ARIC Manuscript Proposal #2344

PC Reviewed: 4/8/14	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Evaluation of whole exome sequence for loss of function variation influencing quantitative traits

b. Abbreviated Title (Length 26 characters): LOF alleles and trait variation

2. Writing Group:

Writing group members: Alexander Li, Alanna Morrison, Narayanan Veeraraghavan, Bing Yu, Linda Polfus, Thomas Lumley, Andrew Carroll, Xiaoming Liu, Donna Muzny, Thomas Mosley, Richard A. Gibbs, Eric Boerwinkle

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

First author: Alexander Li
Address: University of Texas Health Science Center at Houston
1200 Pressler St.; Suite 453E; Houston, TX 77030

Phone: 713.500.9800 Fax: 713.500.0900
E-mail: Alexander.Li@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, PhD
Address: University of Texas Health Science Center at Houston
1200 Pressler St.; Suite 453E; Houston, TX 77030

Phone: 713.500.9800 Fax: 713.500.0900
E-mail: Eric.Boerwinkle@uth.tmc.edu

3. Timeline: Completion of the manuscript is anticipated in April 2014.

4. Rationale:

Human genetics has traditionally ascertained special phenotypes, such as the most severe cases or the earliest age of onset, to identify genes underlying complex traits. An alternative approach would be to first identify variants with likely functional effects and then investigate the role of these variants on health and disease in a sample of deeply phenotyped individuals. Data from the 1,000 genomes project indicates that each individual contains over 100 variants (20 homozygous) predicted to abolish protein function. In ARIC, we have sequenced the exomes of 5,582 individuals who have been measured for many phenotypes related to common chronic diseases. We annotated the predicted LOF variants and these data can be used to identify novel genetic LOF candidates influencing quantitative physiologic traits.

5. Main Hypothesis/Study Questions:

We propose a "bottom-up" approach can be used within the framework of a large and deeply-phenotyped population cohort by first selecting samples with LOF variation within shared genes and then testing for phenotypic abnormality. In an analysis of 20 core phenotypes (Magnesium, phosphorus, calcium, potassium, sodium, AST, ALT, GGT, DBP, SBP, FCV, FEV/FEC, triglycerides, total cholesterol, fasting insulin, fasting glucose, creatinine, lactate, uric acid, while blood cell count), we aim to identify novel statistically significant associations with trait variation.

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).

Whole exome sequence data is available for 3,173 individuals of European ancestry (EA) and 2,409 individuals of African ancestry (AA) sampled from the Atherosclerosis Risk in Communities (ARIC) study (i.e., Freeze 3). DNA sequencing was performed at the Baylor College of Medicine Human Genome Sequencing Center.

Prior to statistical analysis, we will determine whether a variant leads to "full" or "partial" loss-of-function (LOF) affecting all or a subset of gene isoforms. We will develop a gene-based metric (OP ratio) which quantifies the ratio of the number of observed LOF sites in a gene to the number of possible LOF sites in the same gene.

Genotype-phenotype analyses will employ two methods, a single variant Wilcoxon rank sum test and a T5 burden test (B. Li & Leal, 2008), to assess the significance of the effect of the LOF variants on each phenotype. Based on an estimation of the number of genes having LOF variants and the number of phenotypes analyzed, a p-value less than $10x10^{-7}$ may be considered statistically significant.

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

____ Yes _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 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)?

There are several manuscript proposals in place that are specific to particular CHARGE working groups and the analysis of exome sequence data (e.g., #2194 for Hemostasis, #2068 for Blood Pressure, #2084 Metabolome). However, these manuscripts do not focus on a genome-wide detailed evaluation of loss-of-function alleles. The authors of this manuscript are intimately involved (or lead) the other manuscripts so that there is seamless coordination.

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

_X_ A. primarily the result of an ancillary study (list number* 2009.14)

___ 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.

Agree.

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.

Agree.