# EWAS for Incident CVD Events

## 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 [@Bonder2016]. DNA methylation is one such mechanism that can reflect both genetic variation and environmental exposures and potentially drive their effects on CVD outcomes [@Ordovas2010].

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-Andersen2016], acute coronary syndrome [@Li2017], and atherosclerosis [@Nakatochi2017]. 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 [@Dekkers2016;@Wahl2017].

One approach to this problem is to examine epigenetic associations with cardiovascular risk factors. Multiple investigations have explored these relationships genome-wide [@Pfeiffer2015;@Irvin2014], and have even uncovered prognostic CpG sites for incident coronary heart disease in the process [@Hedman2017]. 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 [@Baccarelli2010;@Guarrera2015].

Most epigenetic investigations of CVD to-date have been performed in blood, based on the importance of immune-related processes in its pathogenesis [@Back2015] 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 [@Bacos2016;@Huang2016].

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 sought context for our results using correlation-based modules, and further explored the biological significance of a specific inflammation-related module. Additionally, we discovered informative sets of CVD-related CpG sites using differentially methylated region detection.

## Results

### Weighted correlation network approach finds CVD-related modules

Population is described below.

We set out to find biologically relevant modules (agnostic to CVD) using the WGCNA approach for 422952 CpGs in WHI.

110 modules were uncovered, spanning sizes from 28 to 35361 CpGs (with exclusion of the "grey" set capturing non-assigned CpGs).

Principal component eigenvectors for each module ("eigenCpGs") were calculated in order to examine the characteristics of these modules as a whole. The first principal components of each module tended to explain a large fraction of the variance, suggesting that they are reasonable proxies for the behavior of the modules as

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: Comb-p regions with multiple test-corrected p<0.05 in WHI and Bonferroni p<0.05 in FHS

| Location | # CpGs | Annotated gene | Genomic region | P.whi | Padj.whi | P.fhs |
|---|---|---|---|---|---|---|
| chr1:27440462-27440721 | 3 | SLC9A1 | Body | 1.75e-08 | 2.85e-05 | 0.000103 |
| chr19:47287777-47288263 | 7 | SLC1A5 | CpG shelf near TSS | 5.91e-07 | 0.000514 | 3.8e-10 |
| chr17:76037034-76037562 | 6 | TNRC6C | CpG island in 5' UTR | 1.67e-05 | 0.013300 | 0.000189 |

a whole. Cox proportional hazards models were used to assess relationships between these module eigenCpGs and incident CVD. In partially-adjusted models (adjusted for technical factors and estimated white blood cell proportions), three modules were found to be associated at BH-corrected FDR<0.2. Adjustment for biological covariates (age, BMI, sex/race, and smoking behavior) attenuated these relationships to nominal statistical significance ($0.01 < p < 0.1$). Though the existence of past CVD events (experienced prior to sample collection for DNA methylation measurement) could represent a confounder in the FHS dataset, models adjusted for past events only marginally reduced the strength of these module-trait relationships.

| module | modSize | varExpl | p | GO.terms | islandLocEnrich | geneLocEnrich |
|---|---|---|---|---|---|---|
| blue | 29441 | 0.45 | 0.000274 | Development | N_shore (moderate) | 1stExon/TSS/5' UTR |
| brown4 | 953 | 0.53 | 0.004546 | Immune activation | Open sea | |
| lavenderblush3 | 568 | 0.45 | 0.005003 | T cell activation | Open sea | Body |

These modules, all collected from blood, showed enrichment for different sets of GO terms, ranging from immune activation (myeloid or T cell) to developmental processes.

Of these modules, two (blue and brown4) replicated strongly in FHS, while lavenderblush3 showed nominal replication (p=0.0203) in partially-adjusted models.

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

To investigate more specific epigenetic signals, we performed an EWAS for incident CVD, focusing on differentially methylated regions in order to improve statistical power. Single-site EWAS p-values were used as input to the Comb-p algorithm, which seeks regions enriched for low p-values while accounting for autocorrelation based on genomic distance. It was used separately on WHI and FHS. '

While 206 DMRs were found in WHI (using the Sidak correction for multiple tests, separately for each DMR based on its length), only 3 were both found in FHS and replicated at a Bonferroni level. Of the relevant WGCNA modules identified above, brown4 CpG sites consituted part of 2 DMRs (at SLC9A1 & SLC1A5), while a single CpG from the blue module was also a member of the SLC9A1 DMR.

Of single sites from the EWAS, 3 reached a genome-wide Bonferroni threshold (see supplementary Table ___), but none replicated strongly in FHS.

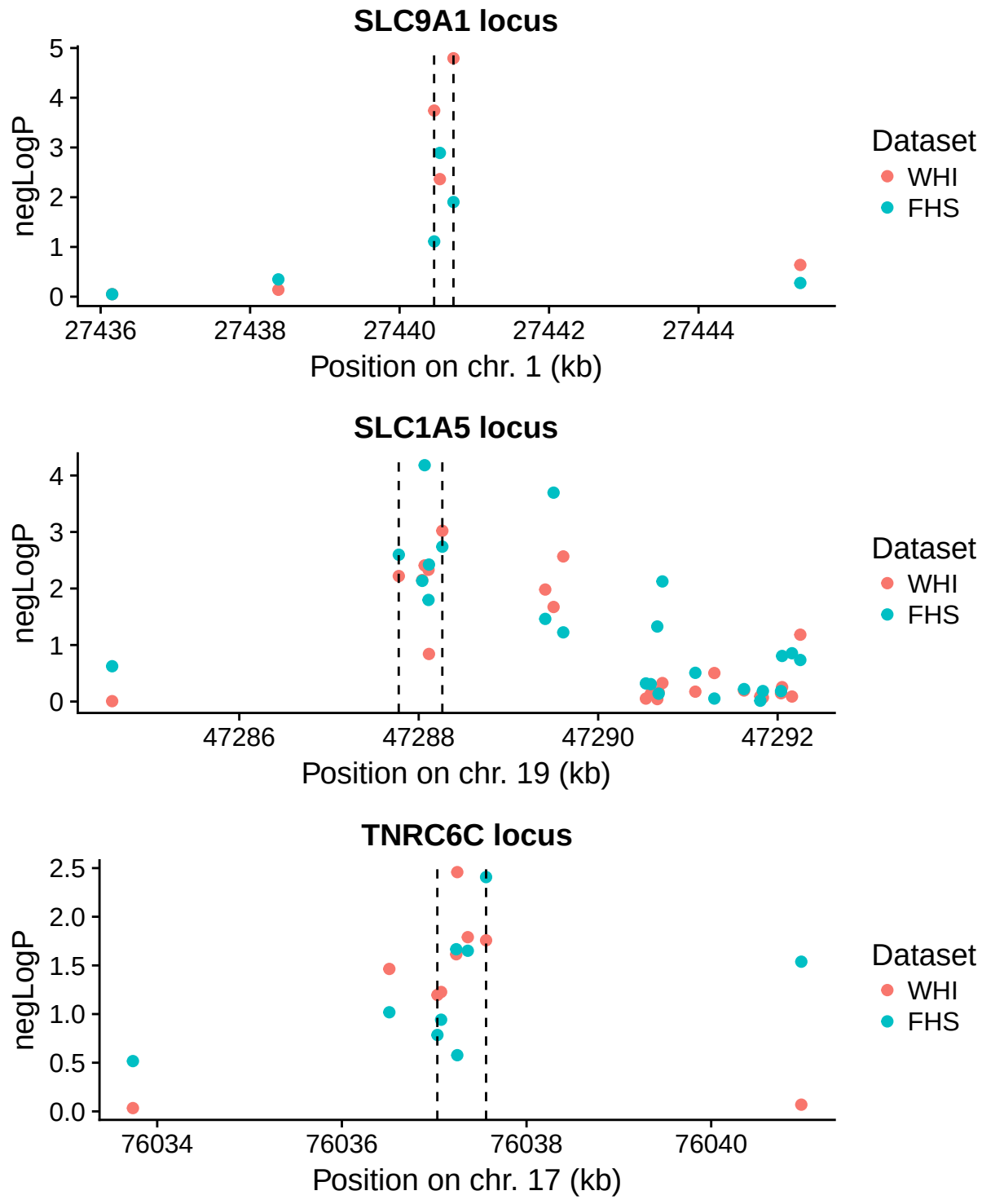


Figure 1: DMRs identified by Comb-p in WHI and validated in FHS at the (a) SLC9A1 and (b) SLC1A5 loci. Negative logarithms of EWAS p-values are shown as a function of the genomic coordinate. EWAS p-values from WHI are in red and FHS in green. Dotted lines demarcate the DMR boundaries.

## Exploration of the brown4 and blue modules

Based on the results from the module- and region-centric analyses, we investigated the brown4 and blue modules further for biological significance.

The brown4 module is associated with immune-related genes as noted above, and is enriched strongly for "open sea" sites (p=5.3e-43) and annotated enhancers (p=8.6e-34). In contrast, the blue module is associated with development-related genes, and is enriched moderately for sites near genic transcription start sites and strongly for CpG islands ($p < 2.2e\text{-}16$).

Given these observations, we examined relative enrichments of enhancer- and promoter-associated histone marks across different blood cell subtypes to better understand any cell type-specificity of this signal. Epigenetic peaks were annotated using data from Roadmap Epigenomics and relative enrichments were calculated as the fraction of module CpGs found in peaks divided by the fraction of all CpGs found in peaks. The eFORGE tool is designed to run a similar cell type specificity analysis, but is currently not able to process CpG sets the size of the blue module.

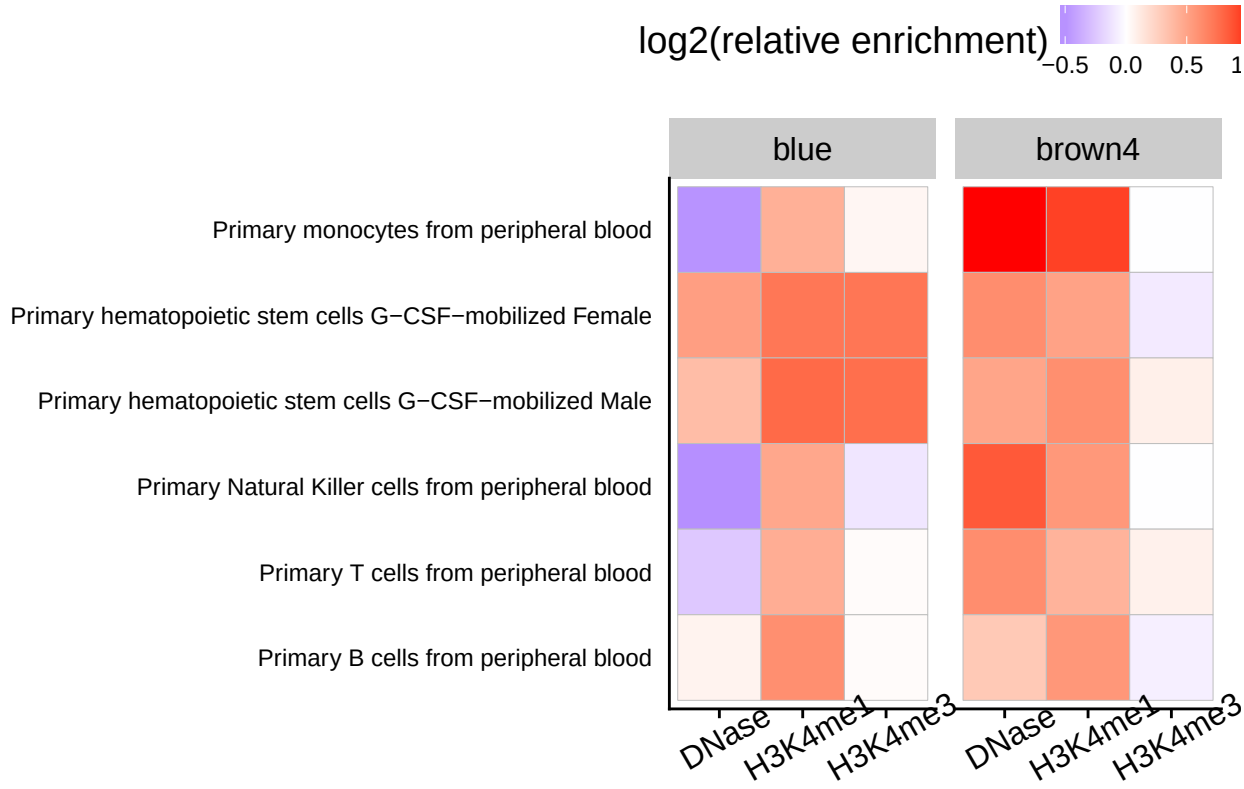


Figure 2: Cell type-specific enrichments based on Roadmap Epigenomics datasets. Shown are relative enrichments of peaks for a given epigenetic mark across many blood cell types, for each of the modules of interest.

We observed the greatest enrichment of brown4 CpGs in enhancer-associated histone peaks from monocytes compared to other blood cell subtypes. This could point towards monocyte-related biology and inflammatory processes as an important shared mechanism for cardiovascular risk between the two cohorts examined here.

### Module-risk factor relationships

```
## Warning in bind_rows_(x, .id): binding factor and character vector,
## coercing into character vector
```

```
## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector
```

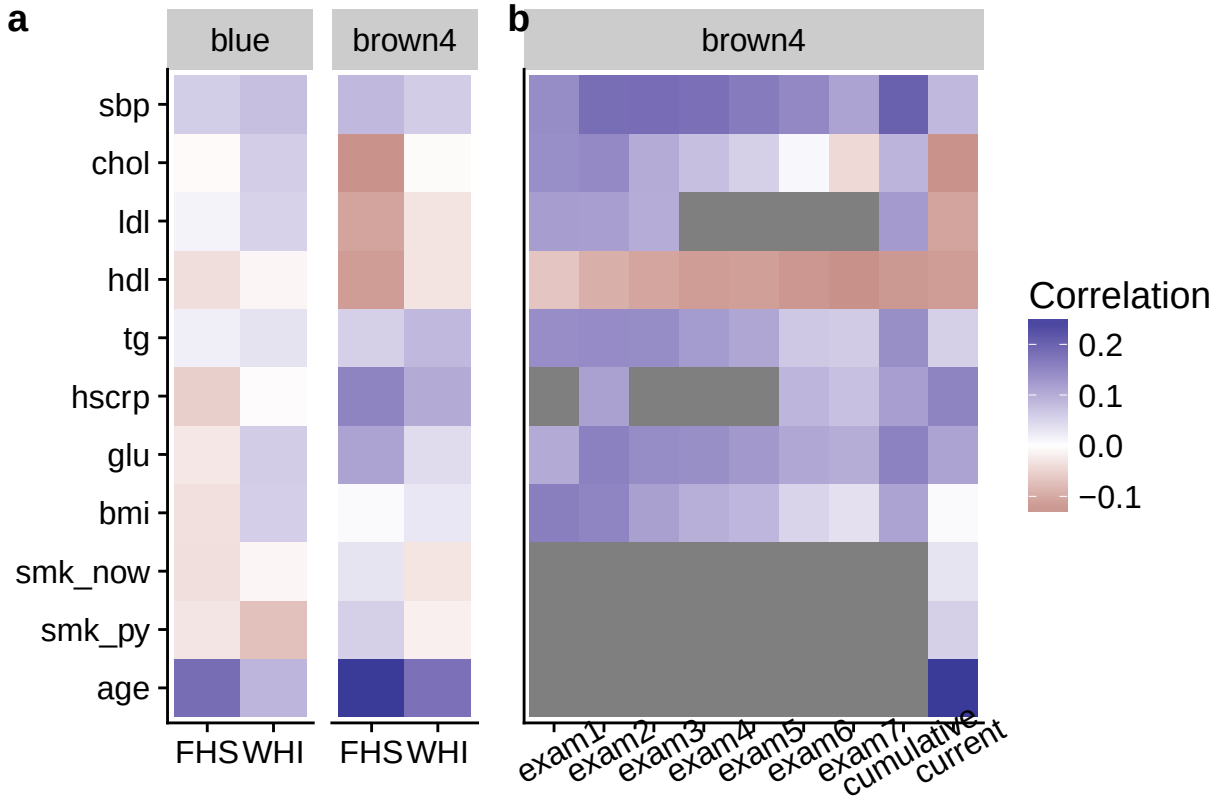


Figure 3: Risk factor-module relationships. (a) Correlations between a series of traditional cardiovascular risk factors and module eigenCpGs (blue and brown4) are shown in each study population. (b) Correlations between historical risk factor levels in FHS (across previous exams, x-axis) and current brown4 module activation are shown.

Next, we examined the association between these module eigenCpGs and traditional cardiovascular risk factors. Though no extremely strong module-risk factor correlations were observed (all $|r|<0.25$), these correlations tended to be stronger for the brown4 module, especially in FHS. Age showed the greatest association, while lipid and glycemic parameters also showed moderate associations. To further probe relationships between brown4 module and risk factors in FHS, we retrieved historical risk factors measured in previous Offspring Cohort exams. Based on visual inspection, there seemed to be a notably stronger correlation between the module eigenCpG and cumulative (mean of all previous exams) compared to current risk factor exposure: this applies for systolic blood pressure (strongly), triglycerides, glucose, BMI, and LDL (which correlated in the "expected" direction cumulatively, but non-intuitively at Exam 8). Interestingly, this pattern did not hold for hsCRP, the most direct marker of inflammation in this group of risk factors.

To better investigate this phenomenon, we tested associations between the brown4 module and each of the

Table 3: Module-risk factor relationships (current and cumulative) after adjustment for covariates

| | brown4 | | blue | |
|---|---|---|---|---|
| Risk factor | p_current | p_cumulative | p_current | p_cumulative |
| bmi | 4.3e-06 | 4.5e-06 | 0.00611 | 0.4915 |
| glu | 0.0112 | 0.047 | 0.94753 | 0.5926 |
| hscrp | 0.0077 | 3.5e-05 | 0.01126 | 0.9345 |
| tg | 0.0205 | 3.7e-06 | 0.02967 | 0.7669 |
| hdl | 0.0152 | 0.945 | 0.00031 | 0.8536 |
| ldl | 0.1280 | 0.183 | 0.94240 | 0.0075 |
| chol | 0.1075 | 0.003 | 0.20461 | 0.0861 |
| sbp | 0.1455 | 0.026 | 0.27159 | 0.7028 |

cumulative risk factors after adjustment for confounding. Specifically, for each risk factor, linear models were used to predict the brown4 eigenCpG value from the cumulative risk factor level while adjusting for the full set of EWAS covariates other than BMI (age/sex/smoking/cell counts/study center/7 ctrl-probe PCs) as well as the current value of the risk factor.

Finally, to assess relationships between cumulative risk factor exposure and brown4 module activation in predicting cardiovascular risk, we created a series of Cox models in FHS in the form of a mediation analysis. Covariates in all models included technical factors, estimated cell counts, age, and sex. Table 5 presents p-values arising from models incorporating 1) current and cumulative risk factors only, 2) module eigenCpGs only, and 3) all 3 quantities together.

| | Risk factors only | | Brown4 only | Full model | |
|---|---|---|---|---|---|
| Risk factor | curr_rf | cum_rf | brown4 | cum_rf.1 | brown4.1 |
| bmi | 0.63612 | 0.00797 | 0.0688 | 0.0143 | 0.1194 |
| hscrp | 0.146595 | 0.000131 | 0.0432 | 0.000272 | 0.099522 |
| tg | 0.312 | 4.25e-07 | 0.0287 | 1.29e-06 | 0.0822 |

# Discussion

- Basic summary of intentions and rationale for EWAS.
- When performing high-dimensional EWAS, it is difficult to separate important signal from noise, as well as to ascribe biological meaning to single-CpG results. We addressed this issue by using WGCNA to create coherent modules that can provide context to the findings.
  - Modules seemed to generally be strongly enriched for one of two types of processes: immune-related (relevant for cell type and for CVD) and development-related (Barker hypothesis?).
- We assessed replication in FHS and found very little. However, the brown4 module showed uniquely successful (nominal) replication, with almost 50% of 25 potential discovery hits from this module showing nominal replication.
- We assessed brown4 module further to understand what it may represent.
  - Enriched for "open sea" and enhancers as annotated by Illumina -> may be playing regulatory roles.
  - CpGs are near genes related to immune processes.
  - Taking the enhancer prevalence into account, enrichment of the brown4 CpG set loci with enhancer-associated histone marks (H3K4me1 & H3K27ac) was assessed in different blood cell subtypes -> notably stronger enrichment for H3K4me1 in monocytes than in any other cell type.
- Brown4 correlation with risk factors – very significant but modest correlation with age and SBP in

Table S1: CpGs with FDR<0.05 in the discovery set (Bonferroni threshold = 1.18e-7)

| CpG | Chromosome | Direction of Association | P-value | Location | Annotated Gene |
|---|---|---|---|---|---|
| cg09155044 | chr16 | 1 | 6.63e-09 | TSS1500;TSS1500 | VKORC1;VKORC1 |
| cg24434800 | chr1 | 1 | 5.04e-08 | | |
| cg11691298 | chr2 | 1 | 1.1e-07 | Body | FAM59B |
| cg02379107 | chr20 | 1 | 4.72e-07 | TSS1500 | KIAA1755 |

WHI. However, we were able to employ a unique strategy in FHS by comparing current methylation values with past risk factor values in order to assess methylation as a "molecular recorder" of risk factor exposure. We used basic correlations (cumulative vs. current) and fully-adjusted models to show that three risk factors (BMI, triglycerides, and hsCRP) showed especially strong cumulative correlations with brown4 "activation" even after adjustment for current levels of these factors (p ~ 10e-6). This presents an intriguing basis for further exploration of methylation as a marker or even mediator (in this immune/inflammation-related case) of cumulative risk factor exposure.

- DMR finding using comb-p on EWAS p-values discovered 3 regions that replicated after Bonferroni adjustment in FHS – 2 are SLC family transporters, and one is a gene involved in miRNA-mediated gene silencing.
  - SLC9A1: involved in cation exchange; also known as NHE-1; CpG site showed up in EWAS for BMI; involved in the (leptin/adrenaline/endothelin-1/glucose)-stimulated increase in monocyte adhesion, migration, CD36 expr., and phagocytosis of oxLDL (DOI: 10.1517/13543784.2010.532123)
  - SLC1A5: neutral amino acid transporter; may be involved in T-cell biology?, look into the regulation by infection and can that relate to CVD microbial burden, companion AA transporter SLC7A5 has been shown to regulate metabolic and inflammatory reprogramming of monocytes in response to LPS
  - TNRC6C: part of the miRNA-mediated silencing cascade; showed up in a GWAS for heart failure; was identified as a potential target gene in monocyte-to-macrophage transition

# Methods

Notes:

- WGCNA was performed using unsigned correlations. For computational feasibility, the module creation was performed "blockwise", with the preliminary CpG clustering step perform in a subset of only 100 individiauls.
- Proportional hazards checks were implemented (cox.zph function in R) for the top EWAS in WHI, and no systematic departure from the Cox regression assumptions were detected.

## Supplementary

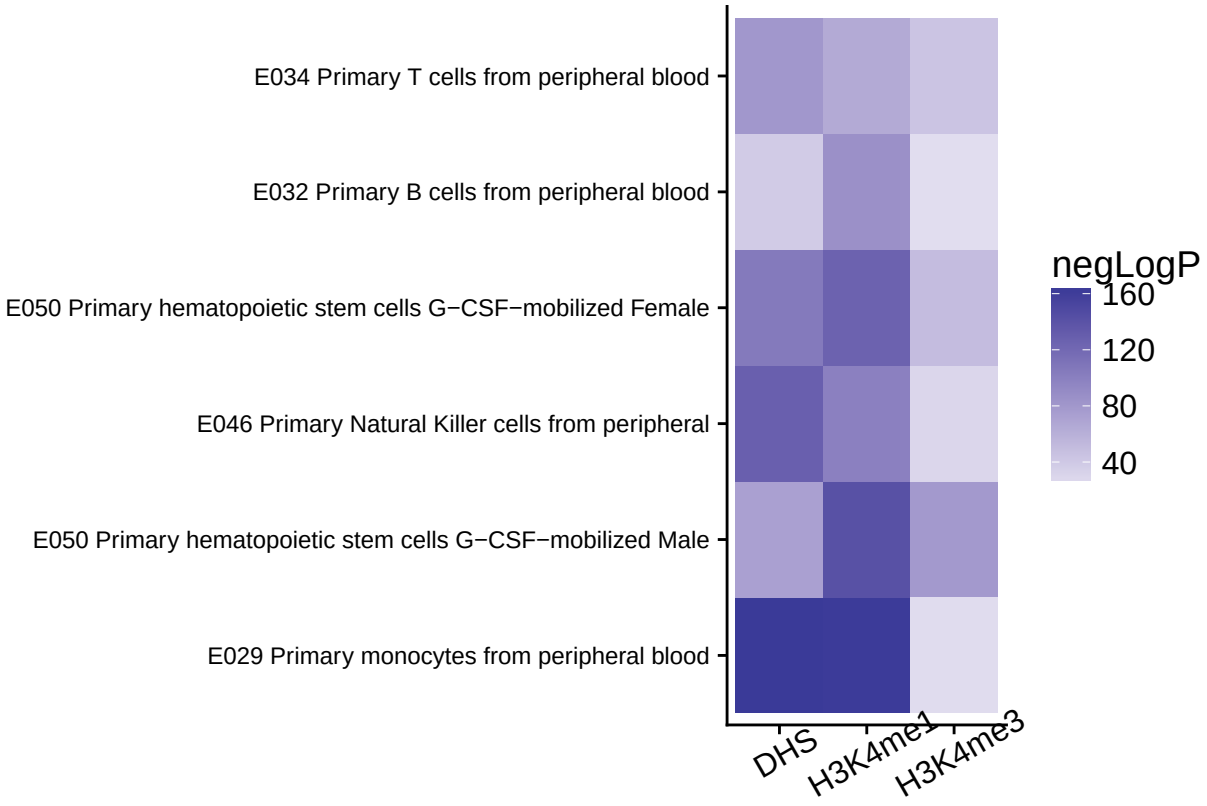


Figure S1: eFORGE cell type-specificity plot for the brown4 module.