN2	-0.0001	$2.7 \times 10^{-6}$	-0.0001	$1.6 \times 10^{-5}$	-0.0001	$1.5 \times 10^{-10}$	-0.0002	$6.4 \times 10^{-8}$	-0.0002	$3.4 \times 10^{-7}$	-0.0002	$1.9 \times 10^{-13}$
cg16246545	1	120255941	PHGDH	-0.0002	$2.4 \times 10^{-10}$	-0.0002	$3.3 \times 10^{-14}$	-0.0002	$1.2 \times 10^{-22}$	-0.0002	$2.2 \times 10^{-4}$	-0.0003	$4.3 \times 10^{-7}$	-0.0002	$1.1 \times 10^{-9}$
cg14476101	1	120255992	PHGDH	-0.0003	$1.5 \times 10^{-16}$	-0.0004	$7.0 \times 10^{-21}$	-0.0003	$2.7 \times 10^{-34}$	-0.0004	$6.0 \times 10^{-11}$	-0.0005	$1.9 \times 10^{-12}$	-0.0004	$2.1 \times 10^{-21}$
cg19693031	1	145441552	TXNIP	-0.0002	$7.7 \times 10^{-13}$	-0.0003	$3.8 \times 10^{-19}$	-0.0002	$3.1 \times 10^{-29}$	-0.0002	$6.0 \times 10^{-7}$	-0.0004	$7.5 \times 10^{-10}$	-0.0003	$1.8 \times 10^{-14}$
cg08035323	2	9843525	-	-0.0001	$4.2 \times 10^{-5}$	-0.0001	$4.1 \times 10^{-3}$	-0.0001	$9.6 \times 10^{-7}$	-0.0003	$1.4 \times 10^{-8}$	-0.0002	$2.6 \times 10^{-4}$	-0.0003	$2.6 \times 10^{-11}$
cg06690548	4	139162808	SLC7A11	-0.0001	$3.4 \times 10^{-16}$	-0.0002	$8.3 \times 10^{-20}$	-0.0002	$1.6 \times 10^{-32}$	-0.0002	$5.5 \times 10^{-14}$	-0.0003	$9.9 \times 10^{-14}$	-0.0003	$7.9 \times 10^{-26}$
cg18120259	6	43894639	LOC100132354	-0.0001	$1.5 \times 10^{-8}$	-0.0002	$9.4 \times 10^{-15}$	-0.0002	$2.2 \times 10^{-21}$	-0.0002	$1.9 \times 10^{-5}$	-0.0003	$6.9 \times 10^{-10}$	-0.0002	$8.9 \times 10^{-14}$
cg00533891	10	80919242	ZMIZ1	-0.0001	$2.4 \times 10^{-7}$	-0.0001	$3.7 \times 10^{-3}$	-0.0001	$5.5 \times 10^{-9}$	-0.0003	$4.4 \times 10^{-9}$	-0.0002	$8.9 \times 10^{-4}$	-0.0002	$2.0 \times 10^{-11}$
cg17061862	11	9590431	-	-0.0001	$6.9 \times 10^{-5}$	-0.0002	$6.6 \times 10^{-9}$	-0.0001	$9.4 \times 10^{-12}$	-0.0003	$5.1 \times 10^{-8}$	-0.0003	$1.2 \times 10^{-6}$	-0.0003	$4.3 \times 10^{-13}$
cg00574958	11	68607622	CPT1A	-0.0001	$1.9 \times 10^{-8}$	$-4.8 \times 10^{-5}$	$1.4 \times 10^{-6}$	-0.0001	$1.2 \times 10^{-13}$	-0.0001	$5.9 \times 10^{-7}$	-0.0001	$2.5 \times 10^{-4}$	-0.0001	$3.0 \times 10^{-10}$
cg10601624	12	6404377	-	-0.0001	$6.6 \times 10^{-8}$	-0.0001	$1.6 \times 10^{-10}$	-0.0001	$2.4 \times 10^{-16}$	-0.0001	$3.5 \times 10^{-7}$	-0.0002	$1.7 \times 10^{-7}$	-0.0002	$4.3 \times 10^{-13}$
cg22304262	19	47287778	SLCIA5	-0.0001	$5.4 \times 10^{-10}$	-0.0001	$8.7 \times 10^{-9}$	-0.0001	$1.4 \times 10^{-17}$	-0.0002	$6.0 \times 10^{-7}$	-0.0002	$4.9 \times 10^{-5}$	-0.0002	$9.6 \times 10^{-11}$
cg02711608	19	47287964	SLCIA5	-0.0001	$3.0 \times 10^{-11}$	-0.0001	$1.1 \times 10^{-11}$	-0.0001	$2.0 \times 10^{-21}$	-0.0002	$3.2 \times 10^{-5}$	-0.0002	$3.0 \times 10^{-6}$	-0.0002	$4.3 \times 10^{-10}$

Position is Hg19. Coefficients give the percent change in DNA methylation for every 1 mmHg change in blood pressure. Abbreviations: BP, blood pressure; Chr, chromosome; Coeff, coefficient; CpG, cytosine-phosphate-guanine; UCSC, University of California Santa Cruz.

**Table 3. Narrow-Sense Heritability Estimated in the FHS for CpG Sites Replicated for Association with BP**

CpG Site	Chr	Position	Gene	CpG $h^2$	(95% CI)
cg23999170	1	115628111	<i>TPAN2</i>	0.45	(0.39, 0.50)
cg16246545	1	120255941	<i>PHGDH</i>	0.47	(0.41, 0.55)
cg14476101	1	120255992	<i>PHGDH</i>	0.53	(0.43, 0.63)
cg19693031	1	145441552	<i>TXNIP</i>	0.55	(0.47, 0.63)
cg08035323	2	9843525	-	0.65	(0.57, 0.73)
cg06690548	4	139162808	<i>SLC7A11</i>	0.35	(0.27, 0.44)
cg18120259	6	43894639	<i>LOC100132354</i>	0.32	(0.26, 0.38)
cg00533891	10	80919242	<i>ZMIZ1</i>	0.54	(0.47, 0.63)
cg17061862	11	9590431	-	0.54	(0.46, 0.62)
cg00574958	11	68607622	<i>CPT1A</i>	1.00	(0.95, 1.05)
cg10601624	12	6404377	-	0.30	(0.27, 0.34)
cg22304262	19	47287778	<i>SLC1A5</i>	0.46	(0.39, 0.52)
cg02711608	19	47287964	<i>SLC1A5</i>	0.31	(0.28, 0.35)

Epigenome-wide average heritability is 0.12. Position is Hg19. Abbreviations: Chr, chromosome; CpG, cytosine-phosphate-guanine.

was also associated with BP and not polymorphic, so we did not exclude cg02711608 from our results. Narrow-sense heritability estimates of the 13 replicated CpG sites are moderate to high ( $h^2 = 30\%–100\%$ ) relative to all epigenome-wide probes (average  $h^2 = 12\%$ ; Table 3). Of the 13 replicated CpG sites, 4 are in DNase I hypersensitivity sites and enhancer regions (Table S2). In *PHGDH* and *SLC1A5*, we identified two nearby CpG sites in each gene associated with BP. We regard cg14476101 as the sentinel CpG site in *PHGDH* and cg02711608 as the sentinel CpG site in *SLC1A5* due to the strength of association p value with BP. Methylation levels at the two CpG sites in *PHGDH* were strongly correlated (AA and EA  $\rho = 0.85$ ), whereas the two CpG sites in *SLC1A5* were only modestly correlated (AA:  $\rho = 0.24$ , EA:  $\rho = 0.37$ ; Figure S5). Heterogeneity (Cochran's Q) that may be attributable to cell type or race was observed in the discovery panel for *SLC7A11* cg06690548 (Table S3); however, estimates in the replication panel for this CpG site were homogeneous with the same direction of effect and similar magnitude of association p value as in the discovery meta-analyses (Table 2). All other reported CpG sites showed homogeneous effects in discovery and replication meta-analyses.

We additionally conducted an overall meta-analysis of the discovery and replication cohorts and identified 126 CpG sites associated with BP after Bonferroni correction ( $p < 1.0 \times 10^{-7}$ ; Table S4). To assess the effects of antihypertensive medication use, we performed epigenome-wide meta-analyses among the 9,894 individuals reporting no concurrent use of antihypertensive medications in the discovery and replication samples. This combined sample free from antihypertensive medication use is of comparable size to the discovery meta-analysis. We did not identify

a large difference in effect estimates among the discovery CpG sites that met our strict replication standards (Figure S6). Many replicated CpGs were also epigenome-wide significant in the non-medicated analysis and three CpG sites on chromosome 10p15.1 were identified that were not significant in the discovery stage (Table S5). These CpG sites map to the first intron of *PFKFB3*, which encodes a glycolytic enzyme.

### Percent Variance Explained

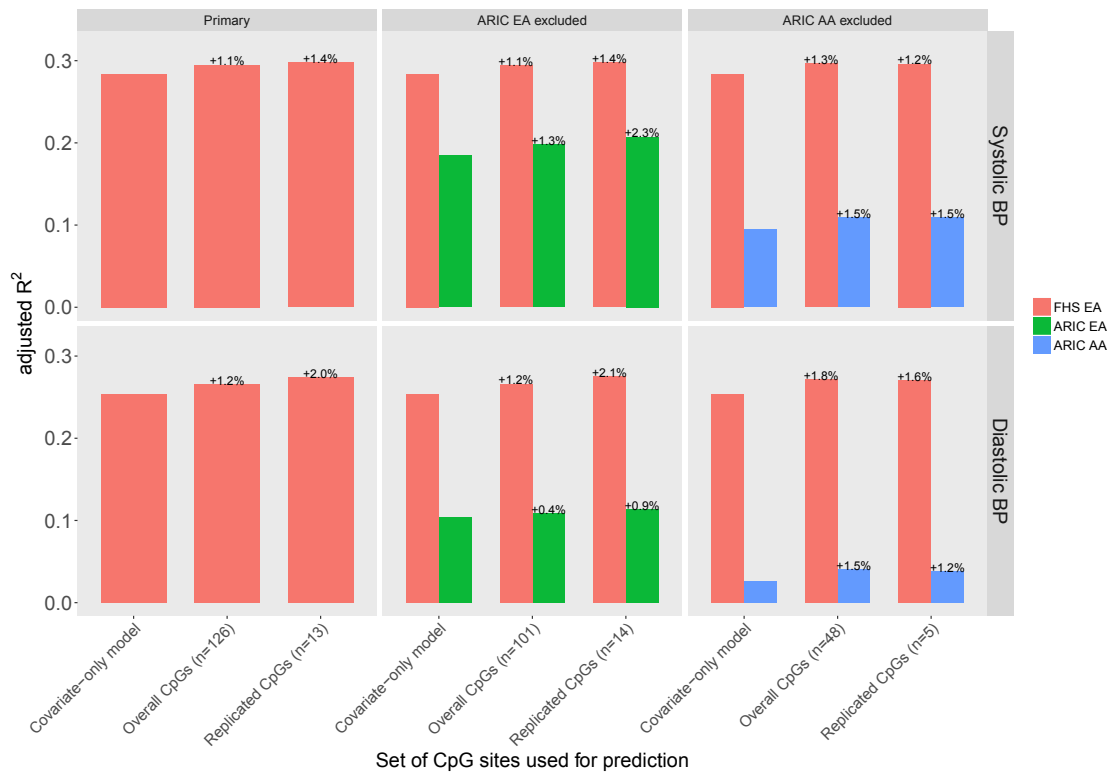
A methylation profile score based on the replicated CpG sites explained an additional 1.4% and 2.0% of the interindividual variation in systolic and diastolic BP, respectively, beyond the traditional BP covariates of age, sex, and BMI in an additional sample set from the FHS (N = 1,516, Third Generation Cohort) not included in the discovery or replication meta-analyses (Figure 1). Expanding the DNA methylation risk score to include the 126 CpG sites that were significant in the overall meta-analysis did not explain additional phenotypic variance in samples of either ancestry. Up to 261 BP-associated genetic variants explained minimal variance in the FHS Third Generation sample set (N = 1,421; PVE = 0.003%–0.1%). We elected to report only percent variances explained for methylation risk scores since our estimates are independent of the distally located known genetic loci.

### Functional Tissue and Gene Set Enrichment Analyses

Tissues enriched for DNase I hypersensitive sites in regions of the replicated CpGs include blood cells, vascular tissues, brain tissues, and cardiac tissues (Figure S7). Gene set enrichment analysis (GSEA) was conducted for intragenic CpG sites identified in the overall meta-analyses for systolic and diastolic BP. DNA methylation associated with either BP trait mapped to genes involved in the transport of neutral amino acids (FDR Q = 0.01; Figure S8). The transport of neutral amino acids was also identified as significantly enriched in the individual meta-analyses for systolic and diastolic BP (FDR Q < 0.05). 43 biological processes reached FDR Q < 0.25 including brain development, hematopoietic or lymphoid organ development, and the transport of amino acids and amines.

### Methylation Quantitative Trait Loci

We assessed genetic determinants of DNA methylation at the 13 replicated CpG sites in 4,036 EA individuals and 2,595 AA individuals in ARIC, FHS, GENOA, and RS. Of the 13 CpG sites, 9 demonstrated substantial evidence for methylation quantitative trait loci (meQTLs) in both ancestries (EA  $p < 3.5 \times 10^{-5}$ , AA  $p < 2.6 \times 10^{-5}$ ), with evidence for weak meQTLs at one additional CpG site in each ancestry (Figure 2). We confirmed our estimated EA meQTLs in an independent EA dataset published by ARIES<sup>34</sup> and found almost all estimated meQTLs were significant or in linkage disequilibrium ( $r^2 > 0.2$  or  $D' = 1$ ) with ARIES meQTLs. We assessed the association of EA meQTLs with BP in 1000 Genomes analysis by the



**Figure 1. Percent Variance Explained by Traditional Covariates and Methylation Profile Scores for Systolic and Diastolic BP**

The plot presents adjusted  $R^2$  values from covariate-adjusted models including a methylation profile risk score based on methylation CpG sites identified to be associated with BP in the overall and replication meta-analyses. The number of CpG sites included in the methylation profile scores is indicated as n. Percent variance explained for the CpG sites identified in the primary replication and overall meta-analyses was calculated among an independent sample from FHS. The two ARIC samples participating in the discovery and replication stages were excluded from meta-analyses used to identify CpGs for their respective methylation risk scores, which caused the sets of methylation sites to differ. Abbreviations: AA, African American; ARIC, Atherosclerosis Risk in Communities; BP, blood pressure; CpG, cytosine-phosphate-guanine; EA, European ancestry; FHS, Framingham Heart Study.

International Consortium for Blood Pressure (ICBP)<sup>3</sup> that is yet to be published. Seven of the ten CpGs demonstrated nominal association with systolic or diastolic BP ( $0.05 > p > 1.0 \times 10^{-3}$ ; Table S6). The strongest association with both systolic and diastolic BP was observed at rs561931 for *PHGDH* cg14476101 and cg16246545 (systolic  $p = 0.007$ ; diastolic  $p = 0.01$ ). Though phenotypic association of exposure SNPs can serve as an indication of causality, we chose to formally test causality using multi-instrument Mendelian randomization, as follows, due to the complex genetic architecture of both DNA methylation and BP.

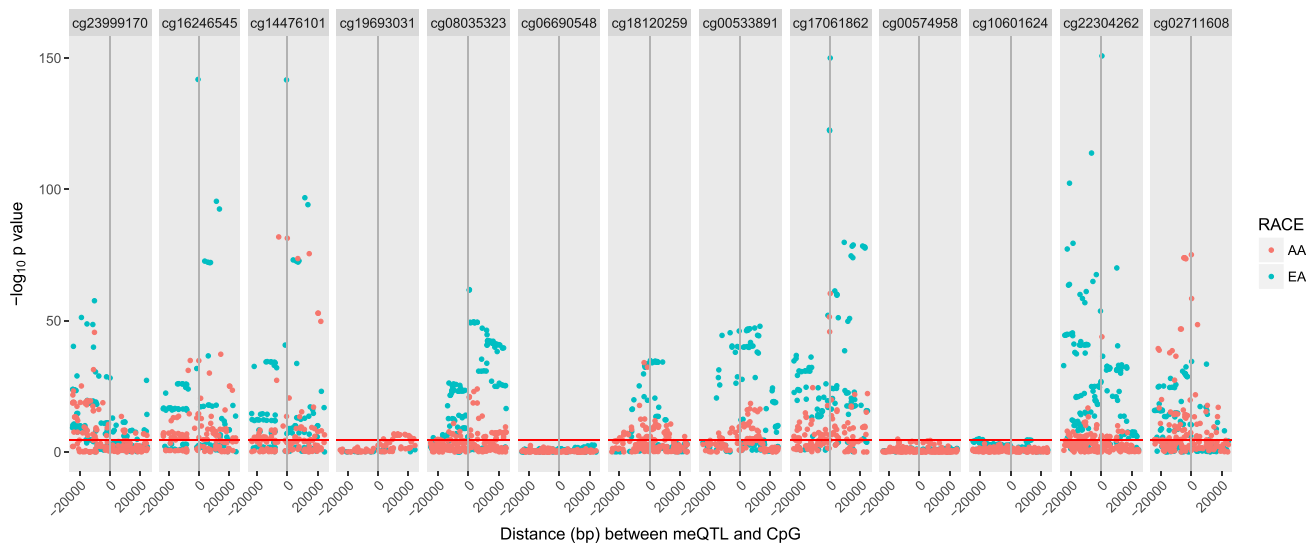
#### Bidirectional Mendelian Randomization

DNA methylation can be the cause or consequence of complex phenotypes. To provide support for causal relationships between DNA methylation and BP, we conducted bidirectional Mendelian randomization among up to 4,513 EA individuals in ARIC, FHS, RS, and WHI-EMPC. We used inverse-variance weighted tests to assess both forward causal roles of DNA methylation on BP and reverse causation where BP influences DNA methylation. For the six sentinel CpG sites with multiple genetic determinants, we were able to test forward causality using independent

meQTLs as the instrumental variables. The mean causal effect estimated across its seven independent meQTLs suggests that methylation at cg08035323 (*TAF1B-YWHAQ*) influences BP (causal effect estimate = 20.9 [11.1] change in systolic BP,  $p$  value = 0.009, and 15.1 [6.4] change in diastolic BP,  $p = 0.01$ , per one-percent change in DNA methylation; Table 4). There is also some evidence for reverse causation at cg08035323 (diastolic BP  $p = 0.02$ ); however, the causal  $p$  values for both BP traits are smaller for, and thus favor, forward causation. We performed an additional Mendelian randomization using BP effect estimates from ICBP and confirmed a causal relationship of methylation at cg08035323 with BP (systolic BP  $p = 0.007$ ; Table S7).

We assessed reverse causation for 11 sentinel CpG sites using 29 independent GWAS loci as instrumental variables to estimate the mean causal effect of BP on DNA methylation. In the absence of pleiotropic effects, inverse-variance weighted tests suggest that DNA methylation at cg00533891 (systolic BP  $p = 0.04$ , diastolic BP  $p = 0.001$ ) and *SLC1A5* cg02711608 (systolic BP  $p = 0.02$ , diastolic BP  $p = 0.0495$ ) is influenced by BP (Table 4). Reverse causation at both cg00533891 and *SLC1A5* cg02711608 is also supported by the lower-powered Egger test for causality (Table S8). Additionally, tests for causality of the second





**Figure 2. Distribution of Unpruned 1000 Genomes Imputed SNPs Assessed for Association with Methylation Relative to the CpG Location ( $\pm 25$  kb)**

SNP position relative to the replicated methylation CpG position ( $X = 0$ ) is plotted against  $-\log_{10}$  of the p value for meQTL meta-analysis in each race. SNPs above the red line are significant after Bonferroni correction for multiple testing ( $p < 3.0 \times 10^{-5}$ ). Abbreviations: AA, African American; bp, base pair; CpG, cytosine-phosphate-guanine; EA, European ancestry; meQTL, methylation quantitative trait locus; SNP, single-nucleotide polymorphism.

CpG in *SLC1A5*, cg22304262, support reverse causality at this locus (diastolic BP  $p = 0.04$ ; Table S8). The significant reverse causal effect estimates are consistent in magnitude and direction with those estimated by our EWAS. We additionally identified significant pleiotropic effects of the instrumental variables with methylation at cg10601624 and diastolic BP ( $p = 0.02$ ; Table S8). Pleiotropy overpowers the inverse-variance weighted test, and we did not identify a causal effect at cg10601624 using Egger regression ( $p = 0.9$ ; Table S8). There was a significant test result for forward causation at cg00533891; however, there also was evidence of pleiotropic effects among the forward instrumental variables and both the inverse-variance weighted and Egger tests favored reverse causality (Table S8). We also identified an effect of diastolic BP on DNA methylation at cg00574958 using Egger regression ( $p = 0.04$ ) in the presence of pleiotropic instrumental variables (Table S8). Using Mendelian randomization, we demonstrate that complex phenotypes can have an effect on DNA methylation among top EWAS signals and that forward causality can be assessed when instrumental variables are available.

### Gene Expression Associations with Replicated CpG Sites and Blood Pressure Traits

In whole-blood gene expression analyses, 4 of the 13 replicated CpG sites were found to have one or more *cis*-located genes (*TSPAN2*, *SLC7A11*, *UNC93B1*, *CPT1A*, *PTMS*, and *LPCAT3*) where transcription levels are associated with both CpG methylation (FHS and RS,  $N = 2,946$ ) and systolic BP, diastolic BP, or hypertension (FHS,  $N = 3,679$ ; Tables 5 and S9). The direction of effects for all six gene transcripts was consistent with the negative associations

of BP with DNA methylation at each CpG (Tables 2 and 5). The mRNA expression of *TSPAN2* showed the strongest associations with both CpG methylation and BP among all transcripts tested. Methylation at cg23999170, located in the first intron of *TSPAN2*, was strongly associated with decreased expression of *TSPAN2* in blood ( $p = 8.6 \times 10^{-14}$ ) and expression levels were associated with systolic BP ( $p = 5.0 \times 10^{-29}$ ), diastolic BP ( $p = 1.3 \times 10^{-16}$ ), and hypertension ( $p = 2.4 \times 10^{-10}$ ).

We identified nominal triangular associations of gene expression levels with methylation at 11 of the replicated CpG sites ( $p < 0.05$ ) and at least 1 BP trait ( $p < 0.05$ ) and present estimates of association and correlation in Table S10. These genes include *YWHAQ* (cg08035323 and diastolic BP), *PPIF* (cg00533891 and diastolic BP), and *GRLF1* (cg02711608/cg22304262 and diastolic BP).

### Two-Step Mendelian Randomization for Genes Mediating the BP-DNA Methylation Association

We used two-step Mendelian randomization to characterize causal mediation by gene transcripts significantly associated with methylation and BP. Using expression data available from the GTEx project and BP GWAS from ICBP, we first sought to establish a directional relationship from DNA methylation to gene expression, then a directional relationship from gene expression to BP. We showed that independent SNPs associated with methylation at cg23999170 are associated with expression of *TSPAN2* and that an additional independent variant associated with *TSPAN2* expression in blood is associated with BP (Table 6). The instrumental variables used for the exposures in each step achieved the associations needed to establish causality without showing strong evidence of association

**Table 4. Bidirectional Mendelian Randomization Results Showing the Inverse-Variance Weighted Effects of Multiple SNPs Used as Instrumental Variables in the Association of DNA Methylation and BP**

CpG	Trait	Forward Mendelian Randomization CpG → BP				Reverse Mendelian Randomization BP → CpG			
		IV SNPs, n	Mean estimate	(SE)	p Value	IV SNPs, n	Mean Estimate	(SE)	p Value
cg00533891	SBP	6	-10.3	(13.5)	-	29	-0.0008	(0.0004)	0.0388
cg00533891	DBP	6	-14.9	(7.8)	0.0405	29	-0.0020	(0.0006)	0.0013
cg00574958	SBP	-	-	-	-	29	0.0001	(0.0001)	0.2301
cg00574958	DBP	-	-	-	-	29	0.0001	(0.0002)	-
cg02711608	SBP	3	-31.1	(30.2)	0.2953	29	-0.0006	(0.0002)	0.0204
cg02711608	DBP	3	-30.2	(17.3)	-	29	-0.0008	(0.0004)	0.0495
cg06690548	SBP	-	-	-	-	29	-0.0004	(0.0003)	0.2267
cg06690548	DBP	-	-	-	-	29	-0.0002	(0.0006)	0.7724
cg08035323	SBP	7	20.9	(11.1)	0.0091	29	-0.0004	(0.0003)	0.2206
cg08035323	DBP	7	15.1	(6.4)	0.0111	29	-0.0012	(0.0006)	0.0226
cg10601624	SBP	-	-	-	-	29	-0.0004	(0.0002)	0.1069
cg10601624	DBP	-	-	-	-	29	-0.0010	(0.0003)	-
cg14476101	SBP	7	-2.5	(14.3)	0.8669	29	-0.0001	(0.0005)	0.7977
cg14476101	DBP	7	1	(5.6)	0.8623	29	-0.0002	(0.0008)	0.7757
cg17061862	SBP	10	-8.6	(10.2)	0.4224	29	0.0002	(0.0004)	0.6574
cg17061862	DBP	10	4.8	(5.1)	0.1112	29	0.0000	(0.0006)	0.9844
cg18120259	SBP	-	-	-	-	29	0.0001	(0.0003)	0.8084
cg18120259	DBP	-	-	-	-	29	-0.0002	(0.0005)	0.6968
cg19693031	SBP	-	-	-	-	29	0.0003	(0.0004)	0.3889
cg19693031	DBP	-	-	-	-	29	0.0006	(0.0006)	0.3509
cg23999170	SBP	5	5.9	(18.4)	0.7547	29	0.0004	(0.0003)	0.1954
cg23999170	DBP	5	-1.2	(10.6)	0.9151	29	0.0003	(0.0005)	0.6080

Causal mean effect and standard error estimates for the 11 sentinel CpGs are shown and causal p values have been omitted when IVs showed significant pleiotropic effects ( $p < 0.05$ ). Mean forward causal estimates are in percent DNA methylation change per mmHg increase in BP; mean reverse causal estimates are mmHg change in BP per 0.01 percent change in DNA methylation. Abbreviations: BP, blood pressure; CpG, cytosine-phosphate-guanine; DBP, diastolic blood pressure; IV, instrumental variable; Pos, position; SBP, systolic blood pressure; SE, standard error; SNP, single-nucleotide polymorphism.

with the outcome that would invalidate the causal test (generally,  $0.05 > p > 1 \times 10^{-5}$ ). Using two-step Mendelian randomization, we demonstrated that *TSPAN2* expression in whole blood is influenced by methylation at cg23999170 and that *TSPAN2* expression affects diastolic BP (Figure 3). Taken together, the direction of causality and our epigenome-wide estimate of association suggest that diastolic BP may increase by 5 mmHg for every 0.1% decrease in DNA methylation at cg23999170.

## Discussion

In a two-stage design of discovery and replication meta-analyses comprising 17,010 individuals, we identified DNA methylation at 13 CpG sites located in 8 intragenic regions and 3 intergenic regions significantly associated with systolic or diastolic BP. These CpGs are heritable, which suggests that DNA methylation that is the cause or conse-

quence of BP may have a transgenerational effect. We identified substantial *cis*-located genetic variation associated with methylation at many of these sites in both EA and AA populations, and these regions have moderate genetic associations with BP but are independent of the top GWAS loci previously reported in either race. Through *cis* gene expression analyses, we identified four CpGs significantly associated with one or more genes that may functionally connect DNA methylation and BP. Mendelian randomization techniques characterized the direction of association for six CpG sites with BP, including the mediation of a causal relationship of cg23999170 with BP through expression of *TSPAN2*. Through the analysis of DNA methylation, we have identified genes that provide insight into the biological mechanisms underlying BP regulation and target genes possibly affected by BP-induced DNA methylation.

We identified expression of *TSPAN2* to influence BP via DNA methylation. *TSPAN2* encodes the tetraspanin 2

**Table 5. Genes in a cis-Region ( $\pm 1$  Mb) of Replicated CpG Sites (1) Associated with DNA Methylation in Meta-analyses of FHS and RS at FDR Q Value  $< 0.05$ , and (2) Associated with BP Traits with at Least One FDR Q Value  $< 0.05$**

CpG Site	Chr	Gene	Gene Expression: DNA Methylation						Gene Expression: Blood Pressure Traits			
			FHS		RS		Meta-Analysis		Trait	Coeff	t Test	FDR Q
			Coeff	p Value	Coeff	p Value	Z-Score	FDR Q				
cg23999170	1	TSPAN2	-1.38	$2.7 \times 10^{-14}$	-1.92	0.0062	-7.32	$2.8 \times 10^{-12}$	SBP	0.0048	11.36	$5.0 \times 10^{-29}$
									DBP	0.0054	8.43	$1.3 \times 10^{-16}$
									HTN	0.1161	6.49	$2.4 \times 10^{-10}$
			-1.38	$2.7 \times 10^{-14}$	-2.72	0.0005	-7.86	$8.6 \times 10^{-14}$	SBP	0.0048	11.36	$5.0 \times 10^{-29}$
									DBP	0.0054	8.43	$1.3 \times 10^{-16}$
									HTN	0.1161	6.49	$2.4 \times 10^{-10}$
cg06690548	4	SLC7A11	-0.62	$2.8 \times 10^{-14}$	NA	NA	-7.61	$2.2 \times 10^{-13}$	SBP	0.0003	1.00	0.3173
									DBP	0.0002	0.32	0.7471
									HTN	0.0304	2.12	0.0338
cg00574958	11	UNC93B1	0.46	0.1375	2.84	0.0130	2.81	0.0376	SBP	-0.0006	-2.52	0.0472
									DBP	-0.0008	-2.06	0.0790
									HTN	-0.0184	-1.69	0.2399
		CPT1A	-2.95	$1.4 \times 10^{-13}$	-2.36	0.0003	-7.79	$2.1 \times 10^{-13}$	SBP	0.0007	2.03	0.0846
									DBP	0.0014	2.65	0.0324
									HTN	0.0225	1.56	0.2399
cg10601624	12	PTMS	-0.78	0.0002	-4.50	0.0020	-4.83	$2.8 \times 10^{-5}$	SBP	0.0009	2.40	0.0807
									DBP	0.0015	2.67	0.0381
									HTN	0.0170	1.10	0.8100
		LPCAT3	0.58	0.0012	1.18	0.2404	3.13	0.0134	SBP	-0.0010	-3.40	0.0069
									DBP	-0.0012	-2.73	0.0381
									HTN	-0.0302	-2.48	0.1321

Abbreviations: Chr, chromosome; Coeff, coefficient; CpG, cytosine-phosphate-guanine; DBP diastolic blood pressure; FDR, false discovery rate; FHS, Framingham Heart Study; HTN, hypertension; RS, Rotterdam Study; SBP, systolic blood pressure.

protein that is involved in signal transduction. *TSPAN2* is highly expressed in vascular tissues and implicated in the contractile ability and differentiation of vascular smooth muscle cells.<sup>35</sup> Sequence variation mapped to *TSPAN2* has previously been associated with large artery atherosclerosis-related stroke<sup>36</sup> and migraine<sup>37,38</sup> and *TSPAN2* suppresses inflammation in the central nervous system.<sup>39</sup> We additionally identified DNA methylation at cg08035323 to affect BP and the transcription of *YWHAQ* has a suggestive triangular relationship. *YWHAQ* encodes a 14-3-3 theta protein involved in signal transduction by binding to phosphoserine-containing proteins. *YWHAQ* has been implicated in phenotypes related to vascular response through transcriptional and DNA methylation changes in human preeclamptic placental tissues,<sup>40</sup> DNA sequence variation associated with exercise heart rate response,<sup>41</sup> and an effect on cardiomyocyte survival in animal models.<sup>42</sup>

We identified an effect of BP on DNA methylation at 4 of the 13 replicated CpG sites: *ZMIZ1* cg00533891, *CPT1A* cg00574958, and *SLC5A1* cg02711608/cg22304262. Previ-

ous epigenome-wide association studies have identified relationships of *CPT1A* cg00574958 and *SLC5A1* cg02711608/cg22304262 with other metabolic phenotypes, particularly lipids and adiposity (Table S11). An effect of triglycerides on methylation at cg00574958 has previously been identified,<sup>43</sup> which supports our hypothesis that an underlying cardiometabolic disease process related to BP and lipids alters DNA methylation within *CPT1A*. Transcriptional changes caused by DNA methylation at these four CpG sites could affect the risk of downstream phenotypes.

A recent GWAS identified 28 of 35 sentinel SNPs to have methylation-mediated associations with BP using a Mendelian randomization technique.<sup>23</sup> In our overall meta-analyses, we assessed the association of the 28 CpG sites reported by Kato et al.,<sup>23</sup> but were not able to confirm the direct association of any of these CpG sites with BP ( $p > 1.0 \times 10^{-5}$ ; Table S12). We further could not confirm the association of the sentinel SNPs with the CpG sites reported to mediate the associations with BP. However, we identified an association of rs12567136 (*CLCN6*) with

**Table 6. Two-Step Mendelian Randomization Results for Genes with Transcription Significantly Associated with DNA Methylation and BP**

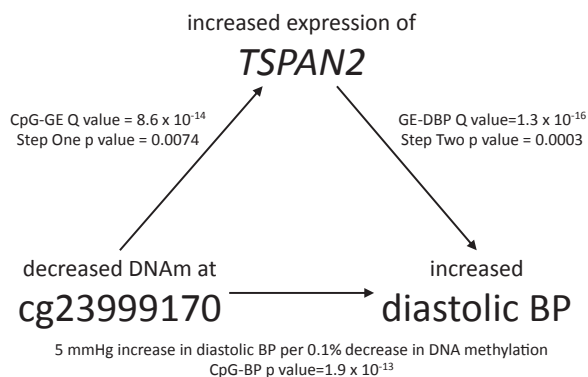
CpG site	Gene	Step One: CpG → GE			Step Two: GE → BP					
		IV: meQTLs			GTEX eQTL	IV: eQTLs			ICBP SBP	ICBP DBP
		SNP	Pos	p Value	p Value	SNP	Pos	p Value	p Value	p Value
cg23999170	TSPAN2	rs4240539	115,603,844	$5.5 \times 10^{-10}$	0.0074	rs12143357	115,321,323	$9.7 \times 10^{-4}$	0.0986	0.0003
		rs72697925	115,604,454	$4.5 \times 10^{-10}$	0.0254	–	–	–	–	–
		rs72697930	115,613,073	$2.2 \times 10^{-20}$	0.0512	–	–	–	–	–
		rs1286366	115,617,856	$2.6 \times 10^{-58}$	0.3148	–	–	–	–	–
		rs10858064	115,625,696	$8.2 \times 10^{-12}$	0.5198	–	–	–	–	–
cg06690548	SLC7A11	NA	NA	NA	NA	rs17050398	139,350,368	$3.6 \times 10^{-3}$	0.0745	0.0852
cg00574958	CPT1A	NA	NA	NA	NA	rs546233	69,311,449	$1.7 \times 10^{-3}$	0.6250	0.8880
	UNC93B1	NA	NA	NA	NA	rs1793252	67,740,366	$1.3 \times 10^{-4}$	0.2410	0.0747
cg10601624	LPCAT3	rs984337	6,383,525	$8.8 \times 10^{-6}$	0.1149	rs2110073	7,075,882	$8.2 \times 10^{-4}$	0.1920	0.2840
		rs4764572	6,417,998	$3.1 \times 10^{-5}$	0.5567	–	–	–	–	–
	PTMS	rs984337	6,383,525	$8.8 \times 10^{-6}$	0.5104	rs9668071	7,123,076	$1.9 \times 10^{-3}$	0.2590	0.6930
		rs4764572	6,417,998	$3.1 \times 10^{-5}$	0.5109	–	–	–	–	–

Position is Hg19. Abbreviations: BP, blood pressure; CpG, cytosine-phosphate-guanine; DBP diastolic blood pressure; eQTL, expression quantitative trait locus; GE, gene expression; GTEx, Genotype-Tissue Expression project; ICBP, International Consortium for Blood Pressure Genome-Wide Association Studies; IV, instrumental variable; meQTL, methylation quantitative trait locus; Pos, position; SBP, systolic blood pressure; SNP, single-nucleotide polymorphism.

methylation at an upstream CpG, cg20946054 (*DRAXIN*). This CpG site has a nominal association with BP (systolic BP  $p = 0.0318$ , diastolic BP  $p = 0.0015$ ) and adjustment for methylation at this locus attenuated the association of rs12567136 with systolic BP but not diastolic BP (Table

S13). Diastolic BP was the primary phenotype identified by Kato et al. Differences between our findings and those reported by Kato et al.<sup>23</sup> may be due to differing linkage disequilibrium structure in the populations examined in the two studies.

The strength of this study is our use of multiple data sources and analytic techniques to characterize functional relationships of DNA methylation and BP. We used strict replication standards to identify 13 CpG sites associated with BP and characterized gene transcriptional changes that could underlie these associations. However, these CpG sites were identified through cross-sectional analyses and we were able to characterize the direction of association only for six CpG sites. Mendelian randomization requires exposure-predictive SNPs to be selected as instrumental variables. For three CpG sites, we estimated no significant *cis*-meQTLs (cg19693031, cg06690548, and cg00574958) and an additional two CpG sites had insufficient independent meQTLs to assess forward causation using multiple instrumental variables (cg18120259 and cg10601624), so we lacked the ability to perform causal tests for select CpG sites. We additionally *a priori* chose to examine gene expression 1 Mb up- and downstream of each CpG site in whole blood, so more distal regulatory effects, and effects in different tissues, could have been missed. However, we did identify at least one biologically plausible gene to have a nominal triangular association with both methylation and BP for 11 of the replicated CpG sites, so it may be that larger sample sizes will uncover the functional and causal relationships of DNA methylation and BP. Despite these limitations, we discovered heritable CpG sites associated with BP among 17,010



**Figure 3. Illustration of the Relationship of Methylation at cg23999170 with Diastolic BP, Mediated by Expression of TSPAN2** Methylation at cg23999170 was identified as associated with diastolic BP in discovery and replication meta-analyses of genome-wide DNA methylation (N = 17,010). Expression of *TSPAN2* was associated with methylation at cg23999170 in meta-analyses of FHS and RS and diastolic BP in FHS. The direction of arrows in the diagram are inferred from significant two-step Mendelian randomization using data from the Genotype-Tissue Expression project and International Consortium for Blood Pressure, which suggests that methylation at cg23999170 influences BP through the expression of *TSPAN2*. The epigenome-wide association of DNA methylation and diastolic BP is interpreted given the evidence of causal direction and based on a 0.1% change in DNA methylation at cg23999170. Abbreviation: DBP, diastolic blood pressure.

individuals and showed that these CpGs explain additional phenotypic variance in an independent sample in which up to 261 BP-associated genetic variants explained minimal trait variance. We characterized the methylation-BP relationship using gene expression analyses and Mendelian randomization techniques to understand both which genes may regulate BP and how BP may affect gene transcription.

In conclusion, our genome-wide analysis of DNA methylation has uncovered loci influencing BP variation independently of underlying genetic variation. We additionally characterized functional and causal relationships connecting methylation at these loci with BP. In particular, we have identified *TSPAN2* as a candidate gene for BP that is regulated by heritable DNA methylation. *TSPAN2* and other methylated regions point to vascular contractility and inflammatory processes that functionally and causally connect DNA methylation and BP, and, thus, may represent new targets for the treatment and prevention of hypertension. Additional studies are needed to provide further mechanistic insights into the environmental exposures and genetic variation that influence DNA methylation and lead to high blood pressure. Nonetheless, our findings suggest that information on DNA methylation is likely to yield additional insights into the pathobiology of complex traits.

### Supplemental Data

Supplemental Data include 8 figures, 13 tables, and Supplemental Material and Methods, including funding acknowledgments and can be found with this article online at <https://doi.org/10.1016/j.ajhg.2017.09.028>.

### Consortia

BIOS consortium members are as follows: Bastiaan T. Heijmans, Peter A.C. 't Hoen, Joyce van Meurs, Aaron Isaacs, Rick Jansen, Lude Franke, Dorret I. Boomsma, René Pool, Jenny van Dongen, Jouke J. Hottenga, Marleen M.J. van Greevenbroek, Coen D.A. Stehouwer, Carla J.H. van der Kallen, Casper G. Schalkwijk, Cisca Wijmenga, Alexandra Zernakova, Etti F. Tigchelaar, P. Eline Slagboom, Marian Beekman, Joris Deelen, Diana van Heemst, Jan H. Veldink, Leonard H. van den Berg, Cornelia M. van Duijn, Albert Hofman, André G. Uitterlinden, P. Mila Jhamai, Michael Verbiest, H. Eka D. Suchiman, Marijn Verkerk, Ruud van der Breggen, Jeroen van Rooij, Nico Lakenberg, Hailiang Mei, Maarten van Iterson, Michiel van Galen, Jan Bot, Peter van 't Hof, Patrick Deelen, Irene Nooren, Matthijs Moed, Martijn Vermaat, Dasha V. Zernakova, René Luijk, Marc Jan Bonder, Freerk van Dijk, Wibowo Arindarto, Szymon M. Kielbasa, Morris A. Swertz, and Erik W. van Zwet.

### Acknowledgments

We thank all participants for contributing to this study. See [Supplemental Data](#) for detailed study acknowledgments and funding.

Received: June 19, 2017

Accepted: September 28, 2017

Published: November 30, 2017

### References

1. Campbell, N.R., Khalsa, T., Lackland, D.T., Niebylski, M.L., Nilsson, P.M., Redburn, K.A., Orias, M., Zhang, X.H., Burrell, L., Horiuchi, M., et al.; World Hypertension League Executive; International Society of Hypertension Executive; World Stroke Organization; International Diabetes Federation; International Council of Cardiovascular Prevention and Rehabilitation; and International Society of Nephrology (2016). High blood pressure 2016: why prevention and control are urgent and important. The World Hypertension League, International Society of Hypertension, World Stroke Organization, International Diabetes Foundation, International Council of Cardiovascular Prevention and Rehabilitation, International Society of Nephrology. *J. Clin. Hypertens. (Greenwich)* *18*, 714–717.
2. Levy, D., DeStefano, A.L., Larson, M.G., O'Donnell, C.J., Lifton, R.P., Gavras, H., Cupples, L.A., and Myers, R.H. (2000). Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the framingham heart study. *Hypertension* *36*, 477–483.
3. Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C., Hwang, S.-J., et al.; International Consortium for Blood Pressure Genome-Wide Association Studies; CARDIoGRAM consortium; CKDGen Consortium; KidneyGen Consortium; EchoGen consortium; and CHARGE-HF consortium (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* *478*, 103–109.
4. Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J., Dehghan, A., Glazer, N.L., Morrison, A.C., Johnson, A.D., Aspelund, T., et al. (2009). Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* *41*, 677–687.
5. Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M.D., Bochud, M., Coin, L., Najjar, S.S., Zhao, J.H., Heath, S.C., Eyheramendy, S., et al.; Wellcome Trust Case Control Consortium (2009). Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* *41*, 666–676.
6. Franceschini, N., Fox, E., Zhang, Z., Edwards, T.L., Nalls, M.A., Sung, Y.J., Tayo, B.O., Sun, Y.V., Gottesman, O., Adeyemo, A., et al.; Asian Genetic Epidemiology Network Consortium (2013). Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. *Am. J. Hum. Genet.* *93*, 545–554.
7. Fox, E.R., Young, J.H., Li, Y., Dreisbach, A.W., Keating, B.J., Musani, S.K., Liu, K., Morrison, A.C., Ganesh, S., Kutlar, A., et al.; International Consortium for Blood Pressure Genome-wide Association Studies (ICBP-GWAS) (2011). Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study. *Hum. Mol. Genet.* *20*, 2273–2284.
8. Ehret, G.B., Ferreira, T., Chasman, D.I., Jackson, A.U., Schmidt, E.M., Johnson, T., Thorleifsson, G., Luan, J., Donnelly, L.A., Kanoni, S., et al.; CHARGE-EchoGen consortium; CHARGE-HF consortium; and Wellcome Trust Case Control Consortium (2016). The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat. Genet.* *48*, 1171–1184.
9. Surendran, P., Drenos, F., Young, R., Warren, H., Cook, J.P., Manning, A.K., Grarup, N., Sim, X., Barnes, D.R., Witkowska,

- K., et al.; CHARGE-Heart Failure Consortium; EchoGen Consortium; METASTROKE Consortium; GIANT Consortium; EPIC-InterAct Consortium; Lifelines Cohort Study; Wellcome Trust Case Control Consortium; Understanding Society Scientific Group; EPIC-CVD Consortium; CHARGE+ Exome Chip Blood Pressure Consortium; T2D-GENES Consortium; GoT2DGenes Consortium; ExomeBP Consortium; and CHD Exome+ Consortium (2016). Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* *48*, 1151–1161.
10. Liu, C., Kraja, A.T., Smith, J.A., Brody, J.A., Franceschini, N., Bis, J.C., Rice, K., Morrison, A.C., Lu, Y., Weiss, S., et al.; CHD Exome+ Consortium; ExomeBP Consortium; GoT2DGenes Consortium; T2D-GENES Consortium; Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia; and CKDGen Consortium (2016). Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat. Genet.* *48*, 1162–1170.
  11. Warren, H.R., Evangelou, E., Cabrera, C.P., Gao, H., Ren, M., Mifsud, B., Ntalla, I., Surendran, P., Liu, C., Cook, J.P., et al.; International Consortium of Blood Pressure (ICBP) 1000G Analyses; BIOS Consortium; Lifelines Cohort Study; Understanding Society Scientific group; CHD Exome+ Consortium; ExomeBP Consortium; T2D-GENES Consortium; GoT2DGenes Consortium; Cohorts for Heart and Ageing Research in Genome Epidemiology (CHARGE) BP Exome Consortium; International Genomics of Blood Pressure (iGEN-BP) Consortium; and UK Biobank CardioMetabolic Consortium BP working group (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* *49*, 403–415.
  12. Niiranen, T.J., Havulinna, A.S., Langén, V.L., Salomaa, V., and Jula, A.M. (2016). Prediction of blood pressure and blood pressure change with a genetic risk score. *J. Clin. Hypertens. (Greenwich)* *18*, 181–186.
  13. Cowley, A.W., Jr., Nadeau, J.H., Baccarelli, A., Berecek, K., Fornage, M., Gibbons, G.H., Harrison, D.G., Liang, M., Nathanielsz, P.W., O'Connor, D.T., et al. (2012). Report of the national heart, lung, and blood institute working group on epigenetics and hypertension. *Hypertension* *59*, 899–905.
  14. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B., Witteman, J.C.M., Boerwinkle, E.; and CHARGE Consortium (2009). Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* *2*, 73–80.
  15. Teschendorff, A.E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., and Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* *29*, 189–196.
  16. Pidsley, R., Y Wong, C.C., Volta, M., Lunnon, K., Mill, J., and Schalkwyk, L.C. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* *14*, 293.
  17. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* *8*, 118–127.
  18. Maksimovic, J., Gordon, L., and Oshlack, A. (2012). SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.* *13*, R44.
  19. Mägi, R., and Morris, A.P. (2010). GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* *11*, 288.
  20. Breeze, C.E., Paul, D.S., van Dongen, J., Butcher, L.M., Ambrose, J.C., Barrett, J.E., Lowe, R., Rakyán, V.K., Iotchkova, V., Frontini, M., et al. (2016). eFORGE: a tool for identifying cell type-specific signal in epigenomic data. *Cell Rep.* *17*, 2137–2150.
  21. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* *102*, 15545–15550.
  22. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* *26*, 2190–2191.
  23. Kato, N., Loh, M., Takeuchi, F., Verweij, N., Wang, X., Zhang, W., Kelly, T.N., Saleheen, D., Lehne, B., Leach, I.M., et al.; BIOS-consortium; CARDIoGRAMplus4C; LifeLines Cohort Study; and InterAct Consortium (2015). Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat. Genet.* *47*, 1282–1293.
  24. Guo, Y., Tomlinson, B., Chu, T., Fang, Y.J., Gui, H., Tang, C.S., Yip, B.H., Cherny, S.S., Hur, Y.-M., Sham, P.C., et al. (2012). A genome-wide linkage and association scan reveals novel loci for hypertension and blood pressure traits. *PLoS ONE* *7*, e31489.
  25. Kato, N., Takeuchi, F., Tabara, Y., Kelly, T.N., Go, M.J., Sim, X., Tay, W.T., Chen, C.-H., Zhang, Y., Yamamoto, K., et al. (2011). Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat. Genet.* *43*, 531–538.
  26. Levy, D., Larson, M.G., Benjamin, E.J., Newton-Cheh, C., Wang, T.J., Hwang, S.-J., Vasan, R.S., and Mitchell, G.F. (2007). Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. *BMC Med. Genet.* *8 (Suppl 1)*, S3.
  27. Org, E., Eyheramendy, S., Juhanson, P., Gieger, C., Lichtner, P., Klopp, N., Veldre, G., Döring, A., Viigimaa, M., Söber, S., et al.; KORA; and BRIGHT (2009). Genome-wide scan identifies CDH13 as a novel susceptibility locus contributing to blood pressure determination in two European populations. *Hum. Mol. Genet.* *18*, 2288–2296.
  28. Padmanabhan, S., Melander, O., Johnson, T., Di Blasio, A.M., Lee, W.K., Gentilini, D., Hastie, C.E., Menni, C., Monti, M.C., Delles, C., et al.; Global BPgen Consortium (2010). Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. *PLoS Genet.* *6*, e1001177.
  29. Wain, L.V., Verwoert, G.C., O'Reilly, P.F., Shi, G., Johnson, T., Johnson, A.D., Bochud, M., Rice, K.M., Henneman, P., Smith, A.V., et al.; LifeLines Cohort Study; EchoGen consortium; AortaGen Consortium; CHARGE Consortium Heart Failure Working Group; KidneyGen consortium; CKDGen consortium; Cardiogenics consortium; and CardioGram (2011). Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat. Genet.* *43*, 1005–1011.

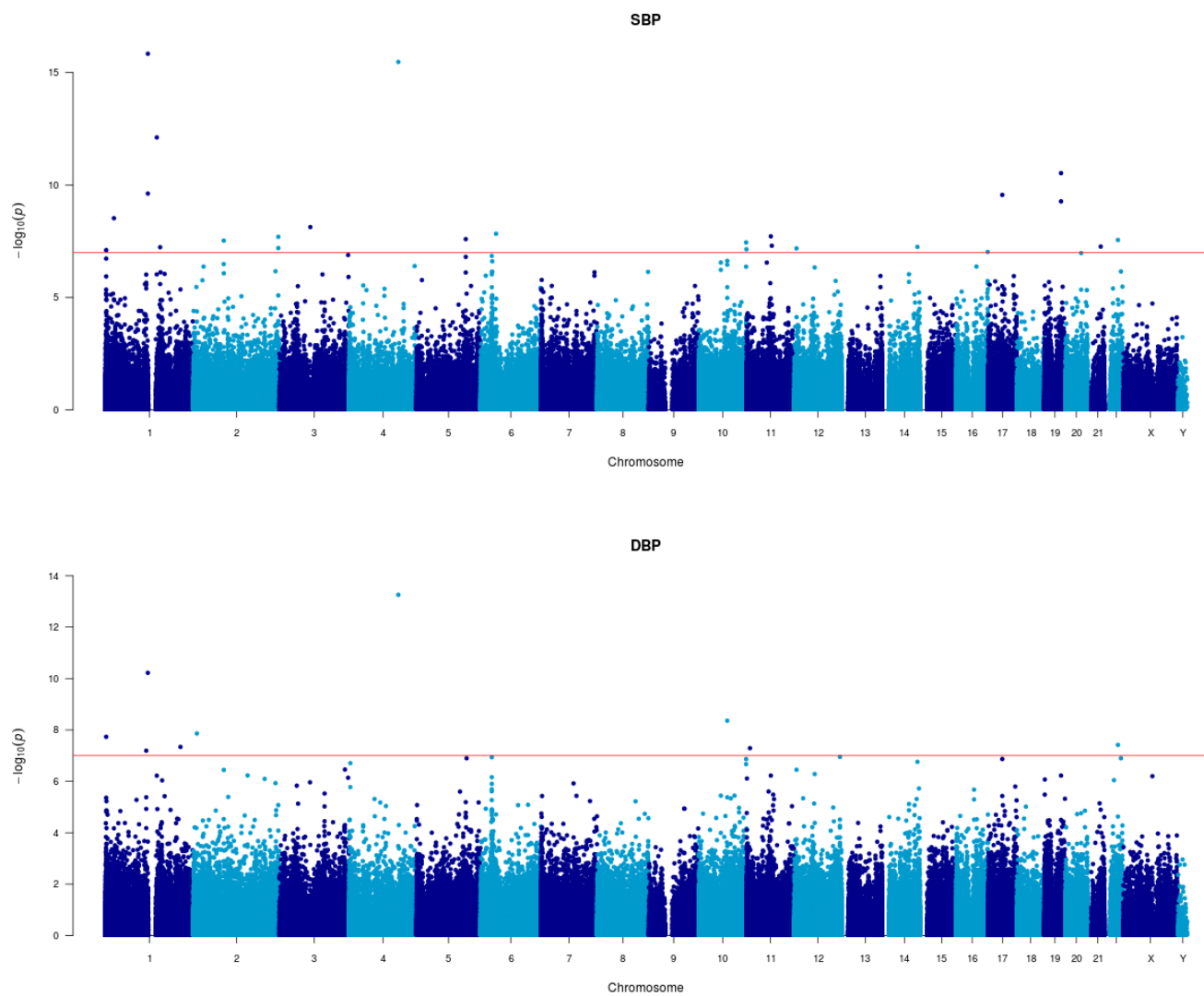
30. Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* *447*, 661–678.
31. Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* *44*, 512–525.
32. Huan, T., Meng, Q., Saleh, M.A., Norlander, A.E., Joehanes, R., Zhu, J., Chen, B.H., Zhang, B., Johnson, A.D., Ying, S., et al.; International Consortium for Blood Pressure GWAS (ICBP) (2015). Integrative network analysis reveals molecular mechanisms of blood pressure regulation. *Mol. Syst. Biol.* *11*, 799.
33. Chen, Y.A., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J., and Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* *8*, 203–209.
34. Gaunt, T.R., Shihab, H.A., Hemani, G., Min, J.L., Woodward, G., Lyttleton, O., Zheng, J., Duggirala, A., McArdle, W.L., Ho, K., et al. (2016). Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* *17*, 61.
35. Zhao, J., Wu, W., Zhang, W., Lu, Y.W., Tou, E., Ye, J., Gao, P., Jourdeuil, D., Singer, H.A., Wu, M., and Long, X. (2017). Selective expression of TSPAN2 in vascular smooth muscle is independently regulated by TGF- $\beta$ 1/SMAD and myocardin/serum response factor. *FASEB J.* *31*, 2576–2591.
36. NINDS Stroke Genetics Network (SiGN); and International Stroke Genetics Consortium (ISGC) (2016). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol.* *15*, 174–184.
37. Esserlind, A.-L., Christensen, A.F., Le, H., Kirchmann, M., Hauge, A.W., Toyserkani, N.M., Hansen, T., Grarup, N., Werge, T., Steinberg, S., et al. (2013). Replication and meta-analysis of common variants identifies a genome-wide significant locus in migraine. *Eur. J. Neurol.* *20*, 765–772.
38. Anttila, V., Winsvold, B.S., Gormley, P., Kurth, T., Bettella, F., McMahon, G., Kallela, M., Malik, R., de Vries, B., Terwindt, G., et al.; North American Brain Expression Consortium; and UK Brain Expression Consortium (2013). Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nat. Genet.* *45*, 912–917.
39. de Monasterio-Schrader, P., Patzig, J., Möbius, W., Barrette, B., Wagner, T.L., Kusch, K., Edgar, J.M., Brophy, P.J., and Werner, H.B. (2013). Uncoupling of neuroinflammation from axonal degeneration in mice lacking the myelin protein tetraspanin-2. *Glia* *61*, 1832–1847.
40. Liu, H., Tang, Y., Liu, X., Zhou, Q., Xiao, X., Lan, F., Li, X., Hu, R., Xiong, Y., and Peng, T. (2014). 14-3-3 tau (YWHAQ) gene promoter hypermethylation in human placenta of preeclampsia. *Placenta* *35*, 981–988.
41. Rankinen, T., Sung, Y.J., Sarzynski, M.A., Rice, T.K., Rao, D.C., and Bouchard, C. (2012). Heritability of submaximal exercise heart rate response to exercise training is accounted for by nine SNPs. *J. Appl. Physiol.* *112*, 892–897.
42. Lau, J.M.C., Jin, X., Ren, J., Avery, J., DeBosch, B.J., Treskov, I., Lupu, T.S., Kovacs, A., Weinheimer, C., and Muslin, A.J. (2007). The 14-3-3tau phosphoserine-binding protein is required for cardiomyocyte survival. *Mol. Cell. Biol.* *27*, 1455–1466.
43. Dekkers, K.F., van Iterson, M., Sliker, R.C., Moed, M.H., Bonder, M.J., van Galen, M., Mei, H., Zernakova, D.V., van den Berg, L.H., Deelen, J., et al.; BIOS Consortium (2016). Blood lipids influence DNA methylation in circulating cells. *Genome Biol.* *17*, 138.

## Supplemental Data

### DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation

Melissa A. Richard, Tianxiao Huan, Symen Ligthart, Rahul Gondalia, Min A. Jhun, Jennifer A. Brody, Marguerite R. Irvin, Riccardo Marioni, Jincheng Shen, Pei-Chien Tsai, May E. Montasser, Yucheng Jia, Catriona Syme, Elias L. Salfati, Eric Boerwinkle, Weihua Guan, Thomas H. Mosley, Jr., Jan Bressler, Alanna C. Morrison, Chunyu Liu, Michael M. Mendelson, André G. Uitterlinden, Joyce B. van Meurs, BIOS Consortium, Oscar H. Franco, Guosheng Zhang, Yun Li, James D. Stewart, Joshua C. Bis, Bruce M. Psaty, Yii-Der Ida Chen, Sharon L.R. Kardia, Wei Zhao, Stephen T. Turner, Devin Absher, Stella Aslibekyan, John M. Starr, Allan F. McRae, Lifang Hou, Allan C. Just, Joel D. Schwartz, Pantel S. Vokonas, Cristina Menni, Tim D. Spector, Alan Shuldiner, Coleen M. Damcott, Jerome I. Rotter, Walter Palmas, Yongmei Liu, Tomáš Paus, Steve Horvath, Jeffrey R. O'Connell, Xiuqing Guo, Zdenka Pausova, Themistocles L. Assimes, Nona Sotoodehnia, Jennifer A. Smith, Donna K. Arnett, Ian J. Deary, Andrea A. Baccarelli, Jordana T. Bell, Eric Whitsel, Abbas Dehghan, Daniel Levy, and Myriam Fornage





**Figure S1** Manhattan plots showing systolic and diastolic BP associations at each CpG site in the discovery meta-analyses. The CpG chromosomal locations are plotted against the  $-\log_{10}(p)$  value. The red line indicates the Bonferroni threshold ( $1 \times 10^{-7}$ ) for epigenome-wide significance.

Abbreviations: DBP, diastolic blood pressure; SBP, systolic blood pressure.

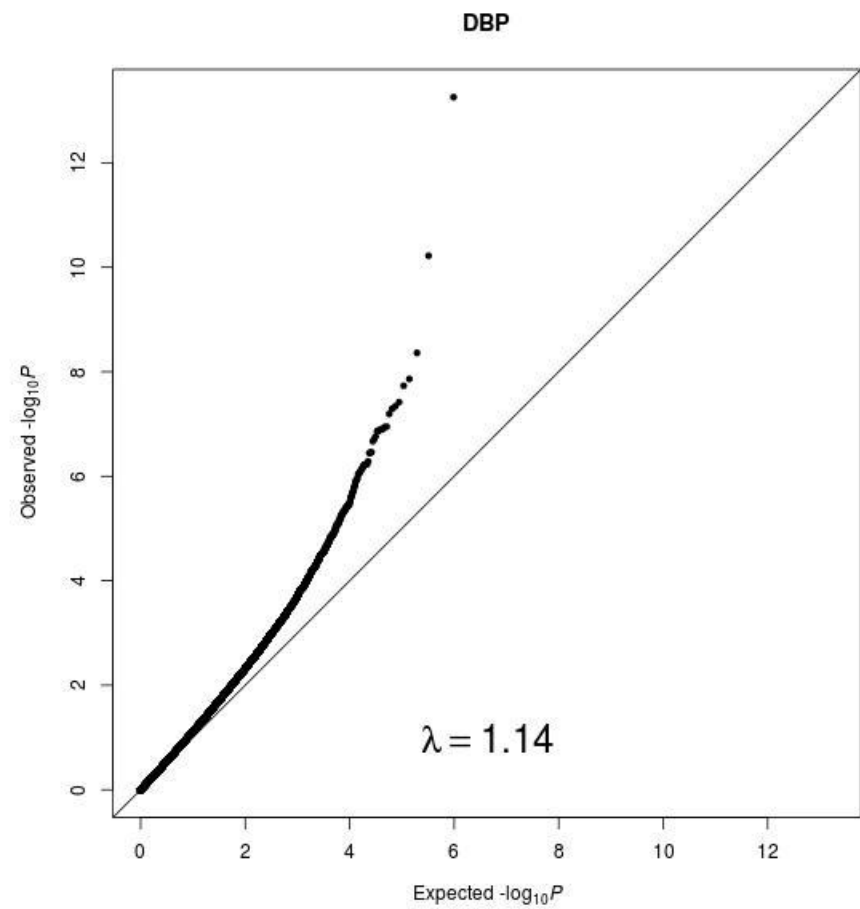
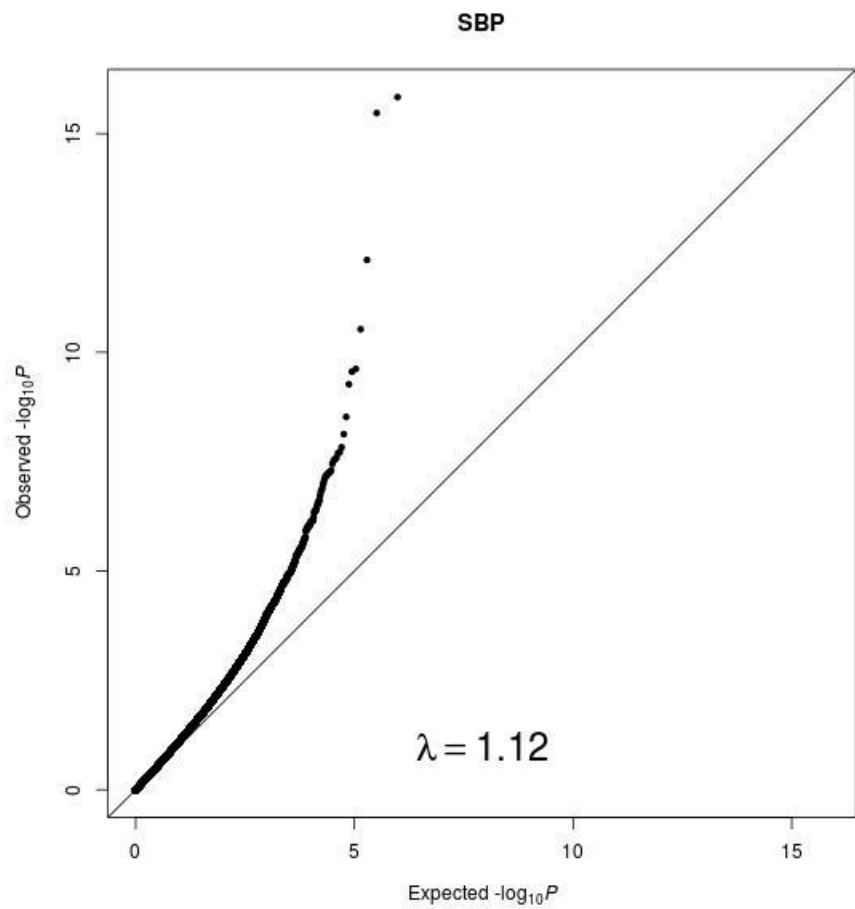
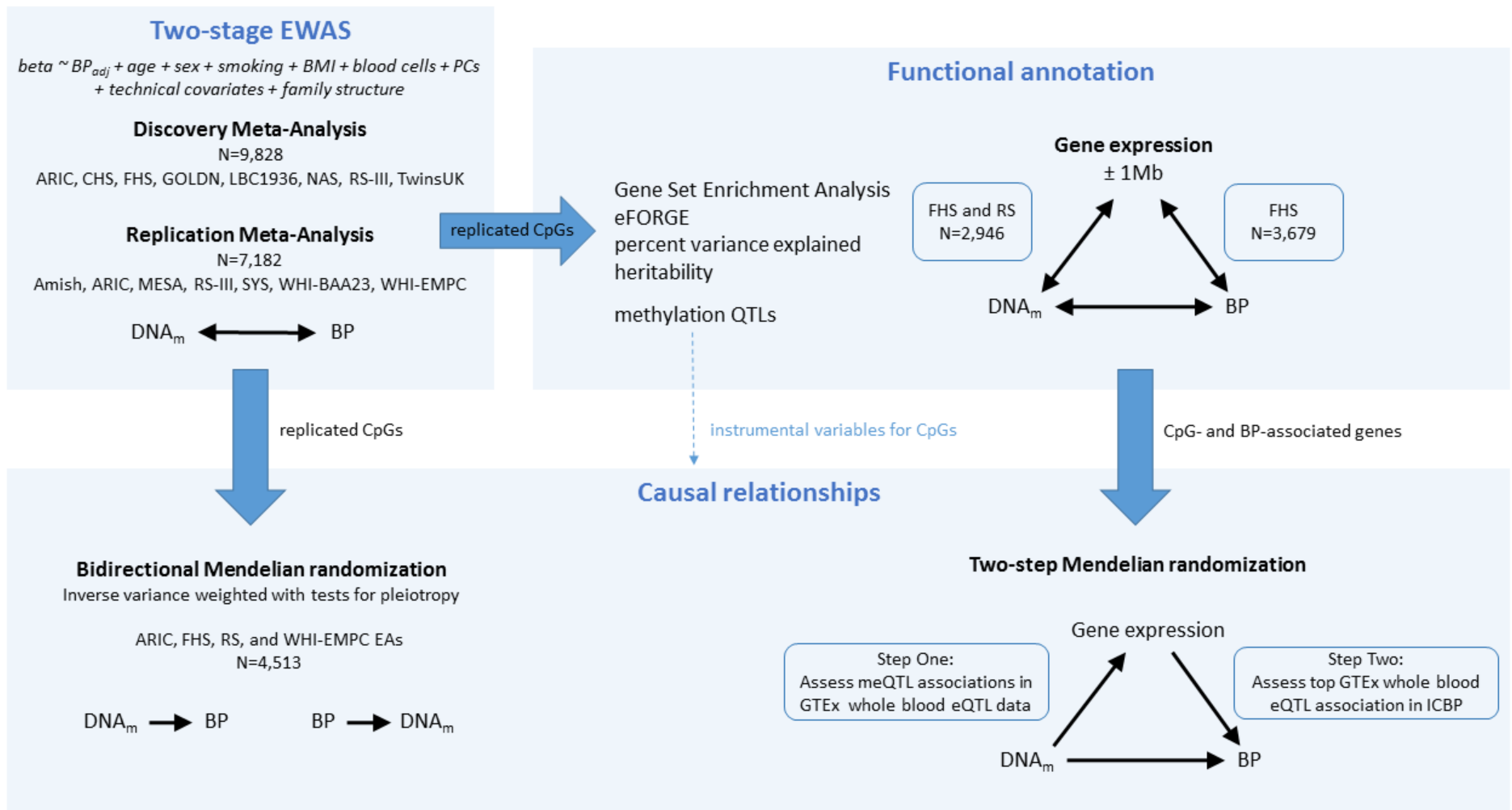


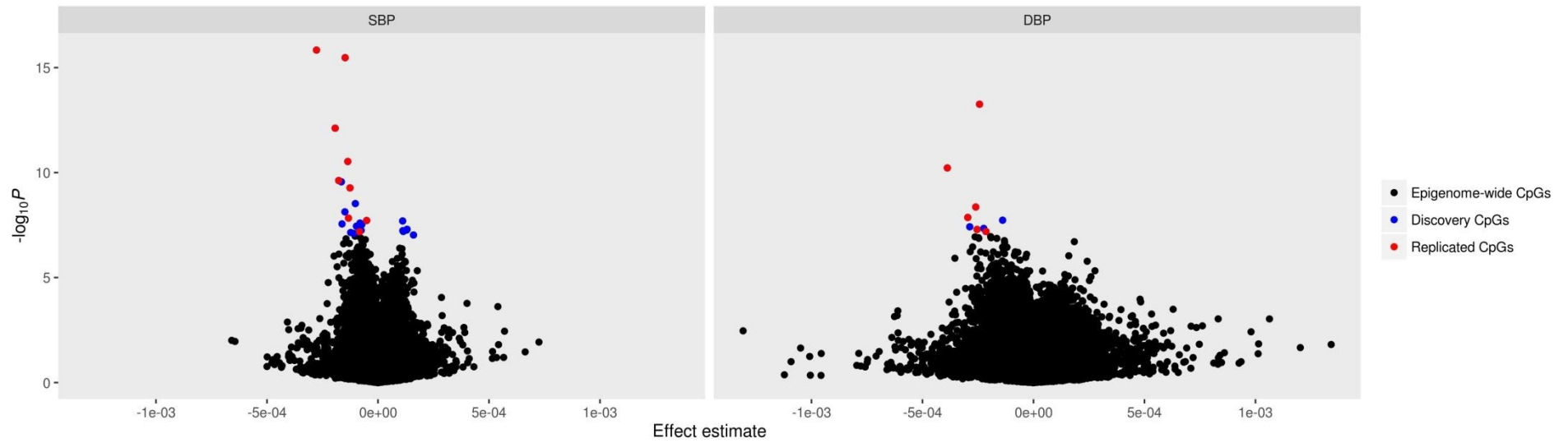
Figure S2 QQ plots for observed vs expected  $-\log_{10}(p \text{ value})$  at each CpG site in the discovery meta-analyses.

Abbreviations: DBP, diastolic blood pressure; SBP, systolic blood pressure.



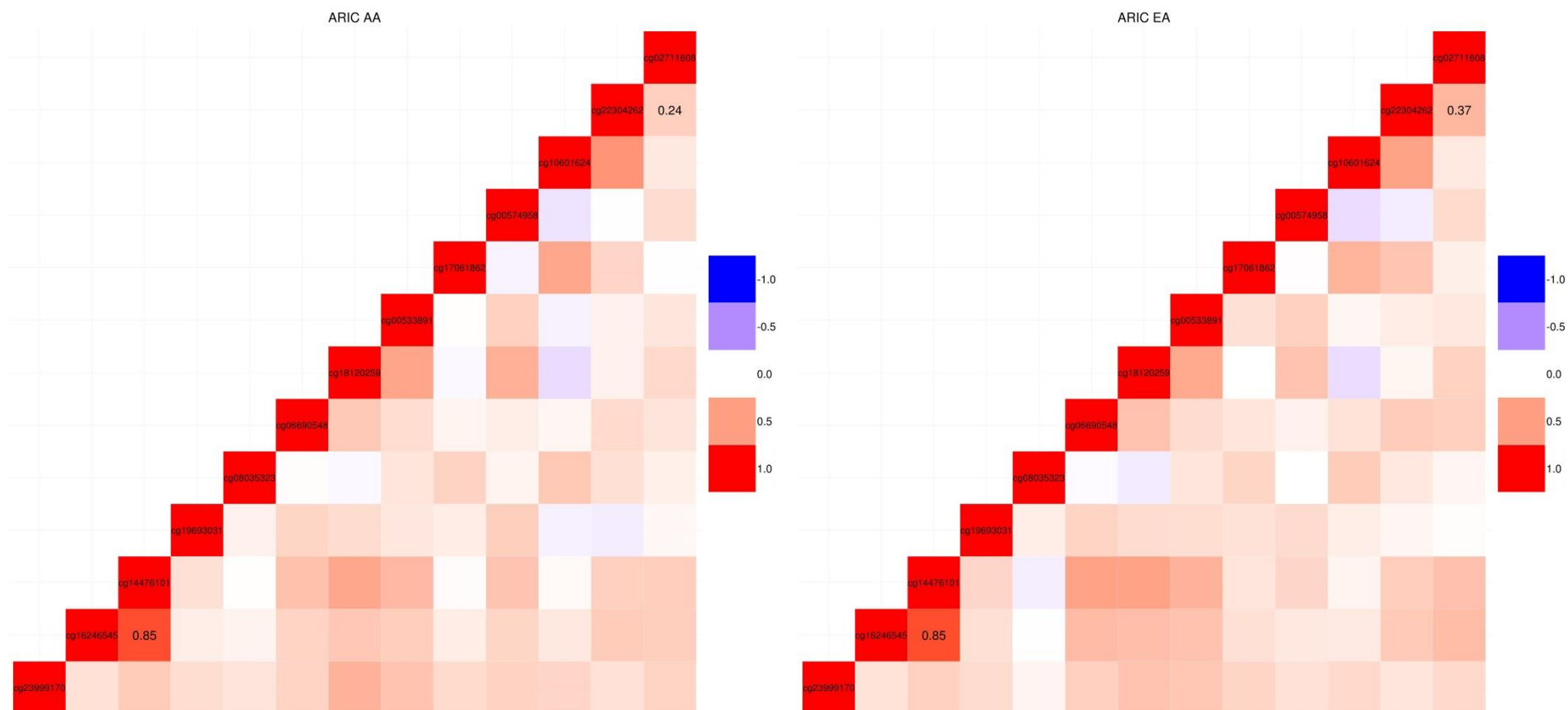
**Figure S3 Schematic of the overall study design.**

Abbreviations: BMI, body mass index; BP, blood pressure; CpG, cytosine-phosphate-guanine; DNA<sub>m</sub>, DNA methylation; EA, European ancestry; eFORGE, Functional element Overlap analysis of the Results of Genome Wide Association Study Experiments for EWAS studies; eQTL, expression quantitative trait locus; EWAS, epigenome-wide association study; GTEx, Genotype-Tissue Expression project; ICBP, International Consortium for Blood Pressure GWAS; Mb, megabase; PC, principal components; QTL, quantitative trait locus.



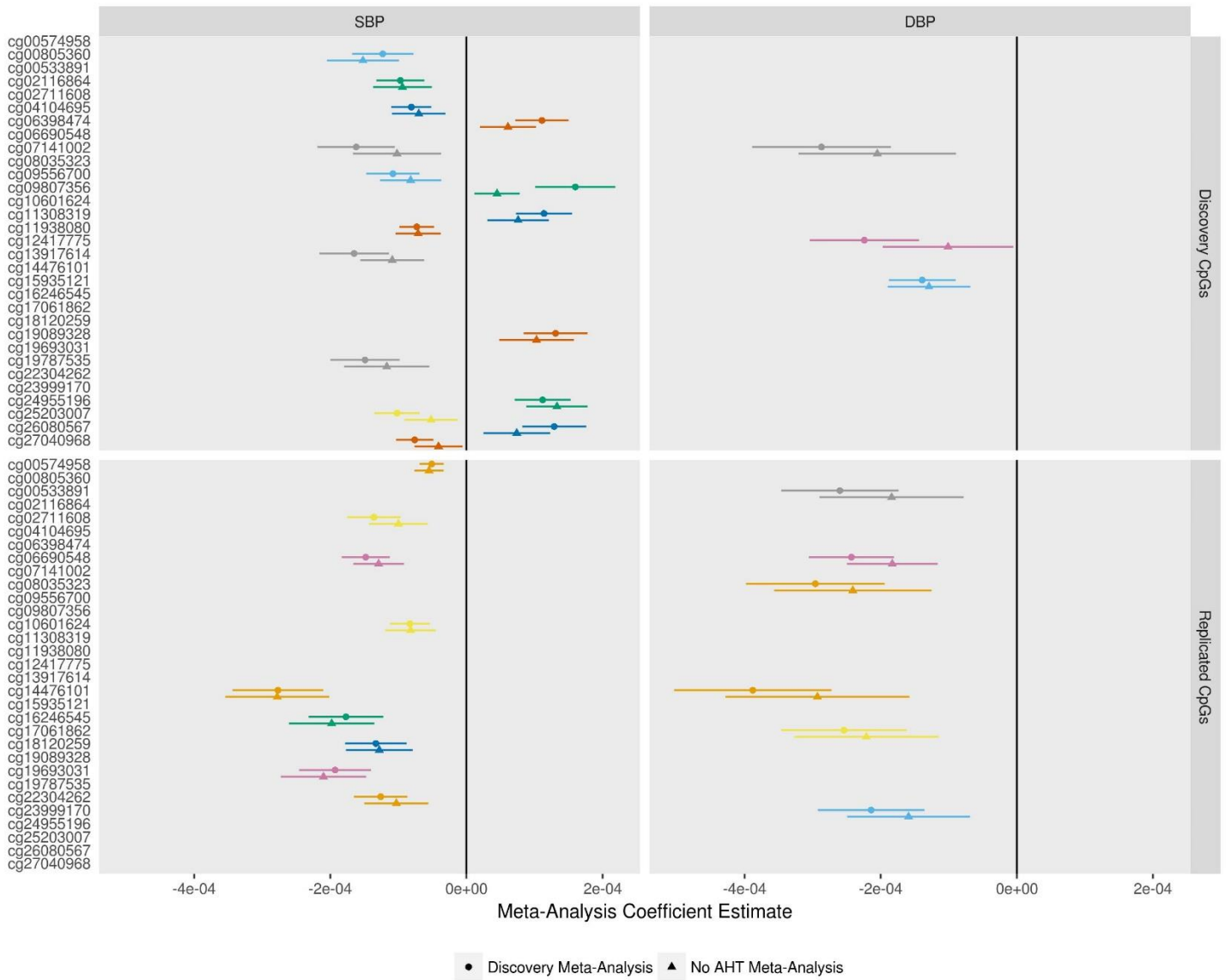
**Figure S4 Volcano plots of the effect sizes plotted against  $-\log_{10}(\text{p value})$  for epigenome-wide discovery meta-analysis associations with systolic and diastolic BP.** Effect size units are percent change in DNA methylation per 1-unit change in blood pressure. CpG sites identified at the discovery stage are colored blue and replicated CpG sites are colored red; all other CpG sites are shown in black.

Abbreviations: CpG, cytosine-phosphate-guanine; DBP, diastolic blood pressure; SBP, systolic blood pressure.



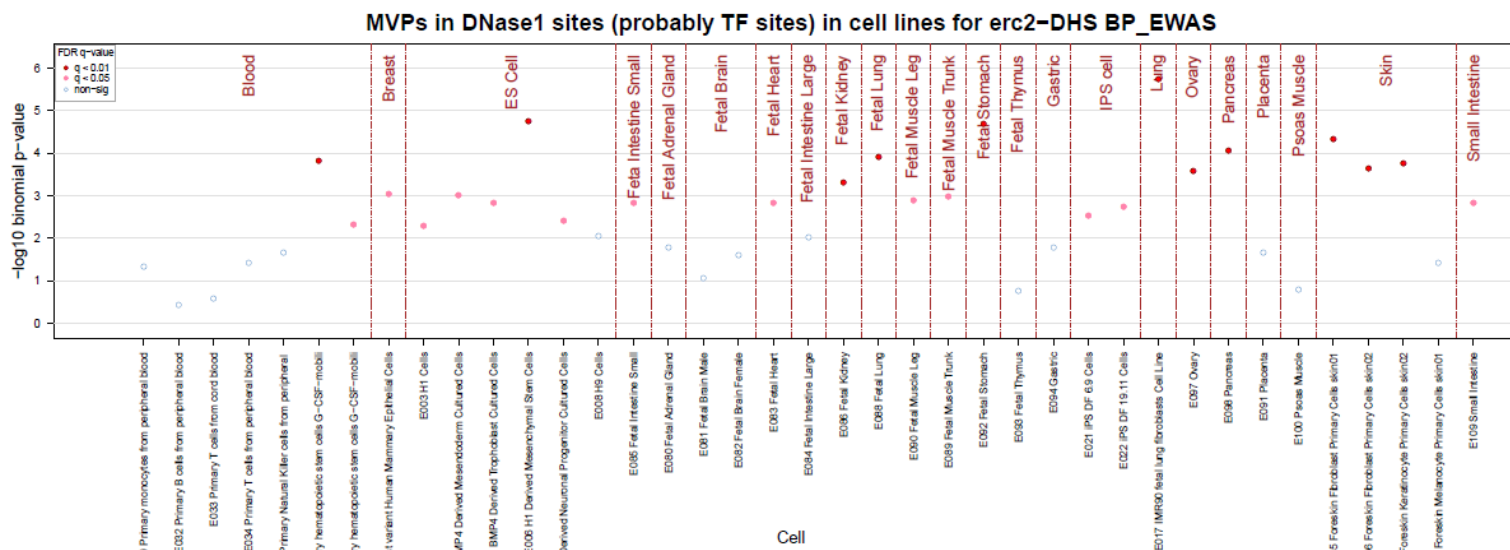
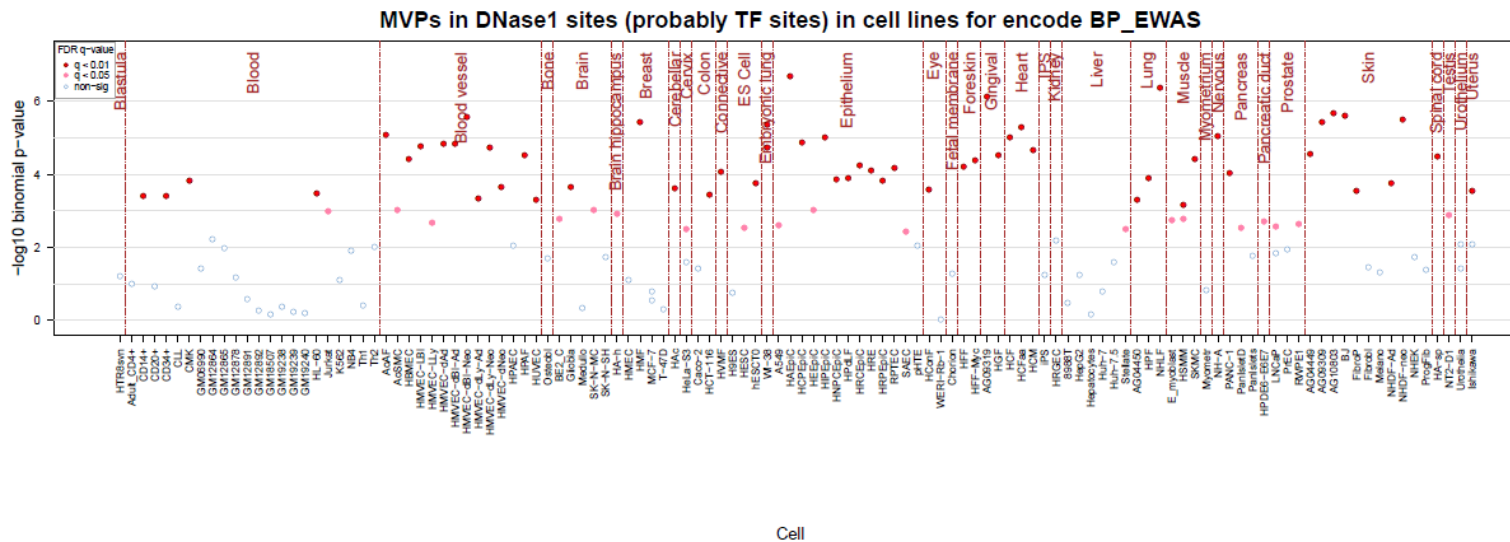
**Figure S5 Spearman correlation of methylation levels at replicated CpG sites for BP among individuals of African American and European ancestry in the ARIC cohort.** For CpG sites in the same gene (*PHGDH* cg16246545 and cg14476101; *SLC1A5* cg22304262 and cg02711608) the estimated correlation coefficients are shown. The color scale corresponds to strength of correlation, where inverse correlations are blue and direct correlations are red.

Abbreviations: AA, African American; ARIC, atherosclerosis Risk in Communities; EA, European Ancestry.



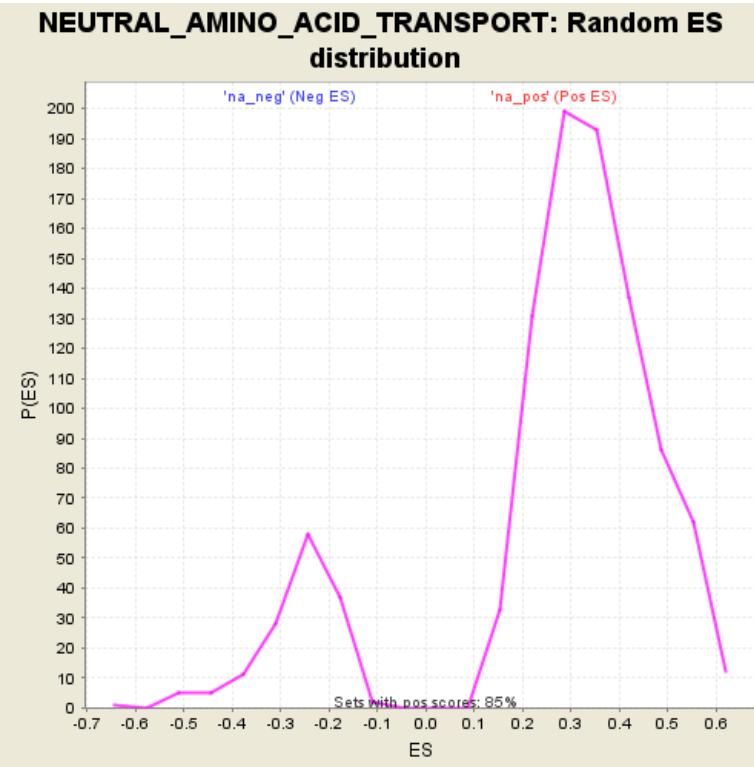
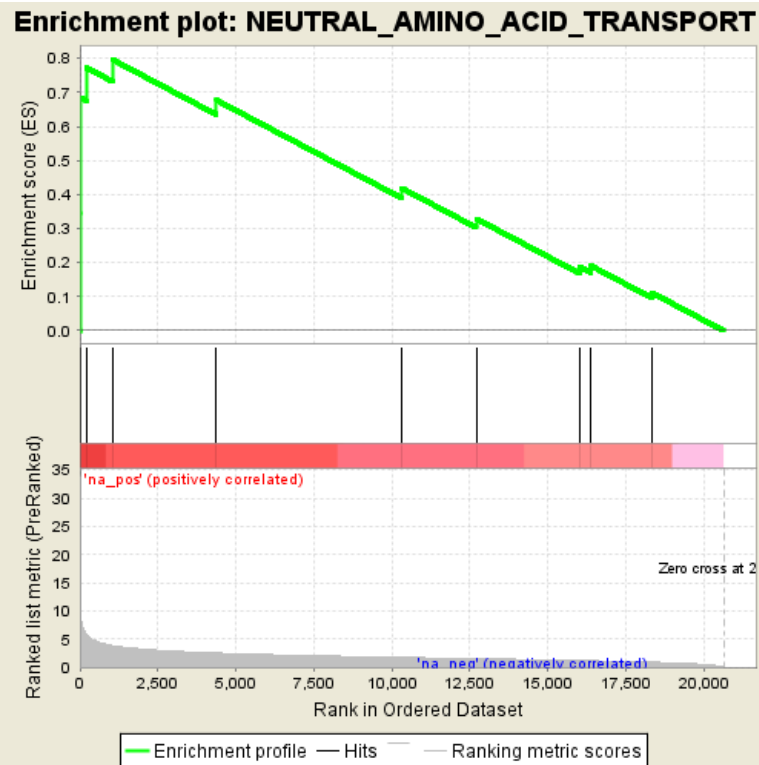
**Figure S6 Forest plots showing coefficient point estimates and 95% confidence intervals for the 31 discovery CpG sites in the discovery meta-analysis (N=9,828; circle) and the meta-analysis among individuals not reporting antihypertensive treatment (N=9,894; triangle). Discovery CpG sites with attenuated coefficient estimates among non-medicated individuals tended to not replicate.**

Abbreviations: AHT, anti-hypertensive treatment; CpG, cytosine-phosphate-guanine; DBP, diastolic blood pressure; SBP, systolic blood pressure.



**Figure S7 Enrichment of replicated CpG sites for DNase I hypersensitive sites across various tissues in the ENCODE (upper panel) and Roadmap Epigenomics (lower panel) projects.** Cell types showing enrichment at FDR-adjusted p value <0.01 are shown in red and FDR-adjusted p value <0.05 are shown in pink.

Abbreviations: BP, blood pressure; EWAS, epigenome-wide association study; FDR, false discovery rate; DHS, DNase I hypersensitive site; MVPs, methylation-varying probe; TF, transcription factor.



**Figure S8 DNA methylation at genes involved in neutral amino acid transport were identified by Gene Set Enrichment Analysis of the overall meta-analyses of systolic and diastolic BP.** Genes belonging to this category are primarily solute carrier (*SLC*) genes. The plots show the distribution of the ranked genes in the overall meta-analysis.

Abbreviations: ES, enrichment score.



CpG site	Chr	Position	UCSC Gene	Systolic BP				Diastolic BP			
				Discovery		Replication		Discovery		Replication	
				Coeff	p value	Coeff	p value	Coeff	p value	Coeff	p value
cg15935121	1	2230601	<i>SKI</i>	-7.2E-05	1.9E-07	-1.0E-05	5.3E-01	-0.0001	1.9E-08	-6.8E-05	2.2E-02
cg09556700	1	2230668	<i>SKI</i>	-0.0001	7.9E-08	-4.1E-05	5.8E-02	-0.0001	7.9E-04	-0.0001	2.4E-03
cg25203007	1	24126017	<i>GALE</i>	-0.0001	3.0E-09	-3.2E-05	1.3E-01	-0.0001	6.7E-05	-2.6E-05	5.0E-01
cg23999170	1	115628111	<i>TSPAN2</i>	-0.0001	2.7E-06	-0.0001	1.6E-05	-0.0002	6.4E-08	-0.0002	3.4E-07
cg16246545	1	120255941	<i>PHGDH</i>	-0.0002	2.4E-10	-0.0002	3.3E-14	-0.0002	2.2E-04	-0.0003	4.3E-07
cg14476101	1	120255992	<i>PHGDH</i>	-0.0003	1.5E-16	-0.0004	7.0E-21	-0.0004	6.0E-11	-0.0005	1.9E-12
cg19693031	1	145441552	<i>TXNIP</i>	-0.0002	7.7E-13	-0.0003	3.8E-19	-0.0002	6.0E-07	-0.0004	7.5E-10
cg24955196	1	154982621	<i>ZBTB7B</i>	0.0001	5.8E-08	5.0E-05	1.5E-02	0.0001	2.6E-03	0.0001	7.5E-04
cg12417775	1	212463238	<i>PPP2R5A</i>	-0.0001	4.4E-06	-4.8E-05	6.7E-02	-0.0002	4.6E-08	-0.0001	1.9E-03
cg08035323	2	9843525		-0.0001	4.2E-05	-8.7E-05	4.1E-03	-0.0003	1.4E-08	-0.0002	2.6E-04
cg11938080	2	85924685	<i>GNLY</i>	-7.3E-05	3.0E-08	-2.8E-05	1.3E-01	-0.0001	3.6E-07	-5.9E-05	7.7E-02
cg11308319	2	240291426	<i>HDAC4</i>	0.0001	6.4E-08	9.0E-06	6.6E-01	0.0001	8.3E-05	1.4E-05	7.0E-01
cg06398474	2	240291509	<i>HDAC4</i>	0.0001	2.0E-08	1.6E-05	3.8E-01	0.0002	8.4E-06	2.2E-05	5.0E-01
cg19787535	3	87843755		-0.0001	7.4E-09	-1.4E-05	6.7E-01	-0.0002	3.0E-04	6.3E-05	2.8E-01
cg06690548	4	139162808	<i>SLC7A11</i>	-0.0001	3.4E-16	-0.0002	8.3E-20	-0.0002	5.5E-14	-0.0003	9.9E-14
cg04104695	5	139058749	<i>CXXC5</i>	-8.1E-05	2.5E-08	-3.7E-05	1.1E-01	-0.0001	1.7E-05	-1.0E-04	1.7E-02
cg18120259	6	43894639	<i>LOC100132354</i>	-0.0001	1.5E-08	-0.0002	9.4E-15	-0.0002	1.9E-05	-0.0003	6.9E-10
cg00533891	10	80919242	<i>ZMIZ1</i>	-0.0001	2.4E-07	-9.2E-05	3.7E-03	-0.0003	4.4E-09	-0.0002	8.9E-04
cg02116864	10	134222453	<i>PWWP2B</i>	-9.7E-05	3.6E-08	-5.5E-05	1.1E-02	-0.0002	1.4E-07	-0.0001	1.1E-02
cg00805360	10	135091210	<i>ADAM8</i>	-0.0001	7.2E-08	-6.0E-05	2.3E-02	-0.0002	6.9E-05	-8.7E-05	7.1E-02
cg17061862	11	9590431		-0.0001	6.9E-05	-0.0002	6.6E-09	-0.0003	5.1E-08	-0.0003	1.2E-06
cg00574958	11	68607622	<i>CPT1A</i>	-5.1E-05	1.9E-08	-4.8E-05	1.4E-06	-8.2E-05	5.9E-07	-6.7E-05	2.5E-04
cg19089328	11	71210210	<i>NADSYN1</i>	0.0001	5.0E-08	1.4E-05	6.4E-01	0.0002	3.2E-05	3.6E-05	5.2E-01
cg10601624	12	6404377		-8.3E-05	6.6E-08	-0.0001	1.6E-10	-0.0001	3.5E-07	-0.0002	1.7E-07
cg27040968	14	99787363		-7.6E-05	5.6E-08	-1.0E-06	9.4E-01	-0.0001	1.7E-07	-5.9E-05	6.9E-02
cg09807356	16	88905700	<i>GALNS</i>	0.0002	9.4E-08	6.0E-06	5.8E-01	0.0002	4.2E-03	1.3E-05	5.6E-01
cg13917614	17	40125660	<i>CNP</i>	-0.0002	2.7E-10	-4.0E-05	3.3E-02	-0.0002	1.4E-07	-7.1E-05	3.9E-02
cg22304262	19	47287778	<i>SLC1A5</i>	-0.0001	5.4E-10	-0.0001	8.7E-09	-0.0002	6.0E-07	-0.0002	4.9E-05
cg02711608	19	47287964	<i>SLC1A5</i>	-0.0001	3.0E-11	-0.0001	1.1E-11	-0.0002	3.2E-05	-0.0002	3.0E-06
cg26080567	21	37536137	<i>DOPEY2</i>	0.0001	5.5E-08	2.7E-05	2.0E-01	0.0002	1.3E-05	-4.0E-05	3.1E-01
cg07141002	22	38201690	<i>H1FO</i>	-0.0002	2.8E-08	-6.3E-05	3.8E-02	-0.0003	3.8E-08	-0.0002	1.6E-03

**Table S1 Discovery and replication meta-analyses results for the thirty-one CpG sites with p value  $<1.0 \times 10^{-7}$  at the discovery stage.** Position is Hg19. Coefficients give the percent change in DNA methylation for every 1-unit change in blood pressure.

Abbreviations: BP, blood pressure; Chr, chromosome; Coeff, coefficient; CpG, cytosine-phosphate-guanine; UCSC, University of California Santa Cruz.

CpG site	Chr	Position	UCSC Gene	Gene Location	CpG Island	DHS	Enhancer
cg23999170	1	115628111	<i>TSPAN2</i>	Body	N Shelf		
cg16246545	1	120255941	<i>PHGDH</i>	Body	S Shore		
cg14476101	1	120255992	<i>PHGDH</i>	Body	S Shore		
cg19693031	1	145441552	<i>TXNIP</i>	3'UTR			
cg08035323	2	9843525				DHS	Enhancer
cg06690548	4	139162808	<i>SLC7A11</i>	Body			
cg18120259	6	43894639	LOC100132354	Body		DHS	Enhancer
cg00533891	10	80919242	<i>ZMIZ1</i>	5'UTR		DHS	Enhancer
cg17061862	11	9590431			N Shelf		
cg00574958	11	68607622	<i>CPT1A</i>	5'UTR	N Shore		
cg10601624	12	6404377			S Shelf	DHS	Enhancer
cg22304262	19	47287778	<i>SLC1A5</i>	Body;5'UTR	N Shelf		
cg02711608	19	47287964	<i>SLC1A5</i>	1stExon;5'UTR;Body	N Shelf		

**Table S2 Illumina annotation for replicated blood pressure CpG sites.** Position is Hg19.

Abbreviations: Chr, chromosome; CpG, cytosine-phosphate-guanine; DHS, DNase I hypersensitive site; UCSC, University of California Santa Cruz; UTR, untranslated region.

CpG site	Chr	Position	Systolic BP					Diastolic BP				
			Discovery			Replication	Overall	Discovery			Replication	Overall
			Het p value	Race p value	Race and cell type p	Het p value	Het p value	Het p value	Race p value	Race and cell type p	Het p value	Het p value
cg23999170	1	115628111	0.8268	0.6674	0.5586	0.0634	0.9521	0.4965	0.8864	0.7823	0.2852	0.7553
cg16246545	1	120255941	0.6613	0.4293	0.1304	0.5039	0.1098	0.1588	0.8459	0.2307	0.6393	0.1271
cg14476101	1	120255992	0.2434	0.5686	0.0744	0.4511	0.0487	0.0947	0.5398	0.0626	0.8208	0.1481
cg19693031	1	145441552	0.2155	0.9776	0.5260	0.2938	0.0081	0.1174	0.4529	0.3516	0.3328	0.0599
cg08035323	2	9843525	0.3762	0.2123	0.1343	0.2910	0.4229	0.1455	0.0245	0.0037	0.6021	0.2192
<b>cg06690548</b>	<b>4</b>	<b>139162808</b>	<b>3.15E-07</b>	<b>0.0008</b>	<b>4.45E-06</b>	<b>0.1621</b>	<b>0.0149</b>	<b>2.20E-05</b>	<b>0.0001</b>	<b>4.39E-06</b>	<b>0.0047</b>	<b>0.0921</b>
cg18120259	6	43894639	0.1549	0.7064	0.4899	0.2389	0.1862	0.3308	0.9610	0.7416	0.2316	0.1954
cg00533891	10	80919242	0.1800	0.5954	0.6065	0.1649	0.3622	0.6205	0.4702	0.8903	0.7653	0.3503
cg17061862	11	9590431	0.0133	0.2222	0.0077	0.1004	0.0779	0.4931	0.6978	0.4094	0.5484	0.8683
cg00574958	11	68607622	0.1491	0.3041	0.1967	0.0047	0.8235	0.0267	0.9703	0.7466	0.3659	0.5334
cg10601624	12	6404377	0.7570	0.0853	0.3723	0.4898	0.0854	0.6368	0.1311	0.5958	0.3886	0.2762
cg22304262	19	47287778	0.3399	0.3279	0.5860	0.3930	0.5529	0.5831	0.2701	0.1851	0.8915	0.9046
cg02711608	19	47287964	0.0763	0.4415	0.8385	0.7613	0.9155	0.0357	0.6652	0.8892	0.8832	0.7833

**Table S3 Tests for heterogeneity in epigenome-wide meta-analyses for systolic and diastolic BP among the replicated CpG sites.** Position is Hg19. Cochran's *Q* statistic was used to assess variation in the discovery, replication, and overall meta-analyses. Differences in the discovery meta-analyses that may be explained by race (EA compared to AA) or cell type (whole blood compared to T cells) were tested using a 1 degree of freedom Chi-square test for differences between strata.

Abbreviations: BP, blood pressure; Chr, chromosome; CpG, cytosine-phosphate-guanine; Het, Cochran's heterogeneity statistic's (*Q*) *P*-value.

Supplemental Excel file provided.

**Table S4 Results of overall epigenome-wide meta-analyses for systolic and diastolic BP (p value  $<1.0 \times 10^{-7}$ ).** Position is Hg19. Coefficients give the percent change in DNA methylation for every 1-unit change in blood pressure.

Abbreviations: BP, blood pressure; Chr, chromosome; Coeff, coefficient; CpG, cytosine-phosphate-guanine; UCSC, University of California Santa Cruz.

CpG site	Chr	Position	UCSC Gene	Systolic BP				Diastolic BP			
				No AHT				No AHT			
				Coeff	p value	Discovery p	Replication p	Coeff	p value	Discovery p	Replication p
cg16246545	1	120,255,941	<i>PHGDH</i>	-0.0002	<b>8.0E-10</b>	2.4E-10	3.3E-14	-0.0002	7.2E-03	2.2E-04	4.3E-07
cg14476101	1	120,255,992	<i>PHGDH</i>	-0.0003	<b>5.8E-13</b>	1.5E-16	7.0E-21	-0.0003	1.9E-05	6.0E-11	1.9E-12
cg19693031	1	145,441,552	<i>TXNIP</i>	-0.0002	<b>5.9E-11</b>	7.7E-13	3.8E-19	-0.0002	8.1E-05	6.0E-07	7.5E-10
cg24955196	1	154,982,621	<i>ZBTB7B</i>	0.0001	<b>8.9E-09</b>	5.8E-08	1.5E-02	0.0003	<b>1.8E-09</b>	2.6E-03	7.5E-04
cg06690548	4	139,162,808	<i>SLC7A11</i>	-0.0001	<b>1.6E-11</b>	3.4E-16	8.3E-20	-0.0002	1.1E-07	5.5E-14	9.9E-14
cg05014727	10	6,214,016	<i>PFKFB3</i>	-0.0002	<b>2.3E-08</b>	3.0E-04	6.8E-04	-0.0001	1.1E-02	5.7E-03	7.0E-02
cg08994060	10	6,214,026	<i>PFKFB3</i>	-0.0002	<b>6.3E-08</b>	2.8E-03	1.7E-03	-0.0002	2.4E-03	4.2E-03	8.0E-02
cg26262157	10	6,214,079	<i>PFKFB3</i>	-0.0002	<b>5.2E-08</b>	3.7E-03	6.1E-04	-0.0001	7.9E-03	7.6E-03	5.9E-02
cg00805360	10	135,091,210	<i>ADAM8</i>	-0.0002	<b>2.4E-08</b>	7.2E-08	2.3E-02	-0.0003	<b>2.6E-08</b>	6.9E-05	7.1E-02

**Table S5 Results of an overall epigenome-wide meta-analyses for systolic and diastolic BP among individuals not reporting the use of antihypertensive medication (p value <1.0 x 10<sup>-7</sup>).** Position is Hg19. Coefficients give the percent change in DNA methylation for every 1 mmHg change in blood pressure.

Abbreviations: AHT, antihypertensive treatment; BP, blood pressure; Chr, chromosome; Coeff, coefficient; CpG, cytosine-phosphate-guanine; UCSC, University of California Santa Cruz.

CpG site	top BP-associated meQTL	meQTL-CpG dist, bp	Trait	ICBP 1000G SBP p value	ICBP 1000G DBP p value
cg00533891	rs737012	13,056	SBP + DBP	0.1967	0.1572
cg02711608	rs56050670	17,333	SBP + DBP	0.0410	0.0813
cg08035323	rs17592239	15,359	SBP	0.0241	0.4423
cg08035323	rs4669419	-12,904	DBP	0.0998	0.0936
cg10601624	rs4764572	13,621	SBP	0.5050	0.6137
cg10601624	rs984337	-20,852	DBP	0.8157	0.3748
cg14476101	rs561931	-1,486	SBP + DBP	0.0074	0.0129
cg16246545	rs561931	-1,435	SBP + DBP	0.0074	0.0129
cg17061862	rs6486429	-2,878	SBP	0.1288	0.8563
cg17061862	rs11042429	9,727	DBP	0.4218	0.0239
cg18120259	rs7745517	456	SBP + DBP	0.7853	0.7433
cg22304262	rs4578775	12,736	SBP + DBP	0.0408	0.1099
cg23999170	rs10858064	-2,415	SBP + DBP	0.0525	0.0195

**Table S6 Association of meQTLs with BP in ICBP 1000Genomes data. The meQTLs for each CpG were pruned to be independent ( $r^2 < 0.2$ ) and the meQTL most associated with systolic and/or diastolic BP is shown.** Distance between SNPs and meQTL is from position in Hg19.

Abbreviations: bp, base pairs; CpG, cytosine-phosphate-guanine; DBP, diastolic blood pressure; dist, distance; meQTL, methylation quantitative trait locus; SBP, systolic blood pressure.

CpG site	Trait	IV SNPs, n	Mean estimate	SE	p value	Pleiotropy p value
cg00533891	<i>SBP</i>	4	-1.5	2.7	0.4972	0.3643
	<i>DBP</i>	4	-0.8	1.8	0.6948	0.1208
cg02711608	<i>SBP</i>	3	-7.2	5.4	0.3156	0.3135
	<i>DBP</i>	3	-6.1	3.0	0.1107	0.5290
cg08035323	<i>SBP</i>	6	6.2	2.1	0.0067	0.0762
	<i>DBP</i>	6	0.9	1.3	0.5016	0.0720
cg14476101	<i>SBP</i>	7	-3.7	2.3	0.1567	0.6649
	<i>DBP</i>	7	-1.8	1.3	0.2291	0.9083
cg16246545	<i>SBP</i>	6	-3.5	3.0	0.2922	0.8693
	<i>DBP</i>	6	-1.5	1.7	0.3964	0.7974
cg17061862	<i>SBP</i>	10	-1.0	1.6	0.4559	0.4309
	<i>DBP</i>	10	-0.5	1.0	0.5896	0.6703
cg22304262	<i>SBP</i>	7	-2.2	2.5	0.4186	0.3419
	<i>DBP</i>	7	-1.9	1.3	0.1465	0.2795
cg23999170	<i>SBP</i>	5	-1.1	4.7	0.8320	0.7656
	<i>DBP</i>	5	1.9	3.0	0.5631	0.8470

**Table S7 Two-sample inverse variance-weighted Mendelian randomization tests for causal relationships of CpG sites with BP (forward causality).** Estimates of meQTL association with CpGs were derived from meta-analyses of ARIC, FHS, RS, and WHI-EMPC. Estimates of meQTL association with BP are from 1000Genomes analysis in the International Consortium for Blood Pressure. All samples are of European ancestry.

Abbreviations: CpG, cytosine-phosphate-guanine; DBP, diastolic blood pressure; IV, instrumental variable; meQTL, methylation quantitative trait locus; SBP, systolic blood pressure; SNP, single nucleotide polymorphism.



		Forward Mendelian Randomization: CpG → BP						Reverse Mendelian Randomization: BP → CpG					
		Inverse-Variance Weighted Test for Causality				Egger Test		Inverse-Variance Weighted Test for Causality				Egger Test	
CpG	Trait	IV SNPs, n	Mean estimate	(SE)	p value	Pleiotropy p	Causal p	IV SNPs, n	Mean estimate	(SE)	p value	Pleiotropy p	Causal p
cg00533891	SBP	6	-10.3	(13.5)	0.3130	0.0318	0.1911	29	-8.3E-04	(3.8E-04)	0.0388	0.3590	0.0358
	DBP	6	-14.9	(7.8)	0.0405	0.1299	0.1135	29	-2.0E-03	(5.7E-04)	0.0013	0.3785	0.0065
cg00574958	SBP	-						29	1.4E-04	(1.2E-04)	0.2301	0.9356	0.4243
	DBP	-						29	1.4E-04	(1.9E-04)	0.4346	0.0254	0.0409
cg02711608	SBP	3	-31.1	(30.2)	0.2953	0.3424	0.7072	29	-5.5E-04	(2.3E-04)	0.0204	0.4964	0.0361
	DBP	3	-30.2	(17.3)	0.1879	2.4E-05	0.5167	29	-7.7E-04	(3.7E-04)	0.0495	0.6032	0.1057
cg06690548	SBP	-						29	-4.3E-04	(3.5E-04)	0.2267	0.4193	-
	DBP	-						29	-1.7E-04	(5.7E-04)	0.7724	0.8800	-
cg08035323	SBP	7	20.9	(11.1)	0.0091	0.5892	-	29	-4.1E-04	(3.5E-04)	0.2206	0.6500	-
	DBP	7	15.1	(6.4)	0.0111	0.5803	-	29	-1.2E-03	(5.7E-04)	0.0226	0.8343	-
cg10601624	SBP	-						29	-3.6E-04	(2.2E-04)	0.1069	0.1527	0.9929
	DBP	-						29	-1.0E-03	(3.5E-04)	0.0033	0.0164	0.9232
cg14476101	SBP	7	-2.5	(14.3)	0.8669	0.5364	-	29	-1.3E-04	(5.0E-04)	0.7977	0.2059	-
	DBP	7	1.0	(5.6)	0.8623	0.7469	-	29	-2.4E-04	(8.2E-04)	0.7757	0.5550	-
cg16246545	SBP	6	-10.7	(17.5)	0.5684	0.2205	-	29	-1.3E-04	(4.1E-04)	0.7618	0.0758	-
	DBP	6	-1.8	(6.7)	0.8042	0.4186	-	29	-5.1E-05	(6.7E-04)	0.9404	0.1424	-
cg17061862	SBP	10	-8.6	(10.2)	0.4224	0.0959	-	29	1.6E-04	(3.5E-04)	0.6574	0.8119	-
	DBP	10	4.8	(5.1)	0.1112	0.9740	-	29	1.1E-05	(5.7E-04)	0.9844	0.4935	-
cg18120259	SBP	-						29	6.8E-05	(3.1E-04)	0.8084	0.6888	-
	DBP	-						29	-1.8E-04	(5.1E-04)	0.6968	0.4091	-
cg19693031	SBP	-						29	3.3E-04	(3.8E-04)	0.3889	0.6004	-
	DBP	-						29	5.9E-04	(6.2E-04)	0.3509	0.6332	-
cg22304262	SBP	9	0.4	(11.4)	0.9696	0.8178	-	29	-5.4E-04	(3.0E-04)	0.0516	0.4670	-
	DBP	9	-8.5	(6.5)	0.2219	0.9145	-	29	-9.3E-04	(4.9E-04)	0.0367	0.2711	-
cg23999170	SBP	5	5.9	(18.4)	0.7547	0.2738	0.3934	29	4.1E-04	(3.1E-04)	0.1954	0.8319	-
	DBP	5	-1.2	(10.6)	0.9151	0.0363	0.1370	29	2.7E-04	(5.1E-04)	0.6080	0.9852	-

**Table S8 Multi-instrument bidirectional Mendelian randomization tests for causality between DNA methylation and BP using inverse-variance weighted effects and Egger regression tests.** Tests for causality based on Egger regression have substantially reduced power compared to inverse-variance weighted tests. Egger causal p values are provided only when pleiotropic effects of IVs are significant for either BP trait ( $p < 0.05$ ) or when both the inverse-variance weighted and Egger tests both support reverse causation.

Abbreviations: BP, blood pressure; CpG, cytosine-phosphate-guanine; DBP, diastolic blood pressure; IV, instrumental variable; SBP, systolic blood pressure; SE, standard error; SNP, single nucleotide polymorphism.

CpG site	Chr	Position	1) Genes tested for GE-Methylation, n	2) Genes tested for GE-BP significant in 1), n	Significant genes in triangular GE analysis, n
cg23999170	1	115628111	22	5	2
cg16246545	1	120255941	20	1	0
cg14476101	1	120255992	21	1	0
cg19693031	1	145441552	32	0	-
cg08035323	2	9843525	22	0	-
cg06690548	4	139162808	8	1	1
cg18120259	6	43894639	50	0	-
cg00533891	10	80919242	14	0	-
cg17061862	11	9590431	26	3	0
cg00574958	11	68607622	30	4	2
cg10601624	12	6404377	61	10	2
cg22304262	19	47287778	52	8	0
cg02711608	19	47287964	52	7	0

**Table S9 Number of gene expression analyses per replicated CpG site used for FDR-correction of p values.** Position is in Hg19.

Abbreviations: BP, blood pressure; Chr, chromosome; CpG, cytosine-phosphate-guanine; GE, gene expression.

Supplemental Excel file provided.

**Table S10 Genes in a *cis*-region (+/- 1Mb) of replicated CpG sites 1) suggestively associated with methylation in meta-analyses of FHS and RS at p value <0.05, and 2) associated with blood pressure traits with at least one p value <0.05.** Start and stop positions are in Hg19.

Abbreviations: Chr, chromosome; Coeff, coefficient; Corr, correlation coefficient for continuous traits; CpG, cytosine-phosphate-guanine; DBP, diastolic blood pressure; HTN, hypertension; SBP, systolic blood pressure; UCSC, University of California Santa Cruz.

CpG site	Chr	Position	UCSC Gene	Phenotype(s)
cg23999170	1	115628111	<i>TSPAN2</i>	
cg16246545	1	120255941	<i>PHGDH</i>	Alcohol intake <sup>1</sup> , BMI <sup>2</sup>
cg14476101	1	120255992	<i>PHGDH</i>	Lipids <sup>3</sup> , A-diol <sup>4</sup> , BMI <sup>2,5</sup> , waist circumference <sup>5</sup>
cg19693031	1	145441552	<i>TXNIP</i>	Lipids <sup>3,4</sup> , type 2 diabetes <sup>6-9</sup>
cg08035323	2	9843525		
cg06690548	4	139162808	<i>SLC7A11</i>	Lipids <sup>3</sup> , A-diol <sup>4</sup> , BMI <sup>2</sup> , IgE <sup>10</sup>
cg18120259	6	43894639	LOC100132354	A-diol <sup>4</sup> , BMI <sup>2</sup>
cg00533891	10	80919242	<i>ZMIZ1</i>	
cg17061862	11	9590431		
cg00574958	11	68607622	<i>CPT1A</i>	Lipids <sup>3,4,11</sup> , BMI <sup>2,5,8,12</sup> , waist circumference <sup>5</sup> , metabolic syndrome <sup>13</sup> , hypertriglyceridemic waist <sup>14</sup> , adiponectin <sup>15</sup>
cg10601624	12	6404377		
cg22304262	19	47287778	<i>SLC1A5</i>	A-diol <sup>4</sup>
cg02711608	19	47287964	<i>SLC1A5</i>	Lipids <sup>3</sup> , BMI <sup>2</sup>

**Table S11 Phenotypes previously associated with the 13 CpG sites replicated for association with BP.** Position is in Hg19.

Abbreviations: A-diol, 4-androsten-3-beta,17-beta-diol disulfate; BMI, body mass index; Chr, chromosome; Coeff, coefficient; CpG, cytosine-phosphate-guanine; DBP, diastolic blood pressure; HTN, hypertension; IgE, immunoglobulin E; SBP, systolic blood pressure; UCSC, University of California Santa Cruz.

SNP	Chr	CpG	CpG position (bp)	SNP-CpG distance (bp)	SNP-CpG		CpG-BP EWAS		Additional GWAS loci in high LD ( $r^2>0.8$ ) with Kato SNP (ARIC SNP-CpG p value)
					ARIC p <sup>a</sup> N=790	Kato p <sup>b</sup> N=6,684	SBP p N=17,010	DBP p N=17,010	
rs880315	1	cg02903756	10,750,680	46,186	NR	4.69E-12	0.7525	0.7593	
rs12567136	1	cg05228408	11,865,352	18,379	NR	3.01E-71	0.1157	0.4805	
		cg20946054	11,761,766	121,965	<b>8.09E-05</b>	NR	0.0318	0.0015	rs17367504 ( <b>1.69E-04</b> )
rs1344653	2	cg13996430	19,741,587	-10,742	NR	5.09E-17	0.6825	0.7047	
		cg17314700	19,548,291	182,554	8.61E-03	NR	0.8128	0.4091	
rs1275988	2	cg02952978	27,165,719	-251,355	<b>1.32E-03</b>	NR	0.2293	0.2035	
		cg19115882	26,919,145	-4,781	NR	2.09E-32	0.0041	0.0174	
rs7629767	3	cg03022575	42,003,672	39,837	2.43E-03	NR	0.4034	0.2647	rs3774372 ( <b>5.21E-04</b> ), rs9815354 ( <b>6.73E-04</b> )
		cg02108620	42,002,230	41,279	NR	2.54E-152	0.9744	0.1441	
rs13149993	4	cg05974274	81,123,369	35,176	8.65E-03	NR	0.1271	0.4145	
		cg05452645	81,117,647	40,898	NR	2.38E-29	0.0010	0.0074	
rs2014912	4	cg01368160	87,187,259	-471,589	9.89E-03	NR	0.0323	0.4454	
		cg20784207	86,597,598	118,072	NR	3.57E-37	0.1274	0.8120	
rs7733331	5	cg24363955	32,788,467	40,379	NR	1.89E-28	0.7480	0.9826	
		cg15198736	32,709,396	119,450	<b>5.82E-04</b>	NR	0.8623	0.1140	rs1173771 ( <b>6.49E-04</b> )
rs13359291	5	cg07849972	122,759,670	-283,213	3.03E-03	NR	0.5951	0.8575	
		cg23290100	122,435,626	40,831	NR	3.34E-189	0.0337	0.7627	
rs9687065	5	cg12302647	148,533,875	-142,735	2.65E-02	NR	0.8404	0.5647	
		cg18129178	148,520,854	-129,714	NR	4.05E-50	0.5191	0.8933	
rs11960210	5	cg22790839	157,883,933	-66,299	NR	6.09E-27	0.8453	0.7847	
		cg12515908	157,413,937	403,697	7.43E-03	NR	0.7405	0.5791	
rs1563788	6	cg12585005	43,655,766	-347,403	<b>1.36E-03</b>	NR	0.4349	0.4140	
		cg00084398	43,249,983	58,380	NR	3.76E-48	0.4597	0.1026	
rs17080102	6	cg01354656	151,411,710	-406,940	9.02E-03	NR	0.9834	0.9821	
		cg02784464	151,121,916	-117,146	NR	2.35E-17	0.5820	0.8370	
rs10260816	7	cg12244052	45,961,469	48,631	NR	1.24E-08	0.5673	0.8391	
		cg13070193	45,613,752	396,348	<b>6.41E-04</b>	NR	0.4437	0.9674	
rs731141	10	cg10751070	96,143,568	-244,887	NR	1.43E-08	0.8070	0.9242	
		cg04143348	95,656,819	241,862	<b>9.18E-04</b>	NR	0.8809	0.9756	rs9663362 ( <b>6.44E-04</b> ), rs932764 (4.18E-03)
rs11191375	10	cg03275084	104,617,031	-152,374	3.56E-02	1.75E-06	0.0016	0.0066	
		cg07119830	104,412,306	52,351	NR	3.E-746	0.6838	0.4193	

		cg12331743	104,153,933	310,724	5.94E-03	NR	0.0484	0.0096	
rs2484294	10	cg20663200	116,163,392	-371,330	<b>1.68E-03</b>	NR	0.8672	0.2613	
		cg05575054	115,804,968	-12,906	NR	4.83E-27	0.0163	0.0283	
rs751984	11	cg00009053	61,283,865	-5,619	NR	2.76E-58	0.7452	0.3231	
		cg26786382	61,277,328	918	5.65E-03	1.69E-06	0.0628	0.2037	
rs2055450	11	cg11637980	101,000,717	-450,300	3.10E-03	NR	0.3830	0.9702	rs633185 ( <b>1.73E-03</b> )
		cg05925497	100,734,094	-183,677	2.07E-02	1.09E-12	0.2339	0.6334	
rs10894192	11	cg03927812	130,271,903	-5,786	NR	1.31E-55	0.1087	0.4554	
		cg04713042	129,912,454	353,663	<b>1.49E-03</b>	NR	0.8072	0.6620	rs11222084 (3.94E-03)
rs11105354	12	cg00757033	89,920,650	105,873	NR	1.91E-194	0.7233	0.8567	
		cg06869160	89,919,815	106,708	2.58E-02	NR	0.5176	0.4349	rs2681472 (2.29E-02), rs2681492 (2.06E-02), rs17249754 (2.24E-02)
rs3184504	12	cg16423624	112,220,997	-336,389	6.35E-03	NR	0.4666	0.2621	rs653178 (2.29E-03)
		cg10833066	111,807,467	96,904	NR	1.29E-48	0.2043	0.4457	
rs1378942	15	cg16377819	75,575,161	-497,794	<b>1.44E-03</b>	NR	0.4735	0.7943	
		cg02696790	75,250,997	-173,630	NR	1.42E-86	0.0262	0.0234	
		cg20668952	75,082,885	-5,518	2.32E-02	6.54E-09	0.0521	0.0665	
rs8032315	15	cg06330618	91,428,456	-10,159	NR	2.01E-167	0.1058	0.4800	
		cg02864248	91,411,838	6,459	1.18E-02	2.91E-16	0.4764	0.4012	
		cg01994513	91,260,603	157,694	2.22E-03	NR	0.7373	0.2635	rs2521501 ( <b>1.58E-03</b> )
rs2301597	17	cg24180402	43,221,464	-48,191	1.96E-02	4.39E-30	0.0039	0.0999	
		cg19407385	43,099,144	74,129	NR	3.51E-283	0.3725	0.6644	
		cg07822074	43,098,904	74,369	8.13E-03	8.36E-27	0.5166	0.7629	
		cg06471905	42,835,688	337,585	<b>6.28E-04</b>	NR	0.9597	0.4451	
rs7405452	17	cg06602723	46,693,336	-18,666	2.00E-02	2.34E-16	0.8131	0.7141	
		cg22053945	46,651,360	23,310	NR	1.05E-151	0.1920	0.3117	
		cg04171235	46,604,393	70,277	4.45E-02	1.75E-14	0.8655	0.4466	
		cg08036188	46,522,560	152,110	3.01E-03	NR	0.3476	0.1459	
rs2240736	17	cg06762332	59,573,001	-87,608	2.25E-03	NR	0.9046	0.5538	
		cg00730441	59,483,863	1,530	NR	6.66E-131	0.8128	0.9765	
rs740406	19	cg14103263	2,727,148	-494,927	<b>1.24E-03</b>	NR	0.4348	0.7746	
		cg04052466	2,251,061	-18,840	NR	2.33E-45	0.7375	0.7853	

**Table S12 Top associations of methylation +/-500 kb of methylation-mediated GWAS SNPs identified by Kato et al.** The meQTL analyses were conducted among ARIC whites (N=790) using methylation data +/-500 kb of each sentinel SNP reported by Kato et al. Genotypes were directly measured or imputed to

1000Genomes. meQTL p values are shown for lead CpGs reported either by Kato et al or in ARIC analyses and any non-lead CpGs (in the online supplement) with p value <0.05 in ARIC. ARIC meQTL analyses were Bonferroni-corrected for the 28 GWAS loci reported by Kato et al (p value <0.0018), with significant p values shown in bold. Lookups in the overall meta-analysis for association of each CpG with systolic and diastolic BP is shown (N=17,058). Position is in Hg19. The 1000Genomes CEU reference population was used to calculate  $r^2$ .

<sup>a</sup>meQTL p values not reported (NR) in ARIC are >0.05. <sup>b</sup>meQTL p values not reported (NR) in Kato et al are >3.8E-6 in discovery (N=1,904). Combined p values for discovery and replication are shown (N=6,684).

Abbreviations: bp, base pairs; BP, blood pressure; Chr, chromosome; Coeff, coefficient; CpG, cytosine-phosphate-guanine; DBP, diastolic blood pressure; GWAS, genome-wide association studies; LD, linkage disequilibrium; NR, not reported; SBP, systolic blood pressure; SNP, single nucleotide polymorphism.



SNP	Chr	CpG	SNP-CpG distance (bp)	Kato CpG?	Systolic BP					Diastolic BP				
					Coeff	p value	CpG		Coeff Percent Change	Coeff	p value	CpG		Coeff Percent Change
							Adj	p value				Adj	p value	
rs12567136	1	cg05228408	18,379	Yes	2.45	0.0366	0.75	0.5817	69%	1.90	0.0039	1.81	0.0178	4%
rs12567136	1	cg20946054	121,965		2.45	0.0366	1.34	0.2908	45%	1.90	0.0039	1.74	0.0145	8%

**Table S13 Adjustment of the association of a GWAS locus (*CLCN6* rs12567136) for DNA methylation at two CpG sites among ARIC participants of European ancestry (N=790). Position is in Hg19.**

Abbreviations: Adj, adjusted; bp, base pairs; BP, blood pressure; Chr, chromosome; Coeff, coefficient; CpG, cytosine-phosphate-guanine; GWAS, genome-wide association studies; SNP, single nucleotide polymorphism.

## 1. Cohort Descriptions and Acknowledgements

### *The Amish Complex Disease Research Studies (Amish)*

The Old Order Amish (OOA) subjects included in this study were participants of several studies of cardiovascular health in relatively healthy volunteers from the OOA community of Lancaster County, PA and their family members. The studies were carried out at the University of Maryland as part of the Amish Complex Disease Research Program (ACDRP); see <http://medschool.umaryland.edu/endocrinology/amish/research-program.asp>. The OOA population of Lancaster County, PA immigrated to the Colonies from Western Europe in the early 1700's. There are now over 30,000 OOA individuals in the Lancaster area, nearly all of whom can trace their ancestry back 12-14 generations to approximately 750 founders. Investigators at the University of Maryland, School of Medicine have been studying the genetic determinants of cardiometabolic health in this population since 1993. To date, over 7,000 Amish adults have participated in one or more of our studies. The subjects on whom the methylation chip was used were participants of the Heredity and Phenotype Interaction (HAPI) heart study<sup>16</sup>, the Pharmacogenomics of Anti-Platelet Intervention (PAPI) study<sup>17</sup>, or the Amish Family Diabetes Study (AFDS)<sup>18</sup>. These studies collected large numbers of variables including demographic and anthropometric information, medical history, clinical characteristics, lifestyle factors, and study specific variables, as well as blood and urine samples. Blood pressure was measured in triplicate using a standard sphygmomanometer in the sitting position after 5 minutes rest, and the average of the 3 measures was used for analysis. All study protocols were approved by the institutional review board at the University of Maryland and participating institutions. Informed consent was obtained from each of the study participants.

Genomic DNA was extracted from whole blood of 432 samples, and quantitated using picogreen. Bisulfite conversion of 500 ng genomic DNA was performed using the EZ DNA Methylation Kit (Zymo Research; Irvine, CA, USA) according to the manufacturer's instructions, and the Illumina Infinium Human Methylation 450K Bead chip array (HM450K)<sup>19</sup> was used for methylation profiling. Sample preparation and hybridization were performed by the same technician, and the arrays were processed using the same scanner to reduce batch effect. Genome Studio (GS) (V2011.1) was used for processing the raw intensities and performing QC for samples and probes. Examining internal control probes identified one array that did not perform well and all 12 samples in it were missing 25% of the data, so we completely excluded this array. GS normalization and background correction were applied according to Illumina's recommendation for the remaining 420 samples. By testing Y chromosome probes we identified 19 potentially mismatched or contaminated samples that were removed. We also removed one from each pair of 18 sets of duplicates leaving 384 samples for analysis all with >95% of probes detected with average p value <0.01. Blood pressure measures were available for only 192 individuals at the same time of blood draw. Probes with average detection p value >0.01 or call rate <95% were removed. Also sex chromosome probes and previously identified cross-reactive probes<sup>20-22</sup> were excluded leaving 323,747 probes for analysis. Blood cell subtype composition were estimated using the Housman method<sup>23,24</sup>.

### *Atherosclerosis Risk in Communities (ARIC)*

The Atherosclerosis Risk in Communities (ARIC) study is a prospective cohort study of cardiovascular disease risk in four U.S. communities<sup>25</sup>. Between 1987 and 1989, men and women aged 45–64 years were recruited from four sites: Forsyth County, North Carolina; Jackson, Mississippi (African Americans only); suburban Minneapolis, Minnesota; and Washington County, Maryland. The ARIC study protocol was approved by the institutional review board of each participating university. After written informed consent was obtained, including consent for genetic studies, participants underwent a baseline clinical examination and four subsequent follow-up clinical examinations (visits 1-5). Systolic and diastolic blood pressures were measured three times using a random zero sphygmomanometer with subjects in a seated position. The average of the second and third readings was used in analyses. DNA methylation data are available for

African American members of the cohort from two study sites (Forsyth County and Jackson) and were included in the discovery meta-analysis. Measurements of DNA methylation for white members of the cohort later became available and were included in the replication meta-analysis. Cross-sectional analyses within each race were conducted using blood pressure, methylation, and covariates measured concurrently at visit 2 or visit 3.

Genomic DNA was extracted from peripheral blood leukocyte samples using the Genra Puregene Blood Kit (Qiagen; Valencia, CA, USA). DNA samples were bisulfite-converted using the EZ-96 DNA Methylation Kit (Deep Well Format) (Zymo Research; Irvine, CA, USA) and hybridized to the Illumina HumanMethylation 450K beadarray (Illumina, Inc., San Diego, CA, USA). Probe intensities were extracted using Illumina GenomeStudio 2011.1, Methylation module 1.9.0 software. Samples were removed based on pass rate <95%, gender mismatch, SNP discordance with previous genotyping, and outliers in principal component analysis (N = 107). Additionally, CpG sites missing in ≥5% of samples were excluded. Methylation values were normalized using the Beta Mixture Quantile dilation (BMIQ) method<sup>26</sup>. Blood cell proportions were imputed using the Houseman method<sup>23</sup> among African Americans and Houseman cell types from the Horvath method<sup>27</sup> among white individuals. Surrogate variables were estimated among African Americans after removing the effects of age, sex, BMI, smoking, and the first four principal components of ancestry.

The Atherosclerosis Risk in Communities (ARIC) study is carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The authors thank the staff and participants of the ARIC study for their important contributions. Funding support for “Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium” was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). This project was funded from R01-NS087541 to Myriam Fornage and Eric Boerwinkle.

#### *Cardiovascular Health Study (CHS)*

The Cardiovascular Health Study (CHS) is a population-based cohort study of risk factors for coronary heart disease and stroke in adults ≥65 years conducted across four field centers. The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists; subsequently, an additional predominantly African-American cohort of 687 persons was enrolled for a total sample of 5,888. DNA methylation was measured on 200 European ancestry and 200 African-American ancestry participants. The samples were randomly selected among participants without presence of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack at study baseline or lack of available DNA at study year 5.

CHS was approved by institutional review committees at each field center and individuals in the present analysis had available DNA and gave informed consent including consent to use of genetic information for the study of cardiovascular disease. Methylation measurements were performed at the Institute for Translational Genomics and Population Sciences at the Harbor-UCLA Medical Center Institute for Translational Genomics and Population Sciences using the Infinium HumanMethylation450 BeadChip (Illumina Inc, San Diego, CA). Quality control was performed in the minfi R package (version 1.12.0, <http://www.bioconductor.org/packages/release/bioc/html/minfi.html>). Samples with low median intensities of below 10.5 (log<sub>2</sub>) across the methylated and unmethylated channels, samples with a proportion of probes falling detection of greater than 0.5%, samples with QC probes falling greater than 3 standard deviation from the mean, sex-check mismatches, failed concordance with prior genotyping or > 0.5% of probes with a detection p-value > 0.01 were removed. Methylation values were normalized using the SWAN<sup>28</sup> quantile normalization method. White blood cell proportions were not directly measured in CHS and were estimated from the methylation data using the Houseman

method<sup>23</sup>.

Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. The CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, U01HL130114, R01HL087652, R01HL092111, R01HL105756, R01HL103612, R01HL111089, R01HL120393, and R01HL130114 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA) as well as Laughlin Family, Alpha Phi Foundation, and Locke Charitable Foundation. A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

### *Framingham Heart Study (FHS)*

The Framingham Heart Study Offspring cohort (FHS-Offspring) was initially recruited in 1971 and included 5,124 offspring of the FHS Original cohort. From 2002 to 2005, the adult children (third generation cohort, N=4,095) of the offspring cohort participants were recruited and examined (FHS-3rd Gen). Detailed descriptions of cohorts have been published<sup>29,30</sup>. A total of 2,836 FHS-Offspring participants who attended the eighth exam cycle from 2005-2008 were included in the DNA methylation meta-analysis study. The 1,549 FHS-3rd Gen participants who attended the second exam cycle from 2005-2008 were used as the validation set for testing the proportion of SBP / DBP phenotypic variation explained by the top BP associated DNA methylation loci identified in this study. All participants provided written consent for genetic research.

DNA methylation measurements of the FHS Offspring were conducted in two laboratories: lab 1 samples included an ongoing case-control study of cardiovascular disease (n=576), and lab 2 included the remainder of the offspring cohort (n=2,270). DNA methylation measurements of the FHS 3rd Generation participants (n=1,549) were conducted in one laboratory. Buffy coat preparations were obtained from the whole blood samples and genomic DNA was extracted using the Gentra Puregene Blood Kit (Qiagen, Venlo, Netherlands) according to the manufacturer's instructions. Subsequently, bisulfite conversion of 1 ug genomic DNA was performed (Zymo Research, Irvine, CA, USA) and bisulfite conversion efficiency was determined by PCR amplification of the converted DNA using Universal Methylated Human DNA Standard and Control Primers (Zymo Research, Irvine, CA, USA). The Infinium HumanMethylation450 array (Illumina, San Diego, CA) was used to quantify genome-wide DNA methylation and Illumina GenomeStudio software was used to estimate  $\beta$  scores.  $\beta$  scores underwent normalization within the two laboratory batches using the DASEN methodology implemented in the watermelon R package<sup>31</sup> which includes background adjustment of the methylated and unmethylated intensities and quantile normalization of the methylated and unmethylated probes within the two types of probes technologies separately. Quality control filters excluded samples with missing rate > 1% at detection P-value <0.01, poor matching to the 65 single nucleotide polymorphism (SNP) control probe locations, and outliers by multi-dimensional scaling techniques. For quality control at the probe level, those with missing rate >20% at detection P-value <0.01, as well as probes previously identified to map to multiple locations or have an underlying SNP (minor allele frequency >5% in European ancestry 1000 genomes project data) at the CpG site or  $\leq$ 10 bp of the single base extension (n=42,251) were excluded, as were probes on the sex chromosomes.

These procedures left 2,377 FHS-Offspring (n=442 for batch 1 and n=1,935 for batch 2) and 1,522 FHS-3rd Gen participants with phenotype data, and 443,252 probes for analyses. Surrogate variable analysis (SVA) was used to adjust for unmeasured technical and batch effects, including cell count composition<sup>23</sup> SVAs were created within laboratory batches and cohorts after removing the signals of interest from BMI, age, sex, smoking status, systolic blood pressure, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, fasting plasma glucose, and exam date. SVAs that were associated with SBP / DBP with a P-value < 0.05 were included in subsequent analysis of testing associations of DNA methylation and BP.

Association testing was performed using linear mixed effect regression models, with DNA methylation  $\beta$  score as the dependent variable, SBP or DBP as the independent variables, age, sex, BMI, and SVAs as fixed effects, and family structure as a random effect in the FHS-Offspring cohort. For individuals receiving antihypertensive treatment, treatment-adjusted SBP and DBP were calculated by adding 10 mm Hg and 5 mm Hg to the measured SBP and DBP values, respectively. Regression models were fit for the two laboratory batches separately. An inverse variances weighted meta-analysis of the two batches were conducted by the *metagen()* function in the R package Meta (<http://cran.r-project.org/web/packages/meta/index.html>). These meta-analysis p values were used to represent the best estimate of the associations of the replication probes in the FHS-Offspring cohort.

The Framingham Heart Study is funded by National Institutes of Health contract N01- HC-25195. The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health and an NIH Director's Challenge Award (D. Levy, Principal Investigator). Additional support was from the NHLBI award K99HL136875 (M. Mendelson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

#### *Genetic Epidemiology Network of Atherosclerosis (GENOA)*

The Genetic Epidemiology Network of Arteriopathy (GENOA) study is a community-based study of hypertensive sibships that was designed to investigate the genetics of hypertension and target organ damage in African Americans from Jackson, Mississippi and non-Hispanic whites from Rochester, Minnesota<sup>32</sup>. In the initial phase of the GENOA study (Phase I: 1996-2001), all members of sibships containing  $\geq 2$  individuals with essential hypertension clinically diagnosed before age 60 were invited to participate, including both hypertensive and normotensive siblings. Exclusion criteria of the GENOA study were secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. Eighty percent of African Americans (1,482 subjects) and 75% of non-Hispanic whites (1,213 subjects) from the initial study population returned for the second examination (Phase II: 2001-2005). Study visits were made in the morning after an overnight fast of at least eight hours. Demographic information, medical history, clinical characteristics, lifestyle factors, and blood samples were collected in each phase. Written informed consent was obtained from all subjects and approval was granted by participating institutional review boards. DNA methylation levels were measured only in African Americans participants, so only African Americans were included in the current analysis. Participants were excluded from this analysis if they were also participants in the ARIC study.

Genomic DNA of 422 participants was extracted from stored peripheral blood leukocytes, bisulfite converted, and then measured for DNA methylation using the Illumina Infinium HumanMethylation450 BeadChip. DNA was extracted from stored blood samples collected during the Phase I examination. The Minfi R package<sup>33</sup> was used to preprocess, normalize (SWAN<sup>28</sup>), and calculate beta values. The proportions of each cell type were estimated using Houseman's method<sup>23</sup>. Detection p-values were calculated for each site, and beta values were set to missing if a site had detection p-value>0.01. In all samples, > 95% of probes had a detection P-value<0.01; thus, no samples were excluded from analysis. A total of 4,070 probes were removed due to detection P-value>0.01 in >1% of samples.

Support for the Genetic Epidemiology Network of Arteriopathy was provided by the National Heart, Lung and Blood Institute (HL054457, HL100185, HL119443, and HL133221) of the National Institutes of Health. We appreciate technical assistance from Jodie L. Van de Rostyne, Pamela I. Hammond, Julie M. Cunningham, and the Mayo Clinic Advanced Genomics Technology Center. We would also like to thank the families that participated in the GENOA study.

#### *Genetics of Lipid Lowering Drugs and Diet Network (GOLDN)*

The National Heart, Lung, and Blood Institute GOLDN study was designed to identify genetic determinants of lipid response to two interventions (a high-fat meal challenge and fenofibrate treatment for 3 weeks). Briefly, the study ascertained and recruited families from the Family Heart Study at two centers, Minneapolis, MN and Salt Lake City, UT, who self-reported to be white. The study protocol was approved by Institutional Review Boards at the University of Minnesota, University of Utah, and Tufts University/New England Medical Center. For the current study we evaluated fasting systolic and diastolic blood pressure among 991 participants for whom epigenetic data were available. Resting blood pressure was measured twice between 7:00 and 11:00 am in a sitting position. The average of the two measurements was used in this analysis.

DNA was extracted from CD4+ T-cells harvested from stored buffy coats using antibody-linked Invitrogen Dynabeads. We lysed cells captured on the beads and extracted DNA using DNeasy kits (Qiagen, Venlo, Netherlands). We used the Illumina Infinium Human Methylation450 Beadchip (Illumina Inc, San Diego, CA) to interrogate ~470,000 autosomal CpG sites across the genome.<sup>10</sup> A description of the array as well as CpG site nomenclature conventions can be found at [http://www.illumina.com/products/methylation\\_450\\_beadchip\\_kits.ilmn](http://www.illumina.com/products/methylation_450_beadchip_kits.ilmn). For each assay, 500ng of DNA was treated with sodium bisulfite (EZ DNA, Zymo Research, Irvine, CA) prior to standard Illumina amplification, hybridization, and imaging steps. The resulting intensity files were analyzed with Illumina's GenomeStudio which generated beta scores (ie, the proportion of total signal from the methylation specific probe or color channel) and "detection *P*-values" ( defined as the 1-*p*-value computed from the background model characterizing the chance that the target sequence signal was distinguishable from the negative controls). Beta scores with an associated detection *P*-value greater than 0.01 were removed and samples with more than 1.5% missing data points were eliminated from further analysis. Furthermore, any CpG probes where more than 10% of samples failed to yield adequate intensity were removed. A total of 58 samples were removed. The filtered beta scores were then subjected to batch normalization with the ComBat package for R software in non-parametric mode (<http://www.bu.edu/jlab/wp-assets/ComBat/Abstract.html>). To correct for probe chemistry we separately normalized probes from the Infinium I and II chemistries and subsequently adjusted the  $\beta$  scores for Infinium II probes using the equation derived from fitting a second order polynomial to the observed methylation values across all pairs of probes located <50bp apart (within-chemistry correlations >0.99), where one probe was Infinium I and one was Infinium II. Principal components (PCs) based on the beta scores of all autosomal CpGs passing QC were generated using the *prcomp* function in R (V 2.12.1) and used to adjust for cell purity in association analysis.

This work has been funded by the NHLBI grant U01HL072524-04 and American Heart Association (AHA) Cardiovascular Genome-Phenome Study (CVGPS) grant 15GPSPG23760006.

#### *Lothian Birth Cohort 1936 (LBC1936)*

The Lothian Birth Cohort of 1936 is a follow-up study of the Scottish Mental Survey of 1947. The survey had tested the intelligence of almost every child born in 1936 and attending school in Scotland in the month of June 1947. DNA methylation and blood pressure concurrently measured at a mean age of 69.5 years was used for analyses. Blood pressure was estimated as the average of three sitting systolic and diastolic readings from an Omron 705IT monitor. DNA methylation was assessed using the Illumina Human Methylation 450k BeadChip from whole blood of consenting

participants (n=1,005). Background correction and quality control were performed to remove probes with a low detection rate, low quality based on manual inspection, and samples with a mismatch between genotypes and SNP control probes, and incorrectly predicted sex. Full details are provided in Shah et al.<sup>34</sup>

We thank the cohort participants and team members who contributed to these studies. This work was supported by numerous funding bodies. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Methylation typing was supported by the Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of Edinburgh, and The University of Queensland. REM and IJD are members of the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE). CCACE is supported by funding from the BBSRC, the Medical Research Council (MRC), and the University of Edinburgh as part of the cross-council Lifelong Health and Wellbeing initiative (MR/K026992/1).

### *Multi-Ethnic Study of Atherosclerosis (MESA)*

The Multi-Ethnic Study of Atherosclerosis (MESA) was designed to investigate the prevalence, correlates, and progression of subclinical cardiovascular disease in a population cohort of 6,814 participants. Since its inception in 2000, five clinic visits collected extensive clinical, socio-demographic, lifestyle, behavior, laboratory, nutrition, and medication data<sup>35</sup>. DNA methylation and gene expression were measured in purified (CD14+) monocyte samples from the April 2010 – February 2012 examination (exam 5) of 1,264 randomly selected MESA participants from four MESA field centers (Baltimore, MD; Forsyth County, NC; New York, NY; and St. Paul, MN) as previously described<sup>36</sup>. The study protocol was approved by the Institutional Review Board at each site. All participants signed informed consent.

As previously described<sup>36</sup>, blood was initially collected in sodium heparin-containing Vacutainer CPT™ cell separation tubes (Becton Dickinson, Rutherford, NJ, USA) to separate peripheral blood mononuclear cells from other elements within 2 h from blood draw. Subsequently, monocytes were isolated with the anti-CD14-coated magnetic beads, using AutoMACs automated magnetic separation unit (Miltenyi Biotec, Bergisch Gladbach, Germany). Based on flow cytometry analysis of 18 specimens, monocyte samples were consistently >90% pure. DNA and RNA were isolated from samples simultaneously using the AllPrep DNA/RNA Mini Kit (Qiagen, Inc., Hilden, Germany). DNA and RNA QC metrics included optical density measurements, using a NanoDrop spectrophotometer and evaluation of the integrity of 18s and 28s ribosomal RNA. Illumina HumanMethylation450 BeadChips and HiScan reader were used to perform the epigenome-wide methylation analysis. Bead-level methylation data were summarized in GenomeStudio. Because a two-channel system and both Infinium I and II assays were used, normalization was performed in several steps using the lumi package. “Smooth quantile normalization” was used to adjust for color bias. Next, the data were background adjusted by subtracting the median intensity value of the negative control probes. Lastly, data were normalized across all samples by standard quantile normalization applied to the bead-type intensities and combined across Infinium I and II assays and both colors. QC measures included checks for sex and race/ethnicity mismatches, and outlier identification by multidimensional scaling plots. To estimate residual sample contamination for data analysis, we generated separate enrichment scores for neutrophils, B cells, T cells, monocytes, and natural killer cells. We implemented a Gene Set Enrichment Analysis<sup>37</sup> as previously described<sup>36</sup> to calculate the enrichment scores using the gene signature of each blood cell type from previously defined lists<sup>38</sup>. To remove technical error in methylation levels associated with batch effects across the multiple chips, positional effects of the sample on the chip, and residual sample contamination with non-monocyte cell types, we adjusted methylation values for chip, sample position on the chip, and estimated residual sample contamination with neutrophils, B cells, T cells, monocytes, and natural killer cells. The final methylation value for each methylation probe was computed as the beta-value, essentially the proportion of the methylated to the total intensity.

MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-001079, UL1-TR-000040, and DK063491. The MESA Epigenomics & Transcriptomics Study was funded by NHLBI grant R01HL101250 to Wake Forest University Health Sciences. Analysis of MESA data reported in this publication was also supported by the NHLBI under Award Number 26 P50HL120163. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

### *The Normative Aging Study (NAS)*

The Normative Aging Study (NAS) is an ongoing longitudinal study of aging established in 1963 by the US Department of Veterans Affairs (VA). It enrolled community-dwelling men living in the Greater Boston area, 21-80 years old and free of known chronic medical conditions at entry<sup>39</sup>. Subsequently participants were invited to medical examinations every three to five years. DNA samples were collected from active participants between 1999-2013<sup>39</sup>. In total, 668 individuals (650 white, 12 black, 5 Hispanic white, 1 Hispanic black; 279 without medication) with cross-sectional data are included in this analysis. The study was reviewed and approved by the Institutional Review Boards (IRBs) of all participating institutions. All participants have provided written informed consent at each visit. At each in-person visit, participants completed questionnaires regarding demography, life-style and medical history. Chronological age and smoking status (never, former, current) were self-reported. High blood pressure was defined as antihypertensive medication use or SBP  $\geq 140$  mmHg or DBP  $\geq 90$  mmHg at study visit. Body Mass Index (BMI) was computed from anthropometric measures, performed with participants in undershorts and socks<sup>40</sup>.

DNA was extracted from buffy coat using the QIAamp DNA Blood Kit (QIAGEN, Valencia, CA). 500 ng of DNA was used to perform bisulfite conversion using the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA). To reduce the chip and plate effects, we used a two-stage age-stratified algorithm to randomize samples and ensure similar age distributions across chips and plates; 12 samples – which were sampled across all age quartiles – were randomized to each chip, then chips were randomized to plates (each housing eight chips). Quality control analysis was performed to remove samples with a detection p-value  $> 0.05$  in more than 1% probes, and probes with a detection p-value  $> 0.05$  in more than 1% samples, respectively. The remaining samples were preprocessed using noob background correction<sup>41</sup> and dye bias adjustment<sup>42</sup>. Quantile normalization was then performed for methylated and unmethylated intensities of both type I and type II probes with nasen method<sup>31</sup>, followed by BMIQ method<sup>26</sup> for probe types on beta methylation values. After preprocessing there were 484,613 CpG probes in the final working set. We adjusted for patient characteristics including age, BMI, smoking status and technical covariates for possible batch effects. Besides, due to the influence of blood cell count on methylation levels, we additionally adjusted each model for white blood cell (WBC) counts<sup>23</sup> and measured cell proportions. In addition, we used first four PCs calculated from the probes within 50bp of SNPs to account for potential population stratification<sup>43</sup>.

The present work on the US Department of Veterans Affairs (VA) Normative Aging Study has been supported by funding from the U.S. National Institute of Environmental Health Sciences (NIEHS) (R01ES015172, R01ES021733). The VA Normative Aging Study is supported by the Cooperative Studies Program/ERIC, US Department of Veterans Affairs, and is a research component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). Additional support to the VA Normative Aging Study was provided by the US Department of Agriculture, Agricultural Research Service (contract 53-K06-510). The views expressed in this paper are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs.

### *Rotterdam Study (RS)*

The Rotterdam Study is a prospective population based cohort study in a well-defined area of Rotterdam, the Netherlands. The design of the Rotterdam Study has been detailed elsewhere<sup>44</sup>. For the current analysis we used data from individuals aged 45 years and older that participated in the third cohort of the Rotterdam Study. In the first visit of



the third cohort, 3,934 participants were examined between February 2006 and December 2008. Whole blood DNA methylation was quantified in a random subset of 750 individuals with genotyping and RNA expression data available.

DNA was extracted from whole peripheral blood (stored in EDTA tubes) by standardized salting out methods. Genome-wide DNA-methylation levels in 750 subjects from the Rotterdam Study-III were determined using the Illumina HumanMethylation 450K beadarray (Illumina, Inc., San Diego, CA, USA). In short, samples (500ng of DNA per sample) were first bisulfite treated using the Zymo EZ-96 DNA-methylation kit (Zymo Research, Irvine, CA, USA). Next, they were hybridized to the arrays according to the manufacturer's protocol. During quality control samples showing incomplete bisulfite treatment were excluded (n=5) as were samples with a low detection rate (<99%) (n=7), or gender swaps (n=4). Probes with a detection p-value>0.01 in >1% samples, were filtered out. A total number of 474,528 probes passed the quality control and the filtered  $\beta$  values were normalized with DASEN implemented in the *wateRmelon* package in R statistical software<sup>31</sup>.

The generation and management of the Illumina 450 K methylation array data (EWAS data) for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. The EWAS data were funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by the Netherlands Organization for Scientific Research (NWO; project number 184021007). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

#### *Saguenay Youth Study (SYS)*

The Saguenay Youth Study (SYS) is a population-based study of adolescents and their middle-aged parents investigating the etiology, early stages and trans-generational trajectories of common cardiometabolic and brain diseases<sup>45</sup>. Half of the adolescents were exposed prenatally to maternal cigarette smoking. Sibships of adolescents (adolescents with one or more siblings aged 12-18 years) and their biological parents were recruited from Saguenay-Lac-St. Jean region of Quebec, Canada. All participants were of French-Canadian origin (i.e., both maternal and paternal grandparents of the adolescents were of French-Canadian ancestry and born in the region). Data collection was conducted in two waves. In Wave 1 (2003 – 2012) adolescents (n=1,029, aged 15.0  $\pm$  1.8 years) were recruited and underwent a 'complete assessment' (cardiovascular and other phenotyping, and genotyping) and parents (n=962, aged 43.3  $\pm$  4.6 years) underwent a 'partial assessment' (questionnaires and genotyping). In Wave 2 (2012 – 2015), a subset of the parents (n=664, aged 49.2  $\pm$  5.0 years) underwent 'complete assessment'. The research Ethics Committee of the Chicoutimi Hospital and the Hospital for Sick Children in Toronto approved the study protocol. Written informed consent was obtained from all participants. In the present study, only data from the parents was considered<sup>45</sup>. A total of 105 parent participants with complete and quality controlled blood-pressure (BP) and DNA-methylation data have been studied here. The cardiovascular assessment involved a 52-minute protocol during which beat-by-beat BP was monitored, at rest and in response to physical and mental challenges, using a Finometer<sup>TM</sup> (FNS Finapres, Amsterdam, The Netherlands). Systolic and diastolic BP values analyzed were 1-minute averages obtained after 5 minutes at rest in a sitting position. Average time interval between BP and DNA-methylation measurements for the participants studied was 3.11 (1.07) years.

Genomic DNA was extracted from peripheral blood cells using the Human610-Quad and HumanOmniExpress BeadChips (Illumina, San Diego, CA, USA), bisulfite converted, and then epityped. Epityping was conducted using the Infinium HumanMethylation450K BeadChip (Illumina, San Diego, CA, USA) at the Montreal Genome Centre (Montreal, Quebec,

Canada). DNA-methylation data was available in 288 parents. A total of 6673 probes were removed due to detection P-value > 0.01 in > 1% of samples. One sample was removed based on pass rate < 95%. Methylation values were normalized using the SWAN method<sup>28</sup>. Blood-cell proportions were imputed using the Houseman method<sup>23</sup>.

### *TwinsUK*

The TwinsUK cohort was established in 1992 as a collection of healthy research volunteers who were monozygotic and dizygotic same-sex twins. The participants are mostly adult female Caucasians and in total there are more than 13,000 participants across the UK<sup>46</sup>. TwinsUK blood methylation profiles were available from 690 female subjects who also had records of blood pressure and blood cell counts (eosinophils, lymphocytes, monocytes, and neutrophils). These subjects are aged from 29 to 82 years old and were not affected with type 2 diabetes or cancer. The fasting and resting blood pressure have been measured three times in a sitting position, and the average of the latter two measurements was used in this analysis.

DNA was extracted from whole blood samples stored in EDTA tubes at 4 degree. Infinium HumanMethylation450 BeadChips (Illumina Inc, San Diego, CA) were used to measure DNA methylation. Details of the experimental approach have been previously described<sup>47</sup>. To correct for technical issues including two Illumina probe types, a normalization using the beta mixture quantile dilation (BMIQ) approach was applied<sup>26</sup>. The DNA methylation probes that mapped incorrectly or to multiple locations in the reference sequence were removed. Probes with detection P-value > 0.01 among > 10% of samples were also removed. All remaining probes were with non-missing values. After quality control, 452,785 probes were used for subsequent analyses. The epigenome-wide association models were adjusted for cell counts and technical covariates that included plate, position on the plate, family structure, and zygosity structure. A linear mixed effect regression model was applied as the data contained twin pairs, and all covariates apart from family and zygosity were taken as fixed-effect term.

The study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

### *Women's Health Initiative – Broad Agency Award 23 (WHI-BAA23)*

Women were selected from one of two large Women's Health Initiative (WHI) sub cohorts that had previously undergone genome-wide genotyping as well as profiling for seven cardiovascular disease related biomarkers including total cholesterol, HDL, LDL, triglycerides, CRP, creatinine, insulin, and glucose through two core WHI ancillary studies. The first cohort is the WHI SNP Health Association Resource (SHARe) cohort of minorities that includes >8000 African American (AA) women and >3500 Hispanic women. These women were genotyped through WHI core study M5-SHARe ([www.whi.org/researchers/data/WHIStudies/StudySites/M5](http://www.whi.org/researchers/data/WHIStudies/StudySites/M5)) and underwent biomarker profile through WHI core study W54-SHARe ([www.whi.org/researchers/data/WHIStudies/StudySites/W54](http://www.whi.org/researchers/data/WHIStudies/StudySites/W54)). The second cohort consists of a combination of European Americans (EA) from the two Hormonal Therapy (HT) trials selected for GWAS and biomarkers in core studies W58 ([www.whi.org/researchers/data/WHIStudies/StudySites/W58](http://www.whi.org/researchers/data/WHIStudies/StudySites/W58)) and W63 ([www.whi.org/researchers/data/WHIStudies/StudySites/W63](http://www.whi.org/researchers/data/WHIStudies/StudySites/W63)). From these two cohorts, two sample sets were formed. The first (sample set 1) is a sample set of 637 CHD cases and 631 non-CHD cases as of Sept 30, 2010. The second sample set (sample set 2) is a non-overlapping sample of 432 cases of coronary heart disease and 472 non-cases as of September 17, 2012. All women with measures of inflammation that passed QC were included in this analysis. Written informed consent was obtained from all participants. All participants with complete and quality controlled blood pressure (BP) and DNA-methylation data have been studied here. Blood pressure was measured in all WHI participants

at baseline and annual clinic visits by certified staff in the WHI clinic using standardized procedures. The average of two baseline readings taken at the same clinic visit will be used for our analyses. Systolic and diastolic BP values analyzed were one-minute averages obtained after 5 minutes at rest in a sitting position.

DNA methylation analysis of blood was performed at HudsonAlpha Institute of Biotechnology using the Illumina Infinium Human-Methylation450 BeadChip. The Illumina BeadChips measures bisulfite-conversion-based, single-CpG resolution DNA methylation levels at 485,577 different CpG sites in the human genome. These data were generated by following the standard protocol of Illumina methylation assays, which quantifies methylation levels by the  $\beta$  value using the ratio of intensities between methylated and un-methylated alleles. Specifically, the  $\beta$  value is calculated from the intensity of the methylated (M corresponding to signal A) and un-methylated (U corresponding to signal B) alleles, as the ratio of fluorescent signals  $\beta = \text{Max}(M,0) / [\text{Max}(M,0) + \text{Max}(U,0) + 100]$ . Thus,  $\beta$  values range from 0 (completely un-methylated) to 1 (completely methylated) (Dunning, 2008).

#### *Women's Health Initiative - Epigenetic Mechanisms of PM-Mediated CVD (WHI-EMPC)*

The Women's Health Initiative – Epigenetic Mechanisms of Particulate Matter-Mediated Cardiovascular Disease (WHI-EMPC) is an ancillary study of epigenetic mechanisms underlying associations between ambient particulate matter air pollution and cardiovascular disease in the Women's Health Initiative clinical trials (CT) cohort, funded by the National Institute of Environmental Health Sciences (R01-ES020836)<sup>48</sup>. The WHI-EMPC study population is a stratified, random sample of 2,200 WHI CT participants who were examined between 1993 and 2001 and had available buffy coat, core analytes, electrocardiograms, and ambient concentrations of PM but were not taking anti-arrhythmic medications at the time. As such, WHI-EMPC is representative of the larger, multiethnic WHI CT population from which it was sampled: n=68,132 participants aged 50-79 years who were randomized to hormone therapy, calcium/vitamin D supplementation, and/or dietary modification in 40 U.S. clinical centers at the baseline exam (1993-1998) and re-examined in the fasting state one, three, six, and nine years later<sup>49,50</sup>. Current analyses involved information collected at the first visit with DNA methylation data available and were stratified by race/ethnicity (black, Hispanic/Latino, and white). On the day of DNA methylation sample collection, BP was measured in the seated position after a five-minute resting period with a conventional mercury sphygmomanometer and an appropriately sized cuff. The mean of two sequential recordings, taken at least 30 seconds apart, of systolic and diastolic BP was used for analyses.

Genome-wide DNA methylation at CpG sites was measured using the Illumina 450K Infinium Methylation BeadChip, quantitatively represented by beta (the percentage of methylated cytosines over the sum of methylated and unmethylated cytosines), and quality controlled using the following filters: detection p-values > 0.01 in > 10% of samples, detection p-values > 0.01 or missing in > 1% of probes, and probes with a coefficient of variation < 5%, yielding values of beta at 293,171 sites. DNAm data was normalized using BMIQ<sup>26</sup>, then stage-adjusted using ComBat<sup>51</sup>. Epigenome-wide association models were adjusted for Houseman-estimated<sup>23</sup> cell subtype proportions (CD8-T, CD4-T, B cell, natural killer, monocyte, and granulocyte) and technical covariates including chip, row and column.

The Women's Health Initiative (WHI) was funded by the NHLBI through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. All contributors to WHI science are listed at

<https://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Long%20List.pdf>.

Within WHI, "*Epigenetic Mechanisms of PM-Mediated CVD Risk*" (WHI-EMPC) was funded by NIEHS R01ES020836 (Whitsel; Baccarelli; Hou); "*Integrative Genomics for Risk of CHD and Related Phenotypes in WHI*" (WHI BAA23) by NHLBI 60442456 (Horvath; Assimes; Absher); "*SNP Health Association Resource*" project (WHI-SHARe) by NHLBI N02HL64278, "*GWAS of Hormone Treatment and CVD and Metabolic Outcomes within the Genomics and Randomized Trials Network*"

(WHI-GARNET) by NHGRI U01HG005152; “GWAS to Identify Genetic Components of Hip Fracture” (WHI-HipFx) by NHLBI N01WH74316; and “Genetics and Epidemiology of Colorectal Cancer Consortium” (WHI-GECCO) by NCI U01CA137088 and R01CA059045. This work was additionally supported by R01HL129132 and R01HG006292 (Li).

## 2. Gene Expression Measurements

### *Rotterdam Study Gene Expression*

Whole-blood was collected (PAXGene Tubes – Becton Dickinson) and total RNA was isolated (PAXGene Blood RNA kits - Qiagen). To ensure a constant high quality of the RNA preparations, all RNA samples were analysed using the Labchip GX (Calliper) according to the manufacturer’s instructions. Samples with an RNA Quality Score more than 7 were amplified and labelled (Ambion TotalPrep RNA), and hybridized to the Illumina HumanHT12v4 Expression Beadchips as described by the manufacturer’s protocol. Processing of the Rotterdam Study RNA samples was performed at the Genetic Laboratory of Internal Medicine, Erasmus University Medical Centre Rotterdam. The RS-III expression dataset is available at GEO (Gene Expression Omnibus) public repository under the accession GSE33828. Illumina gene expression data was quantile-normalized to the median distribution and subsequently log<sub>2</sub>-transformed. The probe and sample means were centered to zero. Genes were declared significantly expressed when the detection p-values calculated by GenomeStudio were less than 0.05 in more than 10% of all discovery samples, which added to a total number of 21,238 probes<sup>52</sup>. Quality control was done using the eQTL-mapping pipeline. We only analyzed probes that uniquely mapped to the human genome build 37<sup>53</sup>.

### *Framingham Heart Study Gene Expression*

Affymetrix Human Exon Array ST 1.0 (Affymetrix, Inc., Santa Clara, CA) was utilized to measure mRNA expression levels genome wide (N~18,000 transcripts). Details of the design, sampling, RNA isolation, and mRNA measurement have been described previously<sup>54</sup>.

## Supplementary References

1. Liu, C. *et al.* A DNA methylation biomarker of alcohol consumption. *Mol. Psychiatry* (2016). doi:10.1038/mp.2016.192
2. Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541**, 81–86 (2017).
3. Dekkers, K. F. *et al.* Blood lipids influence DNA methylation in circulating cells. *Genome Biol.* **17**, 138 (2016).
4. Petersen, A. K. *et al.* Epigenetics meets metabolomics: An epigenome-wide association study with blood serum metabolic traits. *Hum. Mol. Genet.* **23**, 534–545 (2014).
5. Aslibekyan, S. *et al.* Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. *Obesity (Silver Spring)*. **23**, 1493–501 (2015).
6. Soriano-Tárraga, C. *et al.* Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. *Hum. Mol. Genet.* **25**, 1–11 (2015).
7. Chambers, J. C. *et al.* Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: A nested case-control study. *Lancet Diabetes Endocrinol.* **3**, 526–534 (2015).
8. Al Muftah, W. A. *et al.* Epigenetic associations of type 2 diabetes and BMI in an Arab population. *Clin. Epigenetics* **8**, 13 (2016).
9. Florath, I. *et al.* Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. *Diabetologia* **59**, 130–138 (2016).
10. Liang, L. *et al.* An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature* **520**, 670–674 (2015).
11. Irvin, M. R. *et al.* Epigenome-wide association study of fasting blood lipids in the genetics of lipid-lowering drugs and diet network study. *Circulation* **130**, 565–572 (2014).
12. Demerath, E. W. *et al.* Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum. Mol. Genet.* **24**, 4464–79 (2015).
13. Das, M. *et al.* Association of DNA Methylation at CPT1A Locus with Metabolic Syndrome in the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) Study. *PLoS One* **11**, e0145789 (2016).
14. Mamtani, M. *et al.* Genome- and epigenome-wide association study of hypertriglyceridemic waist in Mexican American families. *Clin. Epigenetics* **8**, 6 (2016).
15. Aslibekyan, S. *et al.* CPT1A methylation is associated with plasma adiponectin. *Nutr. Metab. Cardiovasc. Dis.* **27**, 225–233 (2017).
16. Mitchell, B. D. *et al.* The genetic response to short-term interventions affecting cardiovascular function: rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study. *Am. Heart J.* **155**, 823–8 (2008).
17. Shuldiner, A. R. *et al.* Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA* **302**, 849–57 (2009).
18. Hsueh, W. C. *et al.* Diabetes in the Old Order Amish: characterization and heritability analysis of the Amish Family Diabetes Study. *Diabetes Care* **23**, 595–601 (2000).

19. Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288–95 (2011).
20. Dedeurwaerder, S. *et al.* A comprehensive overview of Infinium HumanMethylation450 data processing. *Brief. Bioinform.* **15**, 929–41 (2014).
21. Chen, Y. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8**, 203–9 (2013).
22. Moen, E. L. *et al.* Genome-wide variation of cytosine modifications between European and African populations and the implications for complex traits. *Genetics* **194**, 987–96 (2013).
23. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012).
24. Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* **15**, R31 (2014).
25. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am. J. Epidemiol.* **129**, 687–702 (1989).
26. Teschendorff, A. E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189–96 (2013).
27. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
28. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.* **13**, R44 (2012).
29. Feinleib, M., Kannel, W. B., Garrison, R. J., McNamara, P. M. & Castelli, W. P. The framingham offspring study. Design and preliminary data. *Prev. Med. (Baltim).* **4**, 518–525 (1975).
30. Splansky, G. L. *et al.* The Third Generation Cohort of the National Heart, Lung, and Blood Institute’s Framingham Heart Study: Design, Recruitment, and Initial Examination. *Am. J. Epidemiol.* **165**, 1328–1335 (2007).
31. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).
32. Daniels, P. R. *et al.* Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. *Am. J. Med.* **116**, 676–81 (2004).
33. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–9 (2014).
34. Shah, S. *et al.* Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res.* **24**, 1725–33 (2014).
35. Bild, D. E. *et al.* Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am. J. Epidemiol.* **156**, 871–81 (2002).
36. Liu, Y. *et al.* Methylomics of gene expression in human monocytes. *Hum. Mol. Genet.* **22**, 5065–74 (2013).
37. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. ...* **102**, 15545–15550 (2005).
38. Abbas, A. R. *et al.* Immune response in silico (IRIS): immune-specific genes identified from a compendium of

- microarray expression data. *Genes Immun.* **6**, 319–31 (2005).
39. Bell, B., Rose, C. L. & Damon, A. The Veterans Administration longitudinal study of healthy aging. *Gerontologist* **6**, 179–84 (1966).
  40. Troisi, R. J., Heinold, J. W., Vokonas, P. S. & Weiss, S. T. Cigarette smoking, dietary intake, and physical activity: effects on body fat distribution--the Normative Aging Study. *Am. J. Clin. Nutr.* **53**, 1104–11 (1991).
  41. Triche, T. J., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* **41**, e90 (2013).
  42. Davis, S., Du, P., Bilke, S., Triche, T. J. & Bootwalla, M. methylumi: Handle Illumina methylation data. *R Packag. version 2.14.0* (2015).
  43. Barfield, R. T. *et al.* Accounting for population stratification in DNA methylation studies. *Genet. Epidemiol.* **38**, 231–41 (2014).
  44. Hofman, A. *et al.* The Rotterdam Study: 2016 objectives and design update. *Eur. J. Epidemiol.* **30**, 661–708 (2015).
  45. Pausova, Z. *et al.* Cohort Profile: The Saguenay Youth Study (SYS). *Int. J. Epidemiol.* (2016). doi:10.1093/ije/dyw023
  46. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort Profile: TwinsUK and healthy ageing twin study. *Int. J. Epidemiol.* **42**, 76–85 (2013).
  47. Tsaprouni, L. G. *et al.* Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics* **9**, 1382–96 (2014).
  48. U.S. Department of Health and Human Services. National Institutes of Health. Epigenetic Mechanisms of PM-Mediated CVD. *Research Portfolio Online Reporting Tools* Available at: [https://projectreporter.nih.gov/project\\_info\\_description.cfm?aid=8369149&icde=27485446](https://projectreporter.nih.gov/project_info_description.cfm?aid=8369149&icde=27485446).
  49. Design of the Women’s Health Initiative clinical trial and observational study. The Women’s Health Initiative Study Group. *Control. Clin. Trials* **19**, 61–109 (1998).
  50. Anderson, G. L. *et al.* Implementation of the Women’s Health Initiative study design. *Ann. Epidemiol.* **13**, S5-17 (2003).
  51. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
  52. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–43 (2013).
  53. Schurmann, C. *et al.* Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. *PLoS One* **7**, e50938 (2012).
  54. Joehanes, R. *et al.* Gene Expression Signatures of Coronary Heart Disease. *Arterioscler. Thromb. Vasc. Biol.* **33**, 1418–1426 (2013).