ARIC Manuscript Proposal #3005

PC Reviewed: 7/11/17                      Status: _____                      Priority: 2
SC Reviewed: _________                      Status: _____                      Priority: _____

1.a. Full Title: DNA methylation-based risk score and prediction of all-cause mortality in the Atherosclerosis Risk in Communities Study

b. Abbreviated Title (Length 26 characters):

2. Writing Group:
   Writing group members:

   James Pankow
   Nicholas Roetker
   Myriam Fornage
   Jan Bressler
   Lindsay Fernández-Rhodes
   Ellen Demerath
   Weihua Guan

   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. _JP__ [please confirm with your initials electronically or in writing]

First author: Jim Pankow
Address: Division of Epidemiology and Community Health
         University of Minnesota
         1300 South Second Street, Suite 300
         Minneapolis, MN 55454

   Phone: 612-624-2883                      Fax:
   E-mail: panko001@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:
3. **Timeline:**

Analysis will begin upon approval. We anticipate a draft ready to submit for Publications Committee review by September 2017.

4. **Rationale:**

The epigenome is a heritable set of histone proteins, chemical tags, and microRNA that act to control the structure and functioning of the underlying DNA sequence. The field of epigenetics studies both the factors that modify the epigenome and how these changes affect genetic and cellular functioning. A widely studied epigenetic mechanism is DNA methylation, the process by which methyl groups are added or removed from the DNA sequence, usually at cytosine-guanine dinucleotides (CpGs).

Recent studies have identified CpG sites that can be used to accurately predict chronological age \(^1,^2\). Using methylation data from 82 publicly available datasets consisting of samples of 51 healthy tissue and cell types, Horvath \(^2\) developed a 353 CpG site model to predict age. Using DNA from whole blood of 482 individuals aged 19 to 101 years, Hannum et al. \(^1\) developed a similar age predictor consisting of 71 CpG sites. Both the Horvath and Hannum predictors are associated with all-cause mortality in African Americans from ARIC even after adjusting for chronological age and other risk factors \(^3\). However, the magnitude of the association between these DNA methylation age predictors and mortality is relatively weak.

Using data from a case-cohort sample (N=862) nested within the ESTHER Study, Zhang et al. \(^4\) conducted an epigenome-wide study to identify DNA methylation signatures associated with all-cause mortality. They replicated their findings in a second sample from ESTHER (N=1000, including 231 deaths) and KORA Study (N=1727, including 61 deaths). A total of 58 CpGs were successfully replicated. The authors then used LASSO regression to select 10 independent CpGs from among these 58 replicated CpGs in order to construct a risk score. Hazard ratios for the 10 CpGs ranged from 1.2 to 1.5 per SD of methylation. Interestingly, these 10 CpGs are completely independent of CpGs used in the Horvath and Hannum DNA methylation age predictors. Participants in the highest quartile of methylation for each CpG (or lowest, depending on direction of association) were assigned a score of 1, and scores were then summed across the CpGs (possible range 0 to 10). Comparing individuals with a score of 5 or more to those with a score of 0, the hazard ratios for all-cause mortality were 7.4 (95% CI: 3.7-14.7) in ESTHER and 5.9 (95% CI: 1.5-23.7) in KORA after adjustment for age, sex, smoking status, BMI, physical activity, alcohol consumption, systolic blood pressure, total cholesterol, hypertension, and prevalent CVD, cancer, and diabetes at baseline. There was little attenuation of the relative risks after additional adjustment for the Horvath or Hannum epigenetic age predictors. The risk score was also significantly associated with CVD mortality and cancer mortality in adjusted models, with stronger associations observed for CVD mortality.
The long follow-up and large number of events in ARIC (N=1500 deaths among approximately 4000 with methylation data) provides an opportunity for a well-powered replication of the findings of Zhang et al. Furthermore, it will provide the first data on African Americans, and an opportunity to conduct a more detailed investigation of cause-specific morbidity and mortality.

5. **Main Hypothesis/Study Questions:**

A higher DNA methylation-based risk score developed by Zhang et al. is associated with greater risk of all-cause mortality, independent of potential confounding factors and the Horvath and Hannum DNA methylation age predictors.

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).**

**Design:** Prospective

**Primary endpoint:** All-cause mortality

**Secondary endpoints:** CVD mortality, cancer mortality, non-CVD/non-cancer mortality, incident CVD (aggregate endpoint of CHD, heart failure, and stroke)

**Exposure:** DNA methylation-based risk score constructed using 10 CpGs (cg01612140, cg05575921, cg06126421, cg08362785, cg10321156, cg14975410, cg19572487, cg23665802, cg24704287 and cg25983901)

**Covariates:** Age, sex, race/ethnicity, education, income, BMI, physical activity, smoking (current status and pack-years), current alcohol use, systolic and diastolic blood pressure, use of anti-hypertension medication, LDL cholesterol, HDL cholesterol, triglycerides, use of lipid lowering medication, prevalent diabetes, prevalent cardiovascular disease (CHD, heart failure, or stroke), history of cancer, DNA methylation age predictors (Horvath and Hannum). We will account for potential confounding due to white blood cell heterogeneity by using the modification of the Houseman method to estimate cell distributions.

**Inclusion:** Approximately 2900 African American individuals who had methylation measured on DNA samples collected at visit 2 or visit 3. Approximately 1100 white individuals who had methylation measured on DNA samples collected at visit 2 or visit 3.

**Exclusions:** We will exclude any individuals with a pass rate for all probes on 450K array less than 99% (probes with a detection p-value >0.01/all probes)

**Analysis of primary outcome:** We will run Cox proportional hazards regression models with all-cause mortality as the event of interest and DNA methylation-based risk score as
the independent variable of interest. Follow-up time will be defined as the time from the date of DNA collection to the date of death or December 31, 2014, whichever came first.

**Analysis of secondary outcomes:** We will also analyze cause-specific mortality and incident CVD using Cox regression.

**Subsidiary analyses:**

1. We will analyze all-cause mortality and secondary outcomes stratified by sex. In the study by Zhang et al., associations with all-cause mortality were strong in women than men. Interactions between risk score and sex will be tested by including an interaction term in a combined model.

2. We will analyze all-cause mortality and secondary outcomes stratified by race/ethnicity. Interactions between risk score and race/ethnicity will be tested by including an interaction term in a combined model.

3. We will analyze all-cause mortality and secondary outcomes stratified by socioeconomic position (education and income). Interactions between risk score and socioeconomic position will be tested by including an interaction term in a combined model.

4. To determine if associations are independent of smoking, we will analyze all-cause mortality and secondary outcomes after restricting to never smokers at baseline. Note that of the 10 CpGs used to derive the risk score, 4 were found to be differentially methylated according to smoking exposure.

5. We will repeat analyses of all-cause mortality after excluding individuals with major chronic diseases at baseline (CVD, cancer, diabetes).

6. We will repeat analyses of primary and secondary outcomes using ARIC-derived quartile cutoffs for each CpG site rather than published cutoffs in the ESTHER Study.

7. We will repeat analyses of all-cause mortality and secondary outcomes using a continuous DNA methylation risk score using CpG-specific weights obtained from Zhang et al.

8. To determine whether associations between the DNA methylation-based risk score and all outcomes are dependent on length of follow-up, we will formally test the proportional hazards assumption by including a term reflecting the interaction of the risk score and time in the Cox models.

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?
   _x_ 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”?
   _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)?

1928: Genome-wide methylation analyses of DNA methylation and coronary heart disease (Lead: Jan Bressler)

1928 is an epigenome-wide analysis in collaboration with CHARGE. The new proposal is restricted to a risk score derived from 10 CpG sites.

2827: A prospective study of DNA methylation age acceleration and incidence of coronary heart disease, heart failure, and peripheral arterial disease (Lead: Nick Roetker)

2827 is evaluating the Horvath and Hannum DNA methylation age predictors in relation to CVD outcomes. The new proposal will include the Horvath and Hannum predictors as covariates in Cox models.

2346: Validation of DNA methylation age predictors and association with long-term survival (Lead: Brian Chen; ARIC lead: Jim Pankow)

2346 evaluated the Horvath and Hannum DNA methylation age predictors in relation to total mortality, and has already been published 3. The new proposal will include the Horvath and Hannum predictors as covariates in Cox models.

2346: Validation of DNA methylation age predictors and association with long-term survival (Lead: Brian Chen; ARIC lead: Jim Pankow)
2346 evaluated the Horvath and Hannum DNA methylation age predictors in relation to total mortality, and has already been published ³. The new proposal will include the Horvath and Hannum predictors as covariates in Cox models.

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)* __________  
___x___ 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