ARIC Manuscript Proposal # 3277

1.a. Full Title: Replication Request for: Nutrigenetics & Genomics of Sugar Sweetened Beverage Induced Metabolic Disease

b. Abbreviated Title (Length 26 characters): SSB & Genetics of CMD

2. Writing Group:
   Writing group members: Kristin Young, Danielle Haslam, Nicola McKeown, Kari North, other investigators welcome

   CHARGE Investigators: Gina Peloso, Josée Dupuis, James Meigs, Mark Herman, Caren Smith

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

First author: Kristin Young
Address: 123 West Franklin Street
        Suite 410, CB# 8050
        Department of Epidemiology
        University of North Carolina at Chapel Hill
        Chapel Hill, NC 27516

Phone: 785.318.0774       Fax: 919.966.9800
E-mail: kristin.young@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: Kari North
Address: 123 West Franklin Street
        Suite 410
        Department of Epidemiology
        University of North Carolina at Chapel Hill
        Chapel Hill, NC 27516
        CB# 8050
3. **Timeline:**

1 year (Dec 2018-Dec 2019)

Replication cohort statistical analyses: Dec 2018 Interpretation and meta-analyses: Jan 2019-Jun 2019

Manuscript preparation: July-Nov 2019 Manuscript submission: Dec 2019

4. **Rationale:** Increased consumption of SSBs is one modifiable, dietary factor linked to CVD risk (1). It is estimated that 13 to 14% of energy intake in adults is derived from added sugar, primarily consumed in the form of SSBs (2). Added sugar consumption translates into an 18% increase in CVD mortality risk (3). In epidemiological studies, sugar consumption has been positively related to an atherogenic lipid profile, i.e. hypertriglyceridemia and low HDL-cholesterol (4, 5). Both hypertriglyceridemia and low HDL are highly heritable, with estimated heritability ranging from 35-75% for triglycerides (TGs) and 40-76% for HDL-cholesterol (6, 7). Genome-wide association studies (GWAS) have identified numerous genetic variants contributing to these lipid traits, but, in total, the identified loci account for a small fraction of the estimated total heritability (8). Since current GWAS do not account for the impact of environment, interactions between environmental factors (such as diet) and genetic predisposition are likely important contributors to this missing heritability. Whether SSBs interact with specific genetic variants to increase risk is unknown, and this is the question we aim to address in the following proposal.

SSBs are composed of nearly equal portions of the simple sugars glucose and fructose. Prospective clinical experiments demonstrate that increased fructose ingestion, but not increased glucose ingestion, promotes dyslipidemia, increases visceral adiposity, and impairs insulin sensitivity in obese human subjects (9). One recent metabolic study reported that increasing amounts of SSBs lead to dose-dependent increases of circulating lipid/lipoprotein risk factors (10). Others have found no effect (11). Whether common genetic variants affect an individual’s susceptibility to sugar-induced disease is unknown.

Determining whether SSB consumption interacts with genetic variants to increase cardiometabolic risk will be important for defining mechanisms contributing to risk in order to develop novel diagnostic and therapeutic strategies. However, defining such gene-diet interactions in unbiased GWAS requires prohibitively large study populations. To overcome this barrier, we propose a candidate gene approach which directly tests the interaction of variants in genes in a biological pathway relevant to CVD with SSB intake.

In this proposal, we focus on ChREBP, a key transcriptional regulator of glucose and lipid metabolism (12). We have focused on this factor for several reasons: (1) it is clinically relevant - single nucleotide polymorphisms (SNPs) in the ChREBP locus have been directly linked to lipid
traits in humans (13-15); (2) it is biologically relevant - it senses sugar ingestion and regulates gene programs that contribute to glucose and lipid homeostasis (12); and (3) because it is a transcription factor, a tractable experimental approach can be defined to investigate its contribution to human disease. In this respect, it is notable that the vast majority of causal SNPs associated with complex disease are within enhancer regions indicating that SNPs that modulate transcription factor mediated expression of target genes are major contributors to complex disease (16). Thus, we hypothesize that variants in ChREBP binding sites likely regulate sugar-induced dyslipidemia.

5. Main Hypothesis/Study Questions:

- **Aim 1a**: To examine the cross-sectional and longitudinal associations between SSB consumption and fasting lipids (HDL and triglycerides) in the ARIC as part of the Stage 2 replication of this project in the CHARGE Nutrition working group.
- **Aim 1b**: To examine whether single nucleotide polymorphisms (SNPs) in the ChREBP locus and in ChREBP’s transcriptional targets interact with SSB consumption to regulate lipid traits (HDL cholesterol and triglyceride (TG) levels) in humans in ARIC.
- **Aim 1c**: Using meta-analysis of data from participating cohorts in the CHARGE Nutrition working group, determine the magnitude of association between SSB consumption and lipids and examine whether the effect of SSB consumption on lipids is modified by polymorphisms in ChREBP locus and in ChREBP’s transcriptional targets.

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).

**Outcomes and measures**: The metabolic traits of interest will include fasting TGs (mg/dL) and fasting HDL-cholesterol (mg/dL). To achieve normality for statistical testing with continuous variables, natural logarithmic transformation will be applied to fasting TGs and HDL cholesterol prior to analysis. Variables for capturing fasting status differed across exams.

**Covariates**: Age (y), sex (male/female), energy intake (kcal), smoking status, education status, alcohol intake, physical activity, fruit intake, vegetable intake, whole grain intake, fish intake, and saturated fatty acid intake (% of total energy) will be examined as covariates.

**Estimation of SSB**: Estimates of SSB consumption will include the following categories (1) Coke, Pepsi, or other cola with sugar; (2) caffeine free Coke, Pepsi, or other cola with sugar; (3) other carbonated beverage with sugar (e.g., 7-Up, ginger ale); and (4) Hawaiian Punch, lemonade, or other non-carbonated fruit drinks. One serving of SSB is equivalent to 360 mL (12 fl oz.). Other descriptions may include: regular soda/fruit flavored drink; Tang, Start breakfast drinks; regular/carbonated soft drinks; syrups/soda; Hi-C, Kool Aid, or other drinks with added vitamin C/regular soft drinks. In some cohorts, the FFQ has a line item for sugary sodas or soft drinks...
that includes fruit juice. In these instances, SSB will be estimated including the fruit juice because we cannot isolate it from the other beverage sources.

**Inclusions:**
- European ARIC participants

**Exclusions:**
- Implausible dietary intake (cohort-specific definition)
- Missing age
- Missing sex
- Non-fasting status (NOTE: if a large portion of your cohort has non-fasting sample, we may want to report results separately for fasting/non-fasting – contact)
- Implausible dietary data
- <18 years old
- NOTE: If no exclusions apply, individual will be included in ‘Model 1’ analyses. For subsequent models, individuals missing any covariates should be excluded from that model.

Selection of SNPs: We will employ two distinct strategies to select SNPs, and we will generate a list of SNPs.

1. We will test SNPs in the ChREBP locus because SNPS in the ChREBP locus are associated with serum TG and HDL-cholesterol levels (13, 14). To do this, we will comprehensively test haplotypes within this locus for interactions with SSBs on lipid risk factors. We have already defined these haplotypes by examining a region 100kb upstream and 190kb downstream of the ChREBP (Mlxipl) transcriptional start site. This encompasses the region that includes SNPs that associate with hypertriglyceridemia. We used the following five European populations from the 1000 Genomes Project (http://www.1000genomes.org/home) to generate the list of SNPs within this region: (1) Utah Residents with Northern and Western European Ancestry (CEU), (2) Finnish in Finland (FIN), (3) British in England and Scotland (GBR), (4) Iberian population in Spain (IBS) and (5) Toscani in Italia (TSI). The min and max minor allele frequencies (MAFs) were set at 0.05 and 0.98, respectively. Using this approach, we identified 483 SNPS. We then assessed these SNPS for linkage disequilibrium (LD) in order to reduce the statistical burden arising from multiple testing. We identified SNPs in LD in the CEU population at r²> 0.8 with 1000 Genomes data resulted in 27 LD blocks (2-SNP minimum size), of which 13 contain >3 unique SNPs each. We identified five LD blocks of 20 or more unique SNPs each of which, in total, amount to 199 (41%) of the group of ChREBP SNPs. In summary, there are 27 LD blocks in total. We will select a unique SNP to “tag” each LD block and use this SNP to test for the interaction with SSB on fasting lipid traits. Within an LD block, we will prioritize SNPs that have been linked previously to a specific lipid phenotype.

2. We will select SNPs in ChREBP transcriptional targets. These will be identified in collaboration with Dr. Herman’s research group. We propose to integrate the anti-ChREBP ChIP-sequence data and RNA-sequence data he generates from human liver samples with publicly available polymorphism data from the 1000 Genomes Project to identify genes and
genomic variants that may participate in the pathogenesis of sugar-induced dyslipidemia. We will use SNPs identified via this analysis for further testing in the ARIC/CHARGE analyses.

**Genotyping/Imputation:** We will use the HRC imputed ARIC data for these analyses. SNPs will be excluded on the basis of low call rate (<95%), and departure from Hardy-Weinberg equilibrium (<1E-06), while imputed SNPs will be removed on the basis of low imputation quality (17).

**Statistical Analysis:** Main associations between SSB and metabolic trait (HDL and TG) will be quantified adjusting for:

- (model 1) age, sex and energy intake, and study site
- (model 2) model 1 plus smoking status, education status, physical activity, and alcohol intake
- (model 3) model 2 plus BMI
- (model 4) model 3 plus fruit intake, vegetable intake, whole grain intake, fish intake, and saturated fatty acid intake (% of total energy)

These 4 models will also be used to quantify the longitudinal association between SSB intake and change in metabolic traits (HDL and TG) within ARIC, additionally adjusting for baseline levels of the respective metabolic trait.

**SNP association models:** To examine the effects of our candidate SNPs on lipid traits, we will use linear regression to model the HDL and TG as a function on the SNPs of interest using an additive genetic model (per additional risk allele) adjusted for age and sex and, where relevant, center and/or population substructure.

**SSB-SNP interaction models:** We will investigate SSB-SNP interactions by including a first-order interaction term (SSBxSNP) in:

- model 1: SSB intake, SNP, age, sex, energy intake, center, SSB*SNP
- model 2: model 1 plus BMI

For significant interaction terms, we will include covariates in a final model (model 4), as listed above.

As an exploratory aim, because BMI is an important determinant of circulating lipids, we will further explore associations by stratifying subjects into groups based on BMI. We will repeat the described analyses as outlined above, but within three BMI based categories: normal weight (BMI > 18.5 and <25 kg/m²), overweight (BMI >25 and ≤30 kg/m²), and obese (BMI > 30 kg/m²).

**Meta-Analyses:** We will conduct inverse variance-weighted fixed-effect meta-analyses using the ‘metafor’ R package for (1) main associations of SSB intake on lipid traits; (2) main associations of SNPs with respective outcomes using METAL (version released 2011-03-25)(18); and (3) interactions between SNPs and SSB intake on respective outcomes using METAL. The beta
coefficients and standard errors (SE) will be standardized to a serving of SSB intake. Heterogeneity across studies will be tested using Cochran’s Q statistic and quantified using the I2 statistic(19). Statistical significance for each outcome will be defined based on Bonferroni correction for the total number of interaction tests. Statistical analysis: Model 1 will be the primary model with the other models used to elucidate the impact of lifestyle factors on the SSB-trait associations.

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/mantrack/maintain/search/dtSearch.html

___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)? MP #2786

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 https://www2.cscc.unc.edu/aric/approved-ancillary-studies
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.cscce.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.

References:


