ARIC Manuscript Proposal #3147

PC Reviewed: 4/10/2018       Status: _____       Priority: ___
SC Reviewed: __________       Status: _____       Priority: ___

1.a. Full Title: EWAS on coffee and tea consumption

b. Abbreviated Title (Length 26 characters): EWAS on coffee and tea

2. Writing Group:
Irma Karabegović\textsuperscript{1}
Casey M. Rebholz\textsuperscript{2}
Emily A. Hu\textsuperscript{2}
Myriam Fornage\textsuperscript{3}
Silvana Maas\textsuperscript{1}
Trudy Voortman\textsuperscript{1}
Mohsen Ghanbari\textsuperscript{1}
Athena Vidaki\textsuperscript{1}
Oscar H. Franco\textsuperscript{1}
Manfred Kayser\textsuperscript{1}

1. Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands
2. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
3. The Brown Foundation Institute of Molecular Medicine for the Prevention of Human Diseases at The University of Texas Health Science Center at Houston

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

**First author:** Irma Karabegović  
**Address:** Erasmus University Rotterdam  
Burgemeester Oudlaan 50, 3062 PA  
Rotterdam, Netherlands  
Phone: Fax:  
E-mail: i.karabegovic@erasusmc.nl

**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:** Casey Rebholz  
**Address:** Johns Hopkins Bloomberg School of Public Health  
2024 E. Monument Street, Suite 2-600  
Baltimore, MD 21287
3. Timeline:
Comments or questions on the analytical plan, if any (preferably before April 1\textsuperscript{st}, 2018).
Uploading cohort-specific results by May 1\textsuperscript{st}, 2018.
Meta-analysis in Rotterdam and requesting additional analyses, if needed, by the end of May.
Manuscript shared with co-authors by the end of July.

4. Rationale:
Coffee is the most frequently consumed caffeine-containing beverage around the world [1, 2]. The specific bioactive components in coffee include caffeine and certain phenolic acids, such as chlorogenic acid [3, 4]. Coffee has been researched in association with various phenotypes, including cardiovascular disease [5-7], type 2 diabetes [8-10], inflammatory bowel disease [11-13], and liver-associated diseases [14-18]. Tea, another frequently consumed beverage, shares certain substances with coffee, like caffeine and phenolics, but the levels and bioavailability of phenolics differs between the two beverages [19-21], and sometimes even between different types of tea [22]. In certain phenotypes, coffee and tea have shown a similar effect, for example lowering the incidence of chronic liver disease [17]. Additionally, tea has been associated with cancer preventive effects [23-26], enhanced metabolic and cardiovascular health [27-31], or even protective effect on neurological diseases [32].

There are numerous potential mechanisms which have been proposed as potential mediators in health benefits induced by coffee or tea consumption [33-35]. Besides already proposed mechanisms (e.g., genetics and pharmacokinetics), we hypothesize that coffee and tea consumption have an impact on epigenetic markers. In recent years, an increasing number of studies have provided evidence that lifestyle factors (such as smoking [36], physical activity [37], and diet [38]) may influence DNA methylation [39].

Recently, two studies were performed to investigate the association between coffee consumption and DNA methylation. In one study, conducted by Weronica E. Ek et al. [40] with a sample size of N=3,096 individuals, identified 2 CpGs associated with tea consumption. The association of DNA methylation with tea was observed only in women [40], while no significant association was detected for coffee and DNA methylation. The other study, conducted by Chuang et al. [41] (N=2,315), identified 1 CpG in association with coffee in their meta-analyzed results for the first model (blood cell counts, age, and sex). However, these studies may be underpowered because of insufficient sample size. By increasing the number of study participants, thereby gaining statistical power, we plan to overcome the issue.

5. Main Hypothesis/Study Questions: To find epigenetic modifications related to coffee and tea (black and green) by conducting an epigenome-wide association study (EWAS) on coffee and tea intake.
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).

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identifiers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ptid</td>
<td>Participant cohort ID</td>
<td></td>
</tr>
<tr>
<td>samplename</td>
<td>Sample identifier (from methylation output)</td>
<td></td>
</tr>
<tr>
<td><strong>Exposure Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>coffee</td>
<td>Coffee intake in cups per day</td>
<td>Continuous (cups/day), data collected categorically (for example, 2–3 cups per day) can be converted to cups/day by taking the median value of each category (e.g., 2.5 cups per day).</td>
</tr>
<tr>
<td>Tea</td>
<td>Black and green tea intake combined in cups per day</td>
<td>Continuous (cups/day), data collected categorically can be converted in same manner as coffee.</td>
</tr>
<tr>
<td><strong>Confounding Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>technical covariates</td>
<td>WBC (White blood cell count)</td>
<td>In the same blood sample from which DNA was extracted. If you did not measure white blood percentages in the same sample as used for the DNA methylation measurement, please estimate WBC percentages using a prediction method (e.g. Houseman’s method).</td>
</tr>
<tr>
<td>technical covariates+ cohort specific covariates</td>
<td>Technical and other cohort specific covariates</td>
<td>Please correct for technical (batch) covariates (for example Array-Number, Position-on-Array), and other cohort-specific covariates as you deem necessary. If your cohort includes multiple population ancestries, case-control, or study-specific principal components please take this into account as you deem it is most appropriate for your cohort.</td>
</tr>
<tr>
<td>Age</td>
<td>Age at blood collection</td>
<td>continuous (years)</td>
</tr>
<tr>
<td>Sex</td>
<td>Sex/gender</td>
<td>Factor: men, women</td>
</tr>
<tr>
<td>ethnicity</td>
<td>Race/ethnicity</td>
<td>Factor: for example: &quot;White&quot;, &quot;Black&quot;, &quot;Hispanic&quot;, &quot;Asian/Pacific Islanders&quot;, &quot;Other&quot; (or other categorization if available)</td>
</tr>
<tr>
<td>smoking</td>
<td>Smoking status</td>
<td>Factor: for example: &quot;Never&quot;, &quot;Former&quot;, &quot;Current smoker&quot; (or &quot;Never&quot;, “Ever smoker”)</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
<td>Continuous (kg/m²)</td>
</tr>
</tbody>
</table>
alcohol | Alcohol intake | Continuous (gr/day or glasses/day)
---|---|---

**Analysis: linear regression, mixed effect model.**
Outcome: DNA methylation in adults by using the Illumina Infinium HM450K BeadChip array. We will use untransformed beta values normalized by using the functional normalization method. If this is not possible, other normalization methods are also acceptable.

a) **Association analysis for coffee intake in cups/day:**
M1 → CpG methylation~ technical covariates/cohort specific covariates + age + sex + ethnicity/race + coffee
M2 → CpG methylation~ M1 + smoking
M3 → CpG methylation~ M2 + BMI + alcohol intake

b) **Association analysis for total tea (black AND green combined) intake in cups/day:**
M1 → CpG methylation~ technical covariates/cohort specific covariates + age + sex + ethnicity/race + tea (black AND green)
M2 → CpG methylation~ M1 + smoking
M3 → CpG methylation~ M2 + BMI + alcohol intake

**Sensitivity Analysis**
For both association analyses (coffee and total tea), analyses will be repeated after stratifying by ethnicity/race (without ethnicity adjustment). This is in addition to the main analysis in overall sample (adjusted for ethnicity).

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? __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](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 #1790: Genome-Wide Association Study of Coffee Consumption

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.cscn.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.cscn.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.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes  ___X___ No

References: