ARIC Manuscript Proposal #2345

1.a. Full Title: A prospective study of the association of DNA methylation age with lung function and type 2 diabetes in the Atherosclerosis Risk in Communities Study

b. Abbreviated Title (Length 26 characters):

2. Writing Group:
   Writing group members:
   Nicholas Roetker
   James Pankow
   Heather Nelson
   Weihua Guan
   Chindo Hicks
   Thomas Mosley

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

First author: Nicholas S. Roetker
Address: Division of Epidemiology and Community Health
         University of Minnesota
         1300 South Second Street, Suite 300
         Minneapolis, MN 55454

         Phone: 612-624-9261       Fax:
         E-mail: roetk001@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: 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: 612-624-0315
         E-mail: pankow@umn.edu
3. **Timeline:**

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

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).

Changes in DNA methylation patterns are an integral part of normal development (including genomic imprinting and X chromosome inactivation), tissue-specific gene silencing, and cell differentiation (1) and have more recently been linked with cancer, cardiovascular disease, various neurological, metabolic, and autoimmune disorders, and the aging process in general (1,2). Previous research has also noted that DNA methylation can be modified by diet, smoking, alcohol, physical activity, obesity, stress, genes, and many additional environmental factors (3).

Since changes in DNA methylation patterns occur in response to external and internal stimuli over the life course, it may be possible to characterize these changes as a proxy for the rate of aging (i.e., slow or accelerated). Recent studies have reported sets of CpG sites that can be used to accurately predict chronological age (4,5). Using methylation data from 82 publicly available datasets consisting of samples of 51 healthy tissue and cell types, Horvath (4) developed a 353 CpG site model to predict age and showed that correlation between chronological and predicted age using blood DNA were strong (r=0.95) in datasets of subjects ranging in age from infants to elderly and modest (r=0.6 to 0.8) in datasets of middle- to older-aged adults. The median absolute difference between chronological and predicted age was 3.6 years. Using DNA from whole blood of 482 individuals aged 19 to 101 years, Hannum et al. (5) developed a similar age predictor consisting of 71 CpG sites and demonstrated a high correlation with chronological age (r=0.9) in 174 independent samples. The Horvath and Hannum et al. age predictors share only 6 CpG sites in common.

It is not well studied whether predicted age using DNA methylation better predicts aging related disease compared to chronological age. With its rich covariate composition and many years of follow-up (up to 20 years after collection of DNA samples) of nearly 3,000 African-American participants, the ARIC Study represents a unique setting in which we can examine the association of these predicted age measures with illustrative age-related trait and disease outcomes, including lung function and type 2 diabetes.

5. **Main Hypothesis/Study Questions:**
Higher predicted DNA methylation age is associated with decreased lung function and increased incidence of diabetes, independent of chronological age.

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 (diabetes) and cross-sectional (lung function)

**Endpoints:** lung function (FEV\textsubscript{1} and FEV\textsubscript{1}/FVC), diabetes incidence. Diabetes incidence will be defined based on standard ARIC criteria (elevated fasting or non-fasting glucose, reported physician diagnosis, or reported diabetes medication) applied at later follow-up visits. Secondary analysis will be performed using reported physician diagnosis of diabetes available through the most recently available annual telephone interview.

**Exposure:** DNA methylation age

**Covariates:** chronological age, sex, BMI, height, systolic/diastolic BP, use of anti-hypertension medication, LDL cholesterol, use of lipid lowering medication, physical activity, smoking (current status and pack-years), alcohol use, HRT use, sitting height, and education.

**DNA methylation age:** Methylation status was measured from DNA extracted from whole blood white cells using the Illumina HM450 chip. Degree of methylation was determined using Illumina GenomeStudio 2011.1, Methylation module 1.9.0 software. The methylation score for each CpG was represented as a beta (β) value calculated by dividing the fluorescence intensity of the methylated allele by the sum of the intensities of the methylated allele and unmethylated allele. Background subtraction was conducted with the GenomeStudio software using built-in negative control bead types on the array. An average normalization was applied to minimize scanner-to-scanner variation. We will perform additional normalization and imputation for missing beta values using R code provided by Horvath (4). We will estimate two separate DNA methylation ages using this normalized/imputed dataset, first with R code from Horvath and second with probe-specific coefficient values reported in Table S3 of Hannum et al. (5).

**Inclusion:** 2,905 African American individuals for whom DNA was collected at visit 2 (n=2,504) or visit 3 (n=441) and measured for methylation.

**Exclusions:** Diagnosis of diabetes at or prior to date of DNA collection; pass rate for all probes on 450K array less than 99% (probes with a detection p-value >0.01/all probes)

**Data analysis:** We will run Cox proportional hazards regression and logistic regression models with incident diabetes as the event of interest and DNA methylation age as the main exposure (separate models for Horvath and Hannum et al. predictors). Follow-up time will be defined as the time from the date of DNA collection to the date of disease incidence, death, or loss to follow-up, or December 31, 2011, whichever came first. We
will also run separate linear regression models with visit 2 FEV$_1$ as a continuous outcome and DNA methylation age at visit 2 as the main exposure (excluding those participants with DNA collected at Visit 3). We may also run logistic regression models with airflow obstruction (Yes/no, based on cutoff FEV$_1$/FVC < 0.70 vs. $\geq 0.70$) as the outcome. We will try several approaches to adjust for chronological age, including stratifying by chronological age categories, using residual values from linear regression of DNA methylation age on chronological age in place of DNA methylation age, and simultaneously adjusting for chronological age. Additional models will adjust for potential confounding factors. We will also run additional sensitivity analyses excluding individuals whose correlation with a gold standard (defined by averaging the beta values across the samples from the largest blood dataset analyzed by Horvath (4)) are less than 0.95, whose predicted gender or tissue (according to output from the program by Horvath) is uncertain or in disagreement with actual gender or tissue (whole blood white cells). We will account for potential confounding due to white blood cell heterogeneity by using the method of Houseman (6) to estimate cell distributions using external reference datasets containing methylation profiles of isolated cell types, or through another method that corrects for cell mixture differences without the need for external reference data (7–9).

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.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?  _____ 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


