ARIC Manuscript Proposal #2386

PC Reviewed: 7/8/14  Status: A  Priority: 2
SC Reviewed: ________  Status: _____  Priority: _____

1.a. **Full Title**: Epigenome-wide association study of age at menarche in African American women: The Atherosclerosis Risk in Communities (ARIC) Study

   b. **Abbreviated Title (Length 26 characters)**: Menarche EWAS

2. **Writing Group**: ARIC Epigenetics Working Group

Writing group members:

Ellen Demerath
Weihua Guan
Fan Wang
Jan Bressler
Myriam Fornage
Li-An Li
Chindo Hicks
Kari North
Yun Li
Yi-hui Zhou
Eric Boerwinkle

Others welcome…

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

**First author**: Ellen W. Demerath
Address: Division of Epidemiology and Community Health
University of Minnesota
1300 South Second Street, Suite 300
Minneapolis, MN 55454

Phone: 612-624-8231  Fax: 612-624-0315
E-mail: ewd@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:
3. Timeline:


4. Rationale:

Age at Menarche: Relationship to Disease and Genetics. Earlier age at menarche is associated with increased risk for breast cancer (Peeters et al., 1994; Kotsopoulos et al., 2005; Rockhill et al., 1998), increased risk of obesity (Biro et al., 2001; Freedman et al., 2002; Freedman et al., 2003) and type 2 diabetes (Lakshman et al., 2008), whereas late menarche may be associated with an increased risk of stroke (Cui et al., 2006) and lower fertility (Presser, 1978).

Identification of genomic variation controlling the development of chronic disease risk factors in childhood, such as early menarche, is important because it may point to effective targets and developmental periods for environmental and behavioral intervention in childhood or adolescence, before disease processes are fully entrenched. African American women now experience significantly earlier sexual development (Chumlea et al., 2003) and carry a much higher burden of obesity, CVD, and diabetes than EA women (Ogden et al., 2012; Cowie et al., 2009; Folsom et al., 1998); therefore, the search for genetic and epigenetic determinants of menarche timing may be of particular value in this population. Age at menarche is relatively strong genetic control, with heritability estimated at ~50% (Anderson et al., 2007; Van den Berg et al., 2007; Towne et al., 2005; Kaprio et al., 1995). A large number of genetic variants associated with age at menarche and related pubertal traits have been identified in European ancestry individuals via GWAS (Elks et al., 2010), among other methods, and we recently showed cross-ethnic replication of the majority of these loci in African American women (Demerath et al., 2013). However, effect sizes were small, and indeed our GWAS meta-analysis in over 18,000 African American women in 15 cohort studies did not reveal any SNPs associated with age at menarche after correction for multiple testing (Demerath et al., 2013).

Epigenetics. Epigenetics is the study of mitotically heritable modifications in chromatin structure (i.e., modifications not involving the underlying 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). Although the placement of epigenetic marks is thought to be largely determined early in development to initiate and maintain cell-type specific gene expression (Armstrong, 2012), DNA methylation and other features of the epigenome are modifiable by post-natal environmental factors such as the nutrient content of the diet (Dolinoy et al., 2006), maternal behavior and stress (Weaver et al., 2004), and environmental pollutants (Baccarelli et al., 2009). Understanding epigenetic variation may therefore help to explain, at least in part, the mechanisms by which environmental factors of public health importance influence genetic susceptibility to a variety of diseases.
Available Epigenetic data in ARIC: Of the different forms of epigenetic modification, DNA methylation is the most extensively studied and best understood. 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). In ARIC, the recently released Illumina 450K Infinium Methylation BeadChip has been used to measure DNA methylation in peripheral blood obtained from approximately 3,000 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) and has been validated against pyrosequencing (Sandoval et al., 2011). The ARIC epigenetics working group has developed QC procedures, compared different analytic approaches, and identified CpG (cytosine-guanine dinucleotide) sites that are influenced by sex, age, and other potential confounding factors (see ARIC Ms Proposals #1928 and #1929).

Relevant Preliminary Evidence in ARIC: This ms proposal follows upon an EWAS of adiposity traits nearing completion (ARIC ms #2106, Demerath et al., “Epigenome wide association study (EWAS) of adiposity traits in African American Adults: The Atherosclerosis Risk in Communities study”), which identified 18 loci associated with BMI in ARIC and replicated in the Framingham Heart Study and the GOLDN cohort. Five of these 18 CpG sites were also replicated in adipose tissue DNA. These findings support continued investigations into factors that influence DNA methylation and may explain the mechanisms by which early life and nutritional factors increase cardiovascular disease and diabetes. Because DNA methylation patterns are thought to be strongly influenced by embryological and early post-natal developmental trajectories, age at menarche is an important potential determinant of methylation variation. In addition, early menarche is a known risk factor for obesity and diabetes. Thus, we hypothesize that an EWAS of age at menarche may share many similar results with our EWAS of BMI and waist circumference, but also will reveal novel DNA methylation variants, including both those influenced in particular by early life (as opposed to adulthood) adiposity, and also those influenced by early life hormonal exposures.

Age at menarche comprises one of the phenotype classes specified in the “umbrella”, over-arching ARIC Ms Proposal related to DNA-methylation-phenotype association studies and lead researchers (MS #1928), and reproductive phenotype associations are being lead by Dr. Demerath.

Current State of Literature on Age at Menarche and DNA methylation. Animal models show bidirectional relationships between pubertal development and methylation changes (Lomniczzi et al., 2013; Strakovsky et al., 2014), but there are currently only a handful of human studies examining the relationship of pubertal or menarche timing to DNA methylation. Average global leukocyte DNA methylation was inversely associated with menarche age in 92 adult members of the New York National Collaborative Perinatal Project cohort (Terry et al., 2008). A study that assessed global leukocyte DNA methylation of genomic repeat elements also found an inverse relationship with menarche in 376 members of the EPIC cohort: for each yearly increase in age at menarche there was a 32% increased risk of low (below the median) methylation (OR:1.32,
95% CI: 1.14-1.53) (Demetriou et al., 2013). That study showed no significant genome-wide global methylation association with age at menarche using the same Illumina Infinium HumanMethylation 450 BeadChip as proposed in our study, suggesting platform-specific influences on global DNA methylation results (Demetriou et al., 2013). To our knowledge, that is the only study that has conducted a CpG-site specific EWAS of age at menarche as we intend to perform; one association was found between age at menarche and methylation beta value at cg01339004, located in the body of the SMAD6 gene (p < 1 x 10^-7), although it was not confirmed by pyrosequencing (Demetriou et al., 2013). Finally, a small study published in 2014 in 130 multiethnic girls using saliva DNA collected in pre-puberty reported an inverse association between CPT19A1 methylation and incident risk of early pubic hair development (age 6-8) (Stueve et al., 2014). CYP19A1 is critical for estrogen biosynthesis, and may influence timing of breast development, and a SNP in CYPT19A1 was among the top results in our GWAS meta-analysis of age at menarche in African American women (Demerath et al., 2013). To our knowledge there are as yet no replicated loci with menarche-associated methylation variation.

Preliminary Analyses and Power: The ARIC study has the largest genome-wide DNA methylation database to our knowledge yet assembled for African Americans. There are approximately 1,600 AA women with concurrent DNA methylation and age at menarche data. Our prior work with EWAS in ARIC provides empirical power estimates for the proposed study. In that study, a one SD increase in BMI was associated with up to a 2% higher or lower CpG probe methylation beta value; significant results were typically in the 0.5% - 1.5% range. Given N=1,600, p=1E-07, and detectable slope (regression coefficient) of between 0.002 and 0.015 (0.2% to 1.5% methylation variation), the Figure below shows we will have >80% power for tests of association between methylation level and quantitative traits to detect a contribution to R^2 of at least 0.009 (0.9% variation explained), and >90% power to detect a contribution of 1.5% or greater. Power for weaker associations accounting will be low.

Part of the rationale for this study is to work with interested international groups with similar Illumina 450K array data and age at menarche (including EPIC, UK Twins Study, ALSPAC Study, etc), and thus we expect that while there may be only a small number of cohort-specific loci identified, there may be additional loci detected via meta-analysis in a larger number of women.

5. Main Hypothesis/Study Questions:

1) We will test whether CpG-site methylation variation at each of the probes passing QC filters is associated with age at menarche (continuous, and categorized as early (<=[11
years), average (12-13 years), and late (>=14 years), independent of potential confounders including sex, age, genetic ancestry markers (PC scores), physical activity, socioeconomic status, smoking status and pack-years of smoking, and diabetes at their DNA visit.

Secondary analysis 1: We will look up the 18 replicated BMI CpG probes identified in our BMI EWAS in the Menarche EWAS results above to test whether there is evidence for pleiotropy (methylation probes with association with both menarche timing and adulthood BMI).

Secondary analysis 2: We will add BMI at age 25 and then weight gain from age 25 to the DNA methylation visit to determine whether the methylation associations identified are mediated by early adulthood BMI and/or adulthood weight gain.

2) We will replicate our findings in the UK Twins, ALSPAC, EPIC and other cohorts using meta-analysis.

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).

We will use data from ARIC visit 1 (age at menarche) and ARIC visit 2 or 3 (DNA methylation, and concurrent confounding factors).

Inclusions/Exclusions:
- Only female subjects are included
- Those with missing DNA methylation data are excluded
- Those missing age at menarche will be excluded
- Any missing covariate data (expected to be low) will be imputed using multiple imputation procedures

Sample Size estimate:
Preliminary analyses show we will have approximately 1,600 African American women with both age at menarche and DNA methylation data.

Identifiers/Demographics
Patient ID, Sex, Date of DNA collection visit (visit 2 or 3), Field center, Age, Education, Household income,

Independent Variable
RHXA01: Age at menarche, years, recalled at baseline visit.

Dependent Variable: Methylation values
Methylation level (beta values, ranging from 0 to 1.0) at each of approximately 485,000 CpG sites will be analyzed as continuous variables. The beta value can be interpreted as the percent of
the time that the CpG is methylated in a given DNA sample. Although across the genome, most CpG sites are either highly methylated (e.g., mean beta near 0.80) or are not highly methylated (e.g., mean beta near 0.15), nonetheless at a given CpG site, variation approximates normality, allowing standard linear regression approaches to be used. An alternative is to use the M-value, which although less easily interpretable, provides better performance in terms of Detection Rate (DR) and True Positive Rate (TPR) for both highly methylated and unmethylated CpG sites (Du et al., 2010). However, for relatively large sample sizes as in ARIC, test statistics are similar for M and beta-values (Zhuang et al., 2012). We will explore use of M values for our models.

Statistical Analysis:
We will utilize mixed effects regression techniques implemented in R to model hypothesized relationships, including a random effect for chip number, and fixed effects for all other variables listed above, as we have done in the past for a number of ongoing EWAS studies being conducted at the University of Minnesota (BMI, smoking, methylation age, and others). For the meta-analysis of multiple cohort data, we will use either sample-size weighted meta-analysis on p values (if the large variation in normalization approaches and other analytic methods utilized across cohorts, e.g., use of M vs beta values, would make regression coefficients noncomparable. Otherwise, fixed effect (or perhaps random effects) meta-analysis methods will be used to provide estimates of association.

Covariates
Covariates will include current smoking status, current alcohol consumption, visit 1 physical activity (Baecke questionnaire leisure and sport indices), visit 1 education, visit 1 household income, and batch effects (e.g., plate#, chip #, chip location). White blood cell counts and imputed white blood cell differential using the Houseman method will be included to adjust for the mixed cell type in buffy coat samples from which the DNA is extracted.

General linear regression model: Methylation beta value = Age at Menarche (continuous) + covars (for each of ~487,000 CpG sites)

Threshold for significance: 1 x 10^-7 (Bonferroni)

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 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?  
_____ 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.cscc.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)?

- Genome-wide DNA methylation profiling in peripheral blood: quality control and association with demographic characteristics (MS1929) Pankow, J et al.
- Genome-wide methylation analyses of cardiovascular disease (CVD) and its risk factors (MS1928) Bressler, J. et al
- Epigenome wide association study (EWAS) of adiposity traits in African American Adults: The Atherosclerosis Risk in Communities study (MS2106). Demerath et al.
- Numerous other proposals dealing with DNA methylation are under review or have been recently approved; This proposal has been circulated to all members of the ARIC Epigenetics WG.

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* __________) ___x___ 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)

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:


