ARIC Manuscript Proposal #2577

PC Reviewed: 7/13/15	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. **Full Title**: **Epigenome-wide association study of mitochondrial genetic variation, DNA copy number, and heteroplasmy**

b. **Abbreviated Title (Length 26 characters)**: **Epigenetics and mitochondria**

2. **Writing Group**:
Rebecca Eggebeen, Adrienne Tin, Jan Bressler, Josef Coresh, Jim Pankow, Myriam Fornage, Eric Boerwinkle, and others welcome

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

F**irst author**: **Rebecca Eggebeen**
Address: 733 N Broadway, MRB 420
Baltimore, MD 21205

Phone: (443) 287-0251 Fax: (410) 614-8600
E-mail: reggebe1@jhmi.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, MRB 447
Baltimore, MD 21205

Phone: (410) 502-4867 Fax: (410) 614-8600
E-mail: arking@jhmi.edu

3. **Timeline**:
Data analysis will start immediately. A manuscript is expected to be prepared within 6 months.

4. **Rationale**:
We have previously developed methods to determine mitochondrial DNA copy number (mtDNA-CN) from existing ARIC genotyping arrays, and demonstrated that mtDNA-CN measured in peripheral blood cells declines longitudinally with age and is associated with general health among the elderly, and ultimately, mortality (age- and sex-adjusted

relative risk comparing the lowest to the highest quintiles mtDNA-CN of 1.47 (95% CI 1.33-1.63, P=$4.24x10^{-14}$) [1]. The entire ARIC cohort is currently undergoing whole-exome (WES) and/or whole-genome sequencing (WGS) (WES/WGS is already available on >6,000 samples, and the remaining ~9,000 will be released over the next 6 months). Recent tools have been developed to extract mtDNA sequence from WES/WGS, moreover, with the deep sequencing coverage (>1000x), we are also able to detect heteroplasmy (often due to somatic mutation) down to ~0.5% [2].

DNA methylation is the covalent addition of a methyl group to cytosine at CpG sites and can influence gene transcription. Methylation changes can be inherited or modified by the environment. Recently, it has been suggested that mitochondria play a direct role in the regulation of epigenetics and its role in human disease (for review, see [3]). While some evidence exists for mtDNA copy number influencing methylation, with both hyper- and hypo-methylation observed in cell lines depleted for mitochondria [4], this has not been examined directly in human subjects.

The goals of this study are to whether mitochondrial variation (measured as copy number, heteroplasmy, and inherited genetic variation) is associated with methylation.

5. Main Hypothesis/Study Questions:

Variation in mitochondria will be associated with epigenetic modifications. Secondarily, we propose that these epigenetic modifications will contribute to physical function and aging-related disease (phenotypes associated with reduced mtDNA-CN).

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: cross-sectional for mitochondria/methylation association, prospective cohort study for downstream phenotypes (frailty, successful aging, mortality).

Inclusion criteria: Participants with HM450K data passed quality control (n=2802),[5~~7~~] with data on mitochondrial variation, and values in covariates.

Outcomes (cross-sectional): percent of methylation at 473,788 CpG sites passed quality control

Predictor (cross-sectional): mtDNA-CN, mtDNA SNPs, mtDNA heteroplasmy

Outcomes (prospective): frailty, successful aging, mortality (see MP #2529)

Predictor (prospective): percent of methylation at CpG sites associated with mtDNA-CN, mtDNA SNPs, and/or mtDNA heteroplasmy

Other variable of interest at visit 2: age, gender, race, diabetes, hypertension, BMI, white blood cell count, hsCRP, 10 principal components generated using Affymetric 6.0 autosomal genotype data to control for population substructure.

Data analysis: The methylation data will be first adjusted for batch effect using the Combat approach, an empirical Bayesian method,[6] and then adjusted for the first 10 principal components generated from the percent methylation at each site. We will also explore the use of Surrogate Variable Analysis (SVA),[7,8] in the event that PCA results in over-adjustment. While the principal component adjustment will reduce potential confounding due to differential white blood cell count, we will also explore the use of imputed WBC type distribution, which has been derived using the Houseman algorithm.[9] The standardized residuals from the adjustment will be used as a outcome. The predictors will be mtDNA copy number, mtDNA heteroplasmy, and mtDNA SNPs, together with the covariates. Association between percent methylation and incident frailty/successful aging/mortality will be evaluated using Cox regression.

Significance threshold: the epigenetic-wise significant threshold will be set at $1x10^{-7}$ (=0.05/473,788). For candidate CpG site interrogation of prospective phenotypes, the significant threshold will be set at 0.05/number of genes.

7.a. Will the data be used for non-CVD analysis in this manuscript? __X__ Yes ____ 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)?

#1929 Genome-wide DNA methylation profiling in peripheral blood: quality control and association with demographic characteristics
#2529 The role of mitochondrial heteroplasmy and genetic variation in successful aging

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*)
_X_ 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. Ashar, F. N. *et al.* Association of mitochondrial DNA levels with frailty and all-cause mortality. *J. Mol. Med. Berl. Ger.* **93,** 177–186 (2015).
2. Calabrese, C. *et al.* MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. *Bioinforma. Oxf. Engl.* **30,** 3115–3117 (2014).

3. Minocherhomji, S., Tollefsbol, T. O. & Singh, K. K. Mitochondrial regulation of epigenetics and its role in human diseases. *Epigenetics Off. J. DNA Methylation Soc.* **7,** 326–334 (2012).
4. Smiraglia, D. J., Kulawiec, M., Bistulfi, G. L., Gupta, S. G. & Singh, K. K. A novel role for mitochondria in regulating epigenetic modification in the nucleus. *Cancer Biol. Ther.* **7,** 1182–1190 (2008).
5. Demerath, E. W. *et al.* Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum. Mol. Genet.* (2015). doi:10.1093/hmg/ddv161
6. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostat. Oxf. Engl.* **8,** 118–127 (2007).
7. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3,** 1724–1735 (2007).
8. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinforma. Oxf. Engl.* **28,** 882–883 (2012).
9. Houseman, E. A., Molitor, J. & Marsit, C. J. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinforma. Oxf. Engl.* **30,** 1431–1439 (2014).