ARIC Manuscript Proposal #2572

PC Reviewed: 7/13/11	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Next-generation sequencing for QT Interval: Results from the CHARGE-S Project

b. Abbreviated Title (Length 26 characters): Next-gen sequencing in QT

2.Writing Group: CHARGE-S QT

Writing group members: Dan E. Arking, Ashish Kapoor, Aravinda Chakravarti, Nathan Bihlmeyer, Alvaro Alonso, Alanna Morrison, Eric Boerwinkle (and/or other Houston personnel). Others ARIC authors welcome. Other authors from additional consortium cohorts will be included, with a plan to maintain symmetry across cohorts.

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

First author: Dan E. Arking
Address: Johns Hopkins University School of Medicine
733 N. Broadway, BRB 453
Baltimore, MD 21205
Phone: 410-502-4867 Fax: 410-614-8600
E-mail: arking@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: Alvaro Alonso
Address: Div. Epidemiology & Community Health
Sch Public Health, U of Minnesota
1300 S 2nd St, Suite 300, Minneapolis, MN 55416
Phone: 612-626-8597 Fax: 612-624-0315
E-mail: alonso@umn.edu

3. Timeline: Summer 2015: obtain and analyze ARIC data. Fall 2015: meta-analysis across cohorts. Spring 2016: manuscript to respective P&P committees.

4. Rationale:

The QT interval duration on the surface electrocardiogram (EKG) represents ventricular depolarization and repolarization. Longer QT duration is a predictor of mortality and sudden death not only in select patient populations, but also in the general medical population[1]. Recent genome-wide studies of QT interval, including our own report from the CHARGE consortium (CHARGE QT GWAS)[2], have identified several genetic variants associated with QT interval duration, primarily among those of European descent. That both rare genetic mutations and common genetic variation have been associated with cardiac conduction / QT interval, and that QT duration is associated with clinical outcomes and mortality suggest that genetic variants that influence this trait may have clinical consequence.

The allelic architecture of complex traits such as QT interval is currently unknown, but could involve common variants with modest effects, or an aggregation of rarer variants with strong effects. We have recently published the CHARGE QT GWAS examining ~100K individuals of European descent where we identified 35 loci associated with QT interval duration. With the CHARGE-S QT activity, we have two aims: (1) to fine-map target regions associated with QT interval in the CHARGE QT GWAS to identify the causative variant at each locus examined; and (2) to identify rare variants that influence QT interval duration in a hierarchical fashion, first focusing on the 35 CHARGE QT GWAS loci, then secondarily examining rare variants exome-wide. The goal of CHARGE-S is to follow up on GWAS results exploring genetic regions associated with primarily cardiovascular phenotypes. We propose to perform an analysis of targeted, exome, and whole-genome sequencing data available from subjects in the CHARGE-S QT project.

5. Main Hypothesis/Study Questions:

Whole-exome/whole-genome data will be used to identify rare variants from exome and promoter regions to identify novel variation associated with cardiac conduction (QT interval), and will account for part of the heritability of this EKG phenotype. We will also specifically examine the 35 loci associated with QT interval in CHARGE QT GWAS.

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 samples

The QT analysis in CHARGE-S will employ a cross-sectional study design (though we will explore the use of longitudinal measures of QT interval), focusing on ARIC samples with next-gen sequence data available (~13,000 WES, ~3,600 WGS). We will exclude from the analysis individuals with characteristics that may influence electrical conduction such as those with prevalent heart disease (myocardial infarction or heart failure), or Wolff-Parkinson-White syndrome.

Main Statistical analysis plans & methods

Prior to statistical analyses, rigorous quality control measures will be applied to the data. These measures will follow closely the analysis plan as proposed by the CHARGE-S analysis committee. Each study with available data (e.g. ARIC, CHS, FHS) will conduct its own analyses and summary statistics will then be meta-analyzed across studies.

For both aims, the phenotypes will be adjusted for age, sex, and RR interval (heart rate). Individuals with AF/flutter, QRS>120, LBBB, or RBBB will be excluded. Principal components will be used to adjust for population substructure. We will adjust for sequencing batch effects to account for drift over time.

In addition to analyzing common variants through standard GWAS approaches, we aim to identify novel variants in coding and promoter regions using both regression (SKAT) and burden/collapsing methods (T1).

*Bioinformatics prediction of the identified variants: G*enetic variants identified by sequencing will be categorized. One categorization will be based on the direct impact of the variant on the encoded proteins, such as synonymous, missense, nonsense, frameshift or splice-site alleles. PolyPhen-2[3] and SIFT[4] will be used to predict potential deleterious effects of nonsynonymous mutations on gene function. Deleterious mutations will be further examined in terms of protein functional domains and 3D structures. Evolutionary conservation at variant loci will be studied by multiple sequence alignment of vertebrate genomes. Gene function and pathway information will be retrieved from Gene Ontology database and KEGG pathway database. Variation in regulatory regions will also be interrogated.

Sources of Data to be used

The general goal is to retain as much information as possible without sharing individual level sequence data. Each study will conduct analyses for the data in their study as agreed upon by the working group and then summary statistics from each study will be combined into a single score test across studies to provide a global test of association. For the first pass of analyses using the unweighted approach, the weighted Z approach for combining p values and signs of regression coefficients will be used.

As needed, GWAS data will be used internally in each cohort as a quality control measure for exome sequencing.

Additional analyses will be undertaken to evaluate whether sequence data explain the GWA signal that was the impetus for examining the target. These analyses will condition on the SNP that was the primary signal for the target in the GWA results as an additional covariate in the regression analyses described above.

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? __X__ Yes ____ No
(This file ICTDER03 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)?
MS1396, MS1397, MS1398 which are GWAS for AF, lone AF, and PR, which we lead.
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/

12. 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.

References

1. Dekker, J. M., Crow, R. S., Hannan, P. J., Schouten, E. G. & Folsom, A. R. Heart rate-corrected QT interval prolongation predicts risk of coronary heart disease in black and white middle-aged men and women: the ARIC study. *J Am Coll Cardiol* **43,** 565–71 (2004).
2. Arking, D. E. *et al.* Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat. Genet.* **46,** 826–836 (2014).