ARIC Manuscript Proposal # 1768

PC Reviewed: 3/8/11	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Targeted Sequencing of *ADAM19* (*ADAM metallopeptidase 19*) and *HTR4* (5-hydroxytryptamine (serotonin) receptor 4) to follow-up genome wide association study findings with FEV1/FVC: the CHARGE-S Consortium

b. Abbreviated Title (Length 26 characters): **Sequencing of PFT GWAS hits.**

2. Writing Group:

Writing group members:
Bonnie Joubert
Stephanie London
Alanna Morrison
Laura Loehr
Nora Franceschini
Dana Hancock
Kari North
Eric Boerwinkle
And other interested ARIC authors involved in the sequencing project.
This paper requires analysis of sequencing data from the Framingham (FHS) and Cardiovascular Health (CHS) Studies. Authors from these studies will be included. A pulmonary sequencing committee consisting of representatives of the CHARGE pulmonary group from ARIC, FHS and CHS started meeting by phone in late 2009 and developed and agreed upon the case definition and gene selection plans.

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

ARIC F**irst author**Bonnie Joubert**,** Bonnie Joubert, Postdoctoral Fellow
ARIC **Last author:** Stephanie London, Senior Investigator
Address for both: Epidemiology Branch
NIEHS, PO Box 12233, MD A3-05, RTP NC 27709

Phone: 919-541-5772 (SJL), 919-541-7667 (BRJ) Fax: 919-541-2511
E-mail: london2@niehs.nih.gov

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: Alanna C. Morrison

Address: University of Texas Health Science Center at Houston
1200 Herman Pressler; Suite 453E
Houston TX 77030

Phone: 713.500.9913 Fax: 713.500.0900
E-mail: Alanna.C.Morrison@uth.tmc.edu

3. Timeline: March 1 2011 – preparatory analyses for when the actual targeted sequencing data arrive.

Within 6 months after arrival of targeted sequencing data – complete ARIC analyses of the two genes.

Meta-analysis of targeted sequencing results for FEV1 and FEV1/FVC will be conducted following receipt of sequence analysis results from CHS and FHS investigators. The goal is completed meta-analysis by Dec 15 2011, followed by manuscript preparation. Goal to submit manuscript to ARIC MS committee by Jan 31 2012.

4. Rationale:

I Introduction:

A. Background
- In the CHARGE consortium, we meta-analyzed genome wide association data in relation to two pulmonary function parameters: the forced expiratory volume in one second (FEV1) and its ratio to the forced vital capacity (FEV1/FVC). We identified seven novel loci for FEV1/FVC and one for FEV1.
- The CHARGE-Sequencing project (CHARGE-S, Eric Boerwinkle PI) invited phenotype groups within CHARGE to nominate GWAS findings for targeted sequencing in 200 “cases” and a cohort random sample of 2,000 subjects. The pulmonary group nominated six of our novel loci for follow-up. More genes were nominated across group than were possible to sequence. Two of our six were accepted for targeted resequencing – *ADAM19* and *HTR4*.
- The goal of targeted sequencing is to help identify the causal variant(s) underlying the GWAS hit. Targeted sequencing has the potential to discover novel functional variants and to identify rare variants that are linked to the common variants associated with the top GWAS hits.
- After discussion within the CHARGE-S pulmonary working group, we settled on severe COPD as our “case” group after verifying that we had significant associations for this phenotype for the top GWAS hits with the proposed sample size. Severe COPD is defined based on FEV1 and FEV1/FVC.

B. Why did we choose *ADAM19* and *HTR4* for targeted resequencing?
- For both of these genes, our top GWAS hits (ARIC MS #1357-Hancock et al., Nat Genet 2010,42:45-54) localize well within the gene, as opposed to intragenic regions.

- SNP associations in both genes were replicated in the SpiroMeta consortium (Repapi et al, Nat Genet 2010;42:36-44 – paired submission with Hancock et al. 2010)

- Both genes are expressed in the lung and have high biologic plausibility for causal association with FEV1/FVC.
  - *ADAM19*:
    - member of "a disintegrin and metalloprotease" (ADAM) family of membrane-anchored glycoproteins that control cell-matrix interactions and contribute to the regulation of growth and morphogenesis
  - *HTR4:*
    - Encodes a G-coupled transmembrane receptor that regulates cAMP production in response to 5-hydroxytryptamine (serotonin)
    - Elevated levels of free serotonin have been found in plasma of symptomatic asthmatics
    - Serotonin signaling pathways that incorporate *HTR4* play a role in cholinergic and immune-mediate airway reactivity
    - Upon activation by serotonin, *HTR4* in human airway epithelial cells regulates the release of a proinflammatory cytokine, a signature characteristic of asthma

5. Main Hypothesis/Study Questions:

A. Hypotheses:

1. Upon targeted sequencing we will discover novel variants associated with severe COPD (case-cohort analysis) and/or the FEV1/FVC (quantitative trait analysis)
2. Some novel variants will be predicted to be potentially damaging
3. Novel variants will explain the known GWAS associations.

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).

We are submitting this as a single manuscript. If there are interesting results for both genes, then we would be plan for two separate manuscripts. In that event we would submit a revised manuscript proposal.

II Methods:

A. CHARGE-S study design for targeted sequencing is case-cohort.

- Cohort random sample: no phenotypic exclusions (1000 ARIC, 500 FHS, 500 CHS) with equal proportion of males and females. All individuals are non-Hispanic Whites.
- Selection of cases (n=200): Severe COPD as defined in the CHARGE pulmonary group analyses based on spirometry. Definition: forced expiratory volume in the first second (FEV_1) < 65% predicted and its ratio to forced vital capacity (FEV_1/FVC) < the lower limit of normal based on NHANES III prediction equations. The pulmonary sequencing subcommittee, consisting of representatives from the three cohorts, chose to select 200 cases from ARIC alone. As requested by the CHARGE-S Phenotype Steering Committee, we verified that association signals were observed for our GWAS hits within the

nominated genes were observed in case-control analyses using these 200 cases.

- Power estimates: Because the power to detect associations with an individual rare variant is limited, Price (Price A et al., *Am J Hum Genet* 2010;86:832-8) and others have proposed pooling rare variants and applying one composite test per gene. Under this pooled rare variant allele approach, we performed power analysis for the case-control analysis of severe COPD in *Quanto*. We estimate 83% power to detect a relative risk of 1.8 for a 5% pooled rare variant allele frequency and a two sided alpha of 0.05. For a pooled rare variant allele frequency of 3% we can detect a relative risk of 2.0 with 80% power. Although these effect estimates are higher than those reported in a GWAS of COPD (Pillai et al., *PLoS Genet* 2009;5(3):e1000421), one of the rationales for targeted sequencing is that common variants associated with disease in GWAS may reflect correlation with rare variants with larger effects. We also note that this study is quite large compared to other sequencing efforts, especially if our data is combined with data from other CHARGE-S phenotype groups (see below).

B. Targeted sequencing

- Design of target regions for sequencing *ADAM19*:
  - 1.203 kb region: Regulatory region, CTCF binding site for ADAM19.
  - 101.659 kb region*:* Includes the gene region (98.456 kb) containing the top SNP rs2277027, as well as the regulatory region.
- Design of target region for sequencing *HTR4*:
  - 21.724 kb region: LD block in *HTR4* coding region comprising all of the association signals, including the top SNP rs11168048
  -
- Target capture (Nimblegen) and sequencing protocols established by the lab in Houston.

QC protocol for the sequencing data is being established by the CHARGE-S phenotype steering committee and the CHARGE-S Analysis committee in collaboration with the lab in Houston. Dr. London represents the pulmonary group on the CHARGE-S phenotype steering committee and also attends conference calls for the analysis committee. Other writing group members (Morrison and North and Franceschini) represent other working groups on the phenotype steering committee.

- C. Statistical analyses
  - We will use the approach developed by the CHARGE-S Analysis committee (Thomas Lumley, Josée Dupuis, Adrienne Cupples and others) based on simulations and examination of Type 1 error and power. It was found that a weighted approach alone would lead to a loss in power (i.e. upweighting the controls was lowering our power)
    - Our primary independent variables will include genotypes for variants with MAF > 1% and also a count of the number of rare variants/ person
    - Analysis Step 1: Use unweighted analysis, to determine which associations are significant, acquire p-value
    - Analysis Step 2: Use a weighted analysis to obtain correct estimate of effect size (beta). We will use Thomas Lumley's weighted approach

which is a 2 phase variance weighted analysis). Note our weighted approach will differ from other phenotype groups as 100% of our cases are ARIC participants.
    - We will test for associations with case status (severe COPD) using the case-cohort design. We will also analyze FEV1/FVC and FEV1 as quantitative traits among the combined sample.
    - At present it has been decided for CHARGE-S projects that all analyses will be done separately within each of the three cohorts and then the data will be meta-analyzed. Since our "case" group is ARIC only, the case-cohort analysis would be ARIC only whereas the quantitative trait analysis would combine across all three cohorts.

D. Secondary analyses involve testing for evidence of pleiotropic relationships with related phenotypes such as asthma. Because some of our GWAS hits are also related to height, a determinant of pulmonary function, we will also test for associations with height. These phenotypes are available in the individuals from the case group and cohort random sample.

E. Follow-up of interesting variants
    - If there are interesting findings for any variants, we propose follow-up by either replication or functional studies or both.
    - Replication: Options are *in silico* (look-up) replication in other studies that are getting sequencing data or replication genotyping. For *in silico* replication, we have reached out to the Lung-Exome Sequencing Project. Whether this works out will depend on the timing of when their data are ready, when we have our data analyzed, and their coverage of our two targets in their exome sequencing data. We are collaborating with the SpiroMeta consortium in the UK. They are not getting targeted sequencing data but will be setting some limited whole exome and whole genome data and also have some replication genotyping capability which they could apply to their component cohorts.
    - **7.a. Will the data be use*d for non*-CVD analysis in this manuscript? _x___ Yes ____ No**

b. If Yes, is the author aware that the file ICT*DER03 mus*t 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)?

This proposal follows up our previous work in ARIC and CHARGE for #1357 Genome-Wide Association Study (GWAS) of Pulmonary Function and Chronic Obstructive Pulmonary Disease (COPD) – interaction with intake of fiber and other nutrients in ARIC. S. London. This manuscript proposal was merged with #1360 and resulted in a paper on GWAS of pulmonary function (Hancock *et al*., Nat Genet. 2010;42:45-52.. A separate manuscript on GWAS of COPD is ongoing within the CHARGE Consortium and was split off as separate manuscript and given MS#1357r.

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 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.