1.a. Full Title: Epigenome-wide association of DNA methylation with smoking in the Atherosclerosis Risk in Communities Study

b. Abbreviated Title (Length 26 characters):

2. Writing Group: ARIC Epigenetics Working Group

Weihua Guan
Maitreyee Bose
Chong Wu
James Pankow
Ellen Demerath
Jan Bressler
Myriam Fornage
Megan Grove
Eric Boerwinkle

Other interested investigators are welcome to join the writing group.

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

First author: Weihua Guan
Address: Division of Biostatistics
         University of Minnesota
         A460 Mayo Bldg., MMC 303
         420 Delaware St., S.E.
         Minneapolis, MN 55455

Phone: 612-626-4765                    Fax: 612-624-0660
E-mail: wguan@umn.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:
Address:

Phone:                                   Fax:
E-mail:
3. **Timeline:**

Analysis will begin upon approval. We anticipate a draft ready to submit for Publications Committee review in summer 2014.

4. **Rationale:**

Epigenetics is the study of mitotically heritable modifications in chromatin structure (i.e., modifications not involving the germline DNA sequence), and their impact on the transcriptional control of genes and cellular function. Epigenetic variation includes post-translational modifications of histone proteins, non-coding RNAs, and DNA methylation, the latter primarily occurring at cytosine-guanine dinucleotides (CpGs).

Recent technological advances have provided multiple platforms for systematically interrogating DNA methylation variation across the genome (Laird, 2010). This has paved the way for epigenome-wide association studies (EWASs), analogous to genome-wide association studies, to evaluate regions of the genome in which variation in DNA methylation may influence gene expression and ultimately disease risk (Raykan, 2011). Like GWASs, EWASs are based on an agnostic approach in which epigenetic marks can be investigated across the epigenome without prespecifying the genes or regions in which inter-individual variation in DNA methylation is thought to be important for phenotypic variation. However, unlike inherited changes to the genetic sequence, variation in site-specific methylation varies by tissue, stage of development, disease state, and may be impacted by aging and exposure to environmental factors such as diet or smoking (Raykan, 2011).

Arrays to efficiently profile DNA methylation have only recently become commercially available (Laird, 2010). In ARIC, the Illumina HumanMethylation450 BeadChip (HM450) is being used to measure DNA methylation in peripheral blood obtained from ~3000 African American participants at visit 2 (and a small number at visit 3). The array includes 485,577 assays and provides coverage of 98.9% of RefSeq genes with a global average of 17.2 probes per gene region (Bibikova, 2011; Dedeurwaerder, 2011).

A number of studies have identified association between altered methylation level and smoking and smoking associated illness in multiple populations and cell types (Breitling et al., 2011; Joubert et al., 2012; Monick et al., 2012; Wan et al., 2012; Shenker et al., 2013). Furthermore, there is a well-established and graded relation between cigarette smoking and diabetes, cardiovascular disease, lung disease, and other complex diseases (McBride, 1992; Mikhailidis et al., 1998; Lee et al., 2010). Our preliminary analyses in the ARIC study and other studies indicate that associations between current smoking and methylation level at several CpG sites reach array-wide genome-wide significance, suggesting that epigenetic modifications observed in peripheral blood samples may distinguish participants with greater exposure to cigarette smoke. It is possible that site-specific DNA methylation may mediate the effects of smoking exposure on smoking-related outcomes, such as lung disease.
5. **Main Hypothesis/Study Questions:**

DNA methylation is associated with smoking status.

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 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 were also analyzed on the arrays in accordance with ARIC study policy.

Inclusion/exclusion criteria: A cross sectional selection of African American study participants at either visit 2 or 3 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 genotyping array data available from either the Affymetrix Human SNP Array 6.0, the Illumina HumanCVD BeadChip, the Illumina HumanCardio-MetaboChip, or the Illumina HumanExome BeadChip. 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). Probes on the HM450 array for which the detection p-value is >0.01 will not be analyzed. The methylation measures (beta-values) are normalized using the Subset-quantile Within Array Normalization (SWAN) method (Maksimovic et al., 2012).

We will investigate the cross-sectional association of methylation levels and smoking status in the ARIC Study. We will fit a linear mixed-effects (LMM) model of methylation on self-reported smoking status (never/former smoker vs. current smoker), adjusting for age, sex, body mass index, self-reported alcohol consumption (never/former vs. current drinker), the top principal components of ancestry derived from genotype data, visit number, field center, white blood cell count, cell type proportions, and chip row position.
as fixed covariates. Since the cell type proportions were only measured for 175 samples, we will estimate the proportions for rest of samples using the method by Houseman et al (2012). The values of the covariates were from the specific visit at which the participant's sample was obtained. The batch (chip) effects will be modeled as random effects. Probes with at least one single nucleotide polymorphism (SNP) within 50 base pairs of the CpG site and minor allele frequency (MAF) > .05 based on all samples in the 1000 Genomes project (http://www.1000genomes.org/) will be excluded.

We will also investigate association of methylation level with other related traits, including pack-year, pipe smoking, and second-hand smoking exposure, using the same analysis approach.

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

(This file ICTDER03 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?  _x_ Yes  ____ No

Limited to ancestry information obtained from AIMs or GWAS markers

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”?  _x_ 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.cscce.unc.edu/ARIC/search.php

_ _x_ Yes  ______ No

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)?

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data?  _x_ Yes  ____ 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)* _________ _________)

2007.02 (CARE, genotyping in African Americans)

*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:


Lee, P.N. and J.S. Fry, Systematic review of the evidence relating FEV1 decline to giving up smoking. BMC Med, 2010. 8: p. 84.

