1.a. Full Title: DNA Methylation Related SNPs Interact with Fatty Acids on HDL

b. Abbreviated Title (Length 26 characters): DNA Methylation, fatty acids, and HDL

2. Writing Group:

Writing group members: YiYi Ma, Lyn Steffen, Weihua Guan, Mike Tsai, Brian Steffen, Rozenn Lamaitre, Dary Mozafarrian, Myriam Fornage, and others from the CHARGE fatty acid working group.

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

First author: YiYi Ma
Address: Tufts University, Friedman School of Nutrition and Policy
Phone: 617-858-4006
Fax: none
E-mail: Yiyi.Ma@tufts.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: Lyn M. Steffen
Address: 1300 South Second St, Suite 300; Minneapolis, MN
Phone: 612-625-9307
Fax: 612-624-0315
E-mail: steffen@umn.edu

3. Timeline: 2 years
February, 2013 – January 2014: data analysis at each field center; meta-analysis
February, 2013 – January 2015: manuscript preparation including literature review, describing the methods for each study, methods for meta-analysis, results, and writing the Discussion section.
4. Rationale:

Rationale and Objectives:
The long-term objective of the proposed project is to investigate relationships between fatty acids, epigenetic changes and genetic variants for CVD risk factors. Epigenetic mechanisms have been shown to regulate gene function to alter phenotypes, and epigenetic events may occur as a result of environmental exposures. Evidence demonstrating relationships between epigenetic events and sequence variation is also accumulating, as allele-specific methylation has been documented at loci distributed across the genome\(^1\). We hypothesize that sequence variants that alter the likelihood of epigenetic events provide a potential mechanistic explanation for associations between certain SNPs and phenotypes. In vitro evidence also shows that specific fatty acids (including palmitic, oleic, butyric and arachidonic, and n-3 fatty acids)\(^3\) are correlated with epigenetic changes that may alter gene expression. We further postulate that blood fatty acids interact with SNPs to modulate phenotypes of interest via mechanisms involving epigenetic events.

Although several major epigenetic mechanisms are described (e.g., DNA methylation, histone modification, chromatin remodeling), we will focus on SNPs occurring at predicted DNA methylation sites because of established associations between methylation and multiple CVD risk factors (atherosclerosis\(^5\), \(^7\), dyslipidemia\(^7\) and inflammation\(^8\)). Furthermore, methylation occurs on DNA nucleotides (rather than on chromatin or histone proteins), and is therefore more likely to be sensitive to sequence variation than other epigenetic events.

In the short-term, we seek to investigate the associations and interactions between plasma/red blood cell membrane fatty acids and candidate SNPs in modulating high density lipoprotein cholesterol (HDL), risk factor of CVD. The results from the proposed project will guide future laboratory experiments designed to evaluate SNP functionality that we hypothesize are related to fatty acid-mediated changes in methylation.

In this study, the outcome is HDL. The predictor of interest is those SNPs predicted to be related with DNA methylation or referred as “biomarkers” of DNA methylation change. The effect modifiers are plasma fatty acids. The alternative hypothesis of the study is that plasma fatty acids modify the relationship between SNPs and HDL. Confounders in this study, defined as the factor having associations with both outcome and predictor, are age, sex, BMI, smoking, alcohol, physical activity. Some potential confounders may include population structure or pedigree, education, total fat intake, dietary carbohydrate quality, total energy intake, folate, VitB12, and estrogen therapy. Precision covariates, defined as those reduce standard errors, may include center. Given the extensive gaps in understanding of the relationships between fatty acids, genetic variation and methylation-based mechanisms, we cannot categorize the covariates with complete certainty, and overlaps may exist. However, the proposed models include most standard lifestyle-related factors used for HDL.

SNP Selection:
A flowchart detailing selection of candidate SNPs is listed in Figure 1. The starting list of SNPs was obtained from GWAS\(^10\)-\(^17\) of lipids and candidate genes for lipids. Next, eight selection criteria were developed, covering five characteristics of the SNPs: 1) association with relevant phenotypes, 2) demonstrated association with fatty acids, 3) minimum minor allele frequency (MAF), 4) DNA methylation potential, and 5) potential functionality. As a result, 8 SNPs were selected and listed in Table 1. Of these 8, 7 SNPs will be meta-analyzed due to genotype availability of each cohort and LD between SNPs.

Preliminary Results in the GOLDN Study:
We obtained preliminary evidence in GOLDN that suggest that red blood cell membrane fatty acids interact with genetic variants to modulate HDL. Table 2 and Table 3 illustrate SNP and fatty acids associations for the three phenotypes. Results of SNPs\(^*\) fatty acids interactions are listed in Table 4.
Figure 1 Flow chart of SNP selection

<table>
<thead>
<tr>
<th>Candidate SNP Method</th>
<th>Candidate Gene Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source I</td>
<td>Source II</td>
</tr>
<tr>
<td>SNPs from lipid GWAS</td>
<td>Lipid genes from lipid GWAS;</td>
</tr>
<tr>
<td>(323 SNPs)</td>
<td>(40 genes)</td>
</tr>
</tbody>
</table>

Selection Criteria:
- Association with phenotype *(Required)*
  Criteria 1: genes related with blood lipids and inflammation;
- Fatty acids effect *(Required)*
  Criteria 2: fatty acids affect gene expression or DNA sequence close to SNP predicted to contain PPARα or PPARγ responsive elements;
- Statistical Power *(Required)*
  Criteria 3: Minor allele frequency (MAF) in Hapmap CEU population is greater than 0.01.
- DNA methylation *(Required)*
  Criteria 4: close to CpG;
  Criteria 5: within promoter;
  Criteria 6: within region with tissue differential DNA methylation status;
  Criteria 7: within region with tissue differential chromatin status;
  Criteria 8: has evidence for allele-specific DNA methylation (ASM)
- Functionality of SNP *(Optional)*
  Criteria 9: functional evidence published in previous studies;

Candidate SNP list (8 SNPs)

7 SNPs tend to be meta-analyzed according to availability by each cohort and LD between SNPs

**Table 1 Candidate SNPs selected**

<table>
<thead>
<tr>
<th>Number</th>
<th>SNP</th>
<th>Gene</th>
<th>MAF</th>
<th>Criteria</th>
<th>Genotype Availability*</th>
<th>LD with other SNP</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs405509</td>
<td>APOE</td>
<td>0.491</td>
<td>1,2,3,4,5,6,7,9</td>
<td>3 (2)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>rs2246293</td>
<td>ABCA1</td>
<td>0.412</td>
<td>1,2,3,4,5,6,7</td>
<td>2 (2)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>rs3761740</td>
<td>HMGCR</td>
<td>0.153</td>
<td>1,2,3,4,5,6,7,9</td>
<td>4 (4)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>rs662799</td>
<td>APOA5</td>
<td>0.017</td>
<td>1,2,3,4,5,6,7,9</td>
<td>4 (1)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>rs2479409</td>
<td>PCSK9</td>
<td>0.35</td>
<td>1,2,3,4,5,6,7,9</td>
<td>3 (3)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>rs1169288</td>
<td>HNF1A</td>
<td>0.283</td>
<td>1,2,3,4,5,6,7,9</td>
<td>5 (0)</td>
<td>rs2244608</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>rs2244608</td>
<td>HNF1A</td>
<td>0.283</td>
<td>1,2,3,4,5,6,7</td>
<td>5 (0)</td>
<td>rs1169288</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>rs1169287</td>
<td>HNF1A</td>
<td>0.017</td>
<td>1,2,3,4,5,6,7</td>
<td>4 (0)</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Genotype Availability: Number of cohort with genotype either by chip genotyping or imputation (quality >0.8) (Number of cohort with genotype by chip genotyping)
Table 2 Association of SNPs and phenotypes in GOLDN: (yellow and bold indicate $P<0.05$, pink indicates $0.05<=P<0.1$)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Genotype</th>
<th>Beta</th>
<th>Stderr</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>APOA5_m1123</td>
<td>Genotyped</td>
<td>-0.05</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>APOE_m226</td>
<td>Genotyped</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.34</td>
</tr>
<tr>
<td>ABCA1_rs2246293</td>
<td>Genotyped</td>
<td>0.02</td>
<td>0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>PCSK9_rs2479409</td>
<td>Genotyped</td>
<td>-0.03</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>HMGCR_rs3761740</td>
<td>Genotyped</td>
<td>0.01</td>
<td>0.02</td>
<td>0.56</td>
</tr>
<tr>
<td>HNF1A_rs1169287</td>
<td>Imputed</td>
<td>0.04</td>
<td>0.03</td>
<td>0.20</td>
</tr>
<tr>
<td>HNF1A_rs1169288</td>
<td>Imputed</td>
<td>0.01</td>
<td>0.01</td>
<td>0.68</td>
</tr>
</tbody>
</table>

*Model adjusts for pedigree (assuming exchangeable structure within one pedigree), sex, age, center, BMI, smoking status (categorical: never vs. past vs. current smokers), physical activity* (continuous, based on study-specific metric), alcohol intake (categorical: current vs. former/never), current estrogen therapy (categorical: yes/no), current lipid-lowering medication (categorical: yes/no), education level (categorical: cohort-specific metric), total energy intake (continuous, kcal/day).

Table 3 Association of red blood cell membrane fatty acids and phenotypes in GOLDN: (yellow indicates $<0.05$, pink indicates $0.05<=P<0.1$)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>Stderr</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>fa160</td>
<td>0.00</td>
<td>0.01</td>
<td>0.52</td>
</tr>
<tr>
<td>OleicAcidRbc</td>
<td>0.00</td>
<td>0.01</td>
<td>0.60</td>
</tr>
<tr>
<td>fa182cc</td>
<td>0.01</td>
<td>0.00</td>
<td><strong>0.003</strong></td>
</tr>
<tr>
<td>fa204n6</td>
<td>0.01</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>fa183n3</td>
<td>0.78</td>
<td>0.23</td>
<td><strong>0.0006</strong></td>
</tr>
<tr>
<td>fa205n3</td>
<td>0.09</td>
<td>0.04</td>
<td><strong>0.01</strong></td>
</tr>
<tr>
<td>fa226n3</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.37</td>
</tr>
</tbody>
</table>

*Model adjusts for pedigree (assuming exchangeable structure within one pedigree), sex, age, center, BMI, smoking status (categorical: never vs. past vs. current smokers), physical activity* (continuous, based on study-specific metric), alcohol intake (categorical: current vs. former/never), current estrogen therapy (categorical: yes/no), current lipid-lowering medication (categorical: yes/no), education level (categorical: cohort-specific metric), total energy intake (continuous, kcal/day).

Table 4 Interactions of red blood cell membrane fatty acids and SNPs: (Yellow background with bold red fonts are with $P<0.01$; Yellow background are with $0.01<=P<0.05$. Pink background are with $0.05<=P<=0.10$)*

<table>
<thead>
<tr>
<th>SNP/FA</th>
<th>Fa160 Palmitic acid</th>
<th>OleicAcidRbc Sum of fa18111c, fa181112c, fa18119c</th>
<th>fa182cc e/c linoleic acid</th>
<th>fa204n6 Arachidonic acid</th>
<th>fa18n3 Alphalinolenic acid</th>
<th>fa205n3 EPA</th>
<th>fa226n3 DHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>APOA5</td>
<td>Genotyped</td>
<td>HDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOE</td>
<td>Genotyped</td>
<td>HDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCA1</td>
<td>Genotyped</td>
<td>HDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCSK9</td>
<td>Genotyped</td>
<td>HDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMGCR</td>
<td>Genotyped</td>
<td>HDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HNF1A</td>
<td>Imputed</td>
<td>HDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HNF1A</td>
<td>Imputed</td>
<td>HDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Model adjusts for pedigree (assuming exchangeable structure within one pedigree), sex, age, center, BMI, smoking status (categorical: never vs. past vs. current smokers), physical activity* (continuous, based on study-specific metric), alcohol intake (categorical: current vs. former/never), current estrogen therapy (categorical: yes/no), current lipid-lowering medication (categorical: yes/no), education level (categorical: cohort-specific metric), total energy intake (continuous, kcal/day).
References:


5. Main Hypothesis/Study Questions:
To investigate the associations and interactions between plasma/red blood cell membrane fatty acids and candidate SNPs in modulating high density lipoprotein cholesterol (HDL).

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: Meta-analysis

Study population: ARIC participants (Minnesota) who have baseline phospholipid fatty acid values.

Exclusions: those with missing fatty acids and HDL-cholesterol values; non-white race

ANALYSIS PLAN:
OUTCOMES:
Baseline level of HDL (mg/dL), preferred fasting level but also accept non-fasting level.

ASSOCIATION TEST
(1) ASSOCIATION TEST FOR SNP:
A regression coefficient (β ± robust SE) for the main effect of SNP and outcome will be calculated in each cohort and values meta-analyzed.
Note: an additive genetic model will be used

(2) ASSOCIATION TEST FOR BLOOD FATTY ACIDS:
A regression coefficient (β ± robust SE) for the main effect of plasma or red blood cell membrane fatty acids and outcome will be calculated in each cohort and values meta-analyzed.

(3) ASSOCIATION MODEL COVARIATES:
Model 1:
sex, age (continuous: years), center (if applicable), population structure or pedigree (if applicable).
Model 2:
Covariates from Model 1 plus BMI, smoking status (categorical: never vs. past vs. current smokers), physical activity* (continuous, based on study-specific metric), alcohol intake (categorical: current vs. former/never), current estrogen therapy (categorical: yes/no), current lipid-lowering medication (categorical: yes/no), education level (categorical: cohort-specific metric) , total energy intake (continuous, kcal/day),


dietary total fat intake (continuous, %total energy intake/day), glycemic load (if applicable) (continuous, g/day), dietary total folate intake (if applicable) (continuous, mcg/day), dietary VitB12 intake (if applicable) (continuous, mcg/day).

**INTERACTION TEST:**

(1) **INTERACTION TEST:**

A regression coefficient (β± robust SE) for the interaction term for plasma or red blood cell membrane fatty acids*SNP will be calculated in each cohort and values meta-analyzed.

(2) **INTERACTION MODEL COVARIATES:**

Model 1:
- sex, age (continuous: years), center (if applicable), population structure or pedigree (if applicable).
- physical activity (continuous, %total energy intake/day), alcohol intake (categorical: current vs. former/never), current estrogen therapy (categorical: yes/no), current lipid-lowering medication (categorical: yes/no), education level (categorical: cohort-specific metric), total energy intake (continuous, kcal/day), dietary total fat intake (continuous, %total energy intake/day), glycemic load (if applicable) (continuous, g/day), dietary total folate intake (if applicable) (continuous, mcg/day), dietary VitB12 intake (if applicable) (continuous, mcg/day).

Model 2:
- Covariates from Model 1 plus BMI, smoking status (categorical: never vs. past vs. current smokers), dietary total folate intake (if applicable) (continuous, mcg/day), dietary VitB12 intake (if applicable) (continuous, mcg/day).

(3) **Exposures: Plasma membrane fatty acids**

- Palmitic acid (16:0) (continuous, % of total fatty acids),
- Oleic acid (18:1) (continuous, % of total fatty acids),
- Linoleic acid (18:2n6) (continuous, % of total fatty acids),
- Arachidonic acid (20:4n6) (continuous, % of total fatty acids),
- Alpha-linolenic acid (18:3n3) (continuous, % of total fatty acids),
- EPA (20:5n3) (continuous, % of total fatty acids),
- DHA (22:6n3) (continuous, % of total fatty acids).

FA will be modeled continuously for association and interaction analysis.

**SNPS to be Evaluated:**

<table>
<thead>
<tr>
<th>No.</th>
<th>SNP</th>
<th>NAME</th>
<th>PROTEIN</th>
<th>Function of Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs405509</td>
<td>APOE -219G/T</td>
<td>APOE</td>
<td>Catabolism of TG-rich lipoprotein constituents</td>
</tr>
<tr>
<td>2</td>
<td>rs2246293</td>
<td>ABCA1</td>
<td>ABCA1</td>
<td>Cholesterol efflux from peripheral cells to nascent HDL particles</td>
</tr>
<tr>
<td>3</td>
<td>rs3761740</td>
<td>HMGCGR -911C/A</td>
<td>HMGCGR</td>
<td>Rate limiting enzyme for cholesterol synthesis</td>
</tr>
<tr>
<td>4</td>
<td>rs662799</td>
<td>APOA5 -1131T/C</td>
<td>APOA5</td>
<td>Component of HDL and regulation on TG</td>
</tr>
<tr>
<td>5</td>
<td>rs2479409</td>
<td>PCSK9</td>
<td>PCSK9</td>
<td>Proprotein convertase for cholesterol homeostasis</td>
</tr>
<tr>
<td>6</td>
<td>rs1169287</td>
<td>HNF1A</td>
<td>HNF1 homeobox A</td>
<td>Transcription factor required for the expression of several liver-specific genes</td>
</tr>
<tr>
<td>7</td>
<td>rs1169288</td>
<td>HNF1A</td>
<td>HNF1 homeobox A</td>
<td>Transcription factor required for the expression of several liver-specific genes</td>
</tr>
</tbody>
</table>

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes

_X_ 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.csc.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)?

   a) MS 1928; J Bressler; Genome-wide methylation analyses of cardiovascular disease (CVD) and its risk factors

   b) MS 1929 ; J Pankow; Genome-wide DNA methylation profiling in peripheral blood: quality control and association with demographic characteristics

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data?  

___X__ Yes   _____ No

GWAS via STAMPEDE & GENEVA, #2006.03

11.b. If yes, is the proposal

___X___ 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.csc.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. OK

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.csc.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.