

EWAS for Incident CVD Events

1 Introduction

Genetic approaches to cardiovascular disease (CVD) research have led to important breakthroughs in mechanistic understanding and therapeutic strategies. However, the mechanisms for gene variant-disease relationships are often difficult to determine, and their effects may often be mediated by epigenetic regulation (M. J. Bonder et al. 2016). DNA methylation is one such mechanism that can integrate both genetic variation and environmental exposures and potentially reflect or drive their effects on CVD outcomes (Ordovás and Smith 2010).

A series of recent epigenome-wide association studies (EWAS) have examined relationships between DNA methylation at cytosine-phosphate-guanine (CpG) sites and various CVD phenotypes, including prior myocardial infarction (Rask-Andersen et al. 2016), acute coronary syndrome (J. Li et al. 2017), and atherosclerosis (Nakatouchi et al. 2017). These discoveries may reveal important mechanistic insights, but are susceptible to reverse causation. Mendelian randomization approaches can resolve some of this ambiguity, but are dependent on the existence of appropriate genetic instrumental variables (methylation-quantitative trait loci) for the CpGs of interest. Furthermore, those that have been performed often suggest that reverse causation (methylation being influenced by the phenotype of interest) is more common (Dekkers et al. 2016; Wahl et al. 2017).

One approach to this problem is to examine epigenetic associations with cardiovascular risk factors. Multiple investigations have explored these relationships genome-wide (Pfeiffer et al. 2015; Irvin et al. 2014), and have even uncovered prognostic CpG sites for incident coronary heart disease in the process (Hedman et al. 2017). A few studies looking directly at incident CVD as a binary variable have found relationships with global DNA methylation (as approximated by LINE-1 methylation levels) and with a specific cluster of CpG sites (Baccarelli et al. 2010; Guarnera et al. 2015).

Most epigenetic investigations of CVD to-date have been performed in blood, based on the importance of immune-related processes in its pathogenesis (Bäck and Hansson 2015) as well as its relative ease of collection in comparison to other relevant tissues (e.g. endothelium, myocardium). In addition, several groups have shown a moderate ability of blood-based methylomic signatures to reflect those in disease-relevant tissues (Bacos et al. 2016; Y. T. Huang et al. 2016).

To address the problem of reverse causation while incorporating time-to-event information, we performed a Cox model-based EWAS for incident CVD in a series of two cohorts. We used sensitivity analyses to demonstrate that demographic factors are relevant, but not dominant, in explaining the variation in results across cohorts. Furthermore, we incorporated existing genetic, epigenetic, and transcriptomic datasets to explore the potential mechanistic relevance of different cell types and transcription factors to CVD risk, both within blood and in other cell types of interest. Finally, we assessed relationships between methylation at relevant CpG sites and traditional cardiovascular risk factors.

2 Results

2.1 Genome-wide associations between DNA methylation and incident CVD events

The epigenome-wide association analysis was performed using two datasets: the Women's Health Initiative Study (WHI) for discovery, and the Framingham Heart Study Offspring Cohort (FHS) for validation. Subject

Table 1 Population description.

	FHS	WHI
Sample Size	2587	2023
% female	55 %	100 %
Age	66 (40-92)	65 (50-79)
Mixed race	No	Yes
BMI	27.7 (13.8-54.2)	29.1 (17.2-58.5)
% smoke	9 %	10 %
# prior CVD events	331	0
# incident CVD events	305	1009

Table 2 CpGs with FDR<0.05 in the discovery set.

CpG	Chromosome	Association	P-value	Location	Annotated gene
cg09155044	chr16	+	6.63e-09	TSS1500	VKORC1
cg02379107	chr20	+	8.02e-07	TSS1500	KIAA1755

characteristics are shown in Table 1. The objective of our initial analysis was to identify specific CpG sites whose methylation levels in blood at a given baseline timepoint predict incident cardiovascular events. For each CpG site passing quality control (422952 in WHI and 425326 in FHS), a Cox proportional hazards model was used to predict time-to-event from methylation beta values while adjusting for relevant biological and technical covariates.

Genomic inflation was limited in both datasets across the full set of CpG sites (see Supp. Fig. S3), suggesting adequate correction for substructure and confounding. However, to increase power, a subset of 100833 highly variable sites were taken from a prior analysis of methylation variability in blood cell subtypes (Hachiya et al. 2017) and used during the initial analysis. Though this dataset was gathered in Asian subjects, the variability patterns are highly consistent with those seen in the datasets analyzed here (Supp. Fig. S2).

Sites passing a false discovery rate threshold of 0.05 are shown in Table 2 (a Manhattan plot of the full results can be found in Supp. Fig. S4). 1 CpG site was found to pass a Bonferroni threshold of $p < 5.8 \times 10^{-7}$ in the discovery cohort, though it replicated only marginally in the validation cohort ($p = 0.09$). Despite this lack of validation, there was consistency in the direction of coefficients between cohorts. Of 8 sites reaching a suggestive $\text{FDR} < 0.25$ in both datasets, 100% show consistent direction of effect, and will be referred to as differentially methylated positions (DMPs; Table 3).

The effect of confounding by past cardiovascular events in FHS was minimized in the analysis setup by using FHS as a replication rather than discovery set. However, to further address this possibility, EWAS analyses stratified by past event status were conducted. Of the 10 suggestive sites overlapping at $\text{FDR} < 0.25$, both strata showed perfect consistency of regression coefficient signs with those from WHI. Though power to identify strong signals was reduced by the stratification, 7 of 8 DMPs retained nominal significance ($p < 0.05$)

Table 3 8 CpGs found at FDR<0.25 in both datasets.

CpG	Association	Chr	Location	Gene
cg21917349	+	chr15	5'UTR;1stExon	APBA2
cg06500161	+	chr21	Body	ABCG1
cg06806711	+	chr11	1stExon;5'UTR	MS4A1
cg25354657	+	chr11	Body	APLP2
cg11772801	+	chr14	Body	TMX1
cg06151145	+	chr4	Body	SLAIN2
cg27111756	+	chr19	Body	PPP1R13L
cg19390658	-	chr7	Body	GARS

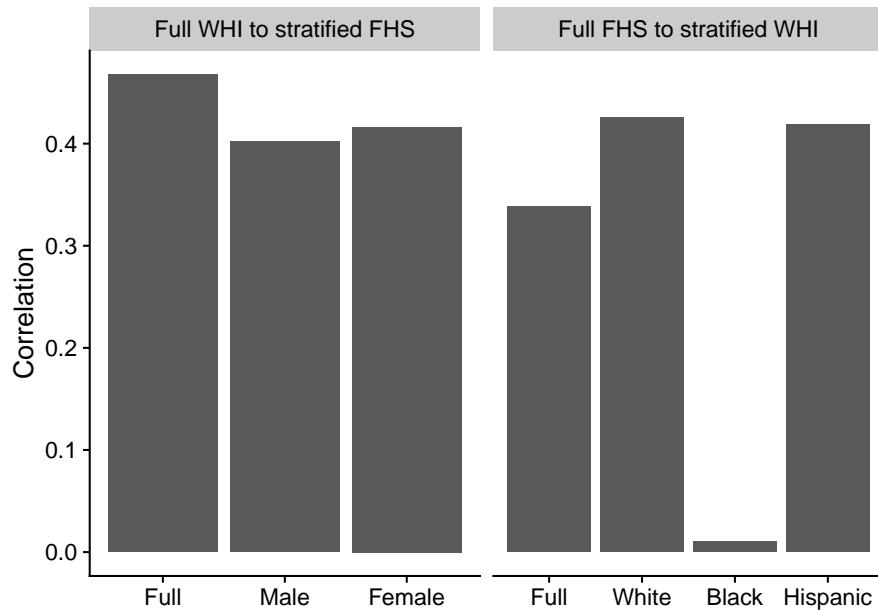


Figure 1 Pearson correlations between EWAS regression coefficients in one dataset with $FDR < 0.25$ and those of the same CpGs in subsets of the other.

in the group without past events, compared to 4 sites in the group with past events. These results indicate that the moderate agreement between cohorts was not being primarily driven by history of CVD.

Because only 8 sites overlapped at $FDR < 0.25$, the union rather than intersection of these sites (1532 CpGs) was used for further investigation of the characteristics of the set. These sites showed no major enrichment for gene-related genomic locations (as compared to a background of only the highly variable subset), but did show a moderate enrichment for presence in CpG islands ($p = 2.8e-05$; hypergeometric test) and shores ($p = 7.1e-15$)

A gene set enrichment analysis of genes annotated to the same set of DMPs showed enrichment for biological processes related to developmental processes (embryonic development, cell fate commitment, and embryonic morphogenesis), cell adhesion, and transcriptional regulation (see Supp. Table S1).

2.2 Demographic stratification partially explains the observed heterogeneity between cohorts

Sex and race were adjusted for in the above analyses in order to remove their direct associations with incident CVD events, but interactions with methylation were not included that would allow loci to have different associations between demographics. To evaluate the potential role of this demographic heterogeneity in explaining the level of agreement between cohorts, stratified analyses were undertaken (separates sexes in FHS, and separate races in WHI).

Fig. 1 shows the correlations between regression coefficients from a main cohort (w/ $FDR < 0.25$) and those of the corresponding sites from the stratified groups. In this dataset, similarity of demographic factors did associate with similarity of regression coefficients: Caucasians in WHI displayed more similar coefficients to FHS and females in FHS displayed more similar coefficients to WHI. However, these comparisons of more similar demographic groups did not uncover an increased number of replicating loci at $FDR < 0.25$.

A clear lack of correlation with FHS was observed in black subjects in comparison to white or hispanic subjects from WHI. Although females from FHS did show a slightly higher correlation with WHI, this difference was

Table 4 Comb-p regions with multiple test-corrected $p < 0.05$ in WHI and uncorrected $p < 0.05$ in FHS

Location	# CpGs	Annotated gene	Genomic region	P.whi	Padj.whi	P.fhs
chr2:21266499-21267334	12	APOB	CpG island at TSS	2.22e-13	7.81e-09	0.000281
chr19:58715250-58716517	8	ZNF274	CpG island in body	4.04e-12	2.13e-07	0.0166
chr1:27440462-27440721	3	SLC9A1	Body	4.85e-08	0.00684	3.53e-05
chr5:171057071-171057339	5		CpG island	4.6e-07	0.0389	0.00892

not large, and neither stratum correlated as well as the full FHS dataset. Notably, however, after stratification of the FHS dataset by sex, the single CpG site passing the Bonferroni threshold in WHI (upstream of the TSS of VKORC1) showed a nominal association in women ($p = 0.026$) but none in men.

2.3 Region-based approach detects additional associations

In order to expand the site-by-site analysis and take advantage of shared information between proximal CpG sites, the Comb-p algorithm was used to identify differentially methylated regions (DMRs). Comb-p seeks regions enriched for low p-values while accounting for autocorrelation based on genomic distance. It was used separately on WHI and FHS (with p-values from the above EWAS used as input), and uses the full set of loci (rather than the variance-based subset) for maximum genome coverage.

In contrast to the site-based approach, 4 regions were identified that passed a Bonferroni correction in the discovery set and showed at least nominal significance in validation (Table 4). Three of these regions are annotated to known genes: ApoB (N shore of island in gene body), ZNF274 (CpG island in gene body) and SLC9A1 (gene body). Fig. 2 shows EWAS associations within these regions, with and without adjustment for traditional cardiovascular risk factors (LDL-C, HDL-C, Tg, Sys. BP, hsCRP). Risk factors seem to explain some, but not all, of the observed relationships in FHS.

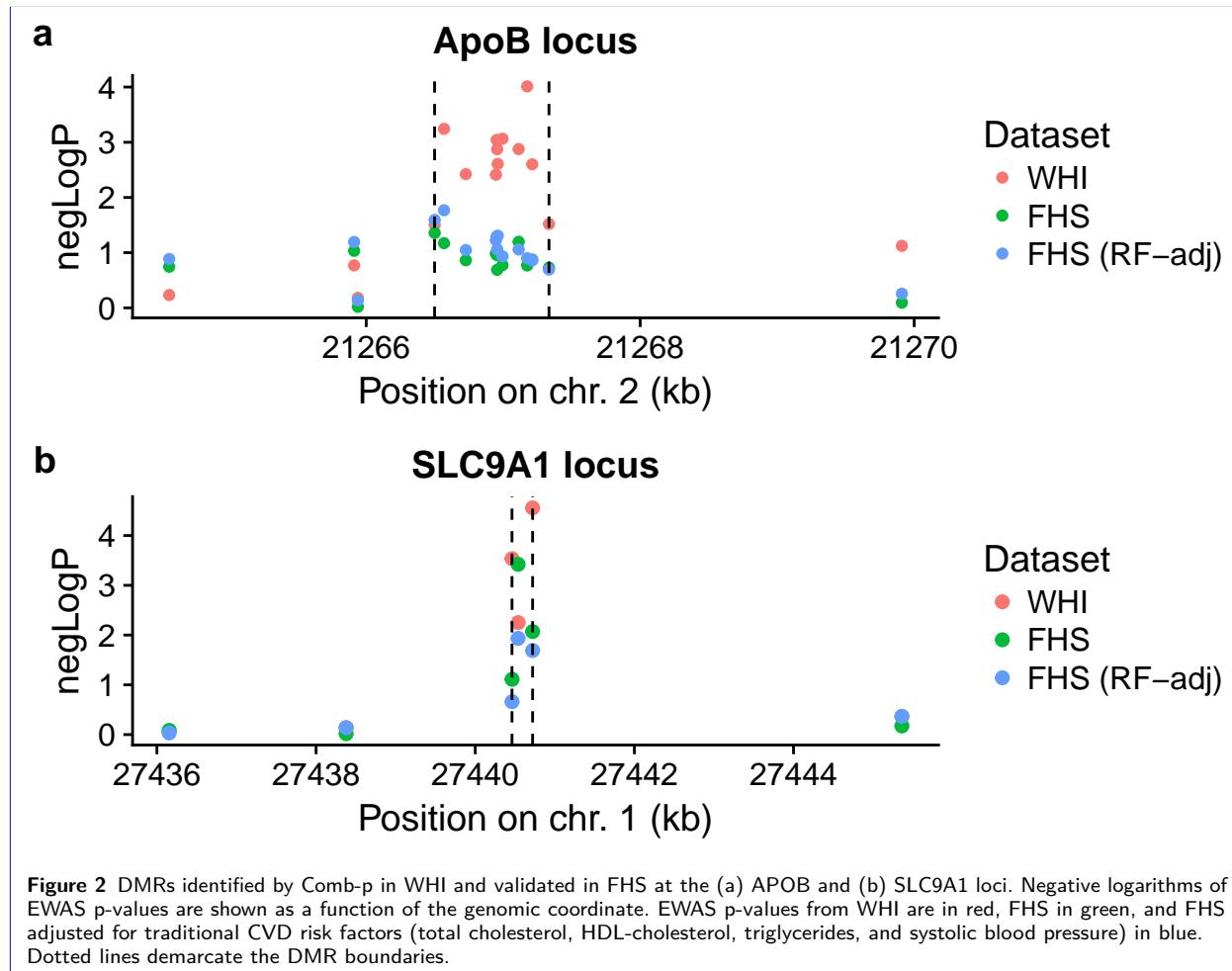
2.4 Weighted correlation network approach

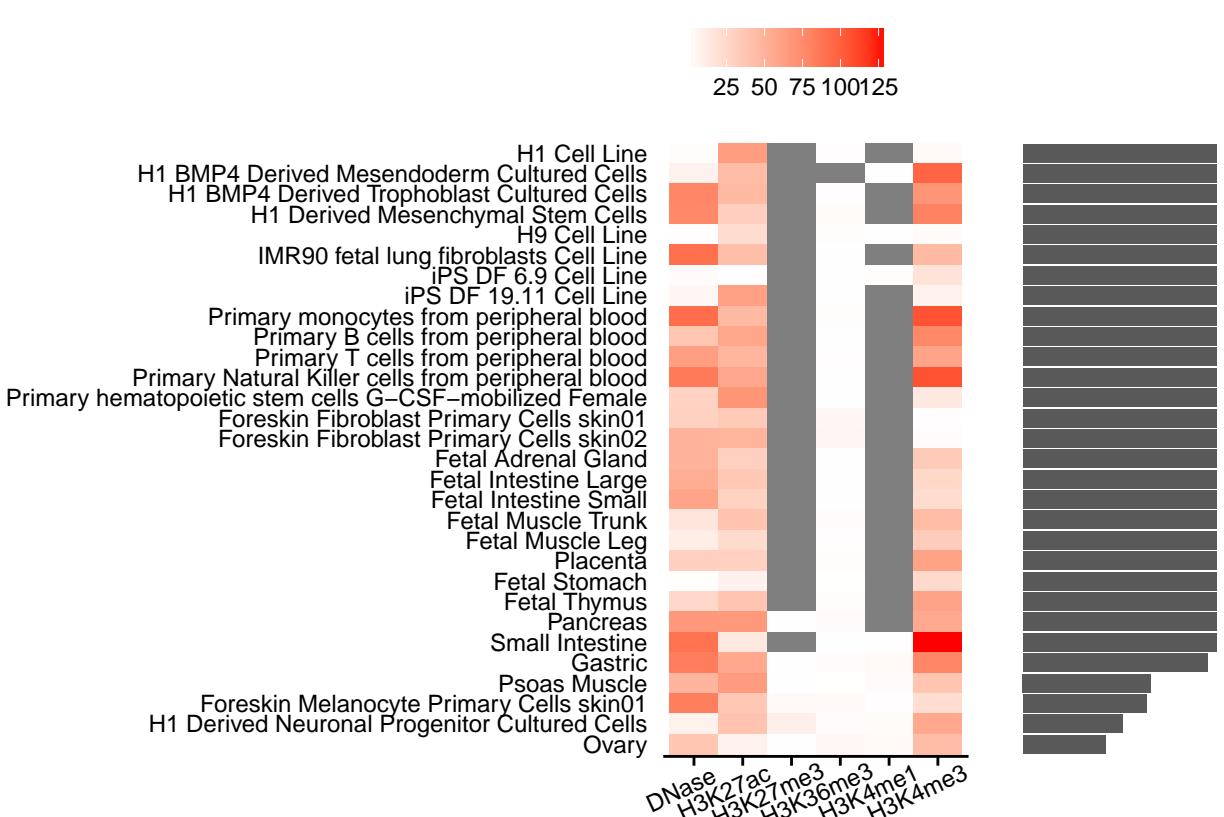
Next, we used weighted gene network correlation analysis (WGCNA) to further assess the association of groups of CpG sites with incident CVD events without requiring their regional proximity. WGCNA uses weighted correlation networks to identify highly correlated submodules that may be more interpretable and more robust as a group than the individual components. In order to focus on sites that are well-annotated, variable, and have better understood directional effects on transcription, only CpGs within 1.5kb of gene transcription start sites ($n=66772$) were included.

The WGCNA analysis in WHI returned 41 modules, 4 of which were at least nominally associated with incident events after adjustment for the same set of covariates as in the initial EWAS. Based on gene enrichment analyses, these modules associated with extracellular matrix function (pink), developmental processes (blue), and cholesterol binding and metabolism (lightyellow), while the fourth (mediumpurple3) consisted of CpGs in a single locus on chromosome 6 near RNF5P1 and AGPAT1. Based on thresholds suggested by the authors, z-scores calculated for these four modules in FHS showed strong preservation (all z-scores >10). However, two of these modules (mediumpurple3 and blue) did not show an association with incident CVD events after adjustment for technical covariates, and the other two (pink and lightyellow) did not show association after additional adjustment for age.

2.5 Incorporation of genetic and epigenetic annotations improves mechanistic interpretation

To probe the cell type specificity of the observed associations, a novel procedure based on epigenomic annotations was used. Peak calls for a series of epigenomic features (DNase hypersensitivity sites and multiple





histone marks) have been determined as previously described (???) and are available from the Roadmap Epigenomics database. These data were retrieved for a broad array of cell types in order to assign CpGs analyzed here to cell- and feature-specific peaks. For each annotation, t-tests were performed to compare the significance of sites ($-\log_{10}(p)$ from EWAS) to their membership in a peak. Overall, this method would be expected to show greater significance in cell types relevant to the observed signal, whether as a specific component of a mixed tissue (e.g. blood) or as a surrogate for the relevant tissue.

In this cross-cohort analysis, monocyte annotations showed the strongest associations with EWAS signal across epigenomic features, with other blood cell types showing many of the other top ranks. Non-blood cell types also showing high ranks, potentially representative of blood acting as a surrogate tissue, included fibroblast, mesenchymal stem cells, and fetal tissues.

The positions of the 8 DMPs and 4 DMRs were assessed for their proximity the set of genome-wide significant ($p < 1e-8$) variants from the CARDIoGRAMplusC4D meta-analysis (Nikpay et al. 2015) (here, within 1MB of a GWAS hit). One DMP, cg27111756 in the gene body of PPP1R13L, was found to be about 500kb downstream of a series of hits in the TOMM40-APOE-C1-C2-C4 cluster on chromosome 19. Hi-C chromatin interaction data visualized using the 3DIV database (D. Yang et al. 2017) indicated that this DMP and SNP cluster mark the approximate respective boundaries of a topologically associating domain (TAD) in the GM12787 lymphoblastoid cell line. In some of the top-ranking cell types from the cell type-specific analysis above, this same TAD did not exist, although specific long-range chromatin contacts between these SNPs and the DMP were found (for example, in lung fibroblasts and mesenchymal stem cells).

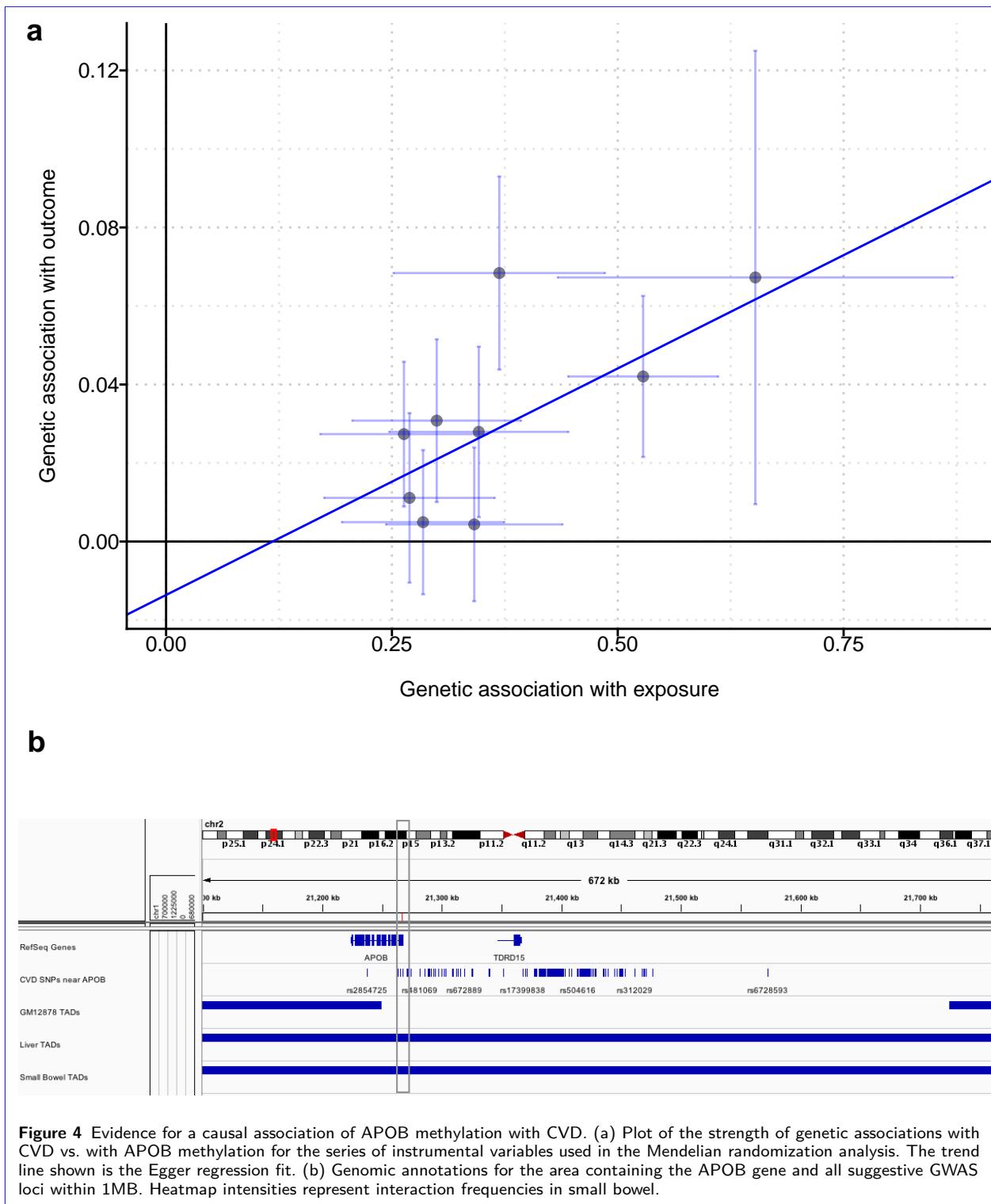
Though the prospective nature of this EWAS analysis precludes reverse causation, associations between DMPs/DMRs and CVD are still subject to confounding. To better examine the causality of these relationships, a Mendelian randomization (MR) approach using single nucleotide polymorphisms (SNPs) as instrumental variables was undertaken through the use of external datasets. SNP-methylation associations were taken from the mQTLdb (ARIES study) (Gaunt et al. 2016) and SNP-CVD associations were taken from the CARDIoGRAMplusC4D meta-analysis as above.

The inverse-variance weighted method was used to assess the causal relationship between differential methylation and potential pleiotropy of the instrumental variables (here, SNPs) for the 8 DMPs and 4 DMRs. The most significant CpG site from each region was chosen as the “lead CpG” to be used in the MR analysis. 7 methylation loci lacked links to any SNPs in the mQTLdb. Of the remaining set of CpGs, one (cg07636176, from the APOB region) demonstrated a consistent and “dose-dependent” association of its IV SNPs with CVD ($p = 3.3e-06$; Fig. 4a). MR-Egger regression (J. Bowden, Smith, and Burgess 2015) was then applied to address potential pleiotropy of the IV SNPs, and showed no direct evidence of pleiotropy (MR-Egger intercept $p = 0.57$), though the causal relationship estimate was attenuated (MR-Egger slope $p = 0.091$).

Based on this result, we expanded the previous CVD GWAS-related analysis to include genome-wide “suggestive” SNPs ($p < 1e-5$) and found that the APOB DMR was within 1MB of 227 of these (non-independent) loci. Because this gene is expressed primarily in liver and small intestine, we examined Hi-C interaction data as above for this locus in these two tissues as well as the GM12787 lymphocyte precursor. This region was contained by a TAD in the two tissues of interest (liver and small bowel) but not in the blood cell precursor (Fig. 4b).

(Note to readers: This regional plot is just a stand-in – recommendations for publication-quality/nice-looking browsers to incorporate annotations, TADs, etc. are welcome.)

To further probe the mechanistic relationship between methylation levels and CVD-related biological activity, the SEPIRA method (Y. Chen, Widschwendter, and Teschendorff 2017) was used to infer blood-specific



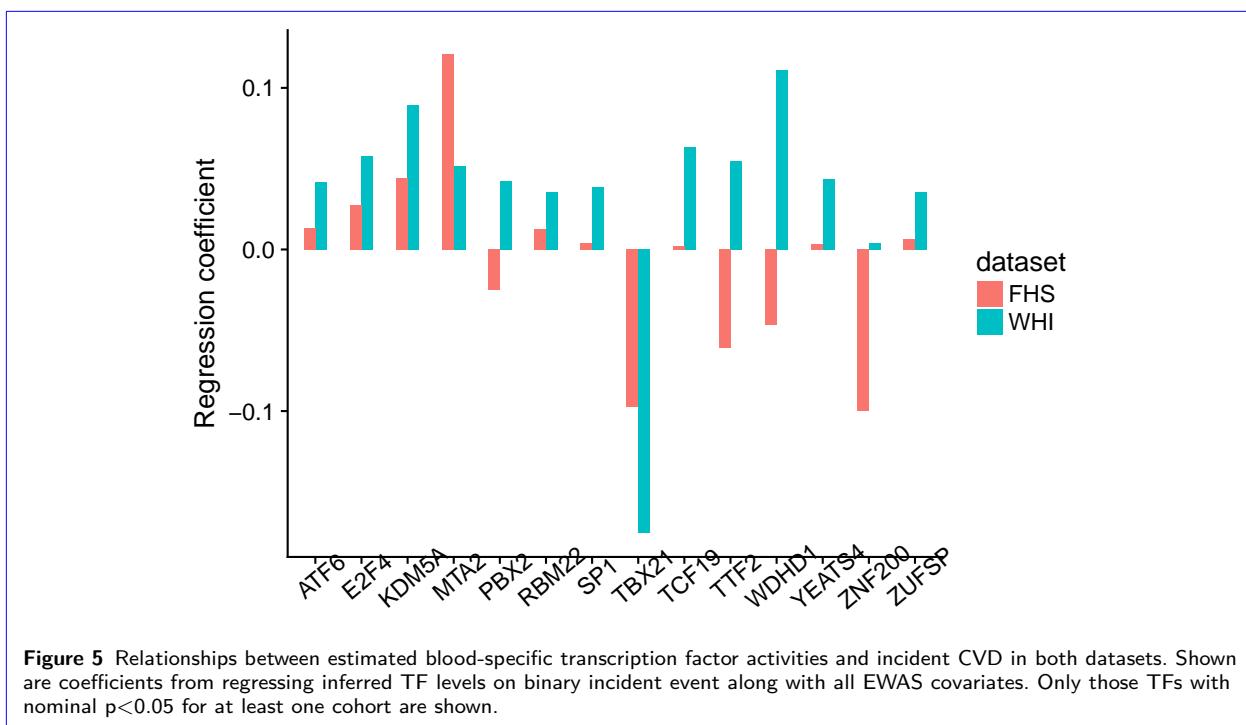


Figure 5 Relationships between estimated blood-specific transcription factor activities and incident CVD in both datasets. Shown are coefficients from regressing inferred TF levels on binary incident event along with all EWAS covariates. Only those TFs with nominal $p < 0.05$ for at least one cohort are shown.

transcription factor activity. First, blood-specific transcription factors are determined using expression data (here, from GTEx), then levels of these TFs in each sample are inferred from DNA methylation data.

A set of 51 blood-specific transcription factors were found based on the input expression data from GTEx. Activity profiles for this set of TFs were then inferred from DNA methylation data for each sample. No TFs reached significance and replication across datasets, but two (TBX21 & MTA2) showed moderate effects and consistency in direction.

2.6 Comparison to known risk factors

To understand how these findings fit into the context of known CVD biology and risk assessment, we evaluated relationships between DMPs and DMRs (represented by their lead CpGs) with traditional blood-based CVD risk factors. First, simple correlations were examined between EWAS sites (single and lead CpGs from DMRs) and log-transformed risk factors (Fig. 6a). A few sites, including cg06500161 (ABCG1 body) and cg02519286 (lead CpG from the SLC9A1 DMR) showed strong correlations with multiple risk factors. Others, such as cg11772801 (TMX1 body), showed little such correlation, indicating that these CpGs may relate to incident CVD through other mechanisms.

Next, a partial correlation network was constructed to enable inspection of putative direct relationships between risk factors and the CpG sites uncovered here (Fig. 6b). The network revealed a general clustering of known blood-based risk factors apart from CpG sites. Unlike the other risk factors, hsCRP demonstrated independent correlations with many CpG sites. Furthermore, a strongly associated cluster of CpGs annotated to genes MS4A1, APBA2, APLP2, and TMX1 was apparent. Triglyceride levels served as a central node in the network, linking to multiple CpG sites, as well as systolic blood pressure, lipid levels, and diabetes status.

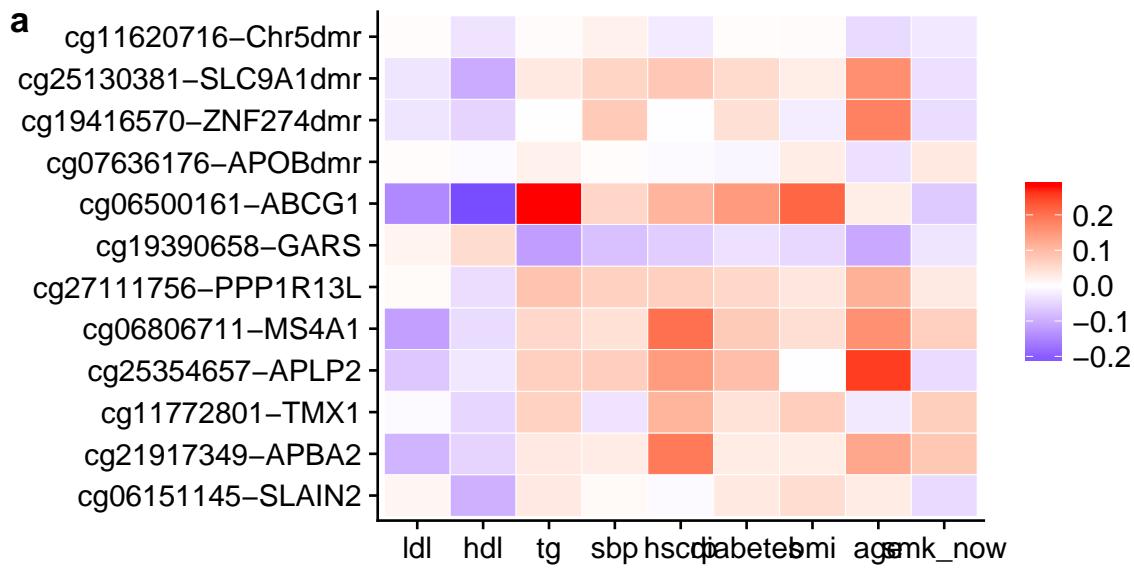
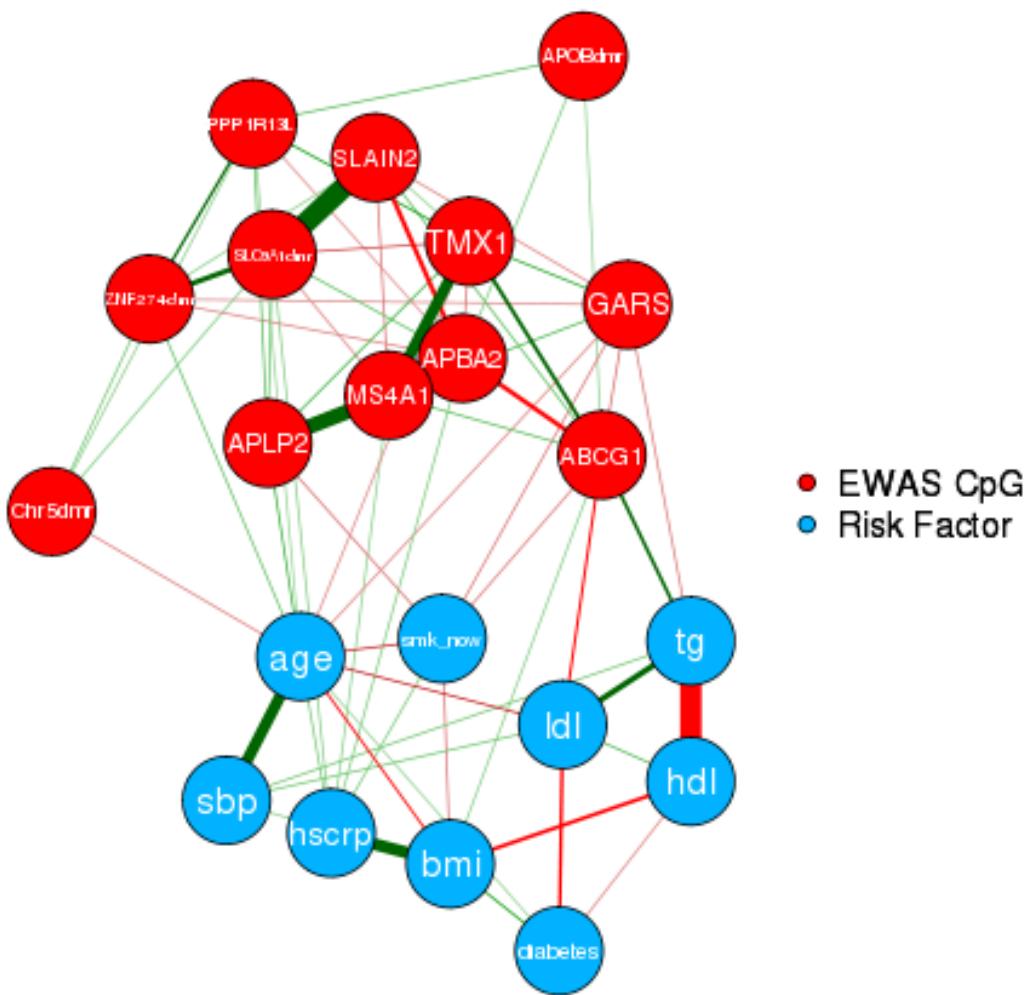
**b**

Figure 6 Relationships between methylation and traditional CVD risk factors. (a) Pearson correlations between DMPs and DMR lead CpGs and risk factors. (b) Partial correlation network between and among the same CpGs and risk factors. Edge widths correspond to partial correlation magnitudes, and edge colors to partial correlation sign (green positive; red negative). CpG sites are labeled based on their associated genes, and CpGs representing DMRs are labeled as such.

3 Discussion

Investigation of DNA methylation data in samples taken before onset of a disease addresses the central difficulty of reverse causation in epigenome analyses. Here, we looked for prospective biomarkers of CVD onset using survival models, and discovered a set of CpG sites and regions that are suggestively associated across the FHS and WHI datasets.

- basic CpG sites rundown – have these been discovered elsewhere?

In order to better understand the effects of sex and race differences across the two datasets, we undertook stratified EWAS analyses. First, the top regression coefficients from WHI were compared to those from the same sites in men and women from FHS. Though women did show somewhat greater correlation of coefficients, neither stratum agreed as well as the full FHS dataset, indicating that sex is not a major driver of the observed heterogeneity in EWAS results. Next, an equivalent comparison was made between FHS coefficients and those from separate races in WHI. Here, coefficients from both white and hispanic subjects showed greater agreement than the full population, while there was essentially no correlation in black subjects. Importantly, the group of hispanic subjects in WHI are mostly of Mexican descent (Zambrana et al. 2014), and thus share a greater genetic similarity to those of European rather than African descent. Previous observations have supported the generalizability of cardiovascular risk functions from the Framingham Heart Study to diverse populations (E. R. Fox et al. 2016), and a reasonable overlap of genetic risk loci across ancestries (Lettre et al. 2011). However, the patterns observed here suggest caution in generalizing epigenetic results from FHS to other populations. Despite this intriguing pattern with respect to race, the improvement in correlation was modest, suggesting that demographic differences were not generally driving the observed differences in epigenetic associations.

- anything explicit about race dominating over sex in explaining epigenetic patterns?

Though no strong patterns were observed at the single-CpG level, we searched for differentially methylated regions, which may be more statistically robust and biologically meaningful (???). Four associations passed a strict multiple test correction threshold and were validated in FHS, the strongest of which was found in the _____ region of the APOB gene. There is existing support for both mechanistic (Goff et al. 2014) and genetic (C. Liu et al. 2015) associations between APOB and cardiovascular disease, related to its function in cholesterol transport. Indeed, a recent Mendelian randomization-based analysis suggested a causal link between _____ the same _____ CpG and ApoB, the APOB protein product and known CVD risk factor (Richardson et al. 2017). APOB methylation has been linked to differential expression in liver cells (El Taghdouini et al. 2015), but whether this translates to increased circulating protein levels is not clear.

To expand on this observation, we undertook a Mendelian randomization analysis to examine potential causal relationships between the regions discovered here and cardiovascular disease, using existing databases for mQTLs (the same as was used in Richardson et al. for CVD risk factors) and CVD GWAS associations. We showed a possible causal relationship between methylation at the lead CpG from the APOB region and CVD. It is particularly interesting that this relationship was uncovered in whole blood samples, as the APOB gene is expressed primarily in small intestine and liver. Thus, in this case, blood may be acting as a surrogate reflecting methylation changes in the primary disease-relevant tissue.

Identification of specific cell types of interest may be valuable for understanding disease mechanisms as well as informing follow-up studies in specific cell types (Bauer 2018). To further probe the cell-type specificity of our EWAS signal, we used a novel method based on the assumption that correlations between epigenomic feature annotations (e.g. histone marks) and EWAS signal strength will be stronger in relevant cell types. By averaging the strengths of this association across features for each cell type, we were able to rank cell types

based on their potential relevance to our overall signal. As expected, blood cell types were all highly ranked. Monocytes show the greatest importance based on this metric, supporting their well-known integral role in CVD pathogenesis (Libby 2006). We also observed high ranks for fibroblasts and adrenal cells, both of which have been linked to CVD (Murtha et al. 2017; Rosmond and Björntorp 2000), and which may suggest additional tissues for which blood is acting as a surrogate.

- WGCNA discussion

The methylation patterns observed here may be acting up- or downstream of known cardiovascular risk pathways (e.g. lipids), or may represent novel disease mechanisms. We looked at correlations between lead CpGs from our identified DMRs and classical CVD risk factors, including BMI, diabetes status, lipid levels, blood pressure, and hsCRP. None of the DMRs correlated strongly with these risk factors. Surprisingly, this included a lack of correlation between APOB methylation and cholesterol levels, which would otherwise be expected based on the biological function of the ApoB protein product.more when have bmi. To move closer to mechanistic or causal relationships, we constructed a partial correlation network to display only Bonferroni-level correlations conditional on all other DMRs and risk factors. Most traditional risk factors showed strong connections to each other, to the exclusion of the DMRs. However, hsCRP stood out as a notable exception, showing stronger conditional correlations with methylation levels than with other risk factors. This suggests that these DMRs mediate or track disease pathways more related to inflammation than lipid levels, which may be intuitive given that they were assessed in blood samples.

There are a few important limitations to this association study. As discussed, only blood samples were analyzed, which represent an important tissue for CVD pathology but do not fully capture epigenetic dynamics of other disease-relevant tissues. Demographic heterogeneity of the two datasets may have contributed to a lack of sensitivity in identifying or validating population-specific epigenetic associations with CVD. Additionally, as this study examines relationships between baseline methylation levels and incident disease, the measurement of DNA methylation at only a single time point precluded an understanding of the stability of these epigenetic signatures over time.

- Summary para.

4 Methods

4.1 Study participants and phenotype collection

Data for the discovery set came from a combined case-control and pseudo case-cohort sampling of 2129 women from the Women's Health Initiative study. Included subjects had no self-reported CVD at baseline, and cases experienced an incident CHD event during follow-up. Data are available in the dbGaP public repository (accession: phs000200.v11.p3).

Data for the validation set came from a substudy of the Framingham Heart Study that measured DNA methylation in 2726 subjects from the Offspring Cohort. Fasting blood samples for both methylation and clinical biochemistry were collected from participants at Exam 8. Data are available in the dbGaP public repository (accession: phs000007.v29.p10).

More details on the cohorts will be requested from collaborators.

CVD events were defined as adjudicated coronary heart disease events, stroke, or death from CVD.

4.2 DNA methylation data processing

In both cohorts, DNA methylation data were collected using the Illumina HumanMethylation450 microarray platform. (reference) Preprocessing was performed using the *minfi* and *wateRmelon* packages for R. As a quality control step, samples were removed if they showed weak overall signal based on visual inspection of an intensity plot, if they had more than 10% of probes undetected at a detection threshold of $p < 1e-16$, or if the reported sex did not match the predicted sex based on methylation patterns. Probes were removed if they met any of the following criteria: more than 10% of samples undetected at a detection threshold of $p < 1e-16$, location in the X or Y chromosomes, non-CpG probes, cross-hybridizing probes, probes measuring SNPs, and probes with an annotated SNP at the CpG site or in the single-base extension region. Samples were normalized using the Noob method for background correction and dye-bias normalization, followed by the BMIQ method for probe type correction. For each dataset, principal components analysis was performed on the set of control probes using code adapted from the CPACOR method of Lehne et al. (Lehne et al. 2015) to account for technical variation. Blood cell counts for 6 blood cell types were estimated using the reference-based method of Houseman et al. (Houseman et al. 2012).

4.3 Association of DNA methylation with incident CVD events

After quality control and filtering steps, 422952 (WHI) and 425326 (FHS) CpG sites remained for downstream analysis. Visualization of the associations between top principal components for each dataset and major biological and technical variables was examined using heatmaps (Supp. Fig. S1). Based on these assessments as well as examination of Scree plots for CPACOR principal components,

For each CpG site, a Cox proportional hazards regression was performed to assess the relationship between methylation beta value as the independent variable and time-to-event measures for incident CVD while adjusting for biological covariates (age, bmi, smoking status, smoking pack-years, and estimated blood cell counts; race in WHI; sex in FHS) and technical covariates (DNA pull batch in WHI; analysis center in FHS). To remove the influence of outliers in the regression, samples were removed from the analysis for each CpG if their beta value was outside of an extended outer fence ($< 25\text{th} \text{ile} - 5 * IQR$ or $> 75\text{th} \text{ile} + 5 * IQR$). QQ plots revealed that genomic inflation was not adequately controlled in FHS, so 7 CPACOR principal components (chosen based on a Scree plot assessment) were additionally adjusted for in FHS.

For the main analysis of single sites, a specific subset of variable sites were selected based on a previous publication that performed bisulfite sequencing in isolated monocytes, CD4+ T-cells, and neutrophils (Hachiya et al. 2017). This study calculated “reference intervals”, defined as the difference in beta value between the 5th and 95th percentile, as a measure of the variability of a given CpG site. Based on their provided recommendation, sites were included here if they demonstrated a reference interval of more than 30 in any cell type.

4.4 Gene set enrichment analysis

Gene ontology-based enrichment analysis of DMPs was performed using the gometh function from the *missMethyl* package for R. In this procedure, gene annotations for CpG sites are retrieved, and enrichment analysis is performed while accounting for the bias in number of CpG probes per gene on the microarray.

4.5 DMR detection

Comb-p (implemented as a Python module) was used to call differentially methylated regions (DMRs) based on p-values from the EWAS, thus removing the requirement for additional covariate adjustment. Comb-p

first calculates an autocorrelation function (ACF), for which a maximum distance of 1kb and a step size of 50 bases were used. Next, it uses the ACF to adjust each p-value using a Stouffer-Liptak-Kechris correction, followed by identification of contiguous regions of sites with adjusted p-values below some threshold (here, $p < 0.1$ with no more than 500 bases between neighboring sites in a region). Finally, the ACF is recalculated out to the maximum region size (a step size of 50 was used here as well) and regional p-values are calculated using the Stouffer-Liptak test.

For multiple testing correction of DMRs, Comb-p calculates the number of effective tests as the number of loci tested divided by the size of the region (typically ~1kb). A more stringent correction was applied here, in which the number of independent tests was calculated as the number of loci tested divided by the number of probes in the region.

4.6 Weighted gene correlation network analysis

Weighted gene correlation network analysis (WGCNA) was used to find highly correlated modules of CpG sites. The set of input CpGs, chosen to be generally the most interpretable and variable, were selected as those upstream of transcription start sites (“TSS200” and “TSS1500” based on Illumina’s annotation file) but not directly in CpG islands (“Island” in the annotation file), resulting in a set of 66772 sites. For computational tractability, blockwise module detection was performed, which treats block of features separately for network creation and module detection, followed by eventual merging of highly similar modules. For the main analysis performed in WHI, parameters included a block size of 20,000, a soft-thresholding power of 8, and a signed network type. Modules were assessed using the same regression framework and covariate set as was used in the initial EWAS, replacing single-site beta values with the first principal component of the module beta value matrix (“eigenCpG”). Enrichment analysis was carried out in each of the nominally associated modules using the method described above. Module preservation assessment in was completed in FHS based on z-scores, followed by calculation of eigenCpGs (according to the principal component weights from WHI) and assessment of associations with incident CVD.

4.7 Inference of transcription factor activity

The SEPIRA method (Y. Chen, Widschwendter, and Teschendorff 2017) was used to infer the activity of blood-specific transcription factors from DNA methylation data (code was adapted from Chen et al.). RNA-Seq data for multiple cell types was obtained from GTEx (release V7; TPM format) and used to infer a blood-specific regulatory network through comparison of transcription factor (TF) expression in blood cells vs. other cell types. Next, activity of these cell type-specific TFs was inferred after estimating cell type-specific gene activity from methylation as introduced in Jiao et al. (Jiao, Widschwendter, and Teschendorff 2014) Briefly, methylation-based gene activities are estimated by assessing average methylation levels near transcription start sites of genes, with the assumption that methylation in these regions consistently associates with decreased activity. TF activity for each sample is then assessed by regressing these gene activity profiles on the TF gene target profile. These TF activities were then assessed in relation to incident CVD using the same Cox regression method and covariate adjustment as was used in the EWAS as well as for the modules from WGCNA above.

4.8 Inference of cell type specificity

Epigenomic annotations were used to test for relative enrichment of strong EWAS signal in CpG sites located in peaks for epigenomic features in specific cell types. Annotations for broad peaks in DNase sensitivity as

well as ChIP-seq signal for H3K4me1, H3K4me3, H3K27ac, H3K27me3, and H3K36me3 were obtained for 30 cell types from the NIH Roadmap Epigenomics Project database [[cite the specific 111 epigenomes paper]]. For each combination of epigenomic feature and cell type, CpGs from the HumanMethylation450 array were classified as to their membership in a peak locus. EWAS results from WHI and FHS were combined by taking the mean of their $-\log_{10}(p)$ values. A t-test was then used to compare these significance values based on peak membership of the associated sites. T-statistics from these t-tests were used to assess the relative strength of the EWAS significance-peak membership associations in across cell types.

4.9 Genetic data integration

Methylation-quantitative trait locus (mQTL) relationships were obtained from the mQTLdb (mqtlDb.org). These relationships were generated by Gaunt et al. (Gaunt et al. 2016) from an analysis of paired genotype and methylation data from the ARIES longitudinal study of mother-child pairs. Only relationships found at the “Middle Age” time point were used here. GWAS summary statistics for coronary artery disease and myocardial infarction (based on imputation using 1000 genomes phase 1 v3) were contributed by CARDIoGRAMplusC4D investigators (Nikpay et al. 2015) and were downloaded from www.cardiogramplusc4d.org.

Mendelian randomization analysis was performed using the R package *MendelianRandomization*. To generate the input for this analysis, SNPs (instrumental variables; IVs) were found in the mQTLdb that associated with the set of top DMPs and DMRs (exposures), with the most significant EWAS CpG from each DMR being used as the representative exposure variable. Only SNPs on the same chromosome as the CpG of interest were included. For CpGs with multiple associated mQTL SNPs, the list of SNPs was first pruned to correlations of $r < 0.2$ using the NCI tool SNPclip (<https://analysistools.nci.nih.gov/LDlink/?tab=snpclip>). The inverse-variance weighted estimator (mr_ivw function), which assumes that all variants are valid instrumental variables, was used initially to calculate causal effect estimates. MR-Egger regression (mr_egger function) was then used to assess this instrumental variable assumption.

4.10 Risk factor comparison

CVD risk factors were log-transformed to reduce the influence of extreme values on correlations. Methylation beta-values in FHS were adjusted for center and 7 control-probe principal components (as in the EWAS) using linear models, and the residuals were taken forward for use in the correlation analysis. Pearson correlations were computed between each risk factor and each CpG of interest (8 suggestive CpGs plus 4 DMR lead CpGs). Next, the full Pearson correlation matrix consisting of risk factors and these CpG sites was transformed into a partial correlation matrix using the R package *qgraph*. The partial correlation network was visualized using a network format after removing edges that did not pass a Bonferroni correction.

5 Supplementary Data

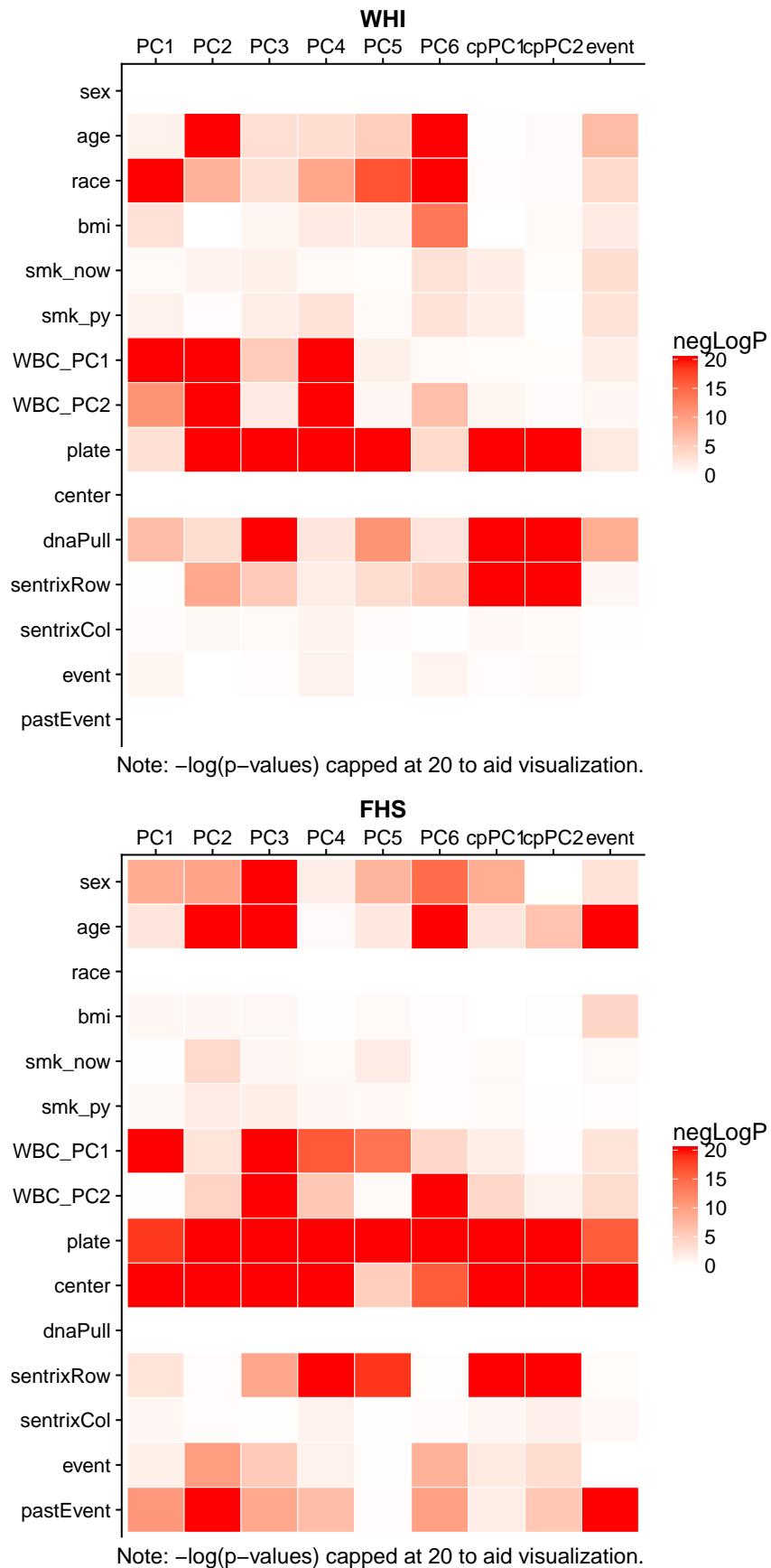
Table S1: GSEA results for CpGs with FDR<0.25 in either dataset.

Term	Ont	N	DE	P.DE	FDR
embryo development	BP	823	104	4.00e-07	0.0078393
cell fate commitment	BP	221	41	1.00e-06	0.0096593
sequence-specific DNA binding	MF	863	98	1.40e-06	0.0096593
transcription from RNA polymerase II promoter	BP	1609	157	3.90e-06	0.0160268

homophilic cell adhesion via plasma membrane adhesion molecules	BP	146	37	3.60e-06	0.0160268
regulation of transcription from RNA polymerase II promoter	BP	1474	148	6.40e-06	0.0217230
nucleobase-containing compound biosynthetic process	BP	3272	260	1.43e-05	0.0285666
embryonic morphogenesis	BP	503	69	1.83e-05	0.0285666
transcription regulatory region sequence-specific DNA binding	MF	559	67	1.75e-05	0.0285666
double-stranded DNA binding	MF	640	73	1.54e-05	0.0285666
transcription factor binding	MF	409	51	1.69e-05	0.0285666
nucleic acid binding transcription factor activity	MF	967	102	1.50e-05	0.0285666
transcription factor activity, sequence-specific DNA binding	MF	966	102	1.45e-05	0.0285666
heterocycle biosynthetic process	BP	3314	261	2.17e-05	0.0315571
RNA polymerase II regulatory region sequence-specific DNA binding	MF	491	61	2.60e-05	0.0330159
RNA polymerase II regulatory region DNA binding	MF	491	61	2.60e-05	0.0330159
aromatic compound biosynthetic process	BP	3325	261	2.94e-05	0.0351657
organic cyclic compound biosynthetic process	BP	3416	266	3.62e-05	0.0398727
smooth muscle cell differentiation	BP	42	12	3.91e-05	0.0398727
sequence-specific double-stranded DNA binding	MF	580	67	3.92e-05	0.0398727
regulation of gene expression	BP	3300	262	4.53e-05	0.0438428

Table S2: Top DMRs from WHI only.

chr	start	end	num_cpgs	p_region_slk	p_adj_region_sidak	p_adj_region	annot_gene
chr2	21266499	21267334	12	0e+00	0.00e+00	0.00e+00	APOB
chr2	219738225	219738732	5	0e+00	0.00e+00	1.00e-07	WNT6
chr1	153762200	153762564	6	0e+00	0.00e+00	1.00e-07	
chr6	29594480	29595661	16	0e+00	0.00e+00	1.00e-07	GABBR1
chr19	58715250	58716517	8	0e+00	0.00e+00	2.00e-07	ZNF274
chr2	241458885	241460002	8	0e+00	0.00e+00	4.00e-07	ANKMY1
chr5	139490894	139492727	7	0e+00	0.00e+00	9.40e-06	
chr22	30901248	30901886	8	0e+00	4.00e-07	3.21e-05	SEC14L4
chr3	185911207	185912486	9	0e+00	6.00e-07	8.33e-05	DGKG
chr6	33083988	33085470	21	0e+00	1.30e-06	9.17e-05	HLA-DPB2
chr10	102806870	102808252	7	0e+00	9.00e-07	1.85e-04	
chr7	27127447	27128169	6	0e+00	4.60e-06	5.59e-04	
chr17	72931568	72932048	12	0e+00	2.78e-05	1.11e-03	OTOP3
chr3	51740740	51741473	7	0e+00	1.10e-05	1.15e-03	GRM2
chr5	131592958	131593812	11	0e+00	2.04e-05	1.58e-03	PDLIM4
chr1	158147445	158147849	3	0e+00	1.29e-05	1.74e-03	
chr15	76631568	76635953	19	1e-07	7.80e-06	1.79e-03	
chr17	36828122	36829758	6	0e+00	7.00e-06	1.92e-03	C17orf96
chr3	350502	351003	6	0e+00	2.36e-05	1.97e-03	CHL1
chr19	19280968	19281559	10	1e-07	8.41e-05	4.97e-03	LOC729991-MEF2

**Figure S1** Significance of regression of top PCs from methylation beta value matrix on covariates.

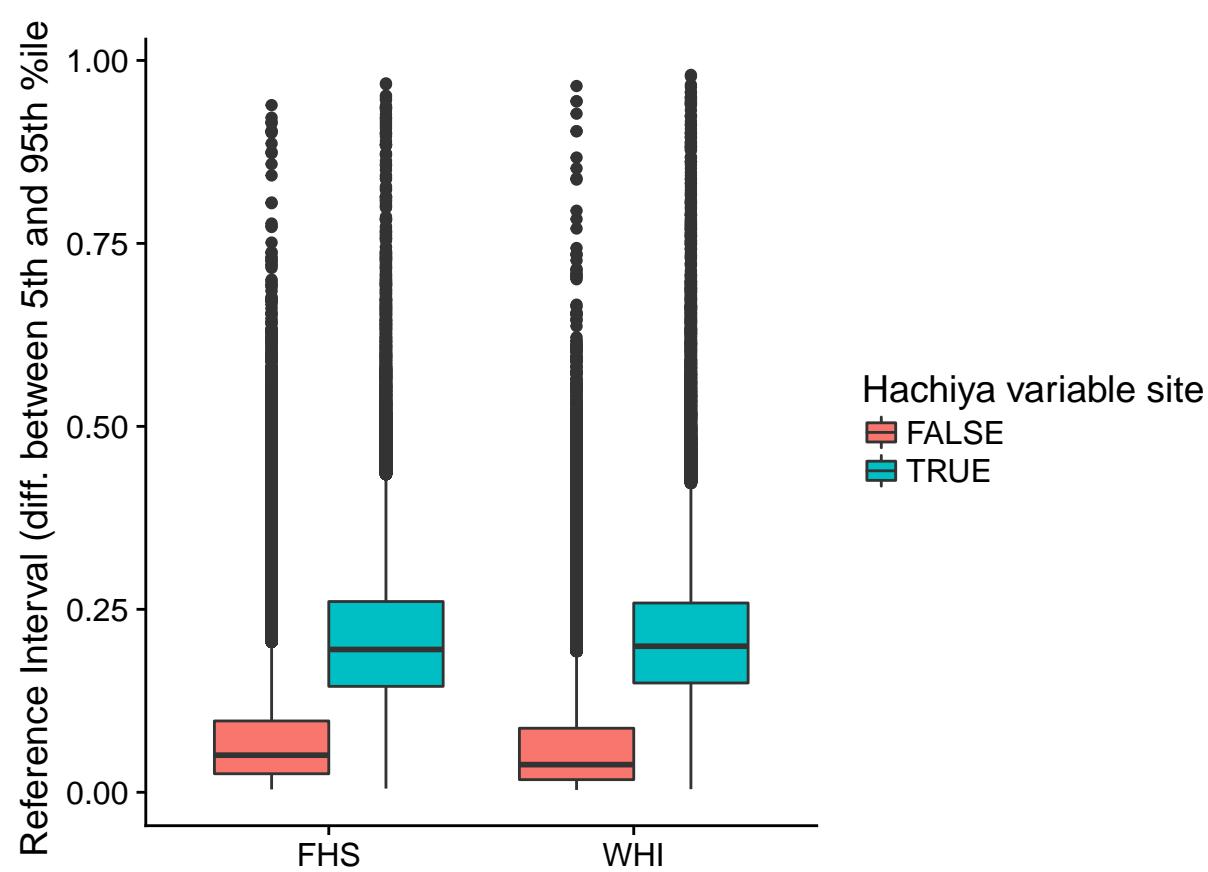


Figure S2 Reference interval comparison between variable sites as taken from Hachiya et al. and all other sites (RI = difference in beta value between 5th and 95th percentiles).

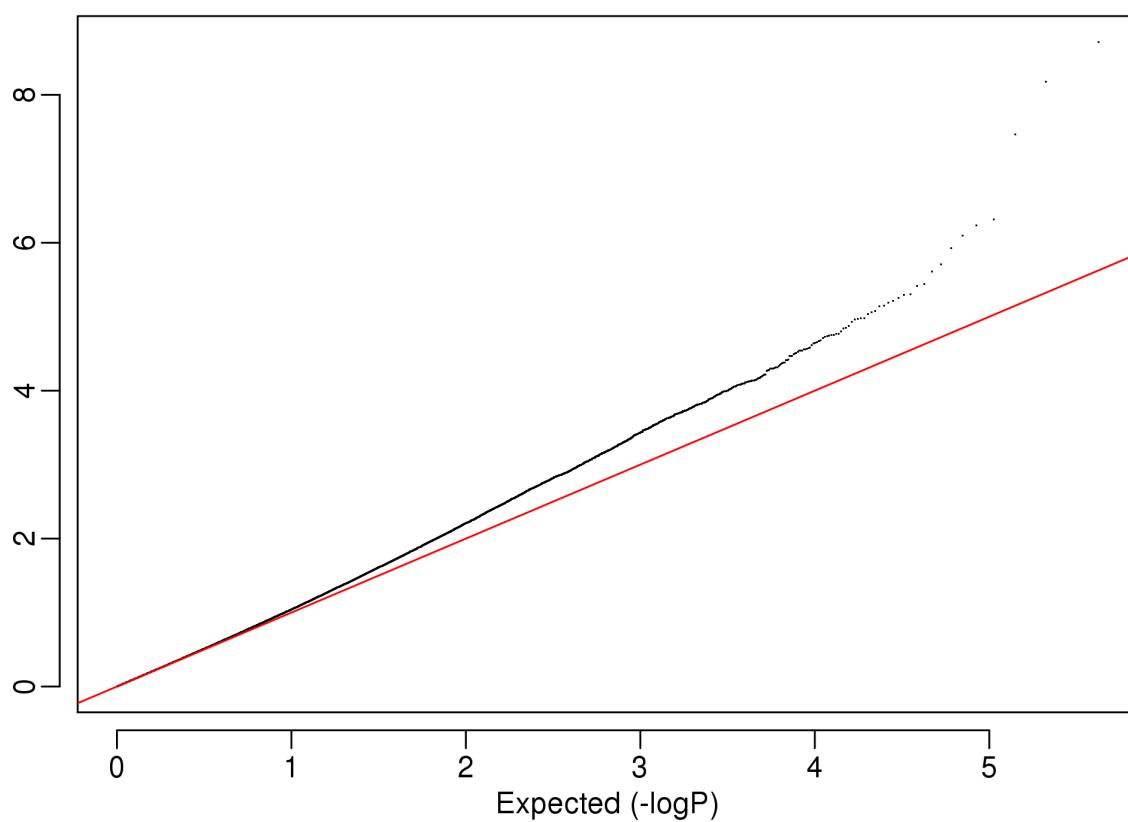
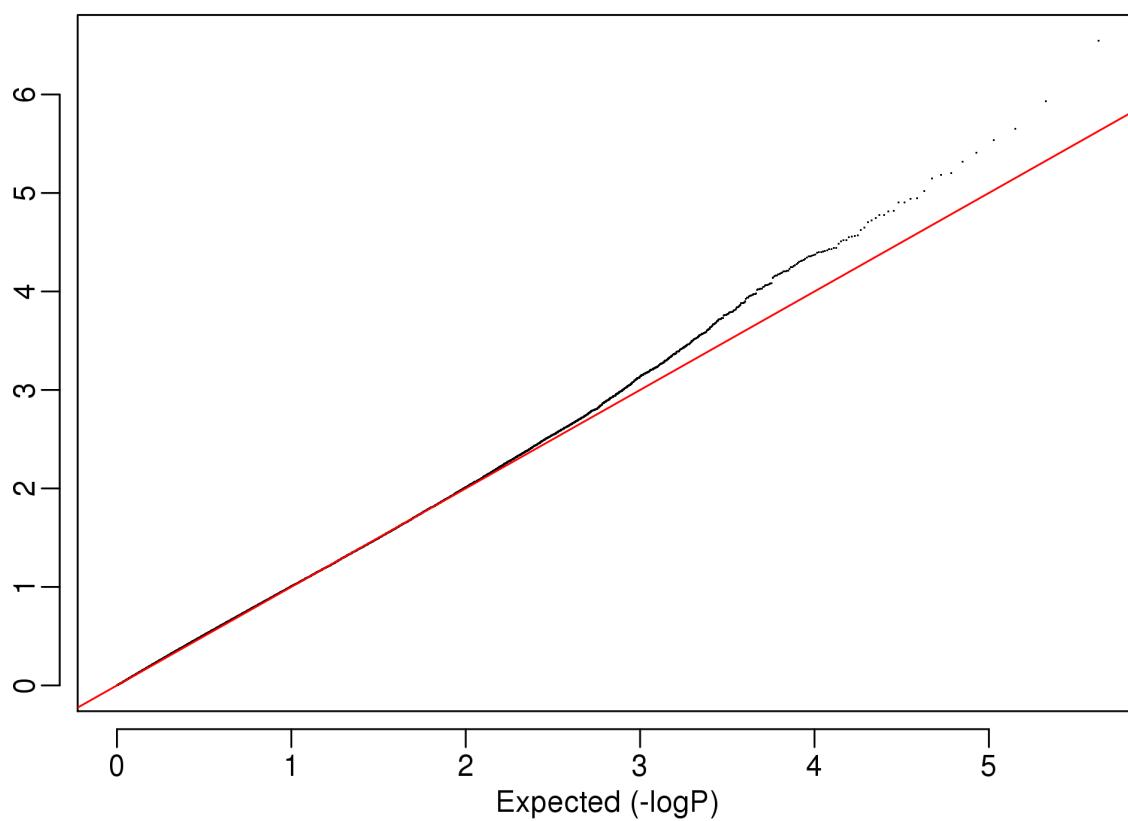
WHI: lambda = 1.02**FHS: lambda = 1.06**

Figure S3 QQ-plots for EWAS results, including the full set of CpG sites passing quality control steps.

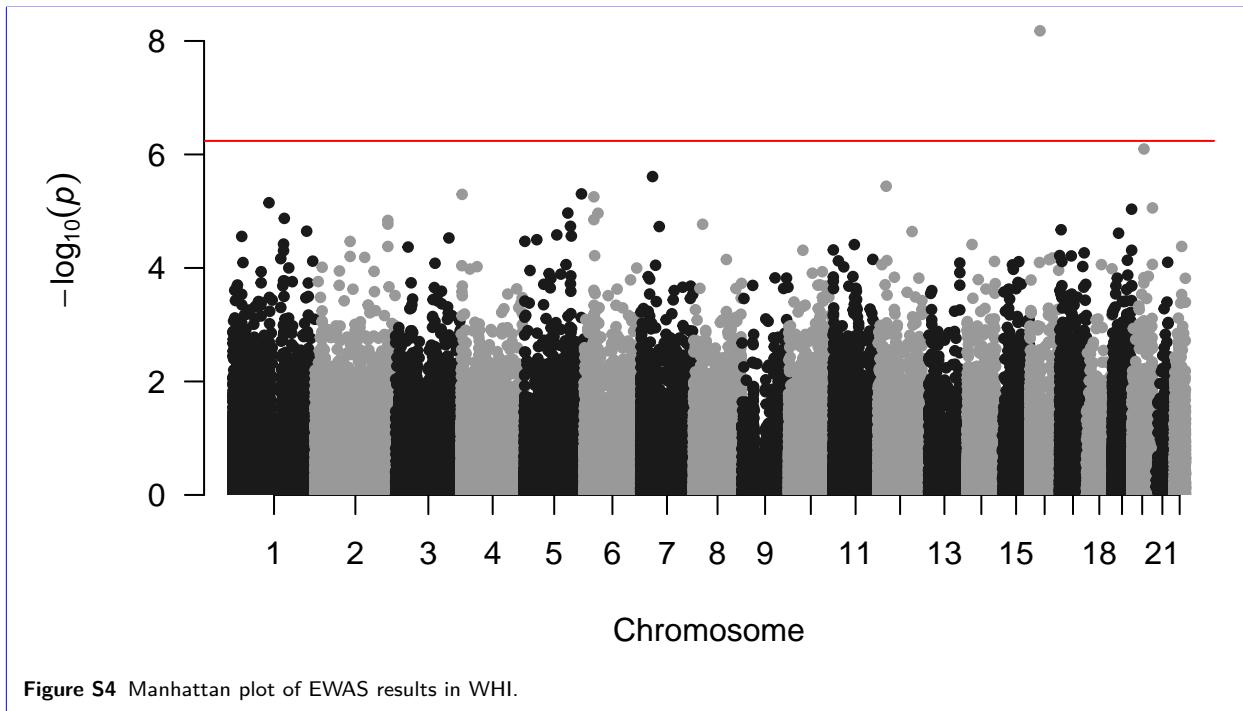


Figure S4 Manhattan plot of EWAS results in WHI.

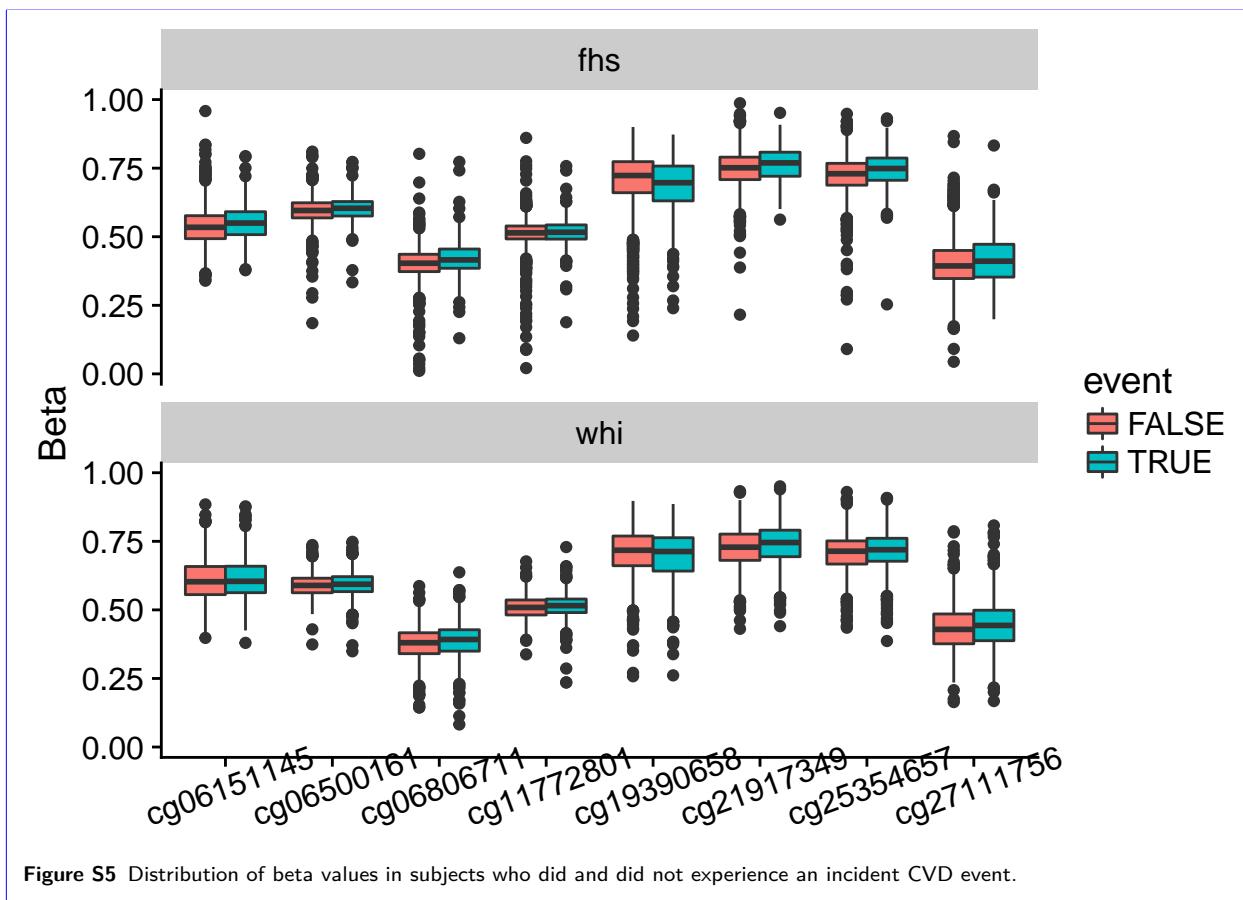


Figure S5 Distribution of beta values in subjects who did and did not experience an incident CVD event.

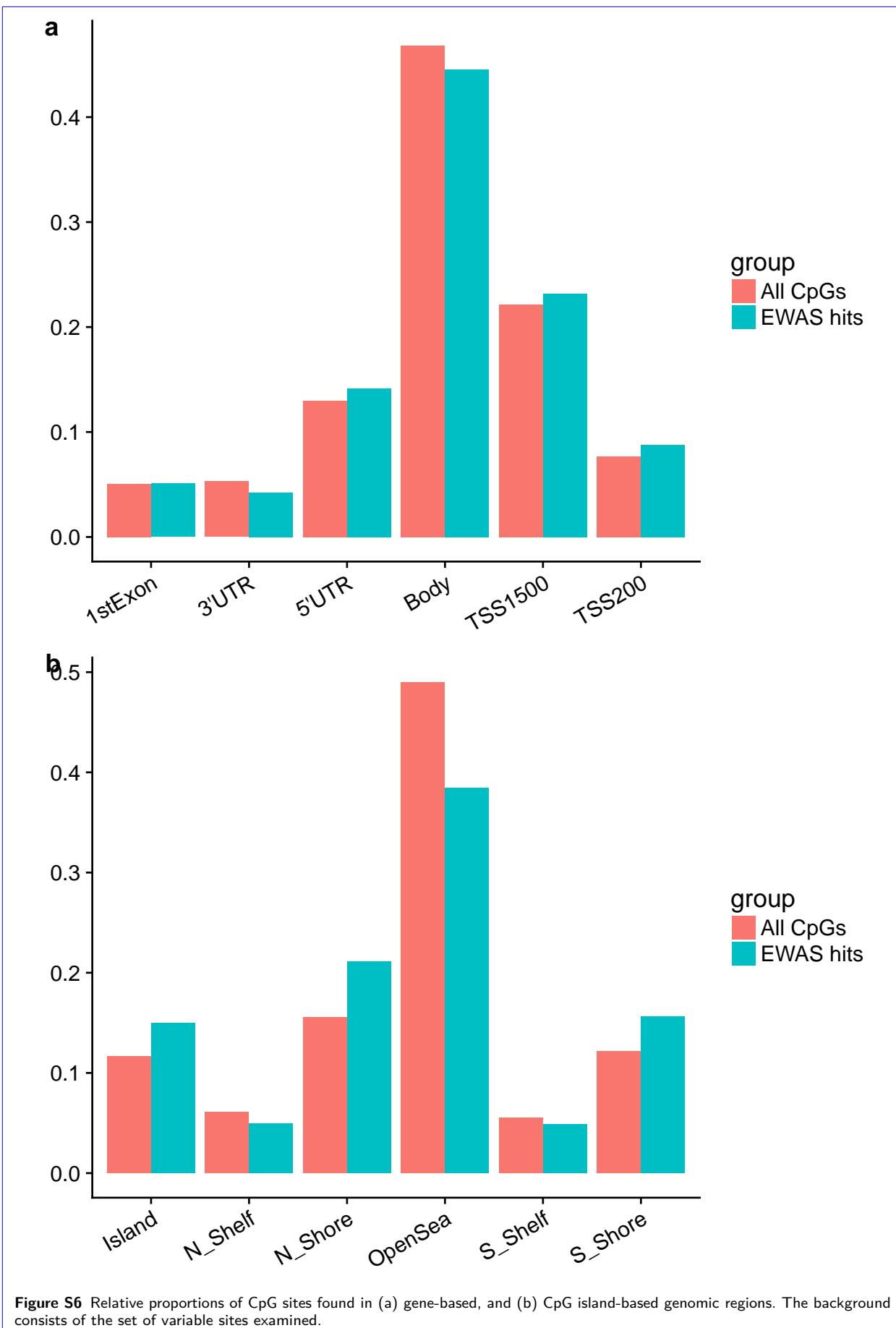


Table S3: Top DMRs from FHS only.

chr	start	end	num_cpgs	p_region_slk	p_adj_region_sidak	p_adj_region	annot_gene
chr19	9473057	9474128	13	0e+00	0.0000001	1.18e-05	ZNF177
chr19	47287777	47288263	6	0e+00	0.0000003	2.59e-05	SLC1A5
chr12	9217078	9217907	11	0e+00	0.0000005	3.46e-05	LOC144571
chr8	41583017	41583523	5	0e+00	0.0000005	5.20e-05	ANK1
chr7	30635213	30636176	6	0e+00	0.0000005	7.37e-05	GARS
chr6	31631637	31632740	16	0e+00	0.0000041	2.81e-04	BAT4
chr19	45899411	45899615	2	0e+00	0.0000032	3.24e-04	PPP1R13L
chr19	37825306	37826008	8	0e+00	0.0000091	7.98e-04	HKR1
chr19	2294886	2295456	4	0e+00	0.0000071	1.01e-03	LINGO3
chr13	20989141	20990243	6	0e+00	0.0000080	1.47e-03	CRYL1
chr1	3339869	3340625	5	0e+00	0.0000181	2.74e-03	PRDM16
chr6	117869026	117869857	4	0e+00	0.0000189	3.92e-03	DCBLD1
chr11	65546987	65547172	4	1e-07	0.0001558	7.21e-03	DKFZp761E198
chr16	75150455	75150880	8	2e-07	0.0001754	9.32e-03	LDHD
chr20	61455861	61456784	4	2e-07	0.0000697	1.61e-02	COL9A3
chr12	113514932	113515969	5	2e-07	0.0000934	1.94e-02	DTX1
chr16	85074867	85075249	5	3e-07	0.0002897	2.21e-02	KIAA0513
chr1	202129191	202131184	11	6e-07	0.0001232	2.23e-02	PTPN7
chr1	247802702	247803166	7	4e-07	0.0003639	2.41e-02	
chr12	10183166	10183364	6	5e-07	0.0011740	3.88e-02	CLEC9A

Table S4: GO enrichment of CVD-related modules from WGCNA.

module	Term	Ont	N	DE	P.DE	FDR
mediumpurple3	CDP-diacylglycerol biosynthetic process	BP	11	1	0.0020567	1.0000000
mediumpurple3	1-acylglycerol-3-phosphate O-acyltransferase activity	MF	16	1	0.0020567	1.0000000
mediumpurple3	lysophosphatidic acid acyltransferase activity	MF	16	1	0.0020567	1.0000000
mediumpurple3	lysophospholipid acyltransferase activity	MF	16	1	0.0020567	1.0000000
mediumpurple3	acylglycerol O-acyltransferase activity	MF	22	1	0.0020567	1.0000000
pink	anatomical structure morphogenesis	BP	2194	335	0.0000039	0.0803439
pink	patterning of blood vessels	BP	26	13	0.0000112	0.1157043
pink	tissue morphogenesis	BP	524	96	0.0000242	0.1662388
pink	animal organ morphogenesis	BP	835	139	0.0000544	0.2777397
pink	morphogenesis of an epithelium	BP	435	80	0.0000876	0.2777397
blue	nervous system development	BP	1791	568	0.0000000	0.0000000
blue	multicellular organism development	BP	4096	1078	0.0000000	0.0000000
blue	system development	BP	3614	973	0.0000000	0.0000000
blue	anatomical structure development	BP	4608	1176	0.0000000	0.0000000

blue	anatomical structure morphogenesis	BP	2194	647	0.0000000	0.0000000
lightyellow	sterol binding	MF	43	9	0.0000452	0.6874038
lightyellow	cholesterol binding	MF	38	8	0.0000996	0.6874038
lightyellow	steroid binding	MF	81	12	0.0001001	0.6874038
lightyellow	oxygen metabolic process	BP	4	3	0.0001551	0.7993077
lightyellow	lipid binding	MF	568	40	0.0003336	1.0000000

6 References

- Baccarelli, A, R Wright, V Bollati, A Litonjua, A Zanobetti, L Tarantini, D Sparrow, P Vokonas, and J Schwartz. 2010. "Ischemic heart disease and stroke in relation to blood DNA methylation." *Epidemiology* 21 (6): 819–28. doi:[10.1097/EDE.0b013e3181f20457](https://doi.org/10.1097/EDE.0b013e3181f20457).
- Bacos, Karl, Linn Gillberg, Petr Volkov, Anders H Olsson, Torben Hansen, Oluf Pedersen, Anette Prior Gjesing, et al. 2016. "Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes." *Nature Communications* 7. Nature Publishing Group: 11089. doi:[10.1038/ncomms11089](https://doi.org/10.1038/ncomms11089).
- Bauer, Mario. 2018. "Cell-type-specific disturbance of DNA methylation pattern: a chance to get more benefit from and to minimize cohorts for epigenome-wide association studies." *International Journal of Epidemiology*, March. doi:[10.1093/ije/dyy029](https://doi.org/10.1093/ije/dyy029).
- Bäck, Magnus, and Göran K. Hansson. 2015. "Anti-inflammatory therapies for atherosclerosis." *Nature Reviews Cardiology*. doi:[10.1038/nrcardio.2015.5](https://doi.org/10.1038/nrcardio.2015.5).
- Bonder, Marc Jan, René Luijk, Daria V Zhernakova, Matthijs Moed, Patrick Deelen, Martijn Vermaat, Maarten van Iterson, et al. 2016. "Disease variants alter transcription factor levels and methylation of their binding sites." *Nature Genetics* 49 (1): 131–38. doi:[10.1038/ng.3721](https://doi.org/10.1038/ng.3721).
- Bowden, Jack, George Davey Smith, and Stephen Burgess. 2015. "Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression." *International Journal of Epidemiology* 44 (2). Oxford University Press: 512–25. doi:[10.1093/ije/dyv080](https://doi.org/10.1093/ije/dyv080).
- Chen, Yuting, Martin Widschwendter, and Andrew E. Teschendorff. 2017. "Systems-epigenomics inference of transcription factor activity implicates aryl-hydrocarbon-receptor inactivation as a key event in lung cancer development." *Genome Biology* 18 (1). doi:[10.1186/s13059-017-1366-0](https://doi.org/10.1186/s13059-017-1366-0).
- Dekkers, Koen F, Maarten van Iterson, Roderick C Slieker, Matthijs H Moed, Marc Jan Bonder, Michiel van Galen, Hailiang Mei, et al. 2016. "Blood lipids influence DNA methylation in circulating cells." *Genome Biology* 17 (1): 138. doi:[10.1186/s13059-016-1000-6](https://doi.org/10.1186/s13059-016-1000-6).
- El Taghdouini, Adil, Anita L Sorensen, Andrew H Reiner, Mar Coll, Stefaan Verhulst, Inge Mannaerts, Cristina I Oie, et al. 2015. "Genome-wide analysis of DNA methylation and gene expression patterns in purified, uncultured human liver cells and activated hepatic stellate cells." *Oncotarget* 6 (29): 26729–45. doi:[10.18632/oncotarget.4925](https://doi.org/10.18632/oncotarget.4925).
- Fox, Ervin R., Tandaw E. Samdarshi, Solomon K. Musani, Michael J. Pencina, Jung Hye Sung, Alain G. Bertoni, Vanessa Xanthakis, et al. 2016. "Development and Validation of Risk Prediction Models for Cardiovascular Events in Black Adults." *JAMA Cardiology* 1 (1): 15. doi:[10.1001/jamacardio.2015.0300](https://doi.org/10.1001/jamacardio.2015.0300).

- Gaunt, Tom R, Hashem A Shihab, Gibran Hemani, Josine L Min, Geoff Woodward, Oliver Lyttleton, Jie Zheng, et al. 2016. "Systematic identification of genetic influences on methylation across the human life course." *Genome Biology* 17 (1): 61. doi:[10.1186/s13059-016-0926-z](https://doi.org/10.1186/s13059-016-0926-z).
- Goff, David C., Donald M. Lloyd-Jones, Glen Bennett, Sean Coady, Ralph B. D'Agostino, Raymond Gibbons, Philip Greenland, et al. 2014. "2013 ACC/AHA guideline on the assessment of cardiovascular risk: A report of the American college of cardiology/American heart association task force on practice guidelines." *Journal of the American College of Cardiology*. doi:[10.1016/j.jacc.2013.11.005](https://doi.org/10.1016/j.jacc.2013.11.005).
- Guarrera, Simonetta, Giovanni Fiorito, N Charlotte Onland-Moret, Alessia Russo, Claudia Agnoli, Alessandra Allione, Cornelia Di Gaetano, et al. 2015. "Gene-specific DNA methylation profiles and LINE-1 hypomethylation are associated with myocardial infarction risk." *Clinical Epigenetics* 7 (1): 133. doi:[10.1186/s13148-015-0164-3](https://doi.org/10.1186/s13148-015-0164-3).
- Hachiya, Tsuyoshi, Ryohei Furukawa, Yuh Shiwa, Hideki Ohmomo, Kanako Ono, Fumiki Katsuoka, Masao Nagasaki, et al. 2017. "Genome-wide identification of inter-individually variable DNA methylation sites improves the efficacy of epigenetic association studies." *Npj Genomic Medicine* 2 (1): 11. doi:[10.1038/s41525-017-0016-5](https://doi.org/10.1038/s41525-017-0016-5).
- Hedman, Åsa K., Michael M. Mendelson, Riccardo E. Marioni, Stefan Gustafsson, Roby Joehanes, Marguerite R. Irvin, Degui Zhi, et al. 2017. "Epigenetic Patterns in Blood Associated With Lipid Traits Predict Incident Coronary Heart Disease Events and Are Enriched for Results From Genome-Wide Association Studies." *Circulation: Cardiovascular Genetics* 10 (1): e001487. doi:[10.1161/CIRCGENETICS.116.001487](https://doi.org/10.1161/CIRCGENETICS.116.001487).
- Houseman, Eugene Andres, William P Accomando, Devin C Koestler, Brock C Christensen, Carmen J Marsit, Heather H Nelson, John K Wiencke, and Karl T Kelsey. 2012. "DNA methylation arrays as surrogate measures of cell mixture distribution." *BMC Bioinformatics* 13 (1): 86. doi:[10.1186/1471-2105-13-86](https://doi.org/10.1186/1471-2105-13-86).
- Huang, Yen Tsung, Su Chu, Eric B Loucks, Chien Ling Lin, Charles B Eaton, Stephen L Buka, and Karl T Kelsey. 2016. "Epigenome-wide profiling of DNA methylation in paired samples of adipose tissue and blood." *Epigenetics* 11 (3): 227–36. doi:[10.1080/15592294.2016.1146853](https://doi.org/10.1080/15592294.2016.1146853).
- Irvin, Marguerite R., Degui Zhi, Roby Joehanes, Michael Mendelson, Stella Aslibekyan, Steven A. Claas, Krista S. Thibault, et al. 2014. "Epigenome-wide association study of fasting blood lipids in the genetics of lipid-lowering drugs and diet network study." *Circulation* 130 (7): 565–72. doi:[10.1161/CIRCULATIONAHA.114.009158](https://doi.org/10.1161/CIRCULATIONAHA.114.009158).
- Jiao, Yinming, Martin Widswendter, and Andrew E. Teschendorff. 2014. "A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control." *Bioinformatics* 30 (16): 2360–66. doi:[10.1093/bioinformatics/btu316](https://doi.org/10.1093/bioinformatics/btu316).
- Lehne, Benjamin, Alexander W. Drong, Marie Loh, Weihua Zhang, William R. Scott, Sian Tsung Tan, Uzma Afzal, et al. 2015. "A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies." *Genome Biology* 16 (1). doi:[10.1186/s13059-015-0600-x](https://doi.org/10.1186/s13059-015-0600-x).
- Lettre, Guillaume, Cameron D. Palmer, Taylor Young, Kenechi G. Ejebu, Hooman Allayee, Emelia J. Benjamin, Franklyn Bennett, et al. 2011. "Genome-Wide association study of coronary heart disease and its risk factors in 8,090 african americans: The nhlbi CARE project." *PLoS Genetics* 7 (2). doi:[10.1371/journal.pgen.1001300](https://doi.org/10.1371/journal.pgen.1001300).
- Li, Jun, Xiaoyan Zhu, Kuai Yu, Haijing Jiang, Yizhi Zhang, Siyun Deng, Longxian Cheng, et al. 2017. "Genome-Wide Analysis of DNA Methylation and Acute Coronary Syndrome." *Circulation Research* 120 (11): 1754–67. doi:[10.1161/CIRCRESAHA.116.310324](https://doi.org/10.1161/CIRCRESAHA.116.310324).

- Libby, Peter. 2006. "Inflammation and cardiovascular disease mechanisms." *The American Journal of Clinical Nutrition* 83 (2): 456S–S.
- Liu, Changqing, Jing Yang, Wei Han, Qi Zhang, Xiaoming Shang, Xia Li, Feng Lu, and Xiaokun Liu. 2015. "Polymorphisms in ApoB gene are associated with risk of myocardial infarction and serum ApoB levels in a Chinese population." *International Journal of Clinical and Experimental Medicine* 8 (9): 16571–77.
- Murtha, Lucy A., Michael J. Schuliga, Nishani S. Mabotuwana, Sean A. Hardy, David W. Waters, Janette K. Burgess, Darryl A. Knight, and Andrew J. Boyle. 2017. "The processes and mechanisms of cardiac and pulmonary fibrosis." *Frontiers in Physiology*. doi:[10.3389/fphys.2017.00777](https://doi.org/10.3389/fphys.2017.00777).
- Nakatouchi, Masahiro, Sahoko Ichihara, Ken Yamamoto, Keiko Naruse, Shigeki Yokota, Hiroyuki Asano, Tatsuaki Matsubara, and Mitsuhiro Yokota. 2017. "Epigenome-wide association of myocardial infarction with DNA methylation sites at loci related to cardiovascular disease." *Clinical Epigenetics* 9 (1). doi:[10.1186/s13148-017-0353-3](https://doi.org/10.1186/s13148-017-0353-3).
- Nikpay, Majid, Anuj Goel, Hong Hee Won, Leanne M. Hall, Christina Willenborg, Stavroula Kanoni, Danish Saleheen, et al. 2015. "A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease." *Nature Genetics* 47 (10): 1121–30. doi:[10.1038/ng.3396](https://doi.org/10.1038/ng.3396).
- Ordovás, José M., and Caren E. Smith. 2010. "Epigenetics and cardiovascular disease." *Nature Reviews Cardiology* 7 (9): 510–19. doi:[10.1038/nrccardio.2010.104](https://doi.org/10.1038/nrccardio.2010.104).
- Pfeiffer, Liliane, Simone Wahl, Luke C. Pilling, Eva Reischl, Johanna K. Sandling, Sonja Kunze, Lesca M. Holdt, et al. 2015. "DNA Methylation of Lipid-Related Genes Affects Blood Lipid Levels." *Circulation: Cardiovascular Genetics* 8 (2): 334–42. doi:[10.1161/CIRCGENETICS.114.000804](https://doi.org/10.1161/CIRCGENETICS.114.000804).
- Rask-Andersen, Mathias, David Martinsson, Muhammad Ahsan, Stefan Enroth, Veronica E Ek, Ulf Gylensten, and Åsa Johansson. 2016. "Epigenome-wide association study reveals differential DNA methylation in individuals with a history of myocardial infarction." *Human Molecular Genetics* 25 (21): 4739–48. doi:[10.1093/hmg/ddw302](https://doi.org/10.1093/hmg/ddw302).
- Richardson, Tom G, Jie Zheng, George Davey Smith, Nicholas J Timpson, Tom R Gaunt, Caroline L Relton, and Gibran Hemani. 2017. "Mendelian Randomization Analysis Identifies CpG Sites as Putative Mediators for Genetic Influences on Cardiovascular Disease Risk." *American Journal of Human Genetics* 101 (4): 590–602. doi:[10.1016/j.ajhg.2017.09.003](https://doi.org/10.1016/j.ajhg.2017.09.003).
- Rosmond, R., and P. Björntorp. 2000. "The hypothalamic-pituitary-adrenal axis activity as a predictor of cardiovascular disease, type 2 diabetes and stroke." *Journal of Internal Medicine* 247 (2): 188–97. doi:[10.1046/j.1365-2796.2000.00603.x](https://doi.org/10.1046/j.1365-2796.2000.00603.x).
- Wahl, Simone, Alexander Drong, Benjamin Lehne, Marie Loh, William R. Scott, Sonja Kunze, Pei-Chien Tsai, et al. 2017. "Epigenome-wide association study of body mass index and the adverse outcomes of adiposity." *Nature* 541 (7635): 81–86. doi:[10.1038/nature20784](https://doi.org/10.1038/nature20784).
- Yang, Dongchan, Insu Jang, Jinhyuk Choi, Min-Seo Kim, Andrew J. Lee, Hyunwoong Kim, Junghyun Eom, Dongsup Kim, Inkyung Jung, and Byungwook Lee. 2017. "3DIV: A 3D-genome Interaction Viewer and database." *Nucleic Acids Research*. doi:[10.1093/nar/gkx1017](https://doi.org/10.1093/nar/gkx1017).
- Zambrana, Ruth E., Lenny López, Gnesha Y. Dinwiddie, Roberta M. Ray, Lawrence S. Phillips, Maurizio Trevisan, and Sylvia Wassertheil-Smoller. 2014. "Prevalence and incident prehypertension and hypertension in postmenopausal hispanic women: Results from the women's health initiative." *American Journal of Hypertension* 27 (3): 372–81. doi:[10.1093/ajh/hpt279](https://doi.org/10.1093/ajh/hpt279).