ARIC Manuscript Proposal # 3233

PC Reviewed: 9/11/18	Status: _____	Priority: 2
SC Reviewed: _____	Status: _____	Priority: ____

1.a. Full Title:

Mitochondrial DNA Copy Number and incident cancer: the Atherosclerosis Risk in Communities (ARIC) Study

b. Abbreviated Title (Length 26 characters):

mtDNA copy number and cancer

2. Writing Group:
Writing group members:

Di Zhao[1], Eliseo Guallar[1], Elizabeth A. Platz,[2] Corinne E. Joshu,[2] Ryan J. Longchamps[3], Christina A. Castellani[3], Yunsoo Hong[1], Brian O'Rourke[5], Dan E. Arking[3]

[1] Departments of Epidemiology and Medicine, and Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD.
[2] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.
[3] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.
[4] Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California, USA.
[5] Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[6] Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis

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

First author:
Name: Di Zhao
Address: 2024 E. Monument Street, Room 2-635, Baltimore, MD 21205.
Phone: 443-287-4678
E-mail: dizhao@jhu.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: Dan E. Arking
Address: 733 N Broadway, BRB 447
Baltimore, MD 21205
Phone: (410) 502-4867 Fax: (410) 614-8600
E-mail: arking@jhmi.edu

3. Timeline:

The manuscript will be complete within 2-3 months upon the approval of this proposal.

4. Rationale:

While significant progresses have been achieved in cancer screening and treatment during the past decades, cancer remains the second leading cause and responsible for 25% of all deaths in the US.[1] Identifying biomarkers for cancer susceptibility or progression in the general population, particularly factors predicting future cancer risk, is crucial in understanding cancer mechanisms, and facilitating the development of novel screening and therapeutic tools .

Mitochondria are critically important in energy production and homeostasis because of their role in adenosine triphosphate (ATP) production via the oxidation phosphorylation (OXPHOS) pathway. Decline in mitochondrial function is associated with aging and multiple pathological changes that increase vulnerability to chronic disease, including initiation and progression of atherosclerosis, cardiovascular disease, and mortality.[2, 3] Mitochondria maintain their own DNA (mtDNA), encoding 37 genes, 22 mitochondrial transfer RNAs, and two ribosomal RNAs.[4] Mitochondrial DNA copy number (mtDNA-CN) can be measured from peripheral blood using a low-cost approach, and has been established as an indirect marker of mitochondrial function.[5] Low mtDNA-CN, indicating mitochondrial dysfunction, was associated with diabetes, chronic kidney disease, sudden cardiac death, CVD, frailty and mortality in general population.[6-10]

Changes in mitochondria may play a role in carcinogenesis. Epidemiological studies suggested that increased mtDNA-CN in peripheral blood was associated with increased risk of multiple malignancies, including breast cancer,[11, 12] non-Hodgkin lymphoma,[13] lung cancer,[14] pancreatic cancer,[15] and colorectal cancer[16]. Other studies have controversial results, showing an inverse association with breast cancer,[17] renal cancer[18] and null association with gastric cancer.[19] It is hypothesized that high mtDNA-CN in peripheral blood reflected a compensatory response to deletion of mutant mtDNA,[20] which predicted poor

cancer prognosis and survival, and this hypothesis was supported by a recent meta-analysis.[21] However, the current studies are mostly conducted in Asian population, and are limited by small sample size, case-control design, and specific subtypes of cancers. The association between mtDNA-CN and cancer in African Americans remains unknown. Additionally, only the association between mtDNA-CN and cancer-free survival was examined, but the association between mtDNA-CN and incident cancer remain unexplored.

Hence, in the present study, we will examine the prospective association between baseline mtDNA-CN and the risk of cancer and cancer-specific mortality among participants from the Atherosclerosis Risk in Communities (ARIC) study.

5. Main Hypothesis/Study Questions:

- To estimate the association between mtDNA-CN with incidence of overall, and each subtype of cancer.

  We hypothesize that mtDNA-CN will be significantly associated with cancer, with variations in subtypes.

- To estimate the association between mtDNA-CN with cancer-specific mortality.

  We hypothesize that mtDNA-CN will be positively associated with cancer-specific mortality.

- To estimate the association between mtDNA-CN with case-fatality.

  We hypothesize that mtDNA-CN will be positively associated with case-fatality.

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: prospective study.

Inclusion/exclusion: We will include Whites and Blacks who have mtDNA-CN available. We will exclude non-whites and non-blacks, Blacks from the Minneapolis or Washington Co. Field center, as well as participants with prevalent cancer at the time of mtDNA-CN measurements, and participants missing cancer data or other covariates.

Primary endpoints:

- Cancer incidence overall and separately by sex
- Cancer mortality
    - Cancer site
    - Lung and bronchus
    - Prostate
    - Breast (female, post-menopausal)
    - Colon and rectum
    - Pancreas
    - Uterine
    - Kidney
    - Any other site with at least 100 cases
- Cancer groupings (obesity associated-post-menopausal breast, colon, pancreas, endometrial, kidney, gall bladder, liver; smoking associated-lung, head/neck, pancreas, kidney; female: breast, endometrial, ovary, cervical, smoking-associated).

The collection of cancer information has been described previously.[22] We will use the ARIC cancer case files, which were developed using data from the MN, NC, MD, and MS state cancer registries, medical records, and hospital discharge codes. We will use first primary cancer cases and cancer deaths occurring after the visit at which mtDNA-CN was measured through 2012. Total cancer incidence is defined as the first primary cancer diagnosis (excluding prostate cancer – prostate cancer does dominate and most of the cases are very early disease that would not have been detected in the pre-PSA era and may never have become clinically apparent, but might be good to look with and without prostate cancer included anyway) in men and women free of any cancer diagnosis at baseline. Total cancer mortality is defined as the death from any cancer, listed as the underlying cause of death on death certificates. Site-specific cancer incidence and mortality will be evaluated for individual cancers that have sufficient number of events.

Primary exposure:

- mtDNA Copy Number

DNA samples were isolated from buffy coat and gynotyped on the Affymetrix Genome-Wide Human SNP Array 6.0 (www.genvisis.org).[6, 7] Mitochondrial SNPs were collected across all samples and were signaled with 25 high-quality mitochondrial probes. Raw mtDNA CN was determined by the median probe intensity difference across all mitochondrial SNPs. To correct for technical artifacts, batch effects, DNA quality, and starting DNA quantity, surrogate variable analysis was applied to probe intensities of 43,316 autosomal SNPs.[23] We calculated residuals using a linear regression model with raw mtDNA CN as the dependent variable and the surrogate variable, age, sex, and

enrollment center as independent variable, The calculated residuals were then used as measurement for mtDNA-CN for all subsequent analyses.

Statistical Analyses

DNA for mtDNA-CN analysis was collected in visit 1 (1987-1989) for 484 participants (4.2%), visit 2 (1990-1992) for 9,112 participants (79.6%), visit 3 (1993-1995) for 1,791 participants (15.6%), and visit 4 (1996-1998) for 66 participants (0.6%). The visit of DNA collection for each participant will be considered the baseline visit and all covariates will be obtained from that visit. Follow-up for events will start from the date of DNA collection and will continue until incident cancer, death due to another cause, or December 31, 2012, whichever comes first.

Baseline characteristics of the study population will be compared across quintiles of mtDNA-CN. We will use Cox proportional hazards regression to estimate hazard ratios (HR) and 95% confidence intervals (CI) for the association between mtDNA-CN and incident cancer and cancer mortality. We will test proportional hazard assumption using Schoenfeld residuals. In the primary analysis, we will categorize mtDNA-CN into quintiles based on the analytic cohort distribution. Tests for linear trend across quintiles will be conducted. In secondary analysis, we will model mtDNA as a continuous variable and estimated the HR for cancer comparing the 10^{th} to the 90^{th} percentile of mtDNA-CN, assuming a linear association on the natural logarithm scale?. Additionally, we will model mtDNA-CN using restricted cubic splines with knots at the 5^{th}, 35^{th}, 65^{th} and 95^{th} percentiles of its distribution to provide a smooth yet flexible description of the dose-response relationship between mtDNA-CN and cancer.

To adjust for potential confounders, we will use 3 multivariate models with progressive degrees of adjustment as described below:

Model 1: age, sex and race/field center groups.

Model 2: model 1 + known or suspected common cancer risk factors, including body mass index, smoking, packyears smoked, alcohol intake, physical inactivity.

Model 3: model 2 + possible cancer risk and protective factors, including total and HDL cholesterol, cholesterol medication, systolic and diastolic blood pressure, hypertensive medication, diabetes, ever use of hormone replacement therapy (women only).

Additionally, we will perform pre-specified subgroup analyses by race, sex, other comorbidities, including diabetes, hypertension, and cardiovascular disease, and will test for potential interactions. All statistical analyses will be performed using STATA version 15 (StataCorp LP, College Station, Texas), with 2-sided test and $p<0.05$ considered to be statistically significant.

Some limitations of this study need to be considered. mtDNA-CN was collected from peripheral blood as a surrogate marker, but was not a direct measurement of mtDNA function in the cardiac myocytes. Additionally, mtDNA-

CN was only measured at single point, and variability due to changes over time was not captured. Furthermore, we don't have good information on cancer treatment, and only have case fatality files for common cancers.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____
XYes _ 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? X____ 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

__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)?

Co-authors include the investigators who measured mtDNA-CN and who are responsible for cancer in ARIC.

No prior ARIC manuscript proposals have looked at the association between mtDNA-CN and cancer. The other proposals related to mtDNA-CN:

3007 Association between Mitochondrial DNA Copy Number and atrial fibrillation: Findings from the Atherosclerosis Risk in Communities Study (ARIC).

3160 Association between Mitochondrial DNA Copy Number and heart failure: Findings from the Atherosclerosis Risk in Communities Study (ARIC).

2975 Enhancing the Infrastructure of the Atherosclerosis Risk in Communities (ARIC) Study for Cancer Epidemiology Research: ARIC Cancer

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study ata?
__X_ Yes ____ No

11.b. If yes, is the proposal

_X A. primarily the result of an ancillary study (list number* __2015.21, 2011.07, 1995.04 [these are the cancer ancillaries that developed the cancer case files]_____**)**

___ 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/

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. Siegel RL, Miller KD and Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68:7-30.
2. Ernster L, Low H, Nordenbrand K and Ernster B. A component promoting oxidative phosphorylation, released from mitochondria during aging. *Exp Cell Res*. 1955;9:348-9.
3. Harman D. The biologic clock: the mitochondria? *J Am Geriatr Soc*. 1972;20:145-7.
4. Chan DC. Mitochondria: dynamic organelles in disease, aging, and development. *Cell*. 2006;125:1241-52.

5. Malik AN and Czajka A. Is mitochondrial DNA content a potential biomarker of mitochondrial dysfunction? *Mitochondrion*. 2013;13:481-92.
6. Tin A, Grams ME, Ashar FN, Lane JA, Rosenberg AZ, Grove ML, Boerwinkle E, Selvin E, Coresh J, Pankratz N and Arking DE. Association between Mitochondrial DNA Copy Number in Peripheral Blood and Incident CKD in the Atherosclerosis Risk in Communities Study. *J Am Soc Nephrol*. 2016;27:2467-73.
7. Ashar FN, Moes A, Moore AZ, Grove ML, Chaves PH, Coresh J, Newman AB, Matteini AM, Bandeen-Roche K, Boerwinkle E, Walston JD and Arking DE. Association of mitochondrial DNA levels with frailty and all-cause mortality. *J Mol Med (Berl)*. 2015;93:177-86.
8. Ashar FN, Zhang Y, Longchamps RJ, Lane J, Moes A, Grove ML, Mychaleckyj JC, Taylor KD, Coresh J, Rotter JI, Boerwinkle E, Pankratz N, Guallar E and Arking DE. Association of Mitochondrial DNA Copy Number With Cardiovascular Disease. *JAMA Cardiol*. 2017;2:1247-1255.
9. Zhang Y, Guallar E, Ashar FN, Longchamps RJ, Castellani CA, Lane J, Grove ML, Coresh J, Sotoodehnia N, Ilkhanoff L, Boerwinkle E, Pankratz N and Arking DE. Association between mitochondrial DNA copy number and sudden cardiac death: findings from the Atherosclerosis Risk in Communities study (ARIC). *Eur Heart J*. 2017;38:3443-3448.
10. Lee HK, Song JH, Shin CS, Park DJ, Park KS, Lee KU and Koh CS. Decreased mitochondrial DNA content in peripheral blood precedes the development of non-insulin-dependent diabetes mellitus. *Diabetes Res Clin Pract*. 1998;42:161-7.
11. Thyagarajan B, Wang R, Nelson H, Barcelo H, Koh WP and Yuan JM. Mitochondrial DNA copy number is associated with breast cancer risk. *PLoS One*. 2013;8:e65968.
12. Shen J, Platek M, Mahasneh A, Ambrosone CB and Zhao H. Mitochondrial copy number and risk of breast cancer: a pilot study. *Mitochondrion*. 2010;10:62-8.
13. Lan Q, Lim U, Liu CS, Weinstein SJ, Chanock S, Bonner MR, Virtamo J, Albanes D and Rothman N. A prospective study of mitochondrial DNA copy number and risk of non-Hodgkin lymphoma. *Blood*. 2008;112:4247-9.
14. Hosgood HD, 3rd, Liu CS, Rothman N, Weinstein SJ, Bonner MR, Shen M, Lim U, Virtamo J, Cheng WL, Albanes D and Lan Q. Mitochondrial DNA copy number and lung cancer risk in a prospective cohort study. *Carcinogenesis*. 2010;31:847-9.
15. Lynch SM, Weinstein SJ, Virtamo J, Lan Q, Liu CS, Cheng WL, Rothman N, Albanes D and Stolzenberg-Solomon RZ. Mitochondrial DNA copy number and pancreatic cancer in the alpha-tocopherol beta-carotene cancer prevention study. *Cancer Prev Res (Phila)*. 2011;4:1912-9.
16. Qu F, Liu X, Zhou F, Yang H, Bao G, He X and Xing J. Association between mitochondrial DNA content in leukocytes and colorectal cancer risk: a case-control analysis. *Cancer*. 2011;117:3148-55.
17. Xia P, An HX, Dang CX, Radpour R, Kohler C, Fokas E, Engenhart-Cabillic R, Holzgreve W and Zhong XY. Decreased mitochondrial DNA content in blood samples of patients with stage I breast cancer. *BMC Cancer*. 2009;9:454.
18. Purdue MP, Hofmann JN, Colt JS, Hoxha M, Ruterbusch JJ, Davis FG, Rothman N, Wacholder S, Schwartz KL, Baccarelli A and Chow WH. A case-control study of

peripheral blood mitochondrial DNA copy number and risk of renal cell carcinoma. *PLoS One*. 2012;7:e43149.
19. Liao LM, Baccarelli A, Shu XO, Gao YT, Ji BT, Yang G, Li HL, Hoxha M, Dioni L, Rothman N, Zheng W and Chow WH. Mitochondrial DNA copy number and risk of gastric cancer: a report from the Shanghai Women's Health Study. *Cancer Epidemiol Biomarkers Prev*. 2011;20:1944-9.
20. Wong LJ, Perng CL, Hsu CH, Bai RK, Schelley S, Vladutiu GD, Vogel H and Enns GM. Compensatory amplification of mtDNA in a patient with a novel deletion/duplication and high mutant load. *J Med Genet*. 2003;40:e125.
21. Chen N, Wen S, Sun X, Fang Q, Huang L, Liu S, Li W and Qiu M. Elevated Mitochondrial DNA Copy Number in Peripheral Blood and Tissue Predict the Opposite Outcome of Cancer: A Meta-Analysis. *Sci Rep*. 2016;6:37404.
22. Joshu CE, Barber JR, Coresh J, Couper DJ, Mosley TH, Vitolins MZ, Butler KR, Nelson HH, Prizment AE, Selvin E, Tooze JA, Visvanathan K, Folsom AR and Platz EA. Enhancing the Infrastructure of the Atherosclerosis Risk in Communities (ARIC) Study for Cancer Epidemiology Research: ARIC Cancer. *Cancer Epidemiol Biomarkers Prev*. 2018;27:295-305.
23. Leek JT, Johnson WE, Parker HS, Jaffe AE and Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28:882-3.