1.a. Full Title: Genome Wide Association Study of interaction with smoking in relation to pulmonary function and COPD.

b. Abbreviated Title (Length 26 characters): GWAS by smoking interaction with PFTs and COPD

2. Writing Group:
Writing group members: Dana Hancock, Stephanie London, Kari North, Laura Loehr, Nora Franceschini, David Couper, Alanna Morrison, Peter Kraft, Yu-Chun Yen plus other interested ARIC investigators.

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

Last:  Stephanie London
Address:  NIEHS, PO Box 12233, MD A3-05, RTP NC 27709

Phone: 919-541-5772   Fax: 919-541-2511
E-mail:  london2@niehs.nih.gov

First:  Dana Hancock
Email:  hancockd@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:  Kari North
Address:  137 E. Franklin St., Suite 306, Campus Box 7435, Chapel Hill, NC

Phone: (919) 966-2148   Fax: (919) 966-9800
E-mail:  kari.north@unc.edu

4. **Rationale:**

This proposal splits out as a separate manuscript, aim number two of the earlier manuscript proposal **1357r** titled: Genome-Wide Association Study (GWAS) of Pulmonary Function and Chronic Obstructive Pulmonary Disease (COPD) – analysis of main effects, interactions with smoking and diet.” From that proposal, we submitted a manuscript to Nature Genetics on GWAS main effects in relation to pulmonary function with the CHARGE consortium. We received generally favorable reviews and resubmitted the revision to Nature Genetics on Sept 11, 2009. The revised manuscript is under review. We waited until we had a disposition on that manuscript before submitting the GWAS by smoking proposal because it was possible that the reviewers would require us to do an interaction analysis as part of that manuscript. Now that we know that we will not be required to do this, we are requesting to split the GWAS by smoking interaction aim into a separate manuscript.

Pulmonary function in adulthood, as has been measured in ARIC, reflects two major processes – maximal growth though young adulthood and the rate of decline with age after maximal growth has been attained. Until recently, the only well established genetic risk factor for accelerated decline in pulmonary function was the rare alpha-1 anti-trypsin deficiency. In 2009, the Framingham Heart Study (FHS) identified associations between single nucleotide polymorphisms (SNPs) in hedgehog interacting protein (HHIP) as related to pulmonary function (FEV1/FVC) in genome wide association analysis (Wilk et al. PLoS Genetics Mar;5(3):e1000429). We then did a CHARGE consortium meta-analysis including ARIC, FHS, Cardiovascular Health Study and Rotterdam Study which confirmed HHIP but also identified seven novel loci related to FEV1/FVC or FEV1. We have reported this finding in the paper submitted to Nature Genetics mentioned above.

While the identification of novel loci related to pulmonary function from our GWAS is important, smoking is the major risk factor for accelerated decline in pulmonary function in adulthood. Examination of gene by smoking interaction in the genome wide setting is likely to lead to identification of novel interacting genes. We therefore propose to examine GWAS by smoking interaction in ARIC in relation to pulmonary function.

5. **Main Hypothesis/Study Questions:**

1. To use genome wide association analysis of interaction with smoking to identify novel loci associated with pulmonary function.

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 would use the ARIC GWAS genotyping data on Caucasians to examine interaction with smoking in relation to pulmonary function. There are few published methods for testing for environmental interaction in the genome wide setting. The standard multiplicative test of interaction presents well known limitations of power. To increase power, a screening method for studying qualitative traits was developed by Murcray et al. (Am J Epidemiol. 2009;169:219-26). We have contacted the senior author (Jim Gauderman) who reports that he has considered the issue at length and concluded that the method cannot be adapted for studying a quantitative trait. The other published method for genome wide interaction testing was proposed by Peter Kraft and colleagues (Hum Hered. 2007;63:111-9). This is a simultaneous test for genetic main effects and interaction which provides greater power than the standard interaction test. We will henceforth refer to this as the G-GE test. While we have already done a separate analysis of main effects, the G-GE method still provides a useful test for genes that operate primarily by interaction with an environmental factor. Although the published G-GE method was for a qualitative trait, Dr. Kraft provided us with PLINK code for analysis of quantitative traits. However, the method was developed for data from genotype calls. In the CHARGE consortium dosages from imputation, rather than genotypes, are meta-analyzed because different genotyping platforms have been used across the cohorts. Dr. Kraft kindly agreed to collaborate to help us figure out solve two practical issues – how to implement the G-GE method for imputed data and how to meta-analyze the results. The meta-analysis issue arises because the G-GE method generates two betas and one needs to take the covariance into account. Dr. Kraft has very recently worked out a solution to these issues, tested it in GWAS data that he works with, and is working on a submission for peer review on the approach.

The primary outcomes for this manuscript proposal are the two that we analyzed in our main effects analysis. These are the forced expiratory volume (FEV1) and the ratio of FEV1 to the forced vital capacity (FEV1/FVC). The primary smoking metric for interaction testing is a dichotomous variable for ever versus never smoking. Although we are interested in quantitative smoking (such as pack-years), Dr Kraft has informed us that his method runs into problems with inflated lambdas when using a continuous environmental variable. He has found that this problem is resolved by dichotomizing the environmental variable. So we will focus on the dichotomous variable. He ascribes this problem to confounding due to mis-specification of the exposure main effect and says that it would also occur with the standard interaction test method. Indeed, several CHARGE groups have noted inflated lambdas in the standard test of interaction. The CHARGE wiki document on gene environment interaction testing does not indicate whether this occurs in the setting of both dichotomous and continuous exposure variables. Within the ARIC genotyped data, we have ascertained that the standard interaction product term analysis does not produce inflated lambda values when using dichotomized smoking status. The CHARGE wiki document suggests that using model-robust (or sandwich) standard errors may resolve this problem in the standard interaction analysis (GxE analysis guidelines From Charge, http://depts.washington.edu/chargeco/wiki/GxEanalysisguidelines, downloaded Sept 8, 2009). Further work is underway by CHARGE biostatisticians. Thus for the standard
interaction test, we may be able to examine a continuous variable such as pack-years in the future.

We propose to do implement both the G-GE method and the standard interaction test analysis. Indeed, we would expect that readers and reviewers would want to know how the results compare between the two methods.

Outcome variables: primary: FEV1, FEV1/FVC
Secondary – COPD* (Defined based on spirometry – see note below).
Interaction variables: primary - ever/never smoking, secondary - pack-years
Adjustment variables: age, gender, height, center, principal components (population stratification adjustment), smoking variables (pack-years in ever/never interaction model; past/current in pack-years interaction model). These were the adjustment variables in the main effects analysis.

* We would prefer NOT to include COPD in the same manuscript with the quantitative traits for a number of reasons, both scientific and logistic. However, we need to prepare for the possibility that journal reviewers or editors might insist that we include COPD in the same paper with the quantitative pulmonary function traits. Therefore we include it as a secondary outcome in this proposal. For the main effects GWAS of pulmonary function manuscript proposal, we had prepared for this possibility but the editor asked instead for analysis excluding COPD rather than analyzing it separately.

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 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.cscu.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)?

Manuscript under review at Nature Genetics Hancock DB, et al. Meta-analyses of genome-wide association studies identify multiple novel loci related to pulmonary function: the CHARGE consortium. – All ARIC co-authors have been contacted regarding participation in the current manuscript proposal and those who have indicated interest have been included.

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.