


ORIGINAL RESEARCH

Epigenomic Assessment of Cardiovascular Disease Risk and Interactions With Traditional Risk Metrics

Kenneth Westerman, PhD; Alba Fernández-Sanlés, PhD; Prasad Patil, PhD; Paola Sebastiani, PhD; Paul Jacques, DSc; John M. Starr, MD, PhD; Ian J. Deary, MD, PhD; Qing Liu, PhD; Simin Liu, MD, ScD; Roberto Elosua, MD, PhD; Dawn L. DeMeo, MD; José M. Ordovás , PhD

BACKGROUND: Epigenome-wide association studies for cardiometabolic risk factors have discovered multiple loci associated with incident cardiovascular disease (CVD). However, few studies have sought to directly optimize a predictor of CVD risk. Furthermore, it is challenging to train multivariate models across multiple studies in the presence of study- or batch effects.

METHODS AND RESULTS: Here, we analyzed existing DNA methylation data collected using the Illumina HumanMethylation450 microarray to create a predictor of CVD risk across 3 cohorts: Women's Health Initiative, Framingham Heart Study Offspring Cohort, and Lothian Birth Cohorts. We trained Cox proportional hazards-based elastic net regressions for incident CVD separately in each cohort and used a recently introduced cross-study learning approach to integrate these individual scores into an ensemble predictor. The methylation-based risk score was associated with CVD time-to-event in a held-out fraction of the Framingham data set (hazard ratio per SD=1.28, 95% CI, 1.10–1.50) and predicted myocardial infarction status in the independent REGICOR (Girona Heart Registry) data set (odds ratio per SD=2.14, 95% CI, 1.58–2.89). These associations remained after adjustment for traditional cardiovascular risk factors and were similar to those from elastic net models trained on a directly merged data set. Additionally, we investigated interactions between the methylation-based risk score and both genetic and biochemical CVD risk, showing preliminary evidence of an enhanced performance in those with less traditional risk factor elevation.

CONCLUSIONS: This investigation provides proof-of-concept for a genome-wide, CVD-specific epigenomic risk score and suggests that DNA methylation data may enable the discovery of high-risk individuals who would be missed by alternative risk metrics.

Key Words: cardiovascular disease ■ DNA methylation ■ epigenomics ■ risk prediction

DNA methylation is an important epigenetic pathway through which genetic variants and environmental exposures impact disease risk.^{1,2} Methylation at specific cytosine-phosphate-guanine (CpG) sites has been associated with disease in epigenome-wide association studies, even showing associations in blood as a convenient but non-target tissue such as for type 2 diabetes mellitus.³ Methylation-based risk scores (MRS) allow genome-wide aggregation of epigenetic information, similarly to the more established

genetic risk scores, and allow for the use of models with arbitrary complexity. These risk scores are often developed initially by using methylation as a proxy for disease risk factors, such as body mass index⁴ and general aging-related morbidity.⁵ Alternatively, given sufficient sample size, epigenetic associations with disease risk can be modeled directly.⁶

Associations between DNA methylation and cardiovascular disease (CVD) have been explored in many different cohorts and using diverse approaches.

Correspondence to: José M. Ordovás, PhD, 711 Washington St., Boston, MA 02111. E-mail: jose.ordovas@tufts.edu

Supplementary Materials for this article are available at <https://www.ahajournals.org/doi/suppl/10.1161/JAHA.119.015299>

For Sources of Funding and Disclosures, see page 12.

© 2020 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- An epigenomic (DNA methylation-based) cardiovascular risk score was developed using a recently introduced statistical approach for combining risk models across cohorts.
- Interactions between an epigenomic risk score and existing genomic and clinical risk scores for cardiovascular disease were assessed.

What Are the Clinical Implications?

- DNA methylation may add a new molecular dimension to the prediction of cardiovascular risk.
- This epigenomic risk score may perform best in individuals with lower Framingham Risk Scores and thus identify high-risk individuals who would otherwise go undetected.

Non-standard Abbreviations and Acronyms

BMI	body mass index
BMIQ	Beta-Mixture Quantile Dilation normalization
CpG	cytosine-phosphate-guanine site
CSL	cross-study learner
CVD	cardiovascular disease
FHS	Framingham Heart Study Offspring Cohort
FHS-JHU	Framingham Heart Study (Johns Hopkins University subset)
FHS-UM	Framingham Heart Study (University of Minnesota subset)
FRS	Framingham Risk Score
GRS	genomic risk score
HDL	high-density lipoprotein
ICC	intraclass correlation coefficient
LBC	Lothian Birth Cohorts 1936
LDL	low-density lipoprotein
MI	myocardial infarction
MRS	methylation-based risk score
REGICOR	Registre Gironí del COR
SNP	single-nucleotide polymorphism
SSL	single-study learner
WHI	Women's Health Initiative

Cross-sectional associations have been found across multiple relevant tissues, namely blood, aorta, and other vascular tissues.⁷ Some investigations aimed at cardiovascular risk factors have discovered CpGs predictive of CVD development,^{8,9} while Mendelian randomization approaches have suggested causality of at least some of these CpG-risk factor associations.¹⁰ A few studies directly modeling incident CVD as a

primary outcome have either been conducted using only global (not locus-specific) methylation levels,¹¹ or have found limited additional predictive power in the presence of known risk factors.¹² A recent large-scale meta-analysis found multiple CpG sites predictive of incident coronary heart disease, but focused on univariate approaches.¹³ We have previously investigated methylation regions and modules associating with incident CVD, generating mechanistic insights but without aggregating these results into a direct predictor of risk.¹⁴ Additionally, it is unclear how the CVD risk tracked by DNA methylation is redundant with or complementary to existing risk metrics, including genetic scores¹⁵ and those based on traditional cardiovascular risk factors (eg, the Framingham Risk Score for generalized CVD).¹⁶

Combining signal across population-scale cohorts can increase sample size while attenuating the effect of study-specific biases and confounding factors, but can be prone to emergent sources of confounding from “batch” effects or other systematic biases in methylation data across cohorts. This is especially problematic when there is notable class imbalance (ie, different outcome frequencies) across cohorts.¹⁷ The most common method for dealing with this heterogeneity is meta-analysis, but standard meta-analysis approaches are restricted to univariate (one CpG site at a time) models. Other approaches include batch effect correction on the input data set (eg, ComBat¹⁸), direct adjustment for batch/study in linear models, or adjustment for derived variables intended to capture technical biases (eg, surrogate variable analysis¹⁹), but these approaches can often lead to over- or underestimates of true biological effects.¹⁷ An alternative approach described recently, cross-study learning, instead trains an ensemble predictor consisting of one or multiple models per cohort.²⁰ This strategy allows the use of arbitrarily complex models while avoiding technical confounding from direct combination of the data sets.

To develop an improved DNA methylation-based cardiovascular risk predictor using multiple heterogeneous training cohorts, we used a cross-study learning method to develop an ensemble of penalized time-to-event regression risk models. The resulting composite risk score performed well in a held-out data subset, associating with survival even in the presence of traditional risk factors, and showing similar performance to models trained on naively merged data sets. External validation was achieved in a case-control study for prevalent myocardial infarction (MI). Further, interactions were assessed between the composite methylation-based risk score and other risk predictors, finding that it is potentially most effective in those with low Framingham Risk Scores.

METHODS

Study Participants and Phenotype Collection

Phenotypes (demographic, anthropometric, biochemical, and clinical), DNA methylation data, and imputed genotypes were available either from publicly available controlled-access databases or upon request from the cohorts. Cohort-specific details are provided in Data S1. Blood-based biochemical markers (total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, fasting glucose, high-sensitivity C-reactive protein, and systolic blood pressure) were log₁₀-transformed for all analyses. In the Lothian Birth Cohort 1936, LDL was estimated from total cholesterol and triglycerides using the Friedewald equation. Diabetes mellitus was defined as either use of diabetes mellitus medication or a measured fasting blood glucose level of >125 mg/dL. Antihypertensive medication use, smoking status, and diabetes mellitus status were assumed to be false where missing, though missing data rates for these variables in the held-out FHS (Framingham Heart Study) subset were low (0.1%, 0.1%, and 7%, respectively). Analysis of these data sets was approved by the Tufts University Health Sciences Institutional Review Board (protocol 12592), and all subjects gave informed consent.

DNA Methylation Data Processing

DNA methylation data for all initial cohorts (Womens Health Initiative [WHI], FHS, and Lothian Birth Cohorts [LBC]) were collected using the Illumina HumanMethylation450 microarray platform²¹ and downloaded as raw intensity files. FHS methylation data were collected in 2 primary batches in 2 centers—1 in subjects from a nested case-control for CVD measured at Johns Hopkins University (FHS-JHU), and the other in a larger set of remaining Framingham Offspring participants measured at the University of Minnesota (FHS-UM). Preprocessing was performed using the *minfi* and *wateRmelon* packages for R.^{22,23} Sample-wise filters were as follows: robust overall signal in the main cluster based on visual inspection of an intensity plot, <10% of probes undetected at a detection threshold of $P < 1e-16$, and a reported sex matching methylation-based sex prediction. Probes were removed using the following criteria: >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 single-nucleotide polymorphism 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.^{24,25} Blood cell fractions for 6 blood cell types (CD4+ T-cells, CD8+ T-cells, B-cells, natural killer cells, monocytes, and granulocytes) were estimated using a common reference-based method,²⁶ and 5 of these (excluding granulocytes) were included in cell count-adjusted statistical models. After quality control and filtering steps, 390 597 CpG sites were shared between the 3 data sets, formatted as beta values (roughly equal to the ratio of methylated signal to total microarray signal, or $\beta = \frac{M}{M+U+100}$).

DNA methylation data for the REGICOR (Registre Gironí del COR) cohort were collected using the Illumina MethylationEPIC microarray platform²⁷ and analyzed using the *wateRmelon*²³ and *methylumi*²⁸ R packages. Samples were excluded based on detection $P > 0.05$ in at least 1% of probes or failure to cluster in the appropriate sex based on X chromosome methylation. Probes were excluded based on detection $P > 0.05$ in at least 1% of samples, a bead count <3 in at least 5% of samples, discarding by Illumina based on underperformance (n=1031) or changes in the manufacturing process (n=977), non-CpG targets, and cross-hybridization (n=43 979). A batch normalization was performed by standardizing beta values to mean zero and unit variance within each bisulfite conversion batch before analysis. After quality control and preprocessing, 811 610 CpG sites across 391 individuals were available for analysis. Participants were further excluded from analysis because of unknown smoking habits (n=10) and unavailable information regarding diabetes mellitus, hypertension, or hyperlipidemia (n=53). Surrogate variable analysis¹⁹ was used to calculate 2 surrogate variables, representing potential technical and biological confounders, for adjustment in MRS replication models.

CVD Risk Modeling

Study-specific CVD risk models were trained using penalized Cox proportional hazards regressions with the elastic net penalty. CVD events were defined as including coronary heart disease, stroke, and death from CVD (see Data S1 for cohort-specific details), and times were right-censored based on the most recent exam available in each cohort. The elastic alpha parameter was initially set at 0.05 (closer to ridge regression) to retain a higher number of CpGs with non-zero weights while still performing feature selection.²⁹ Inner cross-validation loops varying alpha between 0.05 and 0.95 showed negligible differences in model performance (evaluated by mean squared error). The penalty parameter λ was optimized through 5-fold cross-validation (use of 10-fold cross-validation did not meaningfully change the results). For each model, only the most variable 100 000 CpGs according to median

absolute deviation ($\approx 25\%$ of all available sites shared across platforms) were included to decrease the computational burden and ensure that the selected CpGs would have meaningful interindividual variation.

The cross-study learner (CSL) was constructed as an ensemble of study-specific regression models. Scores from each single-study learner (SSL) were combined using the “stacking” approach,²⁰ implemented as follows. First, predictions from each SSL to both itself and the other training data sets were combined into a design matrix (with dimensions $N_{\text{total}} \times \#$ SSLs). This formed the input to an additional penalized Cox regression (ridge regression with λ optimized through 5-fold cross-validation and coefficients restricted to be non-negative) of all training studies at once. Coefficients from this regression, corresponding to input study-specific SSLs, were normalized to sum to 1 to produce the CSL weights. For use in new data sets, SSL scores were each standardized to mean zero and unit variance before calculating their weighted sum (using the “stacking” weights) as the final CSL score.

A series of approaches for combining information across cohorts were tested as alternatives to the CSL. The naive “combined” approach consisted of simply aggregating observations from all training sets into a single data set and training an elastic net regression as described above while adjusting for study as a fixed effect. The ComBat method trained across all studies as with the “combined” approach, but included an empirical Bayes-based preprocessing step to directly remove mean differences across studies that were not associated with the outcome of interest (incident CVD events).¹⁸

MRS evaluation in FHS-UM was performed using Cox proportional hazards models, with a series of models adjusting for covariates including demographics, anthropometrics, biochemical values, and cell subtype estimates. Additional sensitivity models incorporated flexible spline bases for age and cell type fractions (*pspline* function) and an interaction between age and sex. Robust standard errors were used to account for family structure as has been suggested for clustered data³⁰ and used for epigenetic risk models in FHS.³¹ The proportional hazards assumption was assessed using the *cox.zph* R function, and no violation was detected ($P > 0.05$). To compare risk scores generated using different models (combined and ComBat-preprocessed) to the CSL, Cox regressions adjusting for the “basic” covariate set were used to evaluate each MRS alone, the CSL MRS plus the combined MRS, and the CSL MRS plus the ComBat-preprocessed MRS in the held-out FHS-UM data set. Likelihood ratio tests were then used to compare each of the 2-MRS models to that CSL-only model, with the resulting P values indicating whether either of these alternative scores provided additional benefit. MRS

evaluation in the REGICOR case-control used logistic regression models, adjusting for the same sets of covariates where possible, though traditional biochemical risk factors were only available in discrete low versus high categories.

The biology underlying the CSL model was evaluated through a series of enrichment tests using the component CpG loci and annotated genes. Gene ontology-based enrichment analysis of each cohort-specific model was performed using the *gometh* function from the *missMethyl* package for R.³² This procedure uses gene annotations for CpGs from the HumanMethylation450 microarray annotation from Illumina (v1.0 B2). Enrichment analysis is then performed for each gene ontology category using Wallenius’ non-central hypergeometric distribution to account for inconsistent representation of CpG sites across genes. The overall merged set of CpGs included in the final CSL model was then tested for enrichment in transcription factor binding sites using HOMER tool.³³ CpG loci (with respect to genome build hg19) were provided as inputs, with 200 base-pair windows and repeat-masked sequences.

Genomic Risk Score Calculation

Imputed genotype data for WHI were retrieved from dbGaP (accession: phs000746.v2.p3. Variants were filtered for imputation $R^2 > 0.3$, and annotated with rsIDs, loci, and allelic information using the 1000 Genomes Phase 3 download from dbSNP (download date: April 13, 2018). Weights for the genetic risk score calculation (6 630 151 variants) were based on the genome-wide CVD score developed by Khera et al.¹⁵ We note that these scores were developed only for populations of European descent, and thus are not optimized for the mixed-ancestry WHI population. Genomic risk scores (GRS) were then calculated as the weighted sum of allelic dosages, normalized by the number of relevant SNPs available. Genotype data processing and GRS calculation were performed using PLINK 2.0.

Risk Score Interaction Analysis

Interaction analysis was performed using similar Cox regression models to those above, adjusting for the “basic” set of covariates and using robust standard error estimates. To facilitate visual comparisons, main-effect regressions for the MRS were fitted within risk strata defined by the Framingham Risk Score (FRS) or genomic risk score (GRS), both separately in each data set having > 25 events in the group, and after merging these data sets and allowing for stratified baseline hazards (strata() argument to the *coxph* function). To obtain overall interaction effect estimates, an interaction between MRS and either FRS or GRS was introduced into a combined regression including all data sets,

while allowing stratified baseline hazards. We note that main effects in the interaction analysis are biased away from the null since the regression data sets were used for training the MRS. Regressions assessing the GRS excluded non-European ancestry participants to match the ancestry used to develop the CVD score.¹⁵

For quasi-replication of these associations in the REGICOR data set, stratified logistic regressions were used to discriminate MI cases from controls using the MRS, while adjusting for estimated cell count fractions as well as 2 surrogate variable analysis components (as in the main REGICOR models). In the absence of continuous values for blood pressure and lipids, an empirical risk function was generated by first performing a logistic regression on the following cardiovascular risk factors: age, sex, estimated cell count fractions, body mass index, diabetes mellitus, smoking status, hyperlipidemia (binary), and hypertension (binary), along with 2 surrogate variable analysis components. Predicted risks based on this model were then used to stratify subjects into 4 risk groups by evenly splitting the range of predicted risks into 4 segments (thus resulting in strata based on raw risk, rather than percentiles).

RESULTS

Cross-Study Learner Model Development

Epigenomic model development was performed in 3 cohorts, including the WHI, FHS, and LBC 1936. The

FHS data set was divided into 2 functionally separate groups (FHS-JHU and FHS-UM) based on differences in subject selection and geographic location of laboratory methylation analysis (see Methods). Further population details can be found in Table 1.

Figure 1 outlines the computational workflow. Briefly, a cross-study learning (CSL) model was developed by training time-to-event elastic net regressions on 3 of the data sets, while holding out the FHS-UM subset for evaluation. The FHS-UM subset was chosen to hold out as it more closely represents the larger free-living Framingham population. While there is moderate heterogeneity between the included cohorts (for example, in original cohort study designs, details of CVD definitions, and length of follow-up), the intent of the present investigation was to explore the extraction of shared signal across cohorts with recognized heterogeneity. Next, a model re-trained on all 4 data sets were subject to external replication in the REGICOR study. CSL model CpGs were characterized as to their potential biological function, and model performance was assessed across strata of alternative cardiovascular risk metrics.

The initial predictor was developed by training individual penalized Cox proportional hazards regression models (single-study learners, or SSLs) in each of the 3 training cohorts (WHI, FHS-JHU, and LBC). Scores from these models were aggregated through a “stacking” method, in which the outcomes

Table 1. Baseline Parameters of the Populations Used for Model Development

Study/Subset	WHI	FHS-JHU	LBC	FHS-UM
Sample size	2023	484	818	2103
Age, y	65 (59–70)	71 (64–77)	69 (68–70)	64 (59–71)
Sex (women)	2023 (100%)	145 (30%)	406 (50%)	1270 (60%)
Ancestry				
% European	959 (47%)	484 (100%)	818 (100%)	2103 (100%)
% African American	651 (32%)	0 (0%)	0 (0%)	0 (0%)
% Hispanic	413 (20%)	0 (0%)	0 (0%)	0 (0%)
Body mass index, kg/m ²	29.1 (25.5–33.3)	28.2 (25.5–31.3)	27.5 (24.9–30.3)	27.4 (24.3–31)
LDL cholesterol, mg/dL	150 (126–175)	88 (73–107)	118 (89.5–150.3)	107 (87–128)
HDL cholesterol, mg/dL	51 (43–60)	49 (40–60)	56.1 (47.2–68.3)	56 (45.8–69)
Triglycerides, mg/dL	127 (92–177)	101.5 (75–141.2)	128.4 (97.4–171.2)	102 (73–142)
Fasting glucose, mg/dL	96 (88.6–108)	106 (97–116)	Unavailable	100 (94–109)
Systolic blood pressure, mm Hg	131 (120–143)	130 (117–143)	148.7 (137–161.3)	126 (116–138)
No. CVD events				
Prior only	0	127	70	112
Incident only	1009	67	133	146
Prior and incident	0	58	164	34
Total	1009	252	367	292
Follow-up time, y	22	10	14	10

Continuous values shown as: median (interquartile range). CVD indicates cardiovascular disease; FHS-JHU, Framingham Heart Study Offspring Cohort (Johns Hopkins University subset); FHS-UM, Framingham Heart Study Offspring Cohort (University of Minnesota subset); HDL, high-density lipoprotein; LBC, Lothian Birth Cohorts 1936; LDL, low-density lipoprotein; and WHI, Women's Health Initiative.

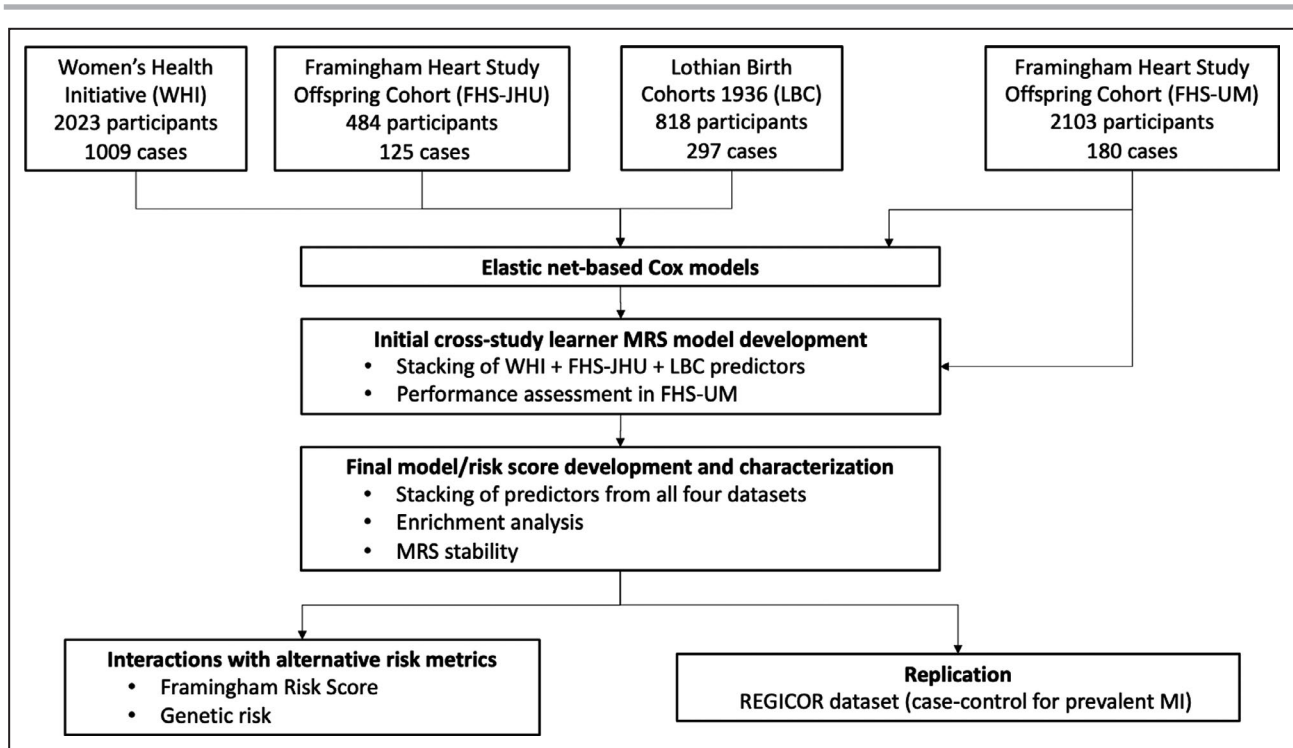


Figure 1. Computational workflow for MRS development and evaluation.

The initial MRS was trained in 3 cohorts with Framingham Heart Study Offspring Cohort (University of Minnesota subset) held out to evaluate performance. The final MRS was then trained using all 4 data sets and examined for biological significance, before testing for prevalent myocardial infarction discrimination in an independent cohort and assessment of interactions with genetic and traditional risk scores. FHS-JHU indicates Framingham Heart Study Offspring Cohort (Johns Hopkins University subset); FHS-UM, Framingham Heart Study Offspring Cohort (University of Minnesota subset); LBC, Lothian Birth Cohorts 1936; MI, myocardial infarction; MRS, methylation-based risk score; and WHI, Women's Health Initiative.

and model predictions from each of the individual data sets are combined, and a regression is used to assign weights to each of the model scores (see Methods). This procedure led to FHS-JHU dropping out of the ensemble model, with weights for this initial predictor as follows: 0.57 (WHI), 0.0 (FHS-JHU), and 0.43 (LBC). This result means that the FHS-JHU score did not transfer to the rest of the data sets (ie, to WHI and LBC) as well as the scores from the other 2 components models.

Assessment in Held-Out FHS Subset

Stacking of the 3 initial predictors resulted in model weights of 0.57, 0, and 0.43 for WHI, FHS-JHU, and LBC, respectively (ie, the FHS-JHU sub-model did not contribute to the initial stacked ensemble model). The resulting ensemble predictor was evaluated using robust Cox proportional hazards models in FHS-UM, showing strong associations with incident CVD in an unadjusted model (hazard ratio [HR]=1.58, 95% CI, 1.37–1.83), which was attenuated partially through adjustment for standard covariates (age, sex, and estimated cell type fractions; HR, 1.28; 95% CI, 1.10–1.50) as well as CVD risk factors (HR, 1.29; 95% CI,

1.09–1.51). Results for the unadjusted model and 3 risk factor-adjusted models are shown in Table 2, and associated Kaplan–Meier curves across epigenetic risk tertiles are shown in Figure 2.

Additional sensitivity analyses were performed to assess the robustness of these results to variations in the model-building or evaluation approach. Hazard ratios in the held-out FHS-UM were no higher using

Table 2. MRS Performance in Held-Out FHS Subset

Model	HR Per SD MRS*	P Value
Unadjusted [†]	1.58 [1.37–1.83]	5.4e-10
Basic [‡]	1.28 [1.10–1.50]	2.0e-03
Plus risk factors [§]	1.29 [1.09–1.51]	2.7e-03
FRS only	1.36 [1.19–1.58]	2.0e-05

FHS indicates Framingham Heart Study; FRS, Framingham Risk Score; HR, hazard ratio; and MRS, methylation-based risk score.

*Estimated hazard ratio per SD of the methylation-based risk score [95% CI].

[†]No covariates.

[‡]Adjusted for age, sex, and estimated cell type fractions.

[§]Additionally adjusted for body mass index, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, systolic blood pressure, diabetes mellitus status, and current smoking.

^{||}Adjusted for Framingham Risk Score (uses all risk factors other than body mass index and cell type fractions).

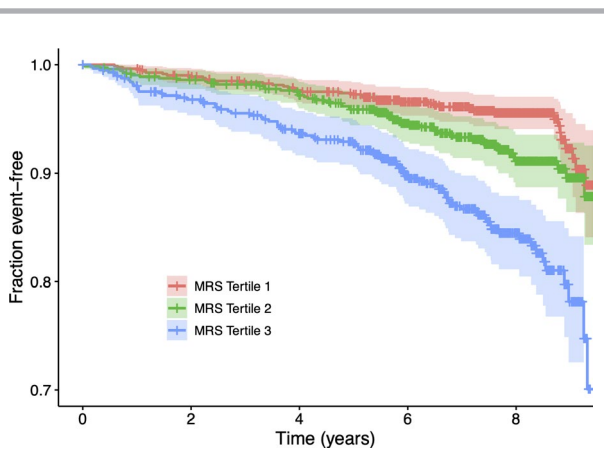


Figure 2. Kaplan–Meier survival curves in the held-out Framingham Heart Study Offspring Cohort (University of Minnesota subset data set).

Individual curves correspond to tertiles of the initial (3-data set) methylation-based risk score. Vertical ticks correspond to censored observations, and colored bands represent 95% CI for tertile-specific survival curves. X-axis is limited to the time span in which at least 50 uncensored observations remained for each tertile (3275 days). MRS indicates methylation-based risk score.

penalized logistic regression in training (unadjusted HR, 1.52; 95% CI, 1.32–1.76), excluding individuals with past events in training (unadjusted HR, 1.55; 95% CI, 1.33–1.81), or adjusting for race in WHI (unadjusted HR, 1.20; 95% CI, 1.03–1.39). Neither were these results affected by training using the full set of 390 597 CpGs. Similarly, variations in the evaluation regressions did not produce meaningfully different results, either when excluding all individuals who experienced prior CVD events (Table S1), analyzing incident CVD as a binary outcome using logistic regression (unadjusted odds ratio per SD=2.15, 95% CI, 1.91–2.42), or stratifying by sex. Adjustment for age and cell type fractions as flexible spline functions as well as an age-sex interaction to assess possible residual confounding did not decrease estimated HRs from the basic model (saturated model HR, 1.31; 95% CI, 1.12–1.52). Use of the MRS for binary incident CVD prediction resulted in a c-statistic of 0.642 (95% CI, 0.599–0.685), compared with 0.691 (95% CI, 0.653–0.729) for the Framingham Risk Score alone and 0.695 (95% CI, 0.655–0.734) using the 2 scores together.

Results from comparison of CSL performance to models trained on combined data sets (either naive combination or including preprocessing using ComBat) are shown in Figure S1. The ComBat-preprocessed model had modestly higher hazard ratios in FHS-UM, while relative differences with the combined model depended on the covariates included. However, likelihood ratio tests using the basic model covariates (age, sex, and cell type fraction-adjusted) did not reveal a

strong added benefit of either the combined ($P=0.58$) or ComBat ($P=0.08$) risk scores over that using only the CSL.

Final CSL Model Characterization

The stacking regression in the final CSL model defining the methylation-based risk score (MRS) gave the most weight to WHI (0.48) and LBC (0.38), while retaining non-zero weights for FHS-JHU (0.06) and FHS-UM (0.08). This result indicates that the WHI and LBC-trained models were better able to transfer across the combined-cohort set of outcomes compared with the other models. There was little overlap of specific CpG sites across cohort-specific models, with a maximum of 13 CpGs shared between 2 models (WHI and FHS-UM) and no CpGs shared between >3 or (Figure 3A). This could result from heterogeneity in the complex relationships between DNA methylation and CVD across populations. However, it may also reflect the tendency of the elastic net regression to select only a single feature from a group of correlated features, where the specific CpGs selected in different data sets varied because of the presence of biological and technical noise. However, even if the SSLs were capturing different biological mechanisms, the CSL model is designed to capture such heterogeneous signal from across cohorts. Despite the lack of site-specific overlap, there was broad agreement for 3 of the 4 component SSL models at the level of enriched biological processes, with all except FHS-JHU enriched most strongly for proximity to genes involved in homophilic cell adhesion (Figure 3B). MRS component CpGs tended to be found in similar genomic loci to the overall set of variable CpGs and were enriched in gene bodies and depleted in CpG islands compared with the full microarray CpG set. However, MRS CpGs did show a modest enrichment in and around CpG islands compared with the set of variable CpGs (Figure 3D). To seek more clarity as to potential biological mechanisms represented by the MRS, the HOMER tool was used to calculate enrichment of transcription factor binding motifs in the MRS component CpG sites. Using the union of all individual SSL CpG sites as input, no strong enrichments were found (all $q>0.5$). To better understand the stability of the risk score over time, intraclass correlation coefficients (ICCs) were calculated for 2 sets of grouped samples: 26 technical replicates from FHS and ≈ 1000 longitudinal samples (across 3 visits, or about 6 years total) from LBC (Table S2). The technical replicates showed an ICC of 0.85, while the longitudinal samples showed an ICC of 0.68. As would be expected, the ICC for samples closer in time (Waves 1 & 2; ICC=0.69) were higher than that for samples more distant in time

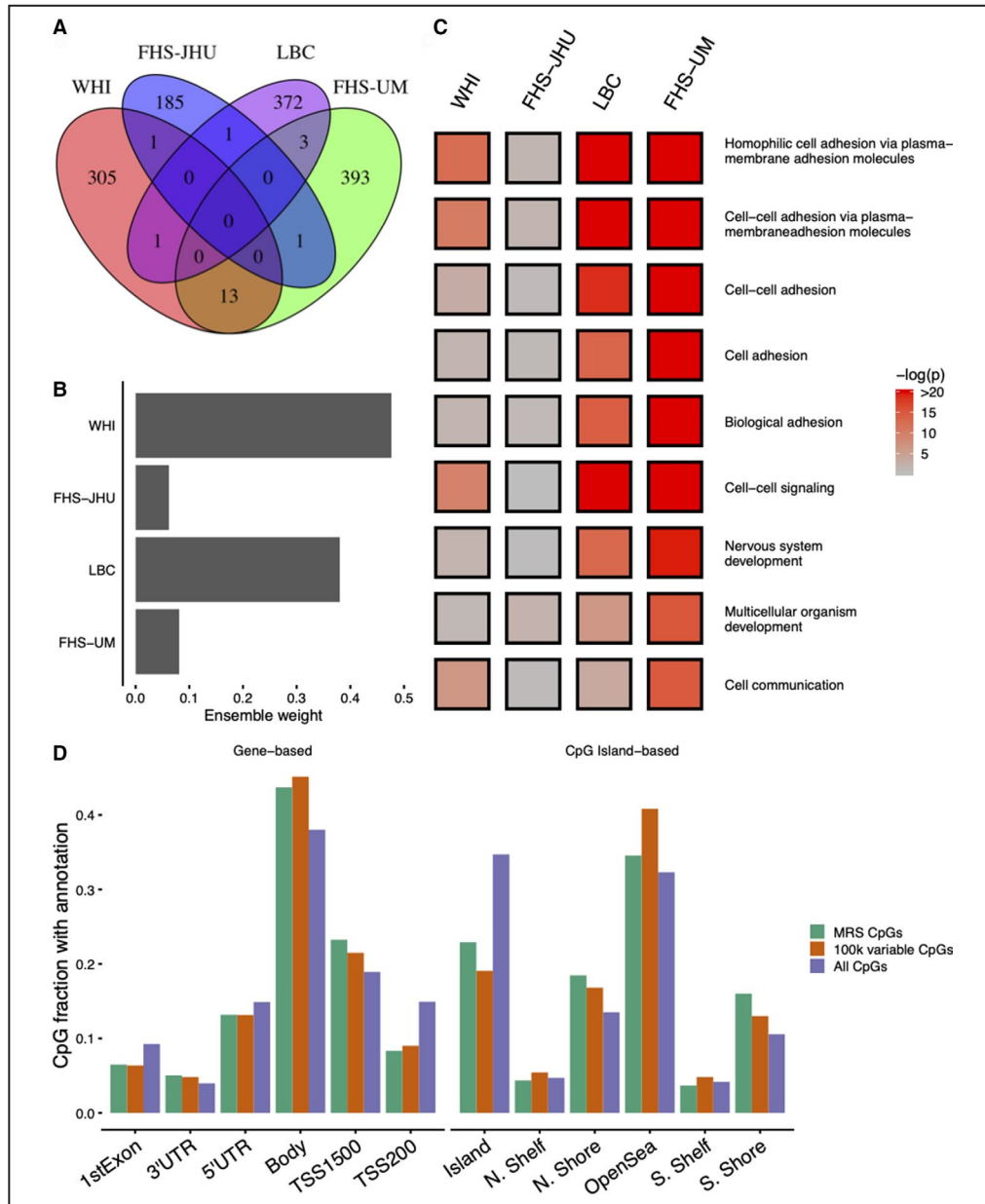


Figure 3. Characterization of the final cross study learner model. **A**, Overlap of cytosine-phosphate-guanine (CpG) sites in the 4 individual predictors constituting the final model. **B**, Study-specific weights for constructing the ensemble model (derived from the “stacking” regression). **C**, Results from Gene Ontology (GO)-based enrichment analysis using genes annotated to single-study learner component CpGs. All GO terms with false discovery rate <0.001 in any cohort are shown and colored according to $-\log(P)$ value for enrichment in each single-study learner. Values were cut at $-\log(P)=20$ for visualization purposes. **D**, Proportion of CpGs in the full set of cross study learner CpGs (union of CpG sets in each component SSL) compared with the 100 000 most variable CpGs (as used in single-study learner model development) and the full set of available CpGs. Groupings according to both gene-based and CpG island-based CpG annotations are shown. CpG indicates cytosine-phosphate-guanine; FHS-JHU, Framingham Heart Study Offspring Cohort (Johns Hopkins University subset); FHS-UM, Framingham Heart Study Offspring Cohort (University of Minnesota subset); LBC, Lothian Birth Cohorts 1936; MRS, methylation-based risk score; WHI, Women’s Health Initiative; UTR, untranslated region; and TSS, transcription start site.

(Waves 1 & 3; ICC=0.61). Based on the observation of imperfect stability of the MRS over time as well as the partial attenuation in held-out hazard

ratios after adjustment for age, its component CpGs (the 1305-element union of all CpGs in any of the 4 individual SSL models) were examined for overlap

with established epigenetic age metrics. While no enrichment was seen for the original cross-tissue DNAm age from Horvath,³⁴ strong enrichment was seen for the morbidity-directed PhenoAge⁵ (9 of 513 CpGs; $P=2.3e-5$) and especially the blood-specific aging marker from Hannum et al³⁵ (13 of 71 CpGs; $P=5.9e-21$). We note that these overlaps do not constitute a major fraction of either CpG set but are nonetheless highly statistically significant. The PhenoAge metric is based on some known cardiovascular risk factors (eg, C-reactive protein) and is known to associate with CVD but is not trained in any of the cohorts included here.

Discrimination in Myocardial Infarction Case-Control Study

As one form of replication, the MRS was investigated for its discriminative performance in a nested case-control for prior myocardial infarction in the REGICOR cohort (Table 3; cohort description in Table S3), which was matched for sex and age and thus free of potential confounding by these variables. We note that this data set contained prevalent (rather than incident) events, and thus provides replication in a similar but not identical biological context. These methylation data were collected on the next generation of Illumina methylation microarray (MethylationEPIC), which does not perfectly overlap with the HumanMethylation450 platform, but contained $\approx 93\%$ of the CpGs input to the MRS model training procedure. The MRS was able to discriminate cases and controls in both unadjusted (odds ratio=1.79, $P=6.33e-6$) and, to a lesser degree, risk factor-adjusted models (odds ratio=1.61, $P=0.019$). Odds ratios were qualitatively similar across modeling strategies (Combined, ComBat, and CSL) for all of the adjustment models (Figure S1B).

Interactions With Alternate Risk Metrics

To understand how the present risk score interacts with other established CVD risk metrics, the performance

Table 3. Results From Replication in REGICOR Myocardial Infarction Case Control

Model	ComBat	Combined	CSL
Minimal [†]	1.79 [1.39–2.31]	1.86 [1.45–2.38]	1.83 [1.41–2.37]
Basic [†]	2.16 [1.58–2.93]	2.12 [1.57–2.87]	2.14 [1.58–2.89]
Plus risk factors [‡]	1.76 [1.22–2.54]	1.66 [1.15–2.4]	1.61 [1.11–2.34]

Results are presented as: odds ratio per SD methylation-based risk score [95% CI]. CSL indicates cross study learner; MRS, methylation-based risk score; and REGICOR, Registre Gironí del COR.

[†]Adjusted for 2 surrogate variable analysis components.

[‡]Additionally adjusted for age, sex, and estimated cell type fractions.

[‡]Further adjusted for body mass index, low-density lipoprotein, high-density lipoprotein cholesterol, systolic blood pressure, diabetes mellitus status, and current smoking.

of the MRS was re-evaluated after stratifying individuals by risk scores reflecting either demographic and biochemical features (Framingham Risk Score), or genetic variants (GRS based on Khera et al¹⁵). First, the marginal effects of these risk scores were confirmed in each population. The FRS was strongly predictive in WHI and FHS, while surprisingly showing no association with CVD incidence in LBC (Table S4). As the data set with the largest number of available events, imputed genotypes were retrieved for WHI and GRS calculated, demonstrating a moderate association with CVD (odds ratio per SD=1.28, $P=1.1e-6$).

In pooled Cox models using study-specific baseline hazards and performed using the final 4-study MRS, it appeared that the MRS was more effective in those in lower “traditional” risk strata (based on models stratified by FRS categories; Figure 4A). As a sensitivity analysis, the cohorts were fully stratified into separate models, in which this pattern was visually clear in WHI and FHS-JHU (Figure S2). The pattern did not appear in LBC, although we note that the Framingham Risk Score also did not show a “main effect” for incident CVD in this cohort. A similar pattern appeared with respect to genetic risk in WHI (European ancestry participants only based on the formulation of the relevant risk score), in which maximum MRS performance was achieved in the lowest alternative risk stratum. Supplementing these visual comparisons, combined Cox regressions across all cohorts (allowing for different baseline hazards across studies) showed a strong MRS-FRS interaction effect (7% reduction in HR for the MRS per 10% increase in FRS; $P=8.27e-05$), while that for the MRS-GRS interaction did not reach nominal statistical significance (2% reduction in HR for the MRS per SD increase in GRS; $P=0.719$).

To explore the clinical potential of these interactions further, we returned to the initial MRS (trained in 3 data sets with FHS-UM held out). The FHS-UM data set was filtered to include only participants with lower CVD risk based on the FRS (<10% estimated 10-year risk). Within this lower-risk subset, participants in the upper MRS quintile had more than double the risk of the remainder of the participants: 7% (12/176) of the upper MRS quintile experienced incident events, while 3% (19/701) of the remaining 4 MRS quintiles experienced incident events.

FRS could not be calculated in the REGICOR data set, as not all risk factors were available as continuous values. However, stratified models replicated the observation of greater MRS discrimination in the lowest alternative risk stratum. An empirical risk function was generated through logistic regression of MI status on cardiovascular risk factors (age, sex, body mass index, diabetes mellitus, smoking status, hyperlipidemia [binary], and hypertension). Predicted MI risk using this

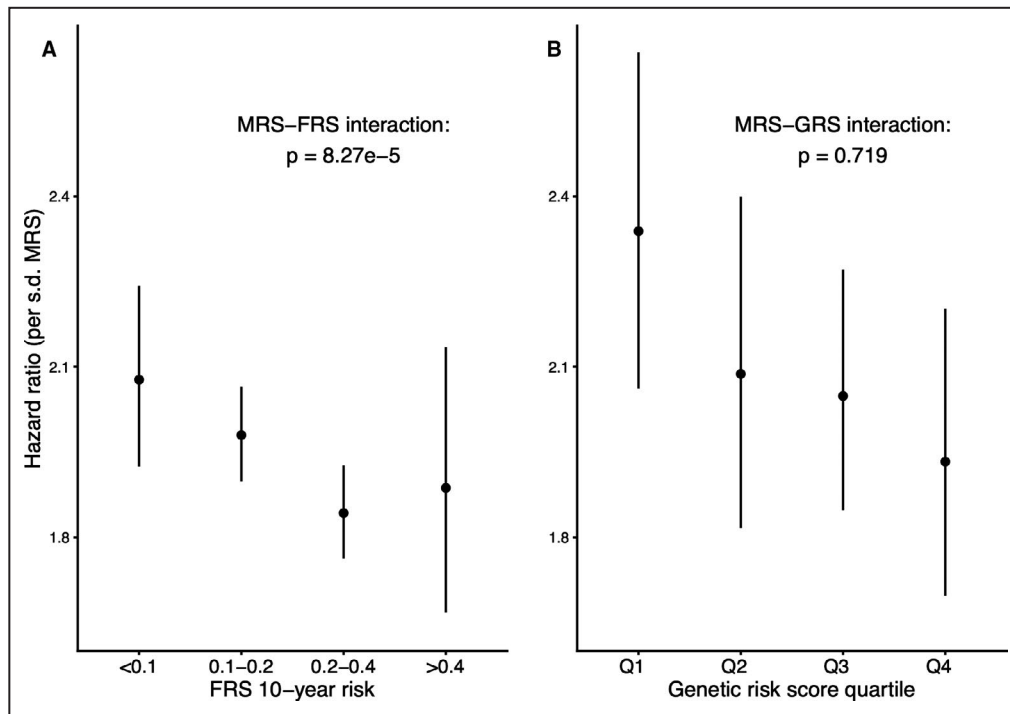


Figure 4. Interactions of MRS with other biomarkers of CVD risk.

A, Hazard ratios for the MRS within subsets of 10-year generalized CVD risk according to the Framingham Risk Score. **B**, Hazard ratios for the MRS within quartiles of a genetic cardiovascular risk score (in European-ancestry WHI participants only). Hazard ratios are estimated using the final MRS, which was trained using each of these data sets. Cox regressions included stratum-specific baseline hazards and were adjusted for age, sex, and estimated cell subtype fractions. Error bars represent standard errors for the hazard ratio estimates. Annotated *P* values describe the test of interaction between the MRS and the alternative risk metric. FRS indicates Framingham Risk Score; GRS, genetic risk score; and MRS, methylation-based risk score.

model was used to stratify subjects into 4 risk groups, with MRS odds ratios (per SD) of 4.49 in the lowest-risk group versus 1.20 in the highest-risk group. More detailed results from these analyses are shown in Table S5.

DISCUSSION

Epigenetic signatures of cardiometabolic diseases and aging in general are being actively explored as biomarkers of disease risk that are potentially modifiable and reveal underlying biological mechanisms. Here, in a novel application of a cross-study ensembling method, we introduce a DNA methylation-based score specific to cardiovascular disease risk. The model performs similarly to one trained on a direct combination of the component data sets and may perform best in individuals predicted to be at lower risk based on traditional risk factors.

We opted to use cross-study learning to train our risk model based on the expectation that differences across cohorts (eg, demographic, behavioral) may contribute to heterogeneity in both the marginal distribution of the CpG features and the conditional

distribution of the CVD outcome. Under these conditions, the generalizability of a single-study predictor is often obscured or overstated.^{36,37} The performance of the CSL model was similar to that of models trained on the merged cohorts with or without batch adjustment via ComBat. This suggests that the assumptions made by these direct combination strategies (ie, that the heterogeneity structure can be captured by variation in the marginal effects of each CpG site) are met. In practice, this underlying structure is unknown, and we highlight that the CSL was able to produce similar gains in accuracy without making specific assumptions.

In assessing the stability of the MRS, we observed reasonable reproducibility between technical replicates (ICC=0.85). ICCs for LBC subjects over time were somewhat lower (ICC=0.68), which is to be expected because of not only changes in environment, but also the known epigenetic evolution with age that we observed to be enriched in the components of our score. Furthermore, this value is at the upper end of the range of single-CpG repeatability measurements over time calculated in the combined Lothian Birth Cohorts (1921 and 1936).³⁸ These ICC

values suggest an imperfect but usable reproducibility of the MRS, and an aggregate marker that is fairly robust considering the low replicability that has been observed for individual sites in technical replicates (general median ICC of 0.3 and mode of 0.75 in a “high reliability” cluster).³⁹

Our observation that different CpGs tended to be selected across studies (Figure 3) is in agreement with the relative lack of replication seen in prior cardiovascular epigenomic studies.⁷ However, the enrichment of the MRS component CpGs for proximity to genes related to cell-cell adhesion (in all subsets except FHS-JHU) is indicative of shared underlying biological mechanisms. As we have previously observed in the WHI and FHS cohorts, it appears that immune activation is central to the prognostic information contained in leukocyte DNA methylation.¹⁴ For example, epigenetic processes have been shown to be involved in the activation and increased adhesion of monocytes in response to environmental insults and metabolic stress, though these have been explored primarily in relationship to histone modifications.⁴⁰ Our results provide preliminary support for an attractive model in which a methylation-based score could act as a monitor of cumulative stress in leukocytes and their corresponding activation towards a more atherogenic state.

Existing epigenetic scores have shown varying strengths of association with incident cardiovascular disease. An early investigation examined blood-based methylation in LINE-1 elements, finding strong associations of global hypomethylation with prevalent and incident ischemic heart disease (LINE-1), though additional reports showed opposite associations of methylation at repetitive elements with CVD.⁴¹ Guarrera et al developed a biomarker for MI based on global LINE-1 and ZBTB12 gene methylation that provided a modest net reclassification index improvement (0.23–0.47) compared with traditional risk factors only. Multiple epigenetic aging metrics, though not developed specifically for CVD, have been shown to predict incident CHD, including PhenoAge (odds ratios from 1.02–1.08) and GrimAge (hazard ratio=1.07, adjusted for age and technical factors).^{5,31} While these associations are statistically significant, they do not represent clinically meaningful improvements in discrimination. Our observed hazard ratio of 1.28 (Basic model in the held-out FHS-UM data set) indicates that this MRS may be closer to clinical relevance. We note that our component CpG sites overlap strongly with those of these established epigenetic metrics including PhenoAge, suggesting that it captures some of the same biological patterns. However, the mechanistic significance of the specific methylation signals captured by these aging-related metrics, whether as markers of epigenetic regulation

breakdown or the work of an “epigenetic maintenance system”, is still unclear.^{34,42}

In examining the potential clinical utility of a novel risk score for CVD, it is important to understand to what extent it is redundant with or complementary to existing risk metrics. We first note that the strength of this epigenetic score in adjusted models is lower than that found for traditional risk scores (Table S4) and some novel biochemical risk measures such as high-sensitivity Troponin I (adjusted HR for global CVD=3.01).⁴³ However, analysis of interactions between different risk metrics can be clinically relevant, as demonstrated for example in a recent investigation exploring the interaction between genetic and lifestyle-based risk prediction for dementia.⁴⁴ Here, we saw a pattern of stronger epigenetic risk associations in individuals whose cardiovascular risk based on traditional metrics (here, the Framingham Risk Score) was low. This pattern replicated in the REGICOR data set (though FRS could not be directly calculated), with improved MRS discrimination in lower-risk subjects based on an empirical risk function. While these associations are preliminary, they suggest that an epigenetic risk score could help identify higher-risk individuals who otherwise would not have been detected by other metrics. While we did not identify any robust patterns of differential MRS performance in strata based on a genetic cardiovascular risk score, there may have been lower power to detect any such patterns from the outset given the modest discriminatory performance of the GRS in WHI.

Multiple limitations should be acknowledged. While lymphocytes are known to be important in CVD pathogenesis, epigenetic signals have been reported in other CVD-relevant tissues, such as aorta and other vascular tissues,⁷ that were not examined here. Additionally, the present definition of CVD was chosen to balance specificity of CVD subtypes with sample size, but this balance could be altered to focus on more specific disease subtypes (eg, myocardial infarction) or a broader definition of CVD (eg, including heart failure). Finally, while the REGICOR data set provided an important age- and sex-matched case-control setting for replication of the MRS, this work would benefit from future replication in an independent cohort enabling assessment of incident disease.

In summary, we have developed an epigenetic risk score for cardiovascular disease that provides additional value beyond existing risk measures and may show improved performance in populations otherwise designated as low risk. Furthermore, we have shown a novel application of a cross-cohort ensembling method that may provide significant value to future investigations in genomic epidemiology.

ARTICLE INFORMATION

Received November 13, 2019; accepted March 10, 2020.

Affiliations

From the JM-USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA (K.W., P.J., J.M.O.); Cardiovascular Epidemiology and Genetics Research Group, REGICOR Study Group, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Catalonia, Spain (A.F.-S., R.E.); Pompeu Fabra University (UPF), Barcelona, Catalonia, Spain (A.F.-S.); Department of Biostatistics, Boston University School of Public Health, Boston, MA (P.P., P.S.); Department of Psychology (J.M.S., I.J.D.) and Centre for Cognitive Ageing and Cognitive Epidemiology (J.M.S., I.J.D.), University of Edinburgh, United Kingdom; Department of Epidemiology, Brown University School of Public Health, Providence, RI (Q.L., S.L.); CIBER Cardiovascular Diseases (CIBERCV), Madrid, Spain (R.E.); Medicine Department, Medical School, University of Vic-Central University of Catalonia (UVic-UCC), Vic, Catalonia, Spain (R.E.); Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA (D.L.D.); IMDEA Alimentación, CEI, UAM, Madrid, Spain (J.M.O.); Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain (J.M.O.).

Acknowledgments

We thank all LBC 1936 study participants and research team members. Code supporting the analyses described here can be found at https://github.com/kwesterman/meth_cvd. Code and instructions related to the original cross-study learning approach can be found at https://github.com/prpatil/csm1_rep.

Sources of Funding

This work was supported by the US Department of Agriculture, Agriculture Research Service (8050–51000-098-00D). Dr. Westerman was additionally supported by National Institutes of Health predoctoral training grant 5T32HL069772-14. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, US Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. This article was prepared in collaboration with investigators of the WHI but has not been reviewed by the WHI and does not necessarily reflect the opinions of the WHI investigators or the National Heart, Lung, and Blood Institute. The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with Boston University (Contract No. N01-HC-25195 and HHSN268201500001). This article was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or National Heart, Lung, and Blood Institute. The LBC 1936 is supported by Age UK (Disconnected Mind program) and the Medical Research Council (MR/M01311/1). Methylation typing in LBC 1936 was supported by 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. LBC 1936 work was conducted in the Centre for Cognitive Ageing and Cognitive Epidemiology, which supported Dr. Deary and is supported by the Medical Research Council and Biotechnology and Biological Sciences Research Council (MR/K026992/1).

Disclosures

None.

Supplementary Materials

Data S1

Tables S1–S5

Figures S1–S2

References 45–49

REFERENCES

- Bonder MJ, Luijk R, Zernakova DV, Moed M, Deelen P, Vermaat M, van Iterson M, van Dijk F, van Galen M, Bot J, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet*. 2016;49:131–138.
- Tobi EW, Slieker RC, Luijk R, Dekkers KF, Stein AD, Xu KM, Slagboom PE, van Zwet EW, Lumey LH, Heijmans BT. DNA methylation as a mediator of the association between prenatal adversity and risk factors for metabolic disease in adulthood. *Sci Adv*. 2018;4:eaa04364.
- Bacos K, Gillberg L, Volkov P, Olsson AH, Hansen T, Pedersen O, Gjesing AP, Eiberg H, Tuomi T, Almgren P, et al. Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. *Nat Commun*. 2016;7:11089.
- Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, Tsai P-C, Ried JS, Zhang W, Yang Y, et al. Epigenome-wide association study of body mass index and the adverse outcomes of adiposity. *Nature*. 2017;541:81–86.
- Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*. 2018;10:573–591.
- Hao X, Luo H, Krawczyk M, Wei W, Wang W, Wang J, Flagg K, Hou J, Zhang H, Yi S, et al. DNA methylation markers for diagnosis and prognosis of common cancers. *Proc Natl Acad Sci USA*. 2017;114:7414–7419.
- Fernández-Sanlés A, Sayols-Baixeras S, Subirana I, Degano IR, Elosua R. Association between DNA methylation and coronary heart disease or other atherosclerotic events: A systematic review. *Atherosclerosis*. 2017;263:325–333.
- Hedman ÅK, Mendelson MM, Marioni RE, Gustafsson S, Joeanes R, Irvin MR, Zhi D, Sandling JK, Yao C, Liu C, et al. Epigenetic patterns in blood associated with lipid traits predict incident coronary heart disease events and are enriched for results from Genome-Wide Association Studies. *Circ Cardiovasc Genet*. 2017;10:e001487.
- Aslibekyan S, Agha G, Colicino E, Do AN, Lahti J, Ligthart S, Marioni RE, Marzi C, Mendelson MM, Tanaka T, et al. Association of methylation signals with incident coronary heart disease in an epigenome-wide assessment of circulating tumor necrosis factor α . *JAMA Cardiol*. 2018;3:463–472.
- Richardson TG, Zheng J, Davey Smith G, Timpson NJ, Gaunt TR, Relton CL, Hemani G. Mendelian randomization analysis identifies CpG sites as putative mediators for genetic influences on cardiovascular disease risk. *Am J Hum Genet*. 2017;101:590–602.
- Baccarelli A, Wright R, Bollati V, Litonjua A, Zanobetti A, Tarantini L, Sparrow D, Vokonas P, Schwartz J. Ischemic heart disease and stroke in relation to blood DNA methylation. *Epidemiology*. 2010;21:819–828.
- Guarrera S, Fiorito G, Onland-Moret NC, Russo A, Agnoli C, Allione A, Di Gaetano C, Mattiello A, Ricceri F, Chiodini P, et al. Gene-specific DNA methylation profiles and LINE-1 hypomethylation are associated with myocardial infarction risk. *Clin Epigenetics*. 2015;7:133.
- Agha G, Mendelson MM, Ward-Caviness CK, Joeanes R, Huan T, Gondalia R, Salfati E, Brody JA, Fiorito G, Bressler J, et al. Blood leukocyte DNA methylation predicts risk of future myocardial infarction and coronary heart disease. *Circulation*. 2019;140:645–657.
- Westerman K, Sebastiani P, Jacques P, Liu S, DeMeo D, Ordovás JM. DNA methylation modules associate with incident cardiovascular disease and cumulative risk factor exposure. *Clin Epigenetics*. 2019;11:142.
- Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50:1219–1224.
- D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*. 2008;117:743–753.
- Goh WWB, Wang W, Wong L. Why batch effects matter in omics data, and how to avoid them. *Trends Biotechnol*. 2017;35:498–507.
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8:118–127.
- Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet*. 2007;3:e161.
- Patil P, Parmigiani G. Training replicable predictors in multiple studies. *Proc Natl Acad Sci USA*. 2018;115:2578–2583.
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011;98:288–295.
- Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Mini: a flexible and comprehensive

- Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30:1363–1369.
23. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013;14:293.
 24. Fortin J-P, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics*. 2016;33:btw691.
 25. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29:189–196.
 26. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.
 27. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Dijk S, Muhlhäuser B, Stirzaker C, Clark SJ, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol*. 2016;17:208.
 28. Davis S, Du P, Bilke S, Triche T, Bootwalla O. methylumi: Handle Illumina methylation data. 2019.
 29. Benton MC, Sutherland HG, Macartney-Coxson D, Haupt LM, Lea RA, Griffiths LR. Methylome-wide association study of whole blood DNA in the Norfolk Island isolate identifies robust loci associated with age. *Aging (Albany NY)*. 2017;9:753–768.
 30. Rogers WH. Regression standard errors in clustered samples. *Stata Tech Bull*. 1993;13:19–23.
 31. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou L, Baccarelli AA, Li Y, Stewart JD, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)*. 2019;11:303–327.
 32. Phipson B, Maksimovic J, Oshlack A. MissMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics*. 2015;32:286–288.
 33. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–589.
 34. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14:R115.
 35. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan JB, Gao Y, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013;49:359–367.
 36. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
 37. Zhang Y, Bernau C, Parmigiani G, Waldron L. The impact of different sources of heterogeneity on loss of accuracy from genomic prediction models. *Biostatistics*. 2018;21:253–268. [Epub ahead of print].
 38. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res*. 2014;24:1725–1733.
 39. Bose M, Wu C, Pankow JS, Demerath EW, Bressler J, Fornage M, Grove ML, Mosley TH, Hicks C, North K, et al. Evaluation of microarray-based DNA methylation measurement using technical replicates: the Atherosclerosis Risk In Communities (ARIC) Study. *BMC Bioinformatics*. 2014;15:312.
 40. Short JD, Tavakoli S, Nguyen HN, Carrera A, Farnen C, Cox LA, Asmussen R. Dyslipidemic diet-induced monocyte “priming” and dysfunction in non-human primates is triggered by elevated plasma cholesterol and accompanied by altered histone acetylation. *Front Immunol*. 2017;8:958.
 41. Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW. DNA methylation as a biomarker for cardiovascular disease risk. *PLoS One*. 2010;5:e9692.
 42. Lund JB, Li S, Baumbach J, Svane AM, Hjelmberg J, Christiansen L, Christensen K, Redmond P, Marioni RE, Deary IJ, et al. DNA methylome profiling of all-cause mortality in comparison with age-associated methylation patterns. *Clin Epigenetics*. 2019;11:23.
 43. Jia X, Sun W, Hoogeveen RC, Nambi V, Matsushita K, Folsom AR, Heiss G, Couper DJ, Solomon SD, Boerwinkle E, et al. High-sensitivity troponin I and incident coronary events, stroke, heart failure hospitalization, and mortality in the ARIC Study. *Circulation*. 2019;139:2642–2653.
 44. Licher S, Ahmad S, Karamujić-Čomić H, Voortman T, Leening MJG, Ikram MA, Ikram MK. Genetic predisposition, modifiable-risk-factor profile and long-term dementia risk in the general population. *Nat Med*. 2019;25:1364–1369.
 45. Anderson GL, Cummings SR, Freedman LS, Furberg C, Henderson MM, Johnson SR, Kuller LH, Manson JE, Oberman A, Prentice RL, et al. Design of the Women's Health Initiative clinical trial and observational study. *Control Clin Trials*. 1998;19:61–109.
 46. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. *Am J Epidemiol*. 1979;110:281–290.
 47. Joeannes R, Ying S, Huan T, Johnson AD, Raghavachari N, Wang R, Liu P, Woodhouse KA, Sen SK, Tanriverdi K, et al. Gene expression signatures of coronary heart disease. *Arterioscler Thromb Vasc Biol*. 2013;33:1418–1426.
 48. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol*. 2012;41:1576–1584.
 49. Taylor AM, Pattie A, Deary IJ. Cohort profile update: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol*. 2018;47:1042–1042r.

SUPPLEMENTAL MATERIAL

Data S1.

Supplemental Methods

Women's Health Initiative

WHI methylation data came from the BA23 ancillary study, a combined case-control and pseudo case-cohort sampling of 2129 women from the Women's Health Initiative study. WHI is a larger prospective cohort beginning in 1993 that included over 160,000 postmenopausal women from across the United States⁴⁵. Included subjects had no self-reported CVD at baseline, and cases were chosen based on incident centrally adjudicated angina, revascularization, or CHD event during follow-up. Inclusion criteria for methylation measurement resulted in an oversampling of African American and Hispanic participants. Blood samples used for measurement of DNA methylation and clinical biochemistry were taken at baseline. Data are available in the dbGaP public repository (accession: phs000200.v11.p3; downloaded on September 27, 2017).

Framingham Heart Study Offspring Cohort

FHS methylation data came from a substudy of the Framingham Heart Study that measured DNA methylation in 2726 subjects from the Offspring Cohort. The Framingham Offspring Cohort was originally established in 1971 to follow 5209 children of the original Framingham Heart Study participants and their spouses⁴⁶. Fasting blood samples for both methylation and clinical biochemistry were collected from participants at Exam 8, which took place from 2005-8. Blood samples were also provided for clinical biochemistry measurements in previous exams, constituting the "past exposures" examined here. Data are available in the dbGaP public repository (accession: phs000007.v29.p10; downloaded on September 27, 2017). Adjudicated cardiovascular event data was collected through 2015, and events were defined here as any of: myocardial infarction, angina pectoris, stroke (approximately 90% being ischemic), or death from CHD (Framingham event codes 1-29). FHS methylation data were collected in two primary batches in two centers – one in subjects from a nested case-control for CVD measured at Johns Hopkins University (FHS-JHU)⁴⁷, and the other in a larger set of remaining Framingham Offspring participants measured at the University of Minnesota (FHS-UM).

Lothian Birth Cohorts

The Lothian Birth Cohorts consist of two birth cohorts (born in 1921 and 1936) established in the Lothian region of Scotland^{48,49}. Only the 1936 cohort was analyzed here. Blood samples were collected in three waves starting in 2004, with our primary analyses here focusing on samples from Wave 1 (2004-2007). Cardiovascular outcomes were defined as general CVD or stroke determined at each wave, and event times for survival models were approximated based on the time between Wave 1 and the wave at which the event was reported. LBC data are accessible through the European Genome-phenome Archive (accession: EGAD00010000604).

REGICOR

The REGICOR dataset analyzed here consisted of a nested case-control for myocardial infarction within the larger REGICOR (REgistre Gironí del COR) cohort from the Girona Province in Catalonia (Spain). Whole blood samples were collected from 391 total participants, with those from cases generally collected within 24 hours of the event. Characteristics for this population are available in Supp. Table S3.

Table S1. MRS performance in held-out FHS subset without past CVD events.

Model	HR per s.d.	MRS	p
Unadjusted ¹		1.55	1.9e-08
Basic ²		1.27	7.3e-03
Plus risk factors ³		1.30	6.0e-03
FRS only ⁴		1.37	7.7e-05

¹ No covariates

² Adjusted for age, sex, and estimated cell type fractions

³ Additionally adjusted for BMI, LDL, HDL, SBP, diabetes status, and current smoking

⁴ Adjusted for Framingham Risk Score only

Table S2. MRS stability as evaluated by using multiple within-subject measurements. Generic ICC heuristics for reference: 0-0.5 = poor, 0.5-0.75 = moderate, 0.75 - 0.9 = good, 0.9-1 = excellent.

Cohort	Group type	# of pairs/groups	ICC
FHS	Duplicates	26	0.85
LBC36	Samples over multiple visits	758	0.68
LBC36	Samples over subsequent visits (Wave 1 & 2)	758	0.69
LBC36	Samples over longer time frame (Wave 1 & 3)	758	0.61

Table S3. Description of REGICOR myocardial infarction nested case-control population (continuous variables presented as: mean (standard deviation))

Sample size	391
Prior myocardial infarction	50.1%
Ancestry (% European)	100%
Age	63.2 (6.9)
Sex (% female)	48.6
Smoking	21.5%
Body mass index	28.5 (4.8)
Hypercholesterolemia	53.0%
Hypertension prevalence	57.1%
Diabetes prevalence	24.7%

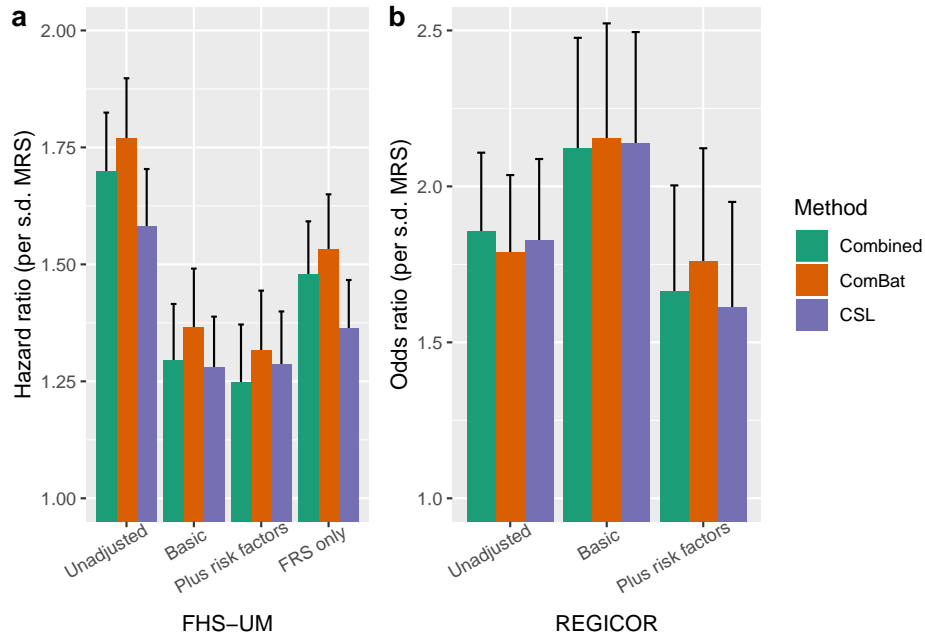
Table S4. Validation of Framingham Risk Score.

Study	HR per s.d. MRS	p
WHI	1.50	4.7e-61
FHS-JHU	1.40	9.9e-06
FHS-UM	1.63	8.5e-22
LBC	1.01	9.2e-01

Table S5. Risk factor-stratified MRS performance in the REGICOR dataset.

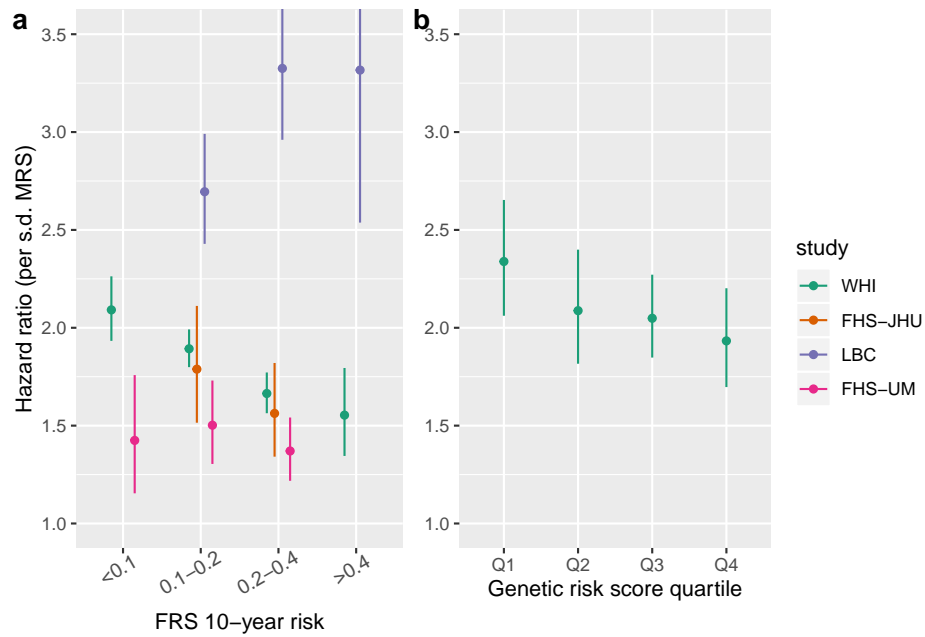
Risk factor group	OR per s.d. MRS [95% CI]	N
Q1	4.49 [1.64-12.28]	119
Q2	1.17 [0.67-2.04]	90
Q3	2.58 [1-6.68]	60
Q4	1.2 [0.31-4.59]	55

Figure S1. Comparison of modeling approaches.



Performance metrics are shown as a function of the test dataset, either FHS-UM (a) or REGICOR (b), and the covariate adjustment. Performance is quantified by either hazard ratio from Cox models (a) or odds ratio from logistic models (b). Covariate sets used for adjustment for models named here are identical to their descriptions for the regression models presented above. Error bars represent standard errors for the hazard ratio or odds ratio estimates.

Figure S2. Interactions of MRS with other biomarkers of CVD risk.



a) Hazard ratios for the MRS within subsets of 10-year generalized CVD risk according to the Framingham Risk Score. b) Hazard ratios for the MRS within quartiles of a genetic cardiovascular risk score (in white participants only for WHI). Hazard ratios are estimated using the final MRS, which was trained using each of these datasets. Stratum-specific Cox regressions were adjusted for age, sex, and estimated cell subtype fractions. Estimates for strata with less than 25 incident events are not shown. Error bars represent standard errors for the hazard ratio estimates (cut off above in panel (a) for ease of visualization of other points).