ARIC Manuscript Proposal #3705

PC Reviewed: 9/8/20	**Status: _____**	**Priority: 20**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Exome Sequence-Based Analysis of Cardiovascular Disease Risk Factor Phenotype Change In African-Americans and European-Americans from the Atherosclerosis Risk in Communities Study

b. Abbreviated Title (Length 26 characters): ARIC WES: CVD change

2. Writing Group:

Writing group members:

Elena V. Feofanova, Elise Lim, Han Chen, MinJae Lee, Ching-Ti Liu, L. Adrienne Cupples, Eric Boerwinkle.

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

First author: **Elena V. Feofanova**

Address: Human Genetics Center
1200 Pressler Street, Suite E-435
Houston, TX 77030

Phone: 713-500-9827 Fax: -
E-mail: Elena.V.Feofanova@uth.tmc.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: **Eric Boerwinkle**

Address: Human Genetics Center
1200 Pressler Street, Suite W114A
Houston, TX 77030

Phone: 713-500-9058 Fax: 713-500-9020
E-mail: Eric.Boerwinkle@uth.tmc.edu

3. Timeline:

All data have already been obtained. The analysis will be performed as soon as approval is obtained. Because of the large and complex nature of the longitudinal data, several manuscripts may emerge from this work and this proposal. The manuscript is to be prepared as soon as approval is available.

4. Rationale:

Cardiovascular disease (CVD), including conditions of the heart, small vessel diseases of the brain and related diseases of the circulatory system, including hypertension, are the leading causes of morbidity and mortality in the world (1). CVD risk factors include, but are not limited to, dyslipidemia (elevated levels of low-density lipoprotein cholesterol [LDL-C] and total cholesterol [TC], decreased levels of high density lipoprotein cholesterol [HDL-C] and increased systolic blood pressure (BP) (2). Multiple genes affecting lipoproteins and BP levels have been identified, but only ~5% of the population variation in lipid profiles (3), and ~2% of the population variation in BP and hypertension (4) are explained by the discovered genetic loci. Levels of CVD risk factors change with age (5-7), and genetic factors influence that change. Understanding the longitudinal change of risk factors can improve CVD risk prediction. To help identify underlying mechanisms of the onset and progression of CVD, my proposed research aims to identify variants and genes associated with the longitudinal change of 7 complex risk factor phenotypes (BP traits: SBP, DBP and PP; lipids traits: TC, HDL, LDL, triglycerides) in European-American and African-American adults participating in the Atherosclerosis Risk in Communities (ARIC) study.

5. Main Hypothesis/Study Questions:

1. To assess associations of common (MAF>5%) nonsynonymous and splice-site variants with levels and, if main effect associations are present, with longitudinal change of 7 risk factor phenotypes (blood pressure traits: SBP, DBP, PP; lipids traits: TG, TC, HDL-C, LDL-C) in EAs and AAs (ages 45-62 at the baseline examination) using WES data from the Atherosclerosis Risk in Communities (ARIC) study.
2. To assess associations of gene-based aggregation of rare (MAF$\leq$5%) nonsynonymous and splice-site variants with levels and, if main effect associations are present, with longitudinal change in the 7 risk factor phenotypes in EAs and AAs.
3. To assess interaction of selected potential time-dependent effect modifiers (BMI, WBC, eGFR) with common single variants and gene-based sets of variants identified in Aim 1 on levels of each of the 7 traits in EAs and AAs using nonsynonymous and splice-site variants from the WES data in a longitudinal setting.
4. To investigate the relationship of the statistically significant variants and genes identified in Aim 1 as well as their interactions with potential effect modifiers identified in Aim 2 and select CVD endpoints (CHD, heart failure, and ischemic stroke).

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: This is a longitudinal study that consists of 15,792 ARIC participants, with data from the first five visits used for this research.

Exclusion criteria: Participants with missing cardiovascular risk factor phenotypes, covariates, or whole exome sequencing measurements might be excluded from the analysis.

Outcome variables: Sitting SBP and DBP was measured by a certified trained technician in accordance with agreed-upon ARIC protocols, 3 times at visits 1-3, twice at visit 4 using a random zero sphygmomanometer, and 3 times using an automatic sphygmomanometer (OMRON HEM-907 XL) in visit 5. TG and TC (mg/dl) were measured by enzymatic procedures. HDL-C (mg/dl) was measured in the fasting state using standard enzymatic and lipoprotein particle precipitation methods. LDL-C was obtained using the Friedwald formula (TC-[TG/5+HDL-C]).
End points of CVD, including heart failure (HF), CHD and ischemic stroke, were obtained every year using telephone interviews and hospital medical record review (8). Individuals were followed up for events from baseline to 31 December 2017, and those who were lost to follow-up were censored at the date of last contact.

Other variables: Serum creatinine was measured on Visits 1, 2, and 4 using a modified kinetic Jaffe method (9-11); at Visit 5, creatinine was measured in serum or urine on a Roche Modular P Chemistry Analyzer (Roche Diagnostics Corporation) using a creatinase enzymatic method (Roche Diagnostics, Indianapolis, IN 46250) (12). Creatinine-based eGFR was subsequently calculated (13).
Standardized anthropomorphic measurements of weight (in kilograms) and height (in centimeters) were obtained, and BMI (kg/m^2) was calculated. Total WBC count was measured using automated hematology analyzers on all 5 visits.

Whole Exome Sequencing: All sequencing was performed at the Baylor College of Medicine Human Genome Sequencing Center (HGSC). Exomes were captured using the HGSC VCRome 2.1 reagent (14) (42Mb, NimbleGen) and all samples were paired-end sequenced using Illumina GAII or HiSeq instruments. Variant calling was completed using the Atlas2 (15) suite.
Whole exome variants were annotated using ANNOVAR (16) and dbNSFP v2.0 (17) according to the reference genome GRCh37 and National Center for Biotechnology Information RefSeq. Coding variants were annotated to a unique gene as well as splicing and nonsynonymous variants categories.

Single variant analyses: To assess associations for each common (MAF>5%) nonsynonymous and splice-site variant with levels of 7 risk factor phenotypes (blood pressure traits: SBP, DBP, PP; lipids traits: TG, TC, HDL-C, LDL-C), for each variant, test was performed using the GEE (https://cran.r-project.org/web/packages/geeM/geeM.pdf) and LGEWIS (https://cran.r-project.org/web/packages/LGEWIS/LGEWIS.pdf) models in R. For each variant with nominal evidence of a main effect (p-value≤0.05) using both analytical methods, association with longitudinal change of the corresponding CVD risk factor phenotype was tested, and test for interaction was performed with each of the time-dependent potential effect modifiers (cBMI, WBC, eGFR) on levels of each of the 7 risk factor phenotypes (BP traits: SBP, DBP, PP; lipids traits: TG, TC, HDL-C, LDL-C).

Set-based Tests: To assess associations of gene-based aggregation of rare (MAF≤5%) nonsynonymous and splice-site variants with levels of CVD risk factor phenotypes, a set-based test was performed using the naïve burden test collapsing the variants within each gene (https://cran.r-project.org/web/packages/geeM/geeM.pdf), and using LGEWIS (https://cran.r-project.org/web/packages/LGEWIS/LGEWIS.pdf) in R. For each gene with nominal evidence of a main effect (p-value≤0.05) using both methods, association with longitudinal change in the

corresponding risk factor phenotype, as well as a test for interaction with each of the time-dependent potential effect modifiers (cBMI, WBC, eGFR) on levels of the corresponding CVD risk factor phenotypes was performed.

Conditional Analysis: To identify whether an observed gene-based test association is due to a single rare or low frequency variant, conditional set-based test was performed using the LGEWIS test (https://cran.r-project.org/web/packages/LGEWIS/LGEWIS.pdf) in R, conditioning on the variant with the lowest p-value within the gene.

Associations with HF, CHD and Ischemic Stroke: To investigate the relationship of the identified statistically significant variants and genes and clinical CVD endpoints (HF, CHD, ischemic stroke), survival analysis was implemented in the 'survival' R package, using single variant and naïve burden tests, respectively (https://cran.r-project.org/web/packages/survival/survival.pdf).

Significance thresholds: Adjustment for multiple testing for the main effect models was performed using Bonferroni correction (to account for the differing number of tests).

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)?

3270 The association of leukocyte telomere length to longitudinal change in lung function and respiratory infection: The NHLBI Pooled Cohorts Study

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study 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)* _____ _____ _____)

*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.

Yes, the lead author is aware that manuscript preparation is expected to be completed in 1-3 years, and if this expectation is not met, 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.

Yes, the lead author is aware of the policy.

References:

1. http://www.who.int/mediacentre/factsheets/fs317/en/ [12/29/2017]. Available from: http://www.who.int/mediacentre/factsheets/fs317/en/.
2. Goff DC, Jr., Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Sr., Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2014;63(25 Pt B):2935-59.
3. Pollin TI, Ordovas JM, Guevara-Cruz M. Genetic Influences on Blood Lipids and Cardiovascular Disease Risk. 2017:571-93.
4. Salfati E, Morrison AC, Boerwinkle E, Chakravarti A. Direct Estimates of the Genomic Contributions to Blood Pressure Heritability within a Population-Based Cohort (ARIC). PLoS One. 2015;10(7):e0133031.
5. Hunter GR, Gower BA, Kane BL. Age Related Shift in Visceral Fat. Int J Body Compos Res. 2010;8(3):103-8.
6. Palmer DB. The effect of age on thymic function. Front Immunol. 2013;4:316.
7. Musso CG, Oreopoulos DG. Aging and Physiological Changes of the Kidneys Including Changes in Glomerular Filtration Rate. Nephron Physiology. 2011;119(s1):p1-p5.
8. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol. 1989;129(4):687-702.
9. Atherosclerosis Risk in Communities Studies Protocol Manual 10 Clinical Chemistry Determinations Version 1.0 1987.
10. Atherosclerosis Risk in Communities Study Protocol Manual 10 Clinical Chemistry Determinations Version 2.0. 1991.

11. Foster MC, Coresh J, Bonventre JV, Sabbisetti VS, Waikar SS, Mifflin TE, et al. Urinary Biomarkers and Risk of ESRD in the Atherosclerosis Risk in Communities Study. Clin J Am Soc Nephrol. 2015;10(11):1956-63.
12. ARIC Manual 8 Laboratory Methods Visit 5 Version 1. 2012.
13. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150(9):604-12.
14. Bainbridge MN, Wang M, Wu Y, Newsham I, Muzny DM, Jefferies JL, et al. Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. Genome Biol. 2011;12(7):R68.
15. Challis D, Yu J, Evani US, Jackson AR, Paithankar S, Coarfa C, et al. An integrative variant analysis suite for whole exome next-generation sequencing data. BMC Bioinformatics. 2012;13:8.
16. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010;38(16):e164.
17. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. Human mutation. 2013;34(9):E2393-402.