ARIC Manuscript Proposal #2399

PC Reviewed: 8/12/14 Status: A Priority: 2
SC Reviewed: _________ Status: _____ Priority: ____

1. a. Full Title: Genome-wide analysis of DNA methylation and coronary heart disease (CHD): the Atherosclerosis Risk in Communities (ARIC) Study

b. Abbreviated Title (Length 26 characters): Methylation and CHD

2. Writing Group: Jan Bressler, Myriam Fornage, Li An Lin, Weihua Guan, James Pankow, Ellen Demerath, Megan Grove, Kari North, Eric Boerwinkle (other investigators welcome)

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. __JB__ [please confirm with your initials electronically or in writing]

First author: Jan Bressler, Ph.D.
Address: Human Genetics Center
School of Public Health
University of Texas Health Science Center at Houston
1200 Pressler Street
Houston, TX 77030

Phone: (713) 500-9919   Fax: (713) 500-0900
E-mail: jan.bressler@uth.tmc.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).
Name: Eric Boerwinkle, Ph.D.
Address: Human Genetics Center
Division of Epidemiology, Human Genetics, and Environmental Sciences
University of Texas School of Public Health
1200 Pressler Street
Houston, TX, 77030

Phone: 713-500-9816   Fax: 713-500-0900
E-mail: Eric.Boerwinkle@uth.tmc.edu
3. **Timeline:**

Statistical analyses: July 2014 – September 2014  
Manuscript preparation: October 2014 – December 2014  
Manuscript revision: January 2015  
Manuscript submission: February 2015

4. **Rationale:**

Heart diseases and cerebrovascular disease were the first and fourth leading causes of death in the United States in 2010, and can develop as a result of the deposition and rupture of atherosclerotic plaques in the coronary and carotid arteries. DNA methylation is a chemical modification of genomic DNA in which a methyl group is added to the carbon-5 position of cytosine when it is directly followed by guanine in the DNA sequence, a configuration known as a CpG dinucleotide. Most often, DNA methylation found in the vicinity of a gene promoter is associated with repression of gene transcription. However, results published by the Encyclopedia of DNA Elements (ENCODE) Project, a large-scale effort to catalog functional elements in the human genome, suggest that regions that are variably methylated between different cell types and tissues are more often found in the gene body. Several lines of evidence suggest that alterations in DNA methylation patterns may be involved in atherosclerosis (for recent reviews see Aslibekyan S et al., Baccarelli A et al., and Ordovas JM and Smith CE et al.). DNA methyltransferases use S-adenosyl methionine (SAM) as a methyl group donor to modify CpG dinucleotides after a series of reactions that depend on the transmethylation of homocysteine to methionine. If the supply of methyl groups is inadequate, the abnormally high level of homocysteine leads to an elevation in the amount of S-adenosyl homocysteine (SAH), a direct inhibitor of methyltransferases. Hyperhomocysteinemia has been associated with an increased risk of cardiovascular disease in two independent meta-analyses.

There have been conflicting results from studies of global DNA methylation patterns and their relationship to homocysteine levels measured in blood. While a higher level of DNA methylation in the peripheral blood leucocytes of patients with prevalent CHD has been observed in the context of elevated homocysteine, lower methylation of blood Long Interspersed Nucleotide Elements (LINE-1) repeats has been shown to predict both baseline and incident ischemic heart disease and stroke but was not associated with homocysteine levels in a cohort of elderly individuals from the Boston-area Normative Aging Study. In a study of male patients with uremia and hyperhomocysteinemia, peripheral blood mononuclear cells showed decreased methylation when compared to those from normal controls. Nearly half of the men showed a shift from monoallelic to biallelic transcription of the maternally expressed H19 gene that was reversed after administration of folate to reduce plasma homocysteine concentration. These results suggest that the methylation status of leucocytes may be responsive to environmental factors such as diet.

Gene specific hypermethylation has also been reported for human atherosclerotic aortas and smooth muscles cells *in vivo*, and in peripheral blood from patients with CHD.
In an early investigation, Post et al. demonstrated that the estrogen receptor-α (ESR1) CpG island was more highly methylated in coronary atherosclerotic plaques when compared to normal samples taken from the proximal aorta. The results of a similar analysis in smooth muscle cells showed that there was a graded increase in methylation of the monocarboxylate transporter 3 (MCT3) gene that was positively correlated with the extent of reduction of MCT3 expression, severity of the atherosclerotic lesions, and impaired lactate transport. More recently, an association between higher methylation of two candidate genes (insulin (INS) and GNAS antisense RNA 1 (GNAS-ASI)) in leucocytes and incident myocardial infarction has been observed among women enrolled in the Prospective Study of Pravastatin in the Elderly at Risk (PROSPER) trial. Methylation of the phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma) (PLA2G7) promoter was also identified as a gender-specific risk factor for CHD using blood DNA samples from Han Chinese women in a cross-sectional analysis.

The availability of the Illumina Infinium HumanMethylation450 BeadChip that can interrogate more than 450,000 CpG sites, and offers coverage of 99% of all RefSeq genes including 5', 3', and promoter regions, CpG islands, and CpG shores, has facilitated epigenome-wide association studies (EWAS) in large population-based cohorts. The aim of the proposed study is to investigate the association between interindividual variation in DNA methylation and prevalent and incident CHD using already collected data from 2,950 African-American participants in the ARIC study. CHD was one of the phenotypes specified in the “umbrella” proposal titled “Genome-wide methylation analyses of cardiovascular disease (CVD) and its risk factors” (MS #1928), and it was anticipated that individual manuscript proposals would be submitted as analysis of each trait was initiated. In addition, opportunities for replication and/or contribution to meta-analyses will be sought in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Epigenetics Working Group.

5. Main Hypothesis/Study Questions:

1. Epigenome-wide variation in DNA methylation is associated with prevalent CHD.
2. Epigenome-wide variation in DNA methylation is associated with incident myocardial infarction and CHD.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Study design: Analysis of DNA methylation by Illumina Infinium HumanMethylation450 (HM450) BeadChip (Illumina Inc., San Diego, CA) has been adapted for methylation profiling by exploiting technology previously developed for SNP genotyping. The assay requires using sodium bisulfite to convert unmethylated cytosine residues to uracil under conditions in which 5-methylcytosine remains unreactive. This difference is then
detected as a C/T nucleotide polymorphism at each CpG site. Data analysis is performed using proprietary Genome Studio software (Illumina Inc.) that includes algorithms to obtain the relative level of methylation as a beta value, a continuous variable ranging between 0 and 1. The beta value is calculated as the ratio of methylated signal intensity to the sum of methylated and unmethylated signals for each probe after first subtracting the background signal intensity of negative controls included on the array. Several different controls were included on each 96-well plate of DNA samples that was processed for hybridization to the HM450 arrays. These consisted of four replicate DNAs, a commercially available completely methylated positive control DNA (Universal Methylated Human DNA Standard, Zymo Research Corporation, Irvine CA), and a whole-genome amplified DNA sample from an ARIC study participant used as an unmethylated negative control. A series of blind duplicates was also analyzed on the arrays in accordance with ARIC study policy.

Inclusion/exclusion criteria: DNA extracted from blood collected from African-American study participants at either visit 2 (n = 2,504) or visit 3 (n = 401) was included on the array if the individual had not restricted use of their DNA, if there was 1 ug or more of DNA available for methylation analysis, and if there was genome-wide genotyping data available either using the Affymetrix Human SNP Array 6.0 the ITMAT-Broad-CARe (IBC) Illumina iSelect custom array.

Quality control: A total of 2,950 study participants had DNA samples hybridized to the HumanMethylation450 arrays. An average detection p-value across all samples <0.01 was used as a measure of performance for each probe on the array, and was determined using an algorithm incorporated in Genome Studio software to assess whether its beta value was above background level defined by negative control probes. Individuals will be excluded from analysis if a pass rate for the DNA sample for the study participant was less than 99% (probes with a detection p-value <0.01/all probes on the array) (n=32). CpG sites on the autosomes and X chromosome will not be analyzed if more than 1% of samples showed a detection p-value >0.01 (n=9,399), or on the Y chromosome if the average detection p-value >0.01 for males (n=370). A total of 473,788 CpG sites remain for methylation analysis. Cross-hybridizing probes and probes containing SNPs will be flagged to aid in interpretation of results.

Normalization: To achieve genome-wide coverage, the HumanMethylation450 BeadChip includes two different assay designs that differ in their chemistry, distribution, and assumptions about the independence of neighboring sites. Infinium II probes (72% of the total) were designed for regions of low CpG density (1-3 CpGs/bead) and measure each CpG dinucleotide individually, while the Infinium I assay (>3 CpGs/bead) assigns the same beta value to all CpG sites included in the probe sequence under the expectation that methylation status will be correlated. The Infinium II assay is less sensitive for the detection of extreme methylation values than the Infinium I assay and has greater average variance between technical replicates, creating bias towards type I probes when ranking differentially methylated regions. Subset quantile within array normalization (SWAN) will be used in these analyses to reduce technical variation within and between arrays.
Statistical analysis: The association between locus-specific genome-wide methylation and prevalent CHD at visit 2 or visit 3 will be analyzed using linear mixed effects models with beta value as the dependent variable. Batch effects (chip) will be specified as random effects, and age, sex, field center, visit, plate, chip row and 10 principal components from the Illumina Infinium HumanExome BeadChip genotype array to account for ancestry will be entered as fixed effects in the primary analysis. A second model will incorporate body mass index, current smoking, diabetes, hypertension, high density lipoprotein cholesterol, and low density lipoprotein cholesterol as additional covariates. All covariates will be chosen from the clinical examination corresponding to the time of measurement of DNA methylation. Cox proportional hazard models will be used to assess the relationship between DNA methylation level and incident myocardial infarction and CHD, with the baseline for assessment of time to event considered as the date of the clinical visit at which the peripheral blood sample was drawn. In the Cox proportional hazards models, methylation level will be analyzed as an independent variable. Since DNA methylation patterns can differ between leukocyte subtypes present in peripheral blood, the association between beta values and CHD could be confounded by differences in the distribution of cell types between samples, a recently reported method to estimate subtype proportions will be implemented, and the resulting data will also be entered as covariates in the regression models. Individuals will be excluded from analysis if they are missing CHD variables and other covariates. An a priori threshold for statistical significance will be set at $1 \times 10^{-7}$ after application of the Bonferroni correction for multiple comparisons (0.05/473,788 probes tested).

7.a. Will the data be used for non-CVD analysis in this manuscript?  ____ Yes  ____ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = “CVD Research” for non-DNA analysis, and for DNA analysis RES_DNA = “CVD Research” would be used?  ____ Yes  ____ No

(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript?  ____ Yes  ____ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = “No use/storage DNA”?  ____ Yes  ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.csc.unc.edu/ARIC/search.php
10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

MS#1928 Bressler J et al. Genome-wide methylation analysis of cardiovascular disease (CVD) and its risk factors

MS#1929 Pankow J et al. Genome-wide DNA methylation profiling in peripheral blood: quality control and association with demographic characteristics.

MS#2106 Demerath E et al. Epigenome-wide association study of obesity traits in African American adults: The Atherosclerosis Risk in Communities (ARIC) Study

MS#2342 Guan W et al. Epigenome-wide association of DNA methylation with smoking in the Atherosclerosis Risk in Communities Study

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data?  _____ Yes  ____x__ No

11.b. If yes, is the proposal

_ _  A. primarily the result of an ancillary study (list number* _________)
_ _  B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* __________ __________ __________)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is your responsibility to upload manuscripts to PUBMED Central whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to Pubmed central.

References: