SUPPLEMENTARY MATERIAL

Supplementary Methods

The Framingham Heart Study (FHS). FHS recruited adults from Framingham, Massachusetts starting in 1948. CVD was defined as a composite of coronary heart disease (coronary death, myocardial infarction, coronary insufficiency, and angina), cerebrovascular events (including ischemic stroke, hemorrhagic stroke, and transient ischemic attack), peripheral artery disease (intermittent claudication), and heart failure.[115] Medical histories, physical examinations during study visits, hospitalization records and personal physician records were used to identify any possible cardiovascular event. A panel of three experienced investigators reviewed the medical records of suspected new events and made final decision about each event. The participants of exam 8 (2005-2008) of FHS offspring cohort were followed through 2014 (average follow-up of 7.7 years; range: 0.04 years – 9.8 years).

DNA methylation was measured from whole-blood samples (Gentra Puregene Blood Kit-Qiagen, Venlo, Netherlands) collected during the eighth examination of the Framingham Offspring Study (2005-2008). DNA methylation was quantified in the bisulfite converted genomic DNA (EZ DNA Methylation Kit-Zymo Research, Irvine, CA) using Illumina Infinium HumanMethylation450K Beadchip array. DASEN methodology in wateRmelon package 10 was used to conduct within laboratory batch normalization of raw data. The exclusion criteria for samples were a missing rate > 1% at detection p-value <0.01, poor matching to the 65 single nucleotide polymorphism (SNP) control probe locations, and identification as outliers using multi-dimensional scaling techniques. In addition, the exclusion criteria for the probes were missing rate > 20 % at detection P-value< 0.01, previously identified to map multiple locations,

underlying SNPs (minor allele frequency > 5% in European ancestry 1000 genomes project data) at the CpG or < 10 bp of the single base extension, and location in sex chromosomes. Finally, we had 408,254 CpGs and 2,631 individuals.

Among 2,631 participants with blood DNA methylation data available in the FHS Offspring, we excluded those with prior CVD (n=316) and those missing information on CVD risk factors (n=325), leaving 1,990 participants.

Women Health Initiative (WHI): WHI enrolled 161,808 women starting in 1993 as part of randomized control trials that were continued as a prospective cohort study. CVD included coronary heart disease (including hospitalized myocardial infarction, definite silent myocardial infarction, and coronary death) and other hospitalized cardiovascular or cerebrovascular events (including stroke, congestive heart failure, angina, peripheral vascular disease, and coronary revascularization).[116] The participants of WHI were followed from baseline (1993-1998) to 2016 with an average follow-up time of 12.18 years (range: 0.003 – 21.3 years). Details about DNA methylation measurement and quality control have been published.[32,117] In brief, standard procedures of 450K Illumina assay were used to measure DNA methylation in peripheral blood. During quality control, we excluded probes with a missing rate > 5 % at detection P value < 0.01, SNPs within 10 base pairs of targeted CpGs, and location on X or Y chromosomes. Finally, 434,113 CpGs and 2,096 individuals were available for analysis. Among those, we excluded individuals with missing information on traditional risk factors of CVD, leaving 1,487 participants.

Multi-Ethnic Study of Atherosclerosis (MESA): MESA was designed to investigate subclinical CVD in 6,814 participants aged 45–84 free of CVD at baseline from July 2000 to July 2002. Participants were enrolled at six different locations: Baltimore, MD; Chicago, IL; Los Angeles, CA; New York, NY; St. Paul, MN, and Winston-Salem, NC. Six clinical visits collected extensive clinical, socio-demographic, lifestyle, behavior, laboratory, nutrition, and medication data. Copies of death certificates and medical records were reviewed by 2 physicians from the MESA Events Committee for classification and assignment of incidence dates. If the reviewers disagreed, the full committee made the final decision. For this study, fatal and non-fatal CHD and stroke events were included. CHD events included myocardial infarction, definite angina followed by revascularization, definite angina not followed by revascularization, probable angina followed by revascularization, and resuscitated cardiac arrest. Stroke included rapid onset of a documented focal neurologic deficit lasting 24h or until death, or having a clinically relevant lesion on brain imaging if <24h. Patients with focal neurologic deficits secondary to brain trauma, tumor, infection, or other non-vascular cause were excluded. MESA study genomic methylation profiling was performed in the Methylation Characterization Center Laboratory at the University of Southern California (USC) under the aegis of the NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, MESA Multi-Omics Program. Genomic DNA was shipped on dry ice to USC for Illumina EPIC array methylation profiling. Samples were randomized in batches of 288 (3 x 96) for array processing. After the initial scan of each beadchip array was completed, the data was processed in the Illumina Genome Studio Methylation module with background correction and normalization to output preliminary beta values. The raw data was processed using R and the minfi package. Probes that were not significant above background signal were filtered at p-value > 0.05. Single sample Noob

(ssNoob) was used for normalization[118] to derive corrected methylated and unmethylated signals, and thence the beta values, which are identical to the original Noob values[119] in the beta scale. Across all race/ethnic groups, samples were dropped based on non-concordance of EPIC SNPs (n=1) and related pairs (n=5). A further six samples were dropped as outliers by visual inspection of the first 10 PCs from within-race group methylation PCA. A total of 108,714 probes were filtered based on Zhou et al recommendations.[120] These included probes that had poor genomic mapping quality; a SNP present in the extension base causing color channnel switch; non-unique 30 bp flanking sequence; extension base inconsistent with specified color channel; a SNP or SNV within 5 bp of the methylation base.[121]

Table S1. Baseline participants' characteristics by cardiovascular disease incidence status for the replication cohorts.

	Framingham Heart Study		Women's Health Initiative		Multi-Ethnic Study of Atherosclerosis	
	Non-incident CVD (N=1792)	Incident CVD (N=198)	Non-incident CVD (N=733)	Incident CVD (N=754)	Non-incident CVD (N=848)	Incident CVD (N=68)
Age (years), median (IQR)	64.0 (59.0, 70.0)	71.0 (64.0, 78.0)	64.0 (58.0, 69.0)	65.0 (60.0, 70.0)	68 (61, 77)	74.5 (65.0, 82.0)
Sex, % Men	41.6	52	-	-	46.7	57.4
Smoking status, %						
Former	-	-	6.68	11.27	49.9	53.7
Current	8.1	6.1	38.61	37.53	8.6	10.4
BMI, median (IQR)	27.3 (24.3, 30.7)	29.0 (25.7, 31.7)	28.6 (25.0, 32.6)	29.4 (25.8, 33.6)	28.3 (25.2, 32.2)	28.0 (24.2, 30.7)
LDL cholesterol (mg/dL), median (IQR)	190 (166, 214)	181.5 (161, 204.8)	139 (118.9, 162.2)	145 (121, 171.1)	105.5 (82.0, 127.0)	110 (80, 125)
HDL cholesterol (mg/dL), median (IQR)	57 (46, 70)	50 (40, 62)	53.5 (46, 63.9)	49 (42.5, 58)	52 (44, 63)	50.0 (40.8, 63.3)
Systolic blood pressure, median (IQR)	126 (115, 137)	132.5 (123, 143.8)	127 (115, 139)	133 (122, 146)	120 (110, 135)	129 (113, 144)
Hypertension, %	42.6	63.6	31.5	47.6	57.9	69.1
Diabetes, %	9.4	22.2	6.1	15.0	19.5	20.6
Albuminuria, %				65.0 (60.0, 70.0)		
Microalbuminuria	0.28	2.0	-	0	11.1	21.2
Macroalbuminuria	5.6	14.7	-		2.4	7.6
Urinary arsenic (μg/g creatinine)*	-	-	-	11.27	2.9 (1.7, 4.9)	3.1 (1.8, 4.5)

CVD: Cardiovascular disease, IQR: interquartile range.

*Urinary arsenic corresponds to the sum of inorganic and methylated species (methylarsonic acid and dimethylarsinic acid) in the urine.

Table S2. Differences in CVD incidence cases per 100000 person-years for the doubling of urinary arsenic levels attributable to changes in DNA methylation for each CpG ('mediated effects') in unweighted models.

CpG	Chr	Gene	Cases attributable to a doubling of urinary As (95 % CI) (direct effect)	Cases attributable to a doubling of urinary As through DNAm (95%CI) (indirect effect)	Percentage of all cases attributable to a doubling of urinary As that is explained by DNAm (95%CI)
cg19693031	1	*TXNIP*	60.8 (-137.9, 259.0)	76.5 (41.6, 118.1)	55.7 (-409.7, 515.1)
cg05779585	8	*LOC286083*	130.8 (-63.9, 325.1)	13.9 (-0.8, 37.0)	9.7 (-57.5, 87.6)
cg03497652	16	*ANKS3*	97.45 (-99.7, 294.3)	45.7 (18.2, 77.6)	31.9 (-224.2, 293.4)
cg01270753	9	*TGFBR1*	131.4 (-61.8, 324.1)	13.5 (-4.8, 34.9)	9.3 (-56.1, 85.9)
cg22294740	19	*LINGO3*	112.3 (-83.9, 308.1)	30.1 (5.9, 58.5)	21.1 (-146.4, 196.7)
cg03362418	22	*TYMP*	122.7 (-71.7, 316.6)	8.9 (-8.7, 28.7)	6.8 (-55.6, 74.6)
cg23027596	9	*UBAC1*	124.8 (-68.4, 317.6)	21.3 (4.1, 42.9)	14.6 (-87.7, 126.1)
cg17608381	6	*HLA-A*	127.5 (-68.8, 323.2)	14.5 (0.8, 33.2)	10.2 (-74.5, 101.1)
cg09956442	19	*ARRDC2*	114.2 (-79.7, 307.7)	32.1 (12.2, 57.6)	21.9 (-136.5, 189.7)
cg06668829	8	*EPPK1*	122.8 (-70.2, 315.3)	22.8 (6.2, 44.5)	15.6 (-98.5, 138.4)
cg14827056	8	*EIF2C2*	109.8 (-85.8, 305.0)	33.6 (10.1, 62.1)	23.4 (-156.6, 212)
cg18032342	3	*NISCH*	123.8 (-71.1, 318.3)	16.0 (1.1, 35.1)	11.4 (-78, 107.3)
cg13092901	22	*TYMP*	130.8 (-63.4, 324.6)	7.6 (-9.3, 26.3)	5.5 (-44.3, 61.8)
cg11946459	6	*HLA-A*	133.3 (-61.9, 328.0)	11.5 (-0.9, 28.5)	7.9 (-54.6, 77.7)
cg06970472	4	*APBB2*	133.9 (-59.8, 327.3)	8.1 (-0.5, 21.7)	5.7 (-35.7, 53.4)
cg06716655	1	*ADAR2*	132.0 (-63.3, 326.9)	10.7 (-1.3, 27.5)	7.5 (-50.6, 72.1)
cg18618815	17	*COL1A1*	124.4 (-69.0, 317.3)	16.2 (3.0, 34.2)	11.5 (-77.2, 105.4)
cg01178924	13	*LMO7*	128.8 (-66.9, 323.9)	13.3 (-1.8, 32.5)	9.3 (-65.8, 91)
cg01542019	19	*TECR*	125.8 (-70.0, 321.2)	14.7 (-0.7, 34.4)	10.5 (-72.8, 100.8)
cg02047803	5	*RELL2*	126.8 (-67.3, 320.5)	14.9 (0.9, 33.9)	10.5 (-71.9, 98.9)
cg16335098	6	*SMOC2*	137.1 (-57.7, 331.4)	6.8 (-1.2, 19.4)	4.7 (-29.4, 45)

Models adjusted for age, sex, smoking status, BMI, LDL cholesterol, study center (Arizona, Oklahoma or North and South Dakota), cell counts (CD8T, CD4T, NK, B cells and monocytes) and genetic PCs.

Table S3. Differences in CVD mortality cases per 100000 person-years for the doubling of urinary arsenic levels attributable to changes in DNA methylation for each CpG ('mediated effects') in unweighted models.

CpG	Chr	Gene	CVD mortality cases attributable to a doubling of iAs dose (95 % CI) (direct effect)	Cases attributable to a doubling of urine As through DNAm (95%CI) (indirect effect)	Percentage of all cases attributable to a doubling of urine As that is explained by DNAm (95%CI)
cg05779585	8	*LOC286083*	79.5 (-16.7, 175.6)	9.9 (-0.6, 26.4)	11.1 (-23.1, 76.0)
cg19693031	1	*TXNIP*	51.9 (-47.5, 150.9)	33.6 (16.9, 53.9)	39.3 (-174.2, 300.8)
cg06716655	1	*ADAR*	73.2 (-23.2, 169.4)	13.0 (3.9, 24.7)	15.1 (-54.0, 110.4)
cg17608381	6	*HLA-A*	76.9 (-20.5, 174.2)	10.3 (2.3, 21.1)	11.7 (-42.4, 90.4)
cg22294740	19	*LINGO3*	70.1 (-26.9, 166.9)	17.7 (3.9, 33.9)	20.1 (-66.0, 146.1)
cg03362418	22	*TYMP*	77.4 (-19.1, 173.7)	4.8 (-4.7, 15.5)	5.9 (-28.2, 53.7)
cg11946459	6	*HLA-A*	81.1 (-15.7, 177.6)	7.8 (0.9, 17.5)	8.8 (-25.6, 65.1)
cg21990700	12	*C1RL*	74.8 (-21.6, 171.0)	10.3 (2.1, 20.6)	12.1 (-44.3, 88.9)
cg06970472	4	*APBB2*	80.2 (-16.4, 176.5)	6.8 (0.2, 15.3)	7.8 (-24.3, 57.5)
cg03026982	11	*NAV2*	80.8 (-16.1, 177.6)	6.1 (-2.2, 17.2)	7.1 (-22.0, 55.8)
cg05527044	2	*EGR4*	79.0 (-17.6, 175.4)	9.5 (1.3, 20.2)	10.7 (-32.3, 78.3)
cg00451635	16	*EMP2*	85.7 (-11.0, 182.2)	2.7 (-3.2, 10.4)	3.1 (-12.3, 28.2)
cg27523527	1	*BARHL2*	84.8 (-11.7, 181.1)	3.5 (-2.8, 11.6)	4.0 (-13.4, 34.1)
cg19301366	6	*HLA-DQB1*	84.6 (-11.6, 180.6)	1.8 (-2.1, 7.5)	2.1 (-8.4, 20.0)

Models adjusted for age, sex, smoking status, BMI, LDL cholesterol, study center (Arizona, Oklahoma or North and South Dakota), cell counts (CD8T, CD4T, NK, B cells and monocytes) and genetic PCs.

Table S4. Other traits associated with CpGs showing significant mediated effects for CVD in our study according to EWAS Catalog.[9]

CpG	Gene	Author	PMID	Trait	N
cg19693031	*TXNIP*	Chambers JC	26095709	Type 2 diabetes	3805
cg03497652	*ANKS3*	Sikdar S	31536415	Smoking	15907
cg22294740	*LINGO3*	Liu C	27843151	Alcohol intake	2423
cg17608381	*HLA-A*	Dugue P-A	31789449	Alcohol intake	5606
cg21990700	*C1RL*	Joehanes R	27651444	Smoking	13474
cg14827056	*AGO2*	Sikdar S	31536415	Smoking	15907
cg13092901	*TYMP*	Marioni R	29311653	Cognitive ability	4794
cg11946459	*HLA-A*	Liu C	27843151	Alcohol intake	2423
cg18618815	*COL1A1*	Sharp GC	29016858	Maternal BMI and offspring DNA methylation	7523
cg01178924	*LMO7*	Kazmi N	31230546	Pregnancy-related hypertension	5242
cg01542019	*TECR*	Singmann P	26500701	Sex (autosomal differences)	1799
cg02047803	*RELL2*	Albao D	31691802	Type 2 diabetes	365
cg06970472	*APBB2*	Sikdar S	31536415	Smoking	15907
cg02145701	*BANP*	Bohlin J	27717397	Gestational age	1068
cg05527044	*EGR4*	Liu J	31197173	Fasting insulin	4808
cg00451635	*EMP2*	Liu C	27843151	Alcohol intake	2423
cg27523527	*BARHL2*	Bonder MJ	25282492	Fetal vs adult liver	195

For those CpGs for which associations with several traits were found in the EWAS catalog, either the most relevant trait for this manuscript or the study with the larger sample size are shown.