1.a. Full Title:
“Association of habitual milk intake with mitochondrial DNA copy number in ARIC.”

b. Abbreviated Title (Length ___ characters): Milk intake and mtDNA copy number.

2. Writing Group:
Natalia Petruski-Ivleva, Dan Arking, Priya Palta, David Couper, Katie Mayer, Ryan Longchamps, Gerardo Heiss, Anna Kucharska-Newton, others welcome.

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

First Author: Natalia Petruski-Ivleva, MS

Address: University of North Carolina – Department of Epidemiology
137 E. Franklin St, Suit 306
Chapel Hill, NC 27514-4145

Phone: 413-230-1522
Email: petruski@email.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: Anna Kucharska-Newton, Ph.D.
Address: Collaboration Studies Coordinating Center
University of North Carolina – Department of Biostatistics
137 E. Franklin St, Suit 306
Chapel Hill, NC 27514-4145

Phone: 919-966-4564 Fax: 919-966-9800
E-mail: anna_newton@unc.edu

3. Timeline: Analysis to be completed immediately. First draft completed by March 2017.
4. Rationale:
Dairy foods such as milk, cheese, and yogurt are consumed by billions of people around the world. Milk plays an important part in childhood development due to its high caloric, fat and protein content. It contributes to absorption of many other nutrients, such as calcium, magnesium, selenium, riboflavin, and vitamin B12. Currently, the USDA recommends 2 daily servings of dairy for children and 3 servings for adults for optimal health. It is estimated that milk accounts for 51% of all dairy intake. Despite the nutritional benefits that milk can provide for the growth and development of children, its effect on health in adults is less well known.

Lactose is the main carbohydrate in milk. It has been suggested that D-galactose, a metabolic derivative of lactose, can contribute to increased reactive oxygen species (ROS) production resulting in increased oxidative stress. Oxidative stress, the state of physiological imbalance between oxidant production and antioxidant defense, is implicated in the development of many pathological conditions such as obesity, type 2 diabetes, aging, neurodegenerative diseases, and cancer initiation and progression. Based on extensive animal studies, a dose of D-galactose, equivalent to 1-2 glasses of milk for humans, is sufficient to induce memory deficits, decreases the number of new neurons, and increases oxidative stress levels in mice when administered for 5-7 weeks. In human studies, high milk intake has been associated with higher mortality in two large prospective studies and with higher fracture incidence in women. Higher milk consumption has also been posited to influence the risk of certain cancers and of cardiovascular disease.

Systemic oxidative stress has been analyzed in serum and blood cells using different biomarkers. Mitochondrial DNA (mtDNA), a circulating, multicopy cytoplasmic DNA, semiautonomously maintained in mitochondria, is known to be more sensitive to oxidative damage than nuclear DNA and has been increasingly used for the assessment of systemic oxidative stress. It has been shown that ROS can cause damage to mitochondrial enzymes, resulting in mtDNA mutations, alterations in mitochondrial membrane permeability, and cell death. These mitochondrial defects have been attributed to reduced mtDNA content, expressed by a lower mtDNA copy number. MtDNA copy number has been examined in relation to many conditions associated with oxidative damage, such as frailty and all-cause mortality, general health among older adult populations, diabetes, several types of cancer, and neurodegenerative diseases. However, no studies have assessed the association of milk intake with mtDNA copy number. We propose to measure this association in the Atherosclerosis Risk in Communities (ARIC) cohort. Milk intake by ARIC participants was measured on two occasions (visit 1 and visit 3) by Food Frequency Questionnaire (FFQ) with 9 response options ranging from “Almost never” to “more than 6 glasses a day”, which will be regrouped into 4 (almost never, less than 1 glass a day, 1 glass a day, more than 1 glass per day). MtDNA copy number was assessed at visit 2. Thus we propose to study the cross-temporal association of milk intake at visit 1 and mtDNA copy number measured at visit 2. Milk intake at visit 3 will be incorporated in a sensitivity analysis to account for reporting error, where only those participants who’s milk intake fell into the same category as the first visit would be included in the analysis.
intake category on two reporting occasions will be included. In addition, a stratified analysis will be performed using data on lactase persistent/non-persistent genotype in order to detect potential differences in the association of interest due to differences in metabolic pathways of milk digestion (lactose digestion by lactase versus colonic adaptation)\(^4\).

5: Main Hypothesis/Study Questions:

Our goal is to examine the association of habitual milk intake with levels of oxidative stress, ascertained by mtDNA copy number, in the overall ARIC cohort. We will also perform analyses stratified by lactase persistent/non-persistent genotype.

**Hypothesis:** Milk intake is inversely associated with mtDNA copy number.

Sensitivity analysis will be performed by excluding those participants whose reported milk intake did not fall into the same category visit 1 and visit 3.

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 methodological limitations or challenges if present).

**Study population**

The study population will include participants of The Atherosclerosis Risk in Communities Study (ARIC) prospective cohort who completed assessments of dietary intake by FFQ at Visit 1 and who have data on mtDNA copy number assessed at visit 2. Participants with a history of cancer, diabetes, myocardial infarction, stroke, or heart failure at the Visit 2 will be excluded because those health conditions are strongly associated with increased oxidative stress and deceased mtDNA copy number. Individuals at the extremes of caloric intake distribution will also be excluded from the analysis.

**Assessment of milk intake**

Milk intake will be measured at Visit 1 and Visit 3 as the reported average number of glasses of milk per day. Intake of both low-fat/skim and whole milk will be treated as a combined exposure. Habitual intake will be estimated by reported intake at Visit 1.

**Assessment of mtDNA copy number in ARIC**

MtDNA copy number (mtDNA-CN) was determined utilizing the Genvisis software package. First, a list of high-quality mitochondrial SNPs were hand-curated removing SNPs which may cross-hybridize to the nuclear genome. The probe intensity of the remaining 25 SNPs was determined using quantile sketch normalization (apt-probeset-summarize) as implemented in the
Affymetrix Power Tools software. The median of the normalized intensity, log R ratio (LRR)(PennCNV-Affy Pipeline) for all homozygous calls was GC corrected and used as an initial estimate of mtDNA-CN for each sample. Principal components were used to correct for DNA quality, DNA quantity, and other technical artifacts. The estimate of mtDNA copy number was validated using quantitative polymerase chain reaction.

Assessment of LP/LNP status
ARIC participants were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 and the IBC chip array (Affymetrix, Santa Clara, CA, USA). Genotypes were excluded for call rates <90%, MAF (minor allele frequency) <1%, Hardy–Weinberg equilibrium deviation <10^-06, and genotype frequency that was different at P<10^-6 from prior genotyped samples. Principle components were generated using the Eigensoft package (http://genepath.med.harvard.edu/~reich/Software.htm) and ancestry outliers were removed. The final sample with genetic data used for imputation was 9713 EA and 2871 AA. Imputation was performed in two steps: (1) Pre-phasing with ShapeIt (v1. r532 ) (2) Imputation with IMPUTE2. After frequency and genotyping pruning, there were 695,783 SNPs in the final set used for the imputation (669,450 autosomal SNPs). Final imputations were performed using IMPUTE2 based on the 1,000 Genomes Phase I integrated variant set release (v3) in NCBI build 37 (hg19) reference panel haplotypes. All 1092 individuals were used for the imputation from the reference panel. The final sample with genetic data used for imputation was 9713 EA and 2871 AA25. In this analysis we will use the imputed genotype LCT-13910 C/T (polymorphism (rs4988235) upstream from the lactase (LCT) gene) in EA and LCT-14010G/C (polymorphism (rs145946881)) in AA26.

Statistical analysis
Characteristics of the study population will be presented by categories of milk intake (none, less then 1 glass per day, 1 glass per day, more than 1 glasses per day). MtDNA copy number will be divided into quintiles. LP/LNP genotype will be defined as LP for TT/CT genotypes in Whites and CC/CG genotypes for Blacks. Genotypes CC in Whites and GG in Blacks will be defined as LNP.

Multinomial logistic regression will be used to ascertain the association of milk intake (almost never, less than 1 glass per day, 1 glass per day, more than 1 glass per day) with quintiles of mtDNA copy number. Covariates considered for inclusion in the model will be age, race, sex, ARIC study center, total energy intake, attained educational level, cigarette smoking (current, former, never), alcohol consumption (grams/week), physical activity (Baecke’s physical activity summary score), body mass index (kg/m²), and diet quality score.

Analyses will be stratified by LP/LNP genotype.
A limitation of the study is the use the FFQ and self-reported intake to assess habitual milk intake. A second limitation is the assumption that milk intake at visit 1 remained the same at the time of mtDNA copy number assessment (at visit 2). These issues will be addressed by using data from visit 3 FFQ in a sensitivity analysis where only those participants who remained in the same milk intake category on both assessments will be included (approximately 70% of participants). There is also the potential for residual confounding by other dietary components, which will be addressed by adjusting for diet quality score.

7.a. Will the data be used for non-CVD analysis in this manuscript? __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 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 AIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)? None noticed

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use ancillary 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)*_______________________________)
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-year from the date of approval, the manuscript proposal will expire.

References


