Population Architecture using Genomics and Epidemiology (PAGE)

Ver. 09/30/16

PAGE Manuscript Proposal Template

Submit proposals by email to the PAGE Coordinating Center at **Rwilliams@biology.rutgers.edu**

All sections must be completed; incomplete applications will be returned.
Do not exceed 3 pages in length (not including references).

PAGE Ms. Number: _122_ Submission Date : _07/26/2019_ [Approval Date: _____]

Title of Proposed MS: Methylation patterns associated with lipid traits and the modifying effect of smoking and alcohol intake in multi-ethnic populations

Abbreviated Title of Proposed MS: _DNA methylation and lipid levels_

I. INVESTIGATOR INFORMATION:

Name of Lead Author: | **Junior Investigator? Y/N**
Yao Hu
Email Address: yhu23@fredhutch.org Y
Telephone Number: 206-667-2312
Authorship model: Authorship will be determined based on contributions.

Name of Corresponding Author (**if different**): Charles Kooperberg, Ulrike Peters
Email Address: clk@fredhutch.org, upeters@fredhutch.org
Telephone Number: 206-667-4142, 206-667-7850

Names, affiliations and email address of PAGE Investigators proposed as co-authors:

N, N	Affiliation in PAGE	Email
Kari North	SOL, ARIC	kari_north@unc.edu

Partner studies in PAGE not collaborating in this ms. proposal:

Study	Contacted? Y/N	Declined? / Other?

Names, affiliations, email address of non-PAGE investigators proposed as co-authors:

II. SCIENTIFIC RATIONALE (Please be specific and concise)

Circulating levels of lipids such as high-density lipoprotein (HDL), low-density lipoprotein (LDL), total cholesterol (TC), and triglycerides (TG), are clinically associated with cardiovascular disease, type 2 diabetes, and fatty liver disease[1-3]. Plasma lipid levels are heritable polygenic traits, with twin studies estimating narrow-sense heritability from 0.48 to 0.76[4]. Genetic association studies have identified over 400 lipids-associated loci[5-11]. However, the phenotypic variance explained by these identified variants is limited (8.8-12.3%), highlighting the importance of searching for additional factors that contribute to interindividual variation of lipids levels beyond genetic sequence variants. In addition, previous studies have reported the modifying effects of non-genetic factors on lipid-gene associations, including smoking[10] and alcohol intake[12].

DNA methylation is an epigenetic modification characterized by the addition of methyl groups predominantly to cytosines at CpG sites and plays a pivotal role in gene expression through promoter silencing[13,14]. DNA methylation has been linked to regulation of lipid levels in previous studies, and a total of 189 CpG sites have been reported for association with lipids through methylome-wide association analyses, 39 of which have shown consistent replications across different studies[15-20]. These findings provided novel insights into the underlying mechanisms of lipid metabolism.

However, all these analyses were performed in European-ancestry populations with moderate sample sizes. It remains unknown whether these identified CpG-lipid associations could be generalized to African Americans, Hispanic/Latinos and other racial/ethnic groups. Previous studies have indicated population-specific effects at several body mass index-associated methylation loci[21] and higher methylation genetic risk scores of type 2 diabetes in Asians compared to Europeans[22], highlighting the importance of examining lipids-associated methylation sites in ancestrally diverse populations. In addition, knowledge on the interaction of methylation patterns and non-genetic factors on lipid profiles is extremely limited.

III. OBJECTIVES AND PLAN (Please be specific and concise)

a. Study Questions/Hypotheses.

- To identify novel CpG sites associated with lipid levels in ancestrally diverse populations.
- To explore the generalization and potential heterogeneity of the previously reported and newly discovered CpG sites by examining their effect sizes and association directions across ethnic groups.
- To explore the modification effects of smoking and alcohol intake on CpG-lipid associations.
- To infer the causality between differential methylation and the change of lipid profiles.

b. **Study populations, study design for each**

The proposed analyses will include ancestrally diverse studies with methylation data measured by Illumina arrays, such as WHI and ARIC.

c. **Variant/SNPs (Specify)**
We propose to use all genomic variants in WHI and ARIC.
d. **Phenotype(s) (Specify)**
High-density lipoprotein, low-density lipoprotein, total cholesterol, and triglycerides.
e. **Covariates (Specify)**
Lipid lowering drugs, age, gender, race/ethnicity, study, center, family structure, white blood cell species and technical covariates.
f. **Main statistical analysis methods**
Before analysis, all four lipid levels will be adjusted for lipid lowering drug intakes and will be inverse-normally transformed within each study. Beta values, which indicate the ratio of methylated to combined intensity from the methylation arrays, will be calculated and normalized using a Beta-MIxture Quantile dilation (BMIQ) approach[23].

In the methylome-wide association analyses, we first plan to perform methylation-lipid association analyses with adjustment for age, sex, race/ethnicity, center (if applicable), proportion of white blood cell species estimated using the Houseman method[24] and technical covariates (chip as random effect, and row and column as fixed effect). The summary statistics from these two studies will be combined through inverse variance-weighted fixed-effect meta-analyses. Bonferroni corrections will be applied to these models where $\alpha=0.05/$(number of CpG sites tested) in order to define significant CpG sites. We will then perform ethnic-specific methylome-wide association analyses in African American and Hispanic/Latino populations, respectively, and explore the generalizability and heterogeneity of the newly identified and previously reported lipids-associated CpG sites across these ethnic groups through examination of their association directions and effect estimates in each ancestral group.

To identify CpG sites that interplay with smoking and alcohol intake, we will perform 2-degree-of-freedom tests that jointly evaluates main effects (CpG sites) and interaction (CpG sites by smoking or by alcohol intake) with the same adjustment applied in the main model.

In the Mendelian randomization analyses, we plan to use genetic variants as instrumental variables to determine whether differential methylation is consequential to the change of lipid profiles or vice versa.

g. **Ancestry information used? No __ Yes_Y_ How is it used in the analyses?**
Ancestry information will be used as a covariate in the methylome-wide association analysis to adjust for population stratification.
h. **Anticipated date of draft manuscript to P&P: _September 2020_**
i. What manuscript proposals listed on *www.pagestudy.org/index.php/manuscripts/* **are most related to the work proposed here? Approved PAGE ms. numbers: ____ ____ ____**

- **If any: Have the lead authors of these proposals been contacted for comments and/or collaboration? Yes __ No __**

IV. SOURCE OF DATA TO BE USED (Provide rationale for any data whose relevance to this manuscript is not obvious): **Check all that apply:**

Aggregate/summary data to be generated by investigators of the study(ies) mentioned:
[ ] ISMMS; **[X]** CALiCO; **[ ]** MEC; **[X]** WHI; **[ ]** CC; **[ ]** Other:_____
If CALiCo please specify:
Included on MEGA Array: **[ ]** SOL
Studies not on MEGA: **[X]** ARIC; **[ ]** CARDIA; **[ ]** SHS-Fam; **[ ]** SHS-Cohort;

I, _YH_, affirm that this proposal has been reviewed and approved by all listed investigators.

V. REFERENCES

1 Emerging Risk Factors, C. *et al.* Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* **302**, 1993-2000, doi:10.1001/jama.2009.1619 (2009).

2 Qi, Q., Liang, L., Doria, A., Hu, F. B. & Qi, L. Genetic predisposition to dyslipidemia and type 2 diabetes risk in two prospective cohorts. *Diabetes* **61**, 745-752, doi:10.2337/db11-1254 (2012).

3 Oresic, M. *et al.* Prediction of non-alcoholic fatty-liver disease and liver fat content by serum molecular lipids. *Diabetologia* **56**, 2266-2274, doi:10.1007/s00125-013-2981-2 (2013).

4 Kettunen, J. *et al.* Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet* **44**, 269-276, doi:10.1038/ng.1073 (2012).

5 Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-713, doi:10.1038/nature09270 (2010).

6 Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-1283, doi:10.1038/ng.2797 (2013).

7 Klarin, D. *et al.* Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet*, doi:10.1038/s41588-018-0222-9 (2018).

8 Liu, D. J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet* **49**, 1758-1766, doi:10.1038/ng.3977 (2017).

9 Hoffmann, T. J. *et al.* A large electronic-health-record-based genome-wide study of serum lipids. *Nat Genet* **50**, 401-413, doi:10.1038/s41588-018-0064-5 (2018).

10 Bentley, A. R. *et al.* Multi-ancestry genome-wide gene-smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. *Nat Genet* **51**, 636-648, doi:10.1038/s41588-019-0378-y (2019).

11 Lu, X. *et al.* Exome chip meta-analysis identifies novel loci and East Asian-specific coding variants that contribute to lipid levels and coronary artery disease. *Nat Genet* **49**, 1722-1730, doi:10.1038/ng.3978 (2017).

12 de Vries, P. S. *et al.* Multiancestry Genome-Wide Association Study of Lipid Levels Incorporating Gene-Alcohol Interactions. *Am J Epidemiol* **188**, 1033-1054, doi:10.1093/aje/kwz005 (2019).

13 van der Harst, P., de Windt, L. J. & Chambers, J. C. Translational Perspective on Epigenetics in Cardiovascular Disease. *J Am Coll Cardiol* **70**, 590-606, doi:10.1016/j.jacc.2017.05.067 (2017).

14 Zhong, J., Agha, G. & Baccarelli, A. A. The Role of DNA Methylation in Cardiovascular Risk and Disease: Methodological Aspects, Study Design, and Data Analysis for Epidemiological Studies. *Circ Res* **118**, 119-131, doi:10.1161/CIRCRESAHA.115.305206 (2016).

15 Irvin, M. R. *et al.* Epigenome-wide association study of fasting blood lipids in the Genetics of Lipid-lowering Drugs and Diet Network study. *Circulation* **130**, 565-572, doi:10.1161/CIRCULATIONAHA.114.009158 (2014).

16 Frazier-Wood, A. C. *et al.* Methylation at CPT1A locus is associated with lipoprotein subfraction profiles. *Journal of Lipid Research* **55**, 1324-1330, doi:10.1194/jlr.M048504 (2014).

17 Pfeiffer, L. *et al.* DNA methylation of lipid-related genes affects blood lipid levels. *Circ Cardiovasc Genet* **8**, 334-342, doi:10.1161/CIRCGENETICS.114.000804 (2015).

18 McClay, J. L. *et al.* High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. *Genome Biol* **16**, 291, doi:10.1186/s13059-015-0842-7 (2015).
19 Sayols-Baixeras, S. *et al.* Identification and validation of seven new loci showing differential DNA methylation related to serum lipid profile: an epigenome-wide approach. The REGICOR study. *Hum Mol Genet* **25**, 4556-4565, doi:10.1093/hmg/ddw285 (2016).
20 Hedman, A. K. *et al.* Epigenetic Patterns in Blood Associated With Lipid Traits Predict Incident Coronary Heart Disease Events and Are Enriched for Results From Genome-Wide Association Studies. *Circ Cardiovasc Genet* **10**, doi:10.1161/CIRCGENETICS.116.001487 (2017).
21 Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541**, 81-86, doi:10.1038/nature20784 (2017).
22 Chambers, J. C. *et al.* Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol* **3**, 526-534, doi:10.1016/S2213-8587(15)00127-8 (2015).
23 Teschendorff, A. E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189-196, doi:10.1093/bioinformatics/bts680 (2013).
24 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *Bmc Bioinformatics* **13**, doi:Artn 86 10.1186/1471-2105-13-86 (2012).

VI. IF USING SOL DATA (Please provide the information below)

a. **Keywords:**
b. **Using biomarker data? Yes __ No _N_**
c. **Where will the SOL data analyses be performed?**

VII. IF USING CHS DATA (Please provide the information below)

a. **Do you propose use of data from a participant's DNA?**