ARIC Manuscript Proposal #4175

PC Reviewed: 12/13/22 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: Heteroplasmy and risk of malignancies: The Atherosclerosis Risk in Communities (ARIC) Study.

b. Abbreviated Title (Length 26 characters): Heteroplasmy and cancer

2. Writing Group:

Writing group members:

Lukasz Gondek
Dan Arking
Elizabeth Platz
Sergiu Pasca
Yun Soo Hong
Anna Prizment

Other interested ARIC investigators:
Corinne Joshu

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

First author: **Lukasz P. Gondek MD, PhD**
Address: Sidney Kimmel Comprehensive Cancer Center
CRB1- 290
1650 Orleans St, Baltimore, MD, 21287
Phone: 410-955-0458
E-mail: lgondek1@jhmi.edu

ARIC author is 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 Arking, PhD**
Address: 733 N. Broadway
Miller Research Building, Room 459
Baltimore, MD 21205
Phone: 410-502-4867
E-mail: arking@jhmi.edu

3. Timeline:

We anticipate completing the analysis, manuscript preparation, submission for ARIC review, and submission for publication within 1-2 years.

4. Rationale:

The role of mitochondria in cancer has been well-recognized [1]. Given their central role in key cellular processes, mitochondrial alterations are key components of several hallmarks of cancer such as cellular energetics, cellular proliferation, and apoptosis [2]. Thus, it is not surprising that somatic alteration in mitochondrial DNA (mtDNA) copy number has been described in a plethora of human cancers, including colon [3], pancreas [4,] and kidney [5]. In addition to copy number changes, somatic mutations in mitochondrial DNA are relatively common. Given that there are multiple copies of mtDNA within each cell, mutations can either affect all of the molecules (homoplasmy that is likely germline) or a proportion of the molecules (heteroplasmy that is mostly somatic). MtDNA is particularly prone to DNA damage due to constant exposure to reactive oxygen species and a lack of effective DNA repair machinery [6]. The accumulation of somatic mtDNA mutations often results in mitochondrial dysfunction and has been associated with several human conditions including aging, cancer, and degenerative diseases [7,8].

Since the alteration in cellular energetics and apoptosis are the hallmark of cancer, it is not surprising that heteroplasmy provides another mechanism of adaptation and selective growth advantage of cancer cells relative to their healthy counterparts [9]. Analyzing the whole genome sequencing (WGS) data from the UK Biobank cohort of 194,872 participants, we demonstrated that heteroplasmy was present in 30% of individuals and was associated with a 1.5-fold increase in all-cause mortality. In addition, applying a novel mitochondrial local constraint (MLC) score sum (MSS), we demonstrated that mutations at highly constrained sites were strongly associated with all-cause mortality (adjusted hazard ration (aHR) for a 1-unit increase in MSS 1.28; 95% CI 1.20, 1.37) and cancer-related mortality (aHR 1.36; 95% CI 1.24, 1.49), particularly lung and breast cancers, lymphoma, and leukemia. Moreover, among individuals with prevalent leukemia, high MSS was strongly associated with leukemia mortality (adjusted HR 4.03; 95% CI 1.34, 12.11). We are proposing to use the available WGS as well as clinic data available from ARIC participants to further define the role of heteroplasmy in the incidence and prognosis of hematologic and solid tumors.

In order for heteroplasmy to be detected by WGS, mtDNA with somatic mutations must represent over 10% of all mtDNA. Such expansion of the mutated mtDNA in hematopoietic cells is only possible if the hematopoietic cells carrying mutated mtDNA expand relatively to unmutated counterparts. This phenomenon is also known as clonal hematopoiesis (CH). CH, similarly to heteroplasmy, has been found to be associated with aging, cardiovascular disease, and an increased risk of hematologic malignancies. Thus, in addition to its effect on mitochondrial function, heteroplasmy is also a marker of CH.

We hypothesize that heteroplasmy may be an independent predictor of cancer development and prognosis.

If our hypothesis is true, we expect that our study will establish heteroplasmy as a novel risk factor for malignancy development and cancer-associated adverse outcome.

5. Main Hypothesis/Study Questions:

Main Hypothesis

Heteroplasmy results in an increased incidence of malignancies, and worse outcomes after the diagnosis of a malignancy.

Study Questions

Is heteroplasmy associated with a higher incidence of overall cancer and the most common cancers?

Is pre-diagnostic heteroplasmy associated with adverse outcomes (overall and cancer-related mortality) in cancer patients?

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 Question 1: Is heteroplasmy associated with a higher incidence of malignancies?

Study design

The current study will be longitudinal, analytic, and observational (prospective cohort). WGS is available for approximately 8,567 participants. Of those, 130 participants had a previous diagnosis of cancer prior to the measurement of WGS and will be excluded. We will analyze 8,437 participants who were at risk for cancer at the time of WGS sampling and were appropriately consented to both genetic and non-CVD outcomes like cancer. The estimated number of participants in the analytic cohort who were diagnosed with cancers after their visit with WGS information through 2015 is presented in Table 1. We expect relinkage with the cancer registries and the availability of the updated cancer case file in mid-2023.

Table 1. Estimated number of first primary cancer cases that occurred after Visit 2 through 12/31/2015 in approximately 8,437 participants with WGS, without prevalent cancer, and with appropriate consent in ARIC

	Estimated number of cases
Cancers	2,347
Common sites	
Prostate	465
Breast	360
Lung and bronchus	344
Colorectal	209
Hematologic	191
Bladder	105
Pancreas	69

Inclusion/exclusion criteria

Inclusion criteria:

- WGS data available from V2 and/or V5

Exclusion criteria:

- did not provide consent for non-cardiovascular disease research
- did not provide consent for genetics research
- cancer diagnosed prior to the WGS visit
- did not pass QC for heteroplasmy or heteroplasmy count ≥ 6

Variables of interest

Outcome data: 2015 cancer case file variables, including malignancy incidence and mortality, follow-up time, and characterizing information.

Covariates at each available visit:

- age
- sex/gender (V1)
- diabetes status and type (at risk for diabetes, undiagnosed diabetes, diagnosed diabetes, normoglycemic)
- body mass index
- physical activity (meeting physical activity guidelines)
- alcohol consumption
- smoking status and packyears smoked
- immunosuppressant use
- hormone replacement therapy (women)
- aspirin and non-aspirin NSAIDs use
- age of onset of menarche (women)
- age at menopause (women)
- family history of cancer

Sequencing data:

- for heteroplasmy evaluation – single nucleotide variants (SNVs) or insertion/deletion (indel) and the variant allele fraction (VAF) of the mutations present within the mitochondrial genome
- for clonal hematopoiesis evaluation – we will use the data from the already approved manuscript proposal (see #11b)

Blood biomarkers:

- complete blood counts from any visit

Study Question 2: Is pre-diagnostic heteroplasmy associated with adverse outcomes in cancer patients?

Among cancer survivors, we will perform a similar analysis for cancer-specific death. The following strategy will be used for follow-up time and specification of heteroplasmy as a time-updated variable:

For participants with WGS assessment:

- Patients with cancers diagnosed before the date of WGS assessment will be excluded from the analysis.
- If cancer was diagnosed after the WGS assessment, then follow-up would start at the date of cancer diagnosis and we would use the heteroplasmy status at the date of the WGS assessment.

Summary of data analysis

Heteroplasmy will be defined as a VAF of the mitochondrial mutations between 5% and 95%.

To assess differences between individuals with and without heteroplasmy, we will do the following: Categorical variables will be represented using category percentage. Chi-square or Fisher's exact test will be used to analyze contingency tables. Shapiro-Wilk's test and histogram visualization will be used to assess the normality of the distribution of continuous variables. Non-normally distributed continuous variables will be represented using median (quartile 1, quartile 3), while normally distributed continuous variables will be represented using mean +/- standard deviation. Differences between two non-normally distributed groups will be assessed using the Mann-Whitney-Wilcoxon rank sum test. Differences between two normally distributed groups will be assessed using the Welch t-test.

To evaluate the association between heteroplasmy and cancer risk, we will use survival analysis with the following outcomes: malignancy risk and mortality in at-risk participants, and cancer-specific death in participants with a malignancy diagnosis. Follow-up time for cancer risk will be defined as the time from heteroplasmy assessment to the date of a first primary cancer diagnosis or the date of cancer death for any malignancy or specified cancer site in the at-risk participants. Follow-up time for cancer-specific death among participants diagnosed with cancer will be defined as the time from cancer diagnosis to the date of death from their cancer. Participants who do not experience these outcomes will be censored at their date of death from another cause or at the end of follow-up on 12/31/2015 for risk and 12/31/2019 for death. In the site-specific analyses of at-risk participants, follow-up time will also be censored at the time that they are diagnosed with a first primary cancer at a different site.

We will estimate the association between heteroplasmy and cancer risk and mortality using Cox proportional hazards regression using age as the time scale and adjusting for sex/gender, race*field center, and then additionally adjusting for shared risk factors for heteroplasmy and cancer (smoking status, pack-years, family history of cancer) and cancer risk factors with unknown influence on heteroplasmy, but with an impact on inflammation (physical activity, immunosuppressant use, hormone replacement therapy (women), aspirin and non-aspirin NSAIDs use, age of onset of menarche (for breast cancer). We will use several expressions of heteroplasmy (e.g., presence vs absence of heteroplasmy, total or maximum VAF continuous or indicator variables for quantiles, number of mitochondrial variants, and pathogenicity of mitochondrial variants). We will confirm the proportional hazards assumption. Analyses will also be performed stratified by (<=median, >median), sex/gender, and race (B, W). Complete blood counts will be used to evaluate the effect of heteroplasmy on clonal hematopoiesis (CH) as both CH and heteroplasmy have been previously demonstrated to increase the risk of hematologic malignancies, including in ARIC (preliminary data from two other approved manuscript proposals). Moreover,

histological type, TNM stage, and cancer stage will be used to explore if heteroplasmy has a different association with cancer once developed (e.g., stratify by early vs late stage at diagnosis) and to act as adjustments for the influence of heteroplasmy on cancer survival. ER/PR/HER2 status in the case of breast cancer and the location of a tumor in the case of colon cancer will only be used to determine if heteroplasmy influences these parameters in the cases that developed these tumors.

Minimum detectable associations:
This study is sufficiently powered for the incidence of total solid cancers and for common solid cancer sites to detect important associations for heteroplasmy (Table 2).

Table 2. Minimum detectable HRs of solid cancers (2-sided test, power=80%, alpha=0.05), analytic cohort in ARIC

	Number of cases	Minimum detectable HR Prevalence of heteroplasmy 30%
Cancers	2,347	1.13
Common sites		
Prostate	465	1.33
Breast	360	1.38
Lung and bronchus	344	1.39
Colorectal	209	1.53
Hematologic	191	1.56
Bladder	105	1.82
Pancreas	69	2.09

Methodologic limitations

Some cancer sites are too uncommon to be investigated separately.

The analysis may be limited by the availability of certain clinical outcome data particularly detailed data regarding tumor features such as receptor status, staging, etc.

Not having more frequent whole exome sequencing data, which would be especially important for estimating heteroplasmy trajectories and for being able to separate associations for heteroplasmy preceding a cancer diagnosis versus following a cancer diagnosis in the association with poor cancer outcome.

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" ? ____ 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)?

#2529 The role of mitochondrial heteroplasmy and genetic variation in successful aging

#3862 Clonal hematopoiesis and risk of solid malignancies: The Atherosclerosis Risk in Communities (ARIC) Study

#4100 Association between baseline clonal hematopoiesis and risk of hematologic malignancy in ARIC

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

___ A. primarily the result of an ancillary study (list number* _____

2011.07 Enhancing ARIC Infrastructure to Yield a New Cancer Epidemiology Cohort

1995.04 Cancer Study

2013.23 The role of mitochondrial copy number and genetic variation in successful aging

2020.10 Mapping associations between CH, environmental risks, and MDS or leukemia (Project 3)

___ 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. Zong, W.-X., Rabinowitz, J. D. & White, E. Mitochondria and Cancer. *Molecular Cell* **61**, 667–676 (2016).
2. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
3. Thyagarajan, B., Wang, R., Barcelo, H., Koh, W.-P. & Yuan, J.-M. Mitochondrial Copy Number is Associated with Colorectal Cancer Risk. *Cancer Epidemiology, Biomarkers & Prevention* **21**, 1574–1581 (2012).
4. Lynch, S. M. *et al.* Mitochondrial DNA Copy Number and Pancreatic Cancer in the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study. *Cancer Prevention Research* **4**, 1912–1919 (2011).
5. Xing, J. *et al.* Mitochondrial DNA Content: Its Genetic Heritability and Association With Renal Cell Carcinoma. *JNCI Journal of the National Cancer Institute* **100**, 1104–1112 (2008).
6. Song, S. *et al.* DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4990–4995 (2005).
7. Picard, M., Wallace, D. C. & Burelle, Y. The rise of mitochondria in medicine. *Mitochondrion* **30**, 105–116 (2016).
8. Wei, W. & Chinnery, P. F. Inheritance of mitochondrial DNA in humans: implications for rare and common diseases. *J Intern Med* **287**, 634–644 (2020).
9. Grandhi, S. *et al.* Heteroplasmic shifts in tumor mitochondrial genomes reveal tissue-specific signals of relaxed and positive selection. *Human Molecular Genetics* **26**, 2912–2922 (2017).