1.a. Full Title: Epigenetics in the House: Spouse Correlations for Chronic Disease Related DNA Methylation

b. Abbreviated Title (Length 26 characters): Spousal Correlation in DNA Methylation

2. Writing Group:
   Writing group members:
   Ellen W. Demerath
   Steven Nguyen
   Weihua Guan
   Jim Pankow
   Kari North
   Jan Bressler
   Megan Grove
   Myriam Fornage
   Eric Boerwinkle

   Others Welcome

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

First author: Steven Nguyen

   Address: 1405 Jessamine Avenue West, Saint Paul, MN 55108
   Phone: 408-807-9552  Fax: 
   E-mail: nguy2295@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: Ellen W. Demerath, PhD

   Address: University of Minnesota School of Public Health
   Division of Epidemiology and Community Health
   1300 S. Second Street, Suite 300
   Minneapolis, MN 55454
   Phone: 612-624-8231  Fax: 
   E-mail: ewd@umn.edu
3. **Timeline:**

There are approximately 3 papers planned under this manuscript proposal:

1) Obesity-related Variants  
2) Smoking-related Variants  
3) Physical Activity-related Variants

The timeline below relates to Paper #1 (Spouse Correlations in Obesity-related DNA Methylation Variants)

- **January 2015:** Draft Introduction
- **February 2015:** Gain IRB approval and access to data and begin Data Analysis
- **March 2015 - April 2015:** Complete Data Analysis and Write Methods and Results
- **May 2015:** Write Discussion
- **June 2015:** Submit to Co-authors for review
- **July 2015:** Submit to ARIC P and P for review

4. **Rationale:**

Epigenetics is the study of heritable modifications in DNA not involving changes in the DNA sequence, such as methylation, which can affect gene expression. DNA methylation from environmental exposures can alter gene expression and is involved in a number of chronic conditions such as those associated with obesity. Research has shown that DNA methylation at multiple sites is associated with obesity, cigarette smoking, and BMI and adiposity. Furthermore, research has shown that married couples have similar BMI, health behaviors, and environmental exposures conducive to BMI gain. Another study observed a small but statistically significant nonzero epigenome wide correlation in DNA methylation. The purpose of this study then is to examine whether married couples also have similar DNA methylation characteristics compared to others possibly as a result of the shared environment. By quantifying spousal (household) correlations in methylation, we will 1) determine whether there is significant non-independence that must be accounted for in analyses on methylation data in ARIC, as it is known that among the ~15,000 subjects in total, some 4,000 spouse pairs exist; and 2) provide a population-based estimate of the effect of shared environments on DNA methylation, a key parameter for understanding to what extent environmental modification in adulthood may alter epigenetic marks.

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

Spouses are hypothesized to share more similar DNA methylation profiles for CpG sites that are identified in EWAS of obesity traits, smoking history, and physical activity history perhaps than would be expected among unrelated individuals, due to decades of shared environmental exposures and health behaviors.
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: Cross Sectional

Inclusion/Exclusion: African-American subjects with DNA methylation data at visit 2/3, and concurrent phenotype information (including sex, age, BMI, smoking status, physical activity level).

Outcome/variables of interest: CpG methylation level (beta values) from Illumina 450K Human Methylation BeadChip; at this time, all of these are at Visit 2/3

Primary exposure/ independent variable: Household ID

Secondary exposures/independent variables: individual level BMI, smoking status, and physical activity; education, income, white blood cell count, differential wbc proportions, technical covariates (chip, chip row, plate).

Data Analysis:

Statistical analyses will be conducted using SAS 9.4. The dataset includes methylation data on approximately 2,800 African American adults from Visit 2/3. Mean methylation and covariate information will be statistically described (means/frequencies) in spouses and in non-spouses by sex. We have determined there are approximately 300 spouse pairs in the ARIC methylation dataset. We will test whether the men and women in spouse pairs are phenotypically different from men and women the rest of the ARIC subjects, to characterize the spouse dataset. Subsequent analyses will include only the 300 spouse pairs (600 individuals).

First, we will estimate the similarity of phenotype (BMI, smoking, physical activity) between spouses. For continuous variables such as BMI and physical activity, we will use a linear mixed effects model:

\[ Y_{ij} = \beta_0 + X_{ij}\beta + u_j + \varepsilon_{ij} \]

where \( Y_{ij} \) is the phenotype of individual \( i \) of household (spouse pair) \( j \), and \( X_{ij} \) the covariates including education, income, white blood cell count, We assume \( u_j \sim N(0, \sigma^2_s) \) with \( \sigma^2_s \) denoting the between household variation, and \( \varepsilon_{ij} \sim N(0, \sigma^2_e) \) the between individual variation. \( u_j \) and \( \varepsilon_{ij} \) are typically assumed to be independent. An intraclass correlation coefficient (ICC) is calculated as \( \frac{\sigma^2_s}{\sigma^2_s + \sigma^2_e} \). A one-side wald test can be applied to test the null hypothesis, \( H_0: \sigma^2_s = 0 \).

For discrete variables such as smoking status, we will use a generalized linear mixed model.

Second, for each phenotype of interest, we will perform a global test of spouse correlations in methylation across the genome. With help from Dr. Weihua Guan, we will calculate for each individual their average methylation level across all ~475,000 CpG sites. That average methylation level will be used as the dependent variable in a mixed effects linear regression model. We will estimate similarity of methylation level within household (spouse-pair) using a
linear mixed-effects model similarly as described above. For individual $i$ of household $j$, the methylation level at a specific CpG site can be modeled as:

$$M_{ijk} = \beta_0 + X_{ij}\beta + u_j + v_k + \varepsilon_{ijk}$$

where $X_{ij}$ denotes individual characteristics which may contribute to the variation of methylation levels, $v_k \sim N(0, \sigma_v^2)$ the random chip effects, $u_j \sim N(0, \sigma_u^2)$ the random household effect, and $\varepsilon_{ijk} \sim N(0, \sigma_e^2)$ the random noise. The ICC can be calculated similarly as $\frac{\sigma_u^2}{(\sigma_v^2 + \sigma_u^2)}$. If the ICC is significantly greater than 0, we will ask the question of whether spousal similarity in BMI (or smoking status, or physical activity) by constructing group variables that specify spouse pairs whose BMI’s are, or are not, similar (i.e., difference < 5 kg/m2 = similar, coded “1” vs. difference >= 5 kg/m2 = dissimilar, coded “0”). The exact definitions of similarity or difference will be determined after closely examining the dataset. This grouping variable is then added to the repeated statement to generate group-specific covariance parameter estimates. As above, we will then calculate the ICC for each group, and compare the -2LL of the two mixed models (with vs. without the group variable) to test the hypothesis that the similarity in spousal BMI’s influence the similarity in global mean methylation. Depending on the trait of interest, the grouping variable could be smoking (both current smokers vs not, but sedentary vs not, etc).

Third, for each phenotype of interest, we will perform CpG-specific tests of spouse correlations in methylation. As an example, we have 38 replicated BMI-related CpG sites that will be examined in this manner. For each identified CpG sites, we will estimate similarity of methylation level within household (spouse-pair) using a linear mixed-effects model similarly as described above. For individual $i$ of household $j$, the methylation level at a specific CpG site can be modeled as:

$$M_{ijk} = \beta_0 + X_{ij}\beta + u_j + v_k + \varepsilon_{ijk}$$

where $X_{ij}$ denotes individual characteristics which may contribute to the variation of methylation levels, $v_k \sim N(0, \sigma_v^2)$ the random chip effects, $u_j \sim N(0, \sigma_u^2)$ the random household effect, and $\varepsilon_{ijk} \sim N(0, \sigma_e^2)$ the random noise. The ICC can be calculated similarly as $\frac{\sigma_u^2}{(\sigma_v^2 + \sigma_u^2)}$. This model will again be run with and without the phenotype of interest as grouping variable, and the corresponding ICC values will be compared.

With 300 spouse pairs, we would have 80% power to detect a spearman correlation of $r=0.10$ or greater, and to detect a difference in correlation between $r1=0.10$ and $r2=0.22$. A global estimate of the covariance in methylation levels across all 450K sites in spouses was $=0.09$. This study is exploratory, and there is very little existing information as yet in the literature (we could find none) on the household-level effect on *site-specific* methylation as proposed here, for CpG sites having known association with important behavioral characteristics.

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 publication 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)?

There are no papers looking at spouse correlations in methylation. The ms by Laura Cobb on spouse correlations in BMI is the one closest to this.

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.