ARIC Manuscript Proposal # 1790

PC Reviewed: 5/10/11  Status: A  Priority: 2
SC Reviewed: _________  Status: _____  Priority: ____

1.a. Full Title: Genome-Wide Association Study of Coffee Consumption

b. Abbreviated Title (Length 26 characters): GWAS of coffee

2. Writing Group:

Writing group members:

Jennifer Nettleton, Ph.D., University of Texas Health Science Center (ARIC author and chair of the CHARGE Nutrition working group)
Keri Monda, Ph.D., University of North Carolina (ARIC author and member of the CHARGE Nutrition working group)
Gerardo Heiss, University of North Carolina (ARIC investigator)
David Couper, University of North Carolina (ARIC investigator)

Lead Author**
**this is likely to be a multi-first authored (starred) effort, but our working lead is listed here:
Marilyn Cornelis, Ph.D., Department of Nutrition, Harvard School of Public Health (Lead author and member of the GENEVA and CHARGE Nutrition working groups)

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

First author:
Name: Marilyn Cornelis
Address: Harvard School of Public Health
655 Huntington Avenue
Building 2 - Room 355
Boston, MA 02115
Office Phone: 617.432.1333
Fax: 617.432.1333
mcorneli@hsph.harvard.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: Jennifer Nettleton
Address: E-641 RAS; 1200 Herman-Pressler; Houston, TX 77030
Phone: 713-500-9367  Fax: 713-500-9264
E-mail: jennifer.a.nettleton@uth.tmc.edu
3. **Timeline:**

- Data sharing deadline = June 1, 2011;
- Meta-Analysis complete = July 1, 2011;
- Additional analyses (if necessary) complete = August 1, 2011;
- Manuscript drafting / submission = September 15, 2011

The WGHS and GENEVA collaboration was recently successful in applying the genome-wide approach to confirming loci implicated in caffeine intake (Cornelis MC, Monda KL, Yu K, Paynter N, Azzato EM, et al. (2011) Genome-Wide Meta-Analysis Identifies Regions on 7p21 (AHR) and 15q24 (CYP1A2) As Determinants of Habitual Caffeine Consumption. PLoS Genet 7(4): e1002033. doi:10.1371/journal.pgen.1002033). The preliminary work (PLoS Gen publication) demonstrated a boost in power when restricting analysis to caffeinated coffee intake. Therefore, new efforts will focus on coffee and are being expanded to include cohorts involved in the CHARGE Nutrition Working Group. Besides caffeine, numerous other compounds are present in this beverage. These may have unique physiological effects but also characteristic taste qualities that may influence one’s choice to consume coffee. The latter might be more effectively captured by investigating total (caffeinated & decaffeinated) coffee or decaffeinated coffee exclusively. The analysis plan given on the following pages seeks to identify genetic loci linked to both aspects of coffee consumption.

4. **Rationale:**

Caffeine is the most widely consumed psychoactive substance in the world with nearly 90% of adults reporting regular consumption of caffeine-containing beverages and foods (1). Caffeine and coffee (the primary source of caffeine for many countries) have been implicated in a number of health conditions (2-4). Coffee contains many other compounds, however, and it remains unclear whether caffeine per se is mediating these effects on health. Identifying the factors contributing to the habitual consumption of caffeine (or coffee) will enable a better understanding of its associations with health.

Although demographic and social factors have been linked to habitual caffeine consumption, twin studies report heritability estimates between 43 and 58% for caffeine use; 77% for heavy use, and 45, 40, and 35%, respectively, for caffeine toxicity, tolerance and withdrawal symptoms (5). In the first genome-wide association study (GWAS) of habitual caffeine intake, we confirmed two loci mapping near AHR and CYP1A2 (PLoS Genetics, in press). Both the AHR and CYP1A2 are biologically plausible candidates as CYP1A2 metabolizes caffeine and AHR regulates CYP1A2. Previous studies suggest that some of the heritability underlying specific caffeine sources (i.e. coffee and tea) may be distinct in relation to total caffeine intake (6). Besides caffeine, numerous other compounds are also present in coffee that may have unique physiological effects but also characteristic taste qualities that may influence one’s choice to consume coffee. In a GWAS restricted to caffeinated coffee, we identified the same loci (AHR and CYP1A2) but with even greater effects. Another GWAS of coffee intake confirmed our findings (7). Coffee and caffeine are strongly correlated in many populations and so deciphering loci unique to each may require a larger sample size and/or different analytical approach.

For additional details regarding this study please contact Marilyn Cornelis (mcorneli@hsph.harvard.edu)

5. Main Hypothesis/Study Questions:

We now propose to extend our previous work to discovery of novel loci. Our preliminary results demonstrate a boost in power to detect caffeine-specific loci when restricting analysis to caffeinated coffee intake. To capture coffee-specific loci we propose investigating total (caffeinated & decaffeinated) coffee or decaffeinated coffee exclusively. Our analysis plan therefore seeks to identify genetic loci linked to both aspects of coffee consumption.

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).

Analytic Plan:

<table>
<thead>
<tr>
<th>Exclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclude individuals with missing information on coffee variables of interest, age, smoking status.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trait Creation</th>
</tr>
</thead>
<tbody>
<tr>
<td>While, the larger meta-analysis will focus on each caffeinated coffee, decaffeinated coffee, and total coffee, ARIC will provide data only for caffeinated coffee intake since only caffeinate coffee consumption was queried by the ARIC FFQ</td>
</tr>
</tbody>
</table>

- **A. caffcoff:** caffeinated coffee (cups/d) -- excluding individuals consuming 0 cups/d.
- **B. caffext:** caffeinated coffee (non/low vs high consumers)
  Using the continuous trait created for A, create a new variable restricted to subjects reporting low intakes (<1 cup/d) and high intakes (≥4 cups/d).

\[
\text{caffext} = 0;
\]

- if **caffcoff** <1 cup/d then 
- if **caffcoff** ≥4 cups/d then 

<table>
<thead>
<tr>
<th>Covariate Creation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. age: continuous</td>
</tr>
<tr>
<td>2. sex: dichotomous, 0: male 1: female</td>
</tr>
<tr>
<td>3. baseline smoking status: categorical (dummy variables)</td>
</tr>
</tbody>
</table>

- **smoke0:** never smokers (reference)
- **smoke1:** former smokers
- **smoke2:** current smokers <15 (or <20) cig/d
- **smoke3:** current smokers ≥15 (or ≥20) cig/d
4. **EV1, EV2, EV3...: if appropriate**, each study should include top eigenvectors (number will vary by study) to adjust for population substructure

5. **Other** study-specific covariates (in ARIC- field center)

<table>
<thead>
<tr>
<th>Running the Analyses</th>
</tr>
</thead>
</table>
| ►Always use an additive genetic model. For imputed data, use allele dosages. Otherwise, use measured genotype. Code reference allele as you wish (each study will document the reference allele in the results file).
| ►Restrict analyses to autosomal SNPs

**A) Continuous trait analysis:** Linear regression of tcoffee on allele dosage/genotype

\[
\text{caffcoff} \sim \mu + \text{SNP} + \text{age} + \text{sex} + \text{smoke1} + \text{smoke2} + \text{smoke3} + \text{EVs} + \text{other}
\]

Label this file: YOURSTUDYNAME_tcoffee_DDMMYYYY

**B) ‘Extreme’ trait analyses:** Logistic regression of tcoffext on allele dosage/genotype

\[
\text{caffcoffext} \sim \text{SNP} + \text{age} + \text{sex} + \text{smoke1} + \text{smoke2} + \text{smoke3} + \text{EVs} + \text{others}
\]

Label this file: YOURSTUDYNAME_tcoffext_DDMMYYYY

<table>
<thead>
<tr>
<th>Results Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable Name</strong></td>
</tr>
<tr>
<td>marker</td>
</tr>
<tr>
<td>chr</td>
</tr>
<tr>
<td>position</td>
</tr>
<tr>
<td>effect_allele</td>
</tr>
<tr>
<td>other_allele</td>
</tr>
<tr>
<td>strand</td>
</tr>
<tr>
<td>beta</td>
</tr>
<tr>
<td>stderr</td>
</tr>
<tr>
<td>pvalue</td>
</tr>
<tr>
<td>freq</td>
</tr>
<tr>
<td>hwe</td>
</tr>
<tr>
<td>n</td>
</tr>
<tr>
<td>imputed</td>
</tr>
<tr>
<td>avpostprob</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?  
NA

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 publications 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)?

- Ms #1686: Genome-wide association study of caffeine intake in the GENEVA and CHARGE Consortiums. The present analysis is an extension of this published work.
- Ms #1375: Coffee intake, lung function, and chronic obstructive pulmonary disease in the ARIC Study. This is not a GWAS study and first author Jennifer Nettleton is a member of this working group.
- Ms #930: Beverage consumption and the risk of type 2 diabetes mellitus. This is not a GWA study.
- Ms #1176B: Cigarette smoking, coffee and alcohol consumption in relation to Parkinson’s Disease in Atherosclerosis Risk in Communities Cohort. This is not a GWA Study.
- Ms #1000: Coffee consumption and the risk of type 2 diabetes in the ARIC Study. This is not a GWA study.
- Ms #470: Coffee intake and homocysteine. This is not a GWA Study.

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

__X__ Yes    ____ No

11.b. If yes, is the proposal

__X__ A. primarily the result of an ancillary study (list number* #2006.03)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

GWAS via STAMPEDE & GENEVA, #2006.03
ARIC is one of ten cohort studies contributing data to the CHARGE Nutrition Working Group -based meta-analysis.
Since this work is a product of CHARGE which utilizes GWA data, ancillaries related to STAMPEDE & GENEVA are also acknowledged.

12. 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.

Understood, and we will meet this deadline