ARIC Manuscript Proposal #4250

PC Reviewed: 5/09/23 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: Genes, Proteins, and Metabolites Associated with Erythritol and Erythronate Levels and their Association to Atherosclerotic Cardiovascular Disease and Heart Failure in the Atherosclerosis Risk in Communities Study (ARIC)

b. Abbreviated Title (Length 26 characters): Erythritol, Erythronate, and ASCVD/HF

2. Writing Group:

Writing group members: Layla A. Abushamat, MD, MPH, Ron C. Hoogeveen, PhD, Caroline Sun, MPH, Chao Cheng, PhD, Sean Hartig, PhD, Mark Herman, MD, Jane EB Reusch, MD, Bing Yu, PhD*, Christie M. Ballantyne, MD*
*co-supervising

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

First author: **Layla Abushamat,**
Address: Baylor College of Medicine
One Baylor Plaza, BCM 285
Houston, TX 77004

Phone: 713-798-4625 Fax: 337-270-2726
E-mail: Layla.abushamat@bcm.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: **Christie M. Ballantyne**
Address: Baylor College of Medicine
One Baylor Plaza, BCM 285
Houston, TX 77030
Phone: Fax:
E-mail: cmb@bcm.edu

3. Timeline: After data is obtained, data analysis and manuscript preparation will be done within 12 months.

4. Rationale:

There has been increased incidence and prevalence of obesity over the last few decades with a concurrent rise in type 2 diabetes (1). With this increase in metabolic disease, artificial

sweeteners have been in widespread use, one of which is erythritol (2). Erythritol was introduced into food production in the 1990s and FDA-approved in the U.S. in 2002 as an artificial sweetener. Along with its excellent tolerability and short-term risk profile, it was also touted as "natural" as it is produced endogenously in human tissue through the pentose phosphate pathway and is a naturally occurring sugar alcohol (3). Metabolomic analysis of ARIC samples collected between 1987 to 1989 (visit 1) has linked erythritol to coronary heart disease (CHD)(4). Artificial sugar sweetener use has also been linked to increased CHD (5). At the same time, it is difficult to distinguish between endogenous production and exogenous intake. Interestingly, endogenous production through the pentose phosphate pathway is induced by glucose intake, is associated with weight gain and higher glycemia, and metabolism of erythritol to erythronate occurs through erythrose-4-phosphate (6). In the ARIC visit 1 analysis, the association with CHD persisted even when adjusting for body mass index (4). Importantly, the metabolomic results from visit 1 would have been from a time prior to erythritol entering U.S. food production or being FDA-approved as a sweetener.

The inspiration for this proposal is a recent study that showed the association of erythritol metabolite levels with increased incident 3-year risk of major adverse cardiovascular events, an association that was confirmed in two validation cohorts of stable patients undergoing elective cardiac evaluation. This same study also showed increases in plasma erythritol levels for 2 days after consumption of foods sweetened with the artificial sweetener, prompting concern that erythritol intake can impact metabolite analysis of samples even in the fasted state (7). Furthermore, preliminary analysis of metabolites associated with adiponectin and the high risk adiponectin/NTproBNP phenotypes in ARIC visit 5 by our group has shown a strong association of erythronate, a downstream product of erythritol metabolism, to a high adiponectin/high NTproBNP phenotype, which has a strong association to future heart failure hospitalization and CVD mortality. Analysis of association of erythronate and erythritol ARIC visit 5 to CV and HF outcomes shows a strong association of both of these metabolites to future events. Further investigation of genes, proteins, and metabolites associated with erythronate and erythritol may shed light on these associations and their causality. Given its production through the pentose phosphate pathway, high levels may be due to increased oxidative stress, rather than due to exogenous intake. Understanding the mechanisms involved will inform study design to further tease out the contributions of exogenous intake vs endogenous production.

5. Main Hypothesis/Study Questions: What genes, proteins, metabolites and pathways are associated with erythritol and erythronate? Is there a causal relationship between erythritol or erythronate to CVD outcomes? Is there a change in erythronate profiles with time (potentially attributable to addition to the diet)?

1. Assess genes associated with erythritol and erythronate levels and use Mendelian Randomization to examine causal relationships between select genes and cardiovascular outcomes .
2. Define proteomic signatures associated with erythritol and erythronate and identify novel pathways involving significant proteins at visit 2 and visit 5.
3. Assess metabolomic signatures associated with erythritol and erythronate and evaluate novel pathways involving significant metabolites at visit 3 and visit 5.

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).

Inclusion: All ARIC participants with proteomic and metabolomic analysis at ARIC visits 2,3, and 5.

Exclusion: Missing information on proteomics/metabolomics, missing information on covariates of interest.

Methods: Erythritol and erythronate concentrations at visit 5 were measured using Metabolon For proteomic analysis, 5000 plasma proteins were measured using a multiplexed modified DNA-based aptamer (SOMAscan) technology at visit 5.(8) Briefly, a slow off-rate modified aptamer (SOMAmer) reagents (SomaLogic, Inc, Boulder, Colorado) captures proteins from blood samples, then the SOMAmer reagents were measured in fluorescent arrays. The relative concentration of proteins was then derived from the concentration of SOMAmer reagents. Metabolomic profiling was performed on visit 5 samples (stored in -80° C since collection) by Metabolon Inc (Durham, NC) using untargeted, gas chromatography mass spectrometry and liquid chromatography-mass spectrometry-based quantification protocol.(9) Metabolite levels were winsorized at the 1st and 99th percentile (with missing/below detection limit values given the lowest detected value of that metabolite) and standardized prior to analysis.

Outcome: Outcome measures include incident heart failure hospitalization, incident coronary heart disease, incident stroke, recurrent cardiovascular events, cardiovascular disease mortality, and all-cause mortality. Incident heart failure hospitalization was ascertained by expert panel adjudication as beginning on January 1, 2005, HF was further defined as HFpEF vs HFrEF. (10) Incident CHD events encompassed fatal CHD and definite or probable MI. (11) Incident stroke events included ischemic strokes and definite or probable hospitalized embolic or thrombotic strokes based on diagnosis codes as well as hospital records and neuroimaging reports. (12) CVD deaths were ascertained by diagnostic codes from hospital discharge records and from death certificates. The cutoff date for administrative censoring for those without events was December 31, 2017.

Covariates: Age, race, sex, education level (data collected at Visit 1), smoking status, BMI, systolic blood pressure (SBP), hypertension treatment, diabetes, total cholesterol, high-density lipoprotein (HDL) cholesterol, lipid treatments, waist-to-hip ratio (WHR), thyroid stimulating hormone (TSH), and estimated glomerular filtration rate (eGFR) by CKD-EPI equations (using creatinine, cystatin C, or both) at visit 5.

Analysis Plan:

<u>Aim 1:</u>

We will obtain genetic summary statistics of erythronate and erythritol from publicly available resources and past ARIC metabolite genome-wide association studies and meta-analyze those summary to produce pooled statistics for erythronate and erythritol. . We will then conduct 2-sample MR analyses to assess causality to outcomes of interest. All statistically independent and significant variants will be assessed for every trait of interest. A causal estimate will be

determined for each independent variant based on a ratio of association to disease compared with published GWAS summary statistics (16, 17, 18).

Aim 2 and 3:
Erythritol and erythronate will be modeled as both a continuous variable and a categorical variable (quartiles). For the categorical analysis, the lowest quartile of erythritol and erythronate will be used as the reference group. Linear regression models will be constructed to assess associations between proteins and metabolites at visit 5 with erythritol and erythronate. As discovery process, we will build a univariate model, a model adjusting for sociodemographic variables (age, sex, race, center, education level) and a full model with all the covariates mentioned above. Interaction will be tested for covariates and stratified analysis will be performed. We will apply Benjamini-Hochberg procedure to all models in order to control the number of false positives during multiple comparisons, with false discovery rate (FDR) controlled at 0.05.(19) Those proteins and metabolites that meet the FDR threshold will be inputted into an ingenuity pathway analysis (IPA) tool that will output associated canonical pathways.
To assess those proteins and metabolites associated with outcomes of interest, we will apply Cox proportional hazard regression models with least absolute shrinkage and selection operator (LASSO) procedures and fit LASSO with 10-fold cross-validation to incorporate proteins/metabolites from Aim 1 and covariates of interest. LASSO-selected proteins and metabolites will then be inputted into Cox models for outcomes of interest that adjust for selected covariates. Associated canonical pathways will be scored based on the set of proteins and metabolites involved with each pathway then modeled to outcomes of interest to elucidate the degree with which these pathways are associated with cardiometabolic risk. IPA will group related proteins and metabolites into networks and identify which networks are independently associated with cardiovascular risk.(20) We will also construct and standardize a continuous metabolomic risk score based on significant metabolites and their direction of effect on outcomes of interest. The continuous metabolomic risk score and metabolomic risk score quartiles will be assessed for association with incident outcomes of interest using Cox regression models.

Limitations: This is a secondary analysis, and residual confounding is a concern. Attempts to overcome this concern include performing the analysis on half the sample and validating using the other half. Additionally, both erythronate and erythritol levels are affected by exogenous intake of erythritol, fructose, and glucose. While the samples used for metabolomic and proteomic analysis are taken while fasting, it would be difficult to elucidate the impact of exogenous intake. We do have food diaries available in ARIC, however possible data that could be obtained from these logs are likely limited and subject to recall bias.

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? ____ Yes __x__ No

b. If Yes, is the author aware that the current derived consent file ICTDER05 must be used to exclude persons with a value RES_OTH and/or RES_DNA = "ARIC only" and/or "Not for Profit" ? __x__ Yes ____ No
(The 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 current derived consent file ICTDER05 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)? none

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
(Metabolite CVD ancillary study AS#2020.21)

___ 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 https://sites.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.cscc.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:

1. National Diabetes Statistics Report. Centers for Disease Control and Prevention: Atlanta, GA, 2017.

2. Sylvetsky AC, Jin Y, Clark EJ, Welsh JA, Rother KI, Talegawkar SA. Consumption of Low-Calorie Sweeteners among Children and Adults in the United States. *J Acad Nutr Diet* 2017;**117:** 441-448 e442.

3. Mazi TA, Stanhope KL. Erythritol: An In-Depth Discussion of Its Potential to Be a Beneficial Dietary Component. *Nutrients* 2023;**15**.

4. Wang Z, Zhu C, Nambi V, Morrison AC, Folsom AR, Ballantyne CM, *et al.* Metabolomic Pattern Predicts Incident Coronary Heart Disease. *Arterioscler Thromb Vasc Biol* 2019;**39:** 1475-1482.

5. Debras C, Chazelas E, Sellem L, Porcher R, Druesne-Pecollo N, Esseddik Y, *et al.* Artificial sweeteners and risk of cardiovascular diseases: results from the prospective NutriNet-Sante cohort. *BMJ* 2022;**378:** e071204.

6. Hootman KC, Trezzi JP, Kraemer L, Burwell LS, Dong X, Guertin KA, *et al.* Erythritol is a pentose-phosphate pathway metabolite and associated with adiposity gain in young adults. *Proc Natl Acad Sci U S A* 2017;**114:** E4233-E4240.

7. Witkowski M, Nemet I, Alamri H, Wilcox J, Gupta N, Nimer N, *et al.* The artificial sweetener erythritol and cardiovascular event risk. *Nat Med* 2023;**29:** 710-718.

8. Tin A, Yu B, Ma J, Masushita K, Daya N, Hoogeveen RC, *et al.* Reproducibility and Variability of Protein Analytes Measured Using a Multiplexed Modified Aptamer Assay. *J Appl Lab Med* 2019;**4:** 30-39.

9. Perkel J. Metabolomics: where seeing is believing. *Biotechniques* 2011;**50:** 285-289.

10. Rosamond WD, Chang PP, Baggett C, Johnson A, Bertoni AG, Shahar E, *et al.* Classification of heart failure in the atherosclerosis risk in communities (ARIC) study: a comparison of diagnostic criteria. *Circ Heart Fail* 2012;**5:** 152-159.

11. Rosamond WD, Chambless LE, Folsom AR, Cooper LS, Conwill DE, Clegg L, *et al.* Trends in the incidence of myocardial infarction and in mortality due to coronary heart disease, 1987 to 1994. *N Engl J Med* 1998;**339:** 861-867.

12. Rosamond WD, Folsom AR, Chambless LE, Wang CH, McGovern PG, Howard G, *et al.* Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. *Stroke* 1999;**30:** 736-743.

13. Cardiovascular Disease Knowledge Portal: A platform for accelerating anlysis of the genetics of cardiovascular disease. Available from: https://cvd.hugeamp.org.

14. Churchhouse C, Neale B. Rapid GWAS of Thousands of Phenotypes for 337000 samples in the UK Biobank. Available from: http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank.

15. American Heart Association Precision Medicine Platform: Diverse Datasets. Data Filtering. Data Access. Available from: https://precision.heart.org.

16. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015;**47:** 1121-1130.

17. Loley C, Alver M, Assimes TL, Bjonnes A, Goel A, Gustafsson S, *et al.* No Association of Coronary Artery Disease with X-Chromosomal Variants in Comprehensive International Meta-Analysis. *Sci Rep* 2016;**6:** 35278.

18. Scott RA, Scott LJ, Magi R, Marullo L, Gaulton KJ, Kaakinen M, *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* 2017;**66:** 2888-2902.

19. Banjamini Y, Hochbery Y, . Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* 1995;**57:** 289-300.

20. Walker KA, Chen JK, Zhang J, Fornage M, Yang Y, Zhou L, *et al.* Large-scale plasma proteomic analysis identifies proteins and pathways associated with dementia risk. *Nature Aging* 2021;**1:** 473-489.