ARIC Manuscript Proposal # 1686

PC Reviewed: 8/10/10	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Genome-wide association study of caffeine intake in the GENEVA and CHARGE Consortiums

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

2. Writing Group:

Writing group members*:
Keri Monda (ARIC)
Jennifer Nettleton (ARIC/MESA)
Gerardo Heiss (ARIC)
David Couper (ARIC)
Marilyn Cornelis (NHS/HPFS)
Neil Caporaso (PLCO)
Rob van Dam (NHS/HPFS)
Yu Kai (PLCO)

*Author group is likely to be larger given the contribution of several cohort studies to the meta-analysis. Abbreviated list noted above.

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

First author: **Keri Monda**
Address: 137 E. Franklin St., Ste. 30, Chapel Hill, NC 27514
Phone: 919-966-8481 Fax: 919-966-9800
E-mail: monda@unc.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: **Gerardo Heiss**
Address: 137 E. Franklin St., Ste. 306, Chapel Hill, NC 27514
Phone: 919-962-3252 Fax: 919-966-9800
E-mail: gerardo_heiss@unc.edu

3. Timeline: Primary data analysis and meta-analysis: 07/2010, Manuscript preparation: 08/2010 – 09/2010, manuscript submission: 09/2010

4. Rationale:

Genetic Determinants of Caffeine Consumption

Caffeine is the most widely consumed stimulant in the world [1]. Caffeine intakes vary widely from country to country and person to person. The pleasurable and reinforcing effects of caffeine have led to some concern that it is a potential drug of dependence [1,2]. However, some individuals experience adverse effects with even small amounts. Identifying the factors contributing to the habitual consumption of caffeine will enable a better understanding of its associations with health. Although demographic, psychosocial, health-related, and environmental factors such as smoking have been linked to habitual caffeine consumption, there is some evidence that genetic factors are also important. Twin studies report heritability estimates of 43-58% for caffeine use [3-5], 77% for its heavy use [3], and 45, 40, and 35%, respectively, for caffeine toxicity, tolerance and withdrawal symptoms [3]. Previous efforts to identify the precise genetic factors have focused on candidate genes involved in the metabolism (*CYP1A2*) of and response (*ADORA2A*) to caffeine [6]. We propose to conduct the first genome-wide association study (GWAS) of habitual caffeine consumption behavior.

References

1. Fredholm, B.B., Battig, K., Holmen, J., Nehlig, A. & Zvartau, E.E. Actions of caffeine in the brain with special reference to factors that contribute to its widespread use. *Pharmacol Rev* **51**, 83-133 (1999).
2. Strain, E.C. & Griffiths, R.R. Caffeine dependence: fact or fiction? *J R Soc Med* **88**, 437-40 (1995).
3. Kendler, K.S. & Prescott, C.A. Caffeine intake, tolerance, and withdrawal in women: a population-based twin study. *Am J Psychiatry* **156**, 223-8 (1999).
4. Hettema, J.M., Corey, L.A. & Kendler, K.S. A multivariate genetic analysis of the use of tobacco, alcohol, and caffeine in a population based sample of male and female twins. *Drug Alcohol Depend* **57**, 69-78 (1999).
5. Luciano, M., Kirk, K.M., Heath, A.C. & Martin, N.G. The genetics of tea and coffee drinking and preference for source of caffeine in a large community sample of Australian twins. *Addiction* **100**, 1510-7 (2005).
6. Cornelis, M.C., El-Sohemy, A. & Campos, H. Genetic polymorphism of the adenosine A2A receptor is associated with habitual caffeine consumption. *Am J Clin Nutr* **86**, 240-4 (2007).

Data Analysis Plan (to be executed by each cohort contributing to the meta-analysis)

GENEVA: Caffeine Working Group Analysis Plan

Exclusions

Exclude non-white individuals and individuals with missing information on caffeine consumption, age and smoking status.

Trait Creation: caffeine consumption

1. Continuous Trait (caffeine**:** caffeine mg/d).
   Normalize distribution by applying the cubic transformation:
   caffeine2=caffeine**(1/3);

Covariate Creation

1. **age**: continuous
2. **sex**: dichotomous, 0: male 1: female
3. **case_status**: dichotomous, 0: diseased 1: non-diseased (if applicable: study-specific)
4. smoke: categorical (dummy variables),
   smoke0: never smokers (reference)
   smoke1: former smokers
   smoke2: current smokers <20 cig/d
   smoke3: current smokers ≥20 cig/d
   Use reported smoking status closest to the time caffeine intakes were measured.
5. **EV1, EV2, EV3**..: *if appropriate*, each study should include top eigenvectors (number will vary by study) to adjust for population substructure

Running the analyses

►Analyses will be restricted to individuals of European descent (self-reported or genetically inferred). Restrict analysis to groups with sample sizes >200.
►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).
►Model: Linear regression of caffeine2 on allele dosage/genotype:

caffeine2 ~ μ + SNP + **age** + **sex** + **case_status** + **EV**s + **smoke1** + **smoke2** + **smoke3**

label this file: YOURSTUDYNAME_caffeine_smoke

Results Documentation

Variable Name	Description
marker	SNP id as rs number.
chr	Chromosome number. Use symbols X and Y for non-autosomal markers.
position	Physical position for the reference sequence (indicate build 35/36 in readme file).
effect_allele	The 'effect-allele', also referred to as 'coded-allele' or 'risk-allele'. For example: AA=0, AG=1 and GG=2, the effect_allele is G.
other_allele	The alternate allele.
strand	+ or -, representing either the positive/forward strand or the negative/reverse strand of the human genome reference sequence.
beta	Beta estimate from genotype-phenotype association --'NA' if not available.
stderr	Standard error of beta estimate --'NA' if not available.
pvalue	p-value for test statistic --'NA' if not available.
freq	Allele frequency for the effect_allele --'NA' if not available.
hwe	Exact test Hardy-Weinberg equilibrium p-value—only directly typed SNPs, 'NA'

	for imputed.
n	Total sample with phenotype and genotype for SNP.
imputed	1=imputed, 0=if directly typed. Use '0' coding when imputation was used to only 'fill-in' the missing genotypes (i.e <10% of subjects/snp).
avpostprob	Average posterior probability for imputed SNP allele dosage. Also referred to as 'Rsq" in MACH. If some other quality measure was used for imputed SNPs, please describe in readme file.

5. Main Hypothesis/Study Questions:

1. Usage of the ARIC freeze 3 GWAS data in participants of Caucasian descent for discovery of markers associated with habitual caffeine consumption. Analyses will be led by the GENEVA Consortium (lead author: Monda or TBD).

2. Follow-up analyses will include GWAS of coffee intake (related analysis to above but limited to coffee consumption) and possibly also interactions between genetic variation and caffeine and/or coffee intake. Analyses will be conducted by the CHARGE Consortium, Nutrition working group, led by Jennifer Nettleton (lead author: Nettleton or TBD).

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: Caffeine consumption was transformed using the cubic root to approximate normality. Outliers were investigated at > or < 4 SD sample mean and none were found. Final model takes the form:
Caff^(1/3) = SNP + age + sex + center + smoking + PCs + e

Inclusion/exclusion: Analyses will be limited to ARIC participants of European descent who were genotyped (freeze 3 data will be used), who passed genotype QC, and who have given consent for DNA analyses. Individuals missing data on analytic variables will be excluded.

Variables of interest: Outcome: caffeine intake in mg/day (calculated using FFQ data on consumption of coffee, tea, and soda). Dependent variable: SNP data from freeze 3. Covariables: age, sex, ARIC field center, smoking status, and principal components.

Anticipated methodologic limitations or challenges: None expected

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 ICTDER03 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

__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 #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* _____)
___ 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/

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 & GENVA 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