ARIC Manuscript Proposal # 3227

PC Reviewed: 9/11/18 Status: _____ Priority: 2
SC Reviewed: _________ Status: _____ Priority: ____

1.a. Full Title: Dairy Consumption and Cardiometabolic biomarkers: Mendelian Randomization and Gene-diet Interaction Analyses

b. Abbreviated Title (Length 26 characters): Diary, cardiometabolic biomarkers, Gene, Interaction

2. Writing Group:
Lu Qi, Tao Huang (lead)
ARIC co-authors and collaborators from other cohorts are listed below.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Collaborators</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIC</td>
<td>Yujie Wang, Misa Graff, Kari North, Gerardo Heiss</td>
</tr>
<tr>
<td>NHS</td>
<td>Tao Huang, Ming Ding</td>
</tr>
<tr>
<td>HPFS</td>
<td>Tao Huang, Ming Ding</td>
</tr>
<tr>
<td>Women Genome Health Initiative</td>
<td>Chu, Audrey Y., Ph.D</td>
</tr>
<tr>
<td>The Cardiovascular Health Study (CHS)</td>
<td>Rozenn Lemaitre, PhD MPH</td>
</tr>
<tr>
<td>The Rotterdam Study</td>
<td>M. Carola, Zillikens, Trudie Voortman</td>
</tr>
<tr>
<td>The Family Heart Study</td>
<td>Mary Wojczynski</td>
</tr>
<tr>
<td>the Malmö Diet and Cancer study</td>
<td>Ulrika Ericson</td>
</tr>
<tr>
<td>Young Finns Study</td>
<td>Mika Helminen</td>
</tr>
<tr>
<td>Framingham</td>
<td>Mary Wojczynski, Adrienne Cupples</td>
</tr>
<tr>
<td>MESA</td>
<td>Lekki Wood</td>
</tr>
<tr>
<td>InCHIANTI</td>
<td>Tosh</td>
</tr>
<tr>
<td>GLACIER</td>
<td>Frida Renstrom</td>
</tr>
<tr>
<td>Raine Study: young birth cohort in Australia</td>
<td>Carol Wang</td>
</tr>
<tr>
<td>the Danish Diet, Cancer and Health cohort</td>
<td>Tuomas Oskari Kilpeläinen</td>
</tr>
<tr>
<td>(Danish part of the EPIC study)</td>
<td>Camilla Sandholt</td>
</tr>
<tr>
<td>Danish cohort called Inter99</td>
<td></td>
</tr>
<tr>
<td>the Copenhagen City Heart Study, CCHS</td>
<td>Christina Ellervik</td>
</tr>
<tr>
<td>the Copenhagen General Population Study, CGPS</td>
<td></td>
</tr>
<tr>
<td>the Danish General Suburban Population Study, GESUS</td>
<td>Christina Ellervik</td>
</tr>
<tr>
<td>DESIR: Epidemiological Study on the Insulin Resistance Syndrome cohort</td>
<td>frederic.fumeron</td>
</tr>
<tr>
<td>the PREDIMED-Valencia study</td>
<td>M. Dolores Corella Piquer</td>
</tr>
<tr>
<td>GOLDN</td>
<td>Smith, Caren E</td>
</tr>
<tr>
<td>BPRHS</td>
<td>Smith, Caren E</td>
</tr>
</tbody>
</table>
Note: More coauthors will be included in this study based on contribution.

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

First author: Ming Ding & Tao Huang
Address: build 2 room 343, 677 Huntington ave, Boston, MA 02115
Phone: 4103401626
Fax:
E-mail: nhtah@channing.harvard.edu & mingding517@gmail.com

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: Gerardo Heiss, M.D., M.Sc., Ph.D.
Address: Gillings School of Global Public Health, UNC-CH
Department of Epidemiology
137 East Franklin Street, Suite 306
Chapel Hill, NC 27514
Phone: (919) 962-3253
Email: gerardo_heiss@unc.edu

Name: Kari E. North, PhD
Address: Professor Department of Epidemiology
Carolina Center for Genome Sciences
University of North Carolina Chapel Hill, NC 27599
919-966-2148 (voice)
919-966-9800 (fax)
Kari_north@unc.edu

3. Timeline:
Each cohort will complete their Mendelian randomization and gene-diet interaction analyses and provide their results to Tao Huang as soon as possible. Subsequently, an analyst from Harvard University will conduct the meta-analysis.

4. Rationale:
Introduction
The prevalence of cardiometabolic diseases rapidly increased over the world. A body of observational epidemiologic studies showed that high dairy intake has been associated with cardiovascular biomarkers such as lipids and glucose. Moreover, dairy intake which may vary over time is difficult to measure in observational study. Possibility of reverse causation and residual confounding is often nonnegligible.
Mendelian randomization (MR) analysis has become widely used to assess potential causal relations of environmental risk factors and diseases. This method is analogous to a RCT where randomization to genotype takes place at conception. Therefore, MR approach overcomes the abovementioned limitations. Our previous large scaled MR analysis demonstrated that high dairy intake associated with higher BMI, but not causally related to hypertension. However, it is
unclear if habitual dairy intake causally associates with body composition and cardiometabolic biomarkers in general populations.

Therefore, in the current study, we further performed a MR analysis using an established dairy intake associated genetic variant near the lactase gene LCT among 182,041 adult participants from 18 cohorts to examine the causal relationship between habitual dairy intake and cardiometabolic traits such as body composition, lipids, glycaemic traits and inflammatory factors in general populations.

5. **Main Hypothesis/Study Questions:**
The main aim of the proposed investigation is to examine the causal effect of dairy consumption on cardiometabolic traits (lipid measures, fasting glucose and insulin, HOMA-IR, HOMA-B, CRP) using two established SNPs (rs4988235) as the instrumental variable.

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 Criteria**
- Sample size: ≥500
- Follow-up time: ≥2 years (*No limitation to maximal follow-up time, please use 10 years as maximal follow-up time if there are repeated measures over time*).

**Part 1: Mendelian Randomization**
In this part, we propose to examine the effect of dairy consumption on blood pressure and risk of hypertension using SNP rs4988235 as an instrumental variable.

**Exposure, Outcome, and Instrumental Variable**
- **Outcome:** Body composition at endpoint (follow-up≥2 years): Body fat percentage (%), waist circumference (cm), Waist:hip ratio (WHR), lean mass (kg), fat mass (kg)
- **Biomarkers at endpoint** (follow-up≥2 years): HDL (mmol/L), LDL (mmol/L), TC (mmol/L), TG (mmol/L), Non-HDL (mmol/L), apoB (mmol/L), Fasting glucose (mmol/L), Fasting insulin (mIU/L), HOMA-IR, HOMA-B, HbA1c (NGSP %), high sensitivity CRP or regular CRP (mg/L).
  - Blood glycemic biomarkers were typically measured after >8 h of fasting.
  - Prior to analysis, the following variables were transformed to the natural logarithmic scale: Fasting insulin, HOMA-IR, HOMA-B, CRP, and TG.
- **Instrumental variable:** SNP (rs4988235) SNP rs4988235 Code: TT+CT=1, CC=0; T allele is associated with lactase persistence. Please treat SNP as continuous variable. Lactase persistence is a dominantly inherited genetic trait.
- **Covariates:** years of follow-up, dyslipidemia (yes), dysglycemia (yes), and baseline covariates including age, sex, ethnicity, region, body mass index (BMI, kg/m²), family
income, education level, smoking status (current vs. former/never), physical activity, total energy intake (kcal), alcohol consumption, and principal components for population stratification (PCA) (if available)

- Dyslipidemia: physician diagnosed-dyslipidemia or with lipid-lowering medication.
- Dysglycemia: physician diagnosed-diabetes or use of insulin/oral hypoglycemic agents.

- Note: for covariate region: if the study includes several countries, or USA study includes several states, please control region.

1) Association between SNP and each Cardiometabolic outcomes

Cardiometabolic outcome ~ SNP + study-specific covariates
- SNP (rs4988235): main model: CC=0, CT=1, TT=2; dominant model: CC=0, CT/TT=1; recessive model: CC/CT=0, TT=1; please treat SNP as continuous variable.
- Study-specific covariates: baseline age, sex, ethnicity, region

2) Association between SNP and baseline dairy intake as an outcome

Total dairy consumption ~ SNP + study-specific covariates
- SNP (rs4988235): main model: CC=0, CT=1, TT=2; dominant model: CC=0, CT/TT=1; recessive model: CC/CT=0, TT=1; please treat SNP as continuous variable.
- Study-specific covariates: baseline age, sex, ethnicity, region
- Dairy consumption: continuous variable; the unit is serving/day

3) Association between dairy consumption and each Cardiometabolic outcome

Cardiometabolic outcome ~ total dairy consumption + covariates
- Dairy consumption: continuous variable; the unit is serving/day
- Covariates: sex, ethnicity, region, years of follow-up, as well as age, BMI, baseline blood pressure/risk of hypertension, smoking status, physical activity, total energy intake, and alcohol consumption (all covariates at baseline).

Part 2: Gene-diet interaction

In this part, we propose to analyze the interaction of rs4988235 and dairy consumption for the outcomes of blood pressure and risk of hypertension.

Step 1. Stratified analysis on the association between dairy consumption and each of the Cardiometabolic outcomes by SNP (rs4988235)

Cardiometabolic outcome ~ Dairy consumption + covariates
- Please split the data into three subsets by SNP rs4988235 (TT, CT, CC), and into two subsets by SNP rs4988235 (CC, TT/CT) or by SNP rs4988235 (CC/CT, TT)
- Dairy consumption: baseline; continuous variable; the unit is serving/day
- Covariates: sex, ethnicity, region, as well as age, BMI, baseline blood pressure/risk of hypertension, smoking status, physical activity, total energy intake, and alcohol consumption (all covariates at baseline).

Step 2. Stratified analysis on the association between SNP and each of the Cardiometabolic outcomes by dairy consumption

Cardiometabolic outcome ~ SNP + covariates
- Please split the data into three subsets based on tertiles of dairy consumption.
• SNP (rs4988235): main model: CC=0, CT=1, TT=2; dominant model: CC=0, CT/TT=1; recessive model: CC/CT=0, TT=1; please treat SNP as continuous variable.
• Covariates: baseline age, sex, ethnicity, region

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 (We use the genotypic data for SNP rs4988235)

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.cscc.unc.edu/ARIC/search.php

❌ 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  ❌ 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.
In one year

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://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to Pubmed central.