Supplementary Info: Cross-study learning for the epigenomic prediction of cardiovascular risk

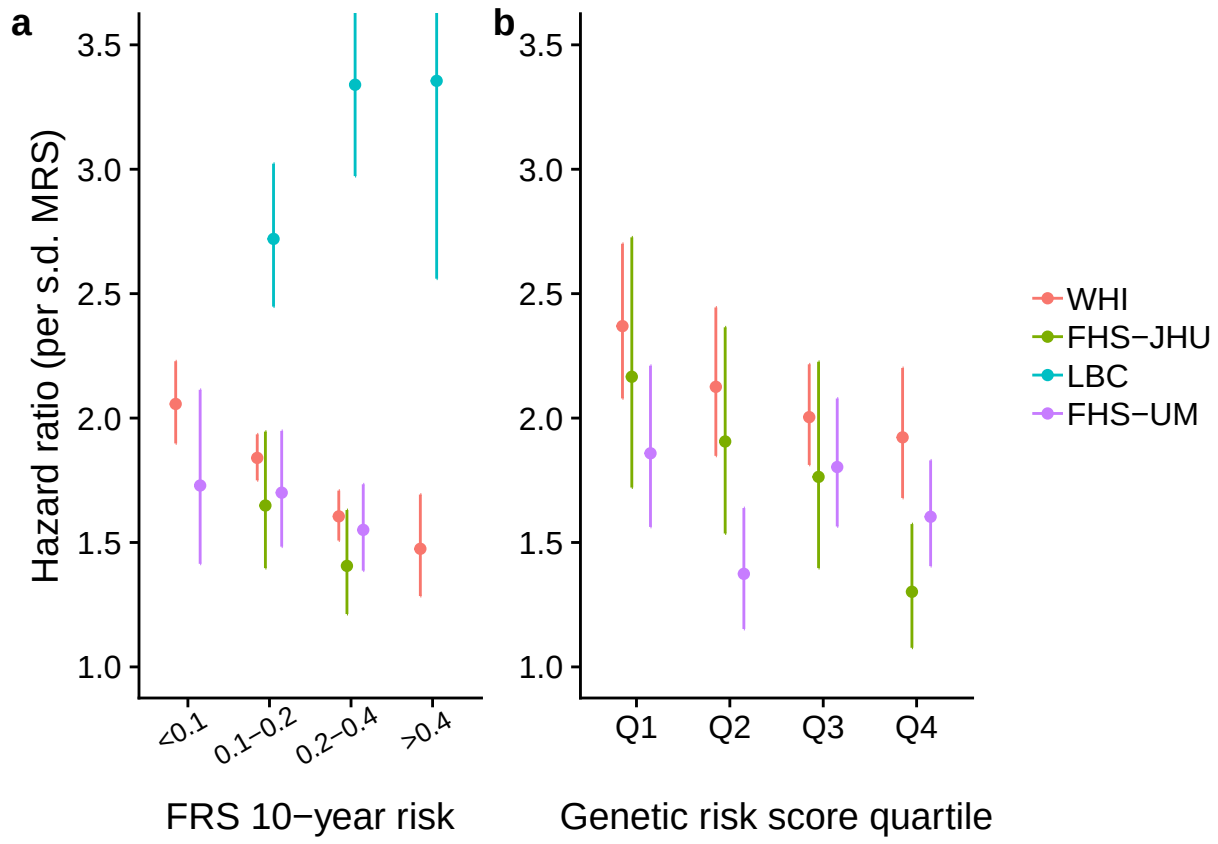

Figure S1: Interactions of MRS with other biomarkers of CVD risk. a) Hazard ratios for the MRS within subsets of 10-year generalized CVD risk according to the Framingham Risk Score. b) Hazard ratios for the MRS within quartiles of a genetic cardiovascular risk score (in white participants only for WHI). 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).

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

Model	HR per s.d. MRS	p
Unadjusted[1]	1.60	4.3e-10
Basic[2]	1.32	8.2e-04
Plus risk factors[3]	1.31	5.0e-03
FRS only[4]	1.41	1.8e-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 enrichment for transcription factor binding motifs: HOMER results

Motif	Consensus	BH q-value	MRS Coverage	Background Coverage
Foxf1(Forkhead)/Lung-Foxf1-ChIP-Seq(GSE77951)/Homer	WWATRTAAACAN	0.169	6.41%	4.27%
Pit1(Homeobox)/GCrat-Pit1-ChIP-Seq(GSE58009)/Homer	ATGMATATDC	0.169	6.16%	4.11%
HEB(bHLH)/mES-Heb-ChIP-Seq(GSE53233)/Homer	VCAGCTGBNN	0.169	28.69%	24.53%
Smad4(MAD)/ESC-SMAD4-ChIP-Seq(GSE29422)/Homer	VBSYGTCTGG	0.169	20.34%	16.75%
c-Myc(bHLH)/LNCAP-cMyc-ChIP-Seq(Unpublished)/Homer	VCCACGTG	0.169	8.44%	6.13%
Ascl1(bHLH)/NeuralTubes-Ascl1-ChIP-Seq(GSE55840)/Homer	NNVVCAGCTGBN	0.169	22.87%	19.36%
CLOCK(bHLH)/Liver-Clock-ChIP-Seq(GSE39860)/Homer	GHCACGTG	0.169	7.76%	5.70%
E2A(bHLH)/proBcell-E2A-ChIP-Seq(GSE21978)/Homer	DNRCAGCTGY	0.169	22.53%	19.17%
NPAS(bHLH)/Liver-NPAS-ChIP-Seq(GSE39860)/Homer	NVCACGTG	0.169	17.22%	14.28%
bHLHE40(bHLH)/HepG2-BHLHE40-ChIP-Seq(GSE31477)/Homer	KCACGTGMCN	0.169	5.06%	3.49%
Max(bHLH)/K562-Max-ChIP-Seq(GSE31477)/Homer	RCCACGTGGYYN	0.169	8.52%	6.48%
Ascl2(bHLH)/ESC-Ascl2-ChIP-Seq(GSE97712)/Homer	SSRGCAGCTGCH	0.169	17.72%	14.87%
Lhx2(Homeobox)/HFSC-Lhx2-ChIP-Seq(GSE48068)/Homer	TAATTAGN	0.169	7.00%	5.20%
FOXK1(Forkhead)/HEK293-FOXK1-ChIP-Seq(GSE51673)/Homer	NVWTGTTTAC	0.175	7.43%	5.63%
bHLHE41(bHLH)/proB-Bhlhe41-ChIP-Seq(GSE93764)/Homer	KCACGTGMCN	0.197	15.61%	13.11%

Table S3: 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	0.81
LBC36	Samples over multiple visits	758	0.68
LBC36	Samples over subsequent visits (Wave 1 & 2)	758	0.68
LBC36	Samples over longer time frame (Wave 1 & 3)	758	0.64

Table S4: Description of REGICOR mycardial infraction nested case-control population

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)
LDL cholesterol	127 (26)
HDL cholesterol	50.0 (10.5)
Systolic blood pressure	135 (18)
Diabetes prevalence	24.7%

Table S5: Validation of Framingham Risk Score

Study	HR_per_SD	p
WHI	1.50	4.7e-61
FHS-JHU	1.42	4.8e-06
FHS-UM	1.62	3.7e-21
LBC	0.97	6.6e-01

Table S6: Validation of genetic risk score

Study	OR_per_SD	p
WHI	1.28	0.0000011
FHS-JHU	1.09	0.3779815
FHS-UM	1.04	0.5575529