Blood Leukocyte DNA Methylation Predicts Risk of Future Myocardial Infarction and Coronary Heart Disease

This is a pre print version of the following article:

Original Citation:

Availability:
This version is available http://hdl.handle.net/2318/1725865 since 2020-02-28T16:12:28Z

Published version:
DOI:10.1161/CIRCULATIONAHA.118.039357

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)
Cohort Supplementary information

1. Description and Methods for each cohort

Atherosclerosis Risk in Communities Study (ARIC)
The ARIC Study is a prospective longitudinal investigation of the development of atherosclerosis and its clinical sequelae in which 15,792 individuals aged 45 to 64 years were enrolled at baseline in 1987-1989 from four communities in the United States: Forsyth County, North Carolina; Jackson, Mississippi (African-Americans only); the suburbs of Minneapolis, Minnesota; and Washington County, Maryland. Four examinations were carried out at three-year intervals (exam 1, 1987-1989; exam 2, 1990-1992; exam 3, 1993-1995; exam 4, 1996-1998.) A fifth clinical examination was completed in 2011-2013. Subjects were contacted annually to update their medical histories between examinations. Written informed consent was provided by all study participants, and the study design and methods were approved by institutional review boards at the collaborating medical institutions: University of Mississippi Medical Center Institutional Review Board (Jackson Field Center); Wake Forest University Health Sciences Institutional Review Board (Forsyth County Field Center); University of Minnesota Institutional Review Board (Minnesota Field Center); and the Johns Hopkins School of Public Health Institutional Review Board (Washington County Field Center).

Incident cardiovascular events occurring between the time of the methylation measurement at either exam 2 or exam 3 and December 31, 2011 were identified during annual phone contact with study participants and by surveillance of local hospital discharge records and death records. Incident CHD was validated by physician review. Bisulfite converted DNA extracted from peripheral blood leukocytes was hybridized to the Illumina HumanMethylation450 BeadChip (Illumina Inc., San Diego, CA), following the Illumina HD Methylation protocol (Illumina Inc., San Diego, CA). Individuals were excluded from the analyses if the pass rate for the DNA sample for the participant was < 99% (probes with a detection p-value <0.01/all probes on the array). CpG sites were not analyzed in this study if more than 5% of the samples showed a detection p-value > 0.01, or if the average detection p-value was > 0.01 on the Y chromosome for males. Methylation values were normalized using the Beta Mixture Quantile dilation (BMIQ) method for type I/type II bias correction. Cox proportional hazards models were adjusted for age, sex, years of education, smoking status (current, former, or never), field center, clinical examination (exam 2 or exam 3), body mass index (BMI), technical variables including plate number, column number, and row number, four principal components from the Illumina Infinium HumanExome BeadChip genotype array to account for potential confounding by genetic ancestry, ten surrogate variables (to control for unmeasured batch effects), and cell type proportion estimates. The proportions of neutrophils, lymphocytes, monocytes, eosinophils, and basophils were imputed using the Houseman method based on the measured differential cell counts available for a subset of ARIC participants at exam 2 (n = 175).

Cardiovascular Health Study (CHS)
The CHS is a population-based cohort study of risk factors for coronary heart disease and stroke in adults ≥65 years conducted across four field centers. The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists; subsequently, an additional predominantly African-American cohort of 687 persons was enrolled for a total sample of 5,888. DNA methylation was measured on 200 European ancestry and 200 African-American ancestry participants. The samples were randomly selected among participants without presence of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack at study baseline or lack of available DNA at study year 5. CHS was approved by institutional review committees at each field center and individuals in the present analysis had available DNA and gave informed consent including consent to use of genetic information for the study of cardiovascular disease.

Between enrollment in the CHS study and 1998–1999, participants were seen in the clinic annually, and contacted by phone at 6-month intervals to collect information about hospitalizations and potential cardiovascular events. Subsequently, telephone calls occurred every 6 months. Hospital records of all potential events were obtained, and all events were adjudicated by a CHS Events Committee. MI was indicated by symptoms of coronary ischemia, elevated serum levels of troponin and cardiac enzymes, and specified electrocardiographic changes. Deaths were identified by a review of obituaries, medical records, death certificates, and the Centers of Medicare and Medicaid Service health care utilization database for hospitalization, and from household contacts; 100% complete follow-up for mortality status was achieved. Deaths from cardiovascular causes included deaths by coronary heart disease,
heart failure, peripheral vascular disease, or cerebrovascular disease. Methylation measurements were performed using the Infinium HumanMethylation450 BeadChip. Quality control was performed in in the minfi R package. Samples with low median intensities of below 10.5 (log2) across the methylated and unmethylated channels, samples with a proportion of probes falling detection of greater than 0.5%, samples with QC probes falling greater than 3 standard deviation from the mean, sex-check mismatches, or failed concordance with prior genotyping were removed. In total, 11 samples were removed for sample QC resulting in a sample of 191 European-ancestry and 198 African-American samples. Methylation values were normalized using the SWAN quantile normalization method. White blood cell proportions were estimated from the methylation data using the Houseman method. Analyses were conducted with penalized cox regression, adjusting for age, sex, years of education, smoking status (never/former/current), BMI, as well as cell count composition as estimated using the Housmann method. Analyses were stratified by European or African-American ancestry. Analyses of African-Americans were additionally adjusted for two genetic principal components.

long-tErm follow-up of antithrombotic management Patterns In acute CORonary syndrome patients (EPICOR)
EPICOR is a case-control study nested within the Italian section of the EPIC cohort (about 50,000 volunteers) EPIC-Italy subjects included in EPICOR were recruited between 1992-1998 in 4 Italian centers, Turin and Varese (Northern Italy), Naples (Central-Southern Italy), and Ragusa (Southern Italy). All subjects were healthy at enrollment and have been followed up for major events (mainly cancer, cardiovascular disease, and other chronic diseases of the adulthood). This current analysis included 584 individuals healthy at the time of recruitment: 292 who developed non-fatal Myocardial Infarction (MI) at follow-up (average Time to Disease (TTD) 6.92 ± 3.77 years), and 292 controls matched to cases by age (± 2.5 years), gender, Center of recruitment (Sampling method: Incidence density method). EPICOR MI cases were identified from hospital discharge databases of clinical reports identifying events based on International Classification of Diseases (ICD)-9 codes. Cases were cross-checked with mortality files to identify fatal and nonfatal cases. Subjects with CHD at EPIC cohort entry were excluded from this study. DNA methylation was measured in bisulfite-converted DNA from white blood cells using the Infinium HumanMethylation450 BeadChip (Illumina). Methylation signals (beta-values) were excluded if they had detection p-value ≥ 0.05; CpG loci were excluded if they had detection p-value ≥ 0.05 in more than 1% of the assayed samples. Samples with a global call rate ≤ 99% were excluded as well. Normalization was performed on raw methylation data, including color bias adjustment, Quantile Normalization, and BMIQ for type I/type II bias correction. Penalized logistic regression model adjusting for matching variables: age (± 2.5 years), gender, and center of recruitment. Additional covariates included in the model are education (numeric in years), BMI (numeric), smoking status (categorical: current, former, never), and cell type proportions estimated using the Housmann method.

The Framingham Heart Study (FHS)
The Framingham Heart Study (FHS) offspring cohort, as previously described, is a community-based cohort recruited in 1971 and included the offspring (and their spouses) of the FHS original cohort. The eligible sample for this investigation was drawn from participants in the FHS offspring cohort who gave consent for genomic studies and had DNA methylation assays completed on whole blood samples collected at the eighth examination cycle (2005-2008). At each examination, participants provided fasting blood samples and had a standardized medical examination, including obtaining smoking history, current medication use, and height and weight. Details are available at http://www.framinghamheartstudy.org/. Participants provided written informed consent at the time of each examination visit. The study protocol was approved by the Institutional Review Board at Boston University Medical Center (Boston, MA). CHD and MI events were prospectively ascertained via periodic health updates with the participants and surveillance of the local medical centers. All events were adjudicated by a panel of three physicians, who reviewed participants' medical records, laboratory findings and clinic examination notes. Offspring cohort participants who had prevalent CHD at the eighth examination cycle (2005-2008) were excluded from the analyses (n=261). In FHS, buffy coat preparations were obtained from peripheral whole blood samples. DNA was extracted using the Gentra Puregene DNA extraction kit (Qiagen, Venlo, Netherlands) and then underwent bisulfite conversion using the EZ DNA Methylation kit (Zymo Research, Irvine, CA). Samples underwent whole genome amplification, fragmentation, array hybridization, and single-base pair extension. DNA methylation arrays were run in two laboratory batches at the Johns Hopkins Center for Inherited Disease Research (lab batch #1) and University
of Minnesota Biomedical Genomics Center (lab batch #2). The first batch included 576 samples from an earlier cardiovascular disease case-control study and the second batch included 2270 samples from the remainder of the offspring cohort participants. DNA methylation results underwent normalization within laboratory batches using the DASEN methodology implemented in the watermelon package in R (version 3.0.2). We excluded samples with a missing rate >1% at p<0.01 (n=10 for batch #1 and n=35 for batch #2), poor single nucleotide polymorphism (SNP) matching to the 65 SNP control probe locations (n=38 for batch #1 and n=41 for batch #2), and outliers by multi-dimensional scaling techniques (n=25 for batch #1 and n=48 for batch #2). We excluded probes with a missing rate >20% at p<0.01 (n=466 from batch #1 and n=366 from batch #2), as well as probes previously identified to map to multiple locations or to have an underlying SNP (minor allele frequency >5% in European ancestry (EUR) 1000 genomes project data) at the CpG site or within 10 bp of the single base extension (n=42,251). Genotype data were obtained from buffy coat samples, assayed using the MIPS 50K and Affymetrix 500K platforms. After quality control for Hardy-Weinberg Equilibrium, excessive Mendelian errors, and low call rate, imputation using the 1000 Genomes reference panel was conducted, yielding approximately 39 million SNPs. Mixed effect Cox proportional hazards regression models were conducted to test the association of genome-wide DNA methylation with time to CHD or MI event using the coxme package in R. The untransformed DNA methylation beta value was specified as the independent variable of interest adjusted for age, sex, body mass index (kg/m²), smoking (current/former/never), educational attainment, imputed cell counts (obtained via the Houseman method, and surrogate variables to account for unmeasured batch effects as fixed effects. The model was also adjusted as a random effect for a covariance matrix that comprises the familial relatedness, which are obtained by self-report and genetic similarity calculated by identity-by-descent probabilities. The methylation quantitative trait loci (meQTL) model utilized a two-step approach. First, the DNA methylation beta value was residualized with adjustment for age, sex, imputed cell count proportions, measured technical covariates (row, chip, column), and the family structure covariance matrix. Second, the residual of the DNA methylation was specified as the dependent variable, SNP genotype dosage as the independent variable of interest and additionally adjusted for 50 methylation SVAs (surrogate variable analysis) to account for unmeasured technical and batch effects.

The Invecchiare in Chianti study (InCHIANTI)
The InCHIANTI study is a population-based epidemiological study aimed at evaluating the factors that influence mobility in the older population living in the Chianti region in Tuscany, Italy. Briefly, 1616 residents were selected from the population registry of Greve in Chianti (a rural area: 11709 residents with 19.3% of the population greater than 65 years of age), and Bagno a Ripoli (Antella village near Florence; 4,704 inhabitants, with 20.3% greater than 65 years of age). The participation rate was 90% (n=1453), and the age of the participants ranged between 21 and 102 years. The study protocol was approved by the Italian National Institute of Research and Care of Aging Institutional Review and Medstar Research Institute (Baltimore, MD). Assessment of DNA methylation has been previously described. Briefly, DNA was extracted from buffy coat samples and bisulfite converted using Zymo EZ-96 DNA Methylation kit (Zymo Research Corp., Irvine, CA) and methylation status of 485,577 CpG sites was assessed with Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc, San Diego, CA). Quality control filtering and normalization accomplished using the watermelon package.8 Markers were excluded if the bead count was less than 3 in ≥ 5% of samples. Samples and markers where also excluded if ≥ 5% of detection p-values were greater than 0.01. A background adjustment and quantile normalization were applied to the filtered data set; followed by the BMIQ method for type I/type II bias correction. Methylation markers on the X and Y chromosome, as well as markers with potentially cross-reactive probes and probes that may be polymorphic in European populations (AF ≥ .01) were excluded from analyses. Analyses were conducted with penalized Cox regression, adjusting for age, sex, smoking status (never/former/current), categorical education, BMI, and CD8T, CD4T, NK, and monocytes cell type proportions estimated with the Houseman method, as well as complete blood count (CBC) neutrophils and monocytes.

The Kooperative Gesundheitsforschung in der Region Augsburg study (KORA)
The Cooperative Health Research in the Region of Augsburg, Germany Survey 4 (KORA S4) is a population based survey of 4,261 individuals recruited from Augsburg, Germany from October, 1999 – April, 2001. This study was
approved by the ethics committee of the Bavarian Medical Association in Munich, Germany. Non-fatal MI events were reported to the Coronary Events Registry and linked to the cohorts using name and date of birth. For individuals who moved out of the study area, mailed questionnaires and general practitioner notes were used to validate MI. In the event that the questionnaires were not returned the date of move was used as the loss to follow-up date. Fatal MI events found in population registries but not the Coronary Events Registry were classified using the general practitioner’s notes, hospital discharge letter, or ICD-9 code of the underlying cause of death.

For the current analysis, all participants with a myocardial infarction (MI) prior to their baseline exam were excluded. Additionally, only participants with blood samples available for DNA methylation assaying were used for this study. Genome-wide DNA methylation measurement at 485,577 genomic sites was performed using the Infinium HumanMethylation450K BeadChip® (Illumina, San Diego, CA) in 1814 KORA F4, and 1535 KORA S4 samples. The laboratory process has been described previously. Briefly, denaturated single-stranded genomic DNA was subjected to bisulfite treatment using the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA, USA). Bisulfite-converted samples were subjected to whole genome amplification, followed by enzymatic fragmentation and application to the BeadChips. The arrays were fluorescently stained and scanned with the Illumina HiScan SQ scanner. Background correction was performed using the R package minfi, version 1.6.0. Detection p-values were defined as the probability of a signal being detected above the background signal level, as estimated from negative control probes. Data were normalized using quantile normalization (QN) on the raw signal intensities using the R package limma, version 3.16.5. Furthermore, beta-mixture quantile normalization (BMIQ) was applied using the R package watermelon, version 1.0.3. Before analyses all samples with detection p-values > 0.05 for 1% of all probes were removed followed by the removal of all probes with detection p-values > 0.05 for 1% or more of the remaining samples.

Proportions of selected cell types (i.e., granulocytes, monocytes, B cells, CD4+ T cells, CD8+ T cells and natural killer cells) were estimated using the method of Houseman et al. Twenty principal components created from the control probes were used as technical covariates. Models were run under penalized Cox regression model (incident MI). Models were additionally adjusted for age, sex, smoking status, and BMI.

Normative Aging Study (NAS)
The ongoing longitudinal US Department of Veterans Affairs (VA) Normative Aging Study (NAS) was established in 1963 and included men, 21-80 years old and free of known chronic medical conditions at entry. Subsequently participants were invited to medical examinations every three to five years. At each visit, men provided information on medical history, lifestyle, and demographic factors, and underwent physical examinations and laboratory tests. DNA samples were collected from 675 active participants between 1999-2007. All subjects were evaluated at the baseline examination for preexisting cardiovascular disease (including nonfatal ischemic heart disease or stroke) based on medical records and physician exams from the current or past study visits. Hospital records and any other clinical documentation for every report of cardiovascular event were reviewed by a board-certified cardiologist. Experienced research staff coded cardiovascular outcomes using ICD-9 codes. DNA was extracted from buffy coat using the QIAamp DNA Blood Kit (QIAGEN, Valencia, CA), bisulfite converted using the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA), and analyzed using the Infinium HumanMethylation450K BeadChip®. Quality control analysis was performed to remove samples and probes, where >1% of probes or samples, respectively, had a detection p-value > 0.05. The remaining samples were preprocessed using the Illumina-type background correction and normalized with the dye-bias and BMIQ adjustments. For the current analysis, participants were excluded if they had previously developed coronary heart disease at the ‘baseline’ examination (i.e. 1st examination from which DNA methylation data was derived). Analyses were performed with Cox Proportional Hazards models, adjusting for age, years of education, smoking status, BMI. Proportions of selected cell types (i.e., granulocytes, monocytes, B cells, CD4+ T cells, CD8+ T cells and natural killer cells) were estimated using the procedure proposed by Houseman et al., and included as adjustment variables in the model. To avoid any batch effect of DNA methylation levels, technical covariates for plate, position of the chip, and column were included as fixed effect in the analyses.

The Women’s Health Initiative Epigenetic Mechanisms of PM-Mediated CVD (WHI-EMPC)
WHI-EMPC is an ancillary study of epigenetic mechanisms underlying associations between ambient particulate matter (PM) air pollution and cardiovascular disease (CVD) in the Women’s Health Initiative clinical trials (CT) cohort. The WHI-EMPC study population is a stratified, random sample of 2,200 WHI CT participants who were examined between 1993 and 2001; had available buffy coat, core analytes, electrocardiograms, and ambient
concentrations of PM; but were not taking anti-arrhythmic medications at the time. As such, WHI-EMPC is representative of the larger, multiethnic WHI CT population from which it was sampled: n = 68,132 participants aged 50-79 years who were randomized to hormone therapy, calcium / vitamin D supplementation, and / or dietary modification in 40 U.S. clinical centers at the baseline exam (1993-1998) and re-examined in the fasting state one, three, six, and nine years later.\textsuperscript{29,30} Self-reported CHD cases were confirmed by physician-review, classification, and local / central adjudication of medical records. Genome-wide DNA methylation at CpG sites was measured using the Illumina 450K Infinium Methylation BeadChip, quality controlled using the following filters: detection p-values > 0.01 in > 10% of samples, detection p-values > 0.01 or missing in > 1% of probes, and probes with a coefficient of variation < 5%. Analyses were conducted with Firth's penalized likelihood Cox (for incident CHD) or mixed-effects Cox (for incident MI) models, adjusted for age, categorical education (vocational or high school; any college / college graduate; post-graduate, professional, or advanced degree), smoking status (never; former; current), BMI, cell type proportions (derived using method of Houseman et al.\textsuperscript{4}), cohort-specific sampling variables, and batch variables including plate, chip, row, and column, which were treated as random effects for incident MI analyses.

**Integrative genomics and risk of CHD and related phenotypes in the Women’s Health Initiative (WHI-BAA23)**

The WHI-BAA23 Ancillary study included a subsample of participants of the Women's Health Initiative (WHI) study, a national study that began in 1993 which enrolled postmenopausal women between the ages of 50-79 years into either one of two three randomized clinical trials (RCTs). None of these women had CHD at baseline but about half of these women had developed CHD by 2010. Women were selected from one of two WHI large sub cohorts that had previously undergone genome wide genotyping as well as profiling for 7 cardiovascular disease related biomarkers including total cholesterol, HDL, LDL, triglycerides, CRP, creatinine, insulin, and glucose through 2 core WHI ancillary studies. The first cohort is the WHI SNP Health Association Resource (SHARe) cohort of minorities that includes >8000 African American (AA) women and >3500 Hispanic women. The second cohort consists of a combination of European Americans (EA) from the two Hormonal Therapy (HT) trials selected for GWAS and biomarkers in core studies W58 and W63 From these two cohorts, two sample sets were formed. The first (sample set 1) is a sample set of 637 CHD cases and 631 non-CHD cases as of Sept 30, 2010. The second sample set (sample set 2) is a non-overlapping sample of 432 cases of coronary heart disease and 472 non-cases as of September 17, 2012. Adjudication methods for cardiovascular disease outcomes in the WHI have been previously described extensively.\textsuperscript{31} Adjudication of these events occurred for all participants through the clinical trial and intervention phase (1993-2005) as well as the first extension (2005-2010). Adjudication of CHD events during the second extension study (2010-2015) was restricted to a subset of participants in the Medical Record Cohort (MRC). DNA methylation data was derived from DNA extracted from buffy coat of whole blood. Methylation analysis was performed at HudsonAlpha Institute of Biotechnology using the Illumina Infinium HumanMethylation450 BeadChip, following the standard protocol of Illumina methylation assays. Any value with a detection p-value above 0.01 was set to missing, and samples with more than 1.5% missing data were removed. Additionally, CpGs with greater than 10% missing data were removed. Batch normalization was carried out using non-parametric empirical Bayes normalization using the Combat\textsuperscript{32} function in R. Probes using Infinium I chemistry were normalized separately from those using Infinium II chemistry, and a chemistry correction was applied after normalization.\textsuperscript{33} Models run under penalized logistic regression, adjusting for age, BMI, 3-category smoking (never/former/current), and categorical education. We additionally included as covariates in the models cell-count estimation using Houseman’s method (i.e., granulocytes, monocytes, B cells, CD4+ T cells, CD8+ T cells and natural killer cells) and 20 principal components created from the control probes.
2. Results for the individual-study epigenome-wide analyses in each cohort.

REFERENCES