ARIC Manuscript Proposal # 1384

1a. Full Title: Genome-Wide Association Study (GWAS) of Pulmonary Function and Chronic Obstructive Pulmonary Disease (COPD) – interaction with traffic exposure in ARIC.

1b. Abbreviated Title (Length 26 characters): GWAS by traffic interaction for PFTS

Writing Group: Stephanie J. London (NIEHS), Dana B. Hancock (Postdoc, NIEHS), Wen Hong Linda Kao (JH), Matthew B. Schabath (UTH), Kari North (UNC) - Plus other interested ARIC investigators

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

First author: Stephanie J. London, M.D., Dr.P.H.
Address: Epidemiology Branch and Laboratory of Respiratory Biology
National Institute of Environmental Health Sciences
MD A3-05
PO Box 12233
RTP, NC 27709

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

ARIC Corresponding/senior author (if different from first author correspondence will be sent to both the first author & the corresponding author):
Name: Kari North, PhD
Address: Department of Epidemiology
Department of Epidemiology, University of North Carolina Chapel Hill,
Bank of America Center, 137 E.Franklin St., Suite 306,
Chapel Hill, NC 27514, USA.

Phone: 919-962-2148
E-mail: kari_north@unc.edu

Corresponding/senior author – non-ARIC
Name: Stephanie London, MD, DrPH
Address: NIEHS, PO Box 12233, MD A3-05, RTP NC 27709
Phone: 919-541-5772
E-mail: london2@niehs.nih.gov

Manuscript Prep: December 2008 to February 2009
Manuscript Revision: March 2009
Manuscript Submission: April 2009
4. **Rationale:**

Pulmonary function is an easily measurable index of the state of the lungs and the airways with a high degree of hereditability (Wilk et al., 2003). Genetic factors appear to influence both the maximal attained pulmonary function in early adulthood as well as the rate of decline after that – pulmonary function at given point in adulthood reflects both processes and different sets of genes may contribute to each (Wilk et al., 2003). The well-described decline in pulmonary function with age is influenced by both environmental and genetic factors. Smoking is a major determinant of the rate of decline in lung function. The relatively uncommon genetic deficiency of alpha-1-antitrypsin has long been known to lead to premature decline in lung function, especially in smokers (Demeo and Silverman 2003). Candidate gene studies also suggest that common variants in other genes influence decline in lung function in relation to smoking (Demeo and Silverman 2003). Air pollution has a smaller, but established effect on lung function in children (Sandstrom and Brunekreef 2007) and adults (Downs et al., 2007). Some limited data in humans along with studies of inbred mouse strains suggest that genetic factors play a role in susceptibility to inhaled pollutants (London 2007). In the ARIC cohort, we have found pulmonary function, in cross-sectional analyses, to be influenced by two indices of traffic-related air pollution – distance to major roads and traffic density (Kan et al. 2007).

Another manuscript proposal on which the first author and several co-authors of this proposal are involved would examine genetic main effects on pulmonary function and COPD phenotypes in the ARIC study using the 1,000,000 single nucleotide polymorphism (SNP) genome wide association study (GWAS) data. In the current proposal, we propose to examine interactions between SNPs in the GWAS data and the two indices of exposure to traffic related air pollution in that were related to pulmonary function in our recent analyses in ARIC (Kan et al., 2007). These two metrics are distance to major roads and traffic density. The goal is both to examine hits in the genetic main effects analysis as interacting factors with traffic but also to use the whole genome analysis to identify SNP associations that may work only by interaction with traffic exposure.

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

1. Genetic variants interact with exposure to traffic related air pollution in relation to lung function in the ARIC cohort. We recently found that two metrics of traffic related air pollution, distance to major roads and traffic density, were related to pulmonary function in ARIC. To identify genetic variants that interact with traffic exposure, we will use approaches consistent with other ARIC whole genome analyses of qualitative and quantitative traits to examine interactions between the dietary factors and whole genome SNPs data in relation to pulmonary function test parameters (quantitative phenotypes) and chronic obstructive pulmonary disease (qualitative phenotype based on pulmonary function).

2. We plan to test for interaction between SNPs that show main affects in the manuscript proposal of Schabath (in which most of us are also involved) as well as to use the whole genome data to identify SNPs that may have effects only in interaction with traffic exposure. We anticipate that the interaction analysis may identify genes that were not identified in the main effects analysis.
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:** The analysis will consider two types of outcomes. Our primary analysis will be based on the quantitative traits:

1. Quantitative traits based on the continuous pulmonary function parameters.

The measures of primary a priori interest are the FEV1 (forced expiratory volume in one second), FVC (forced vital capacity), FEV1/FVC, and FEF25-75 (forced expiratory flow during the middle half of the forced vital capacity, also called MMEF in some older literature). These are the parameters that have been most consistently related to both genetic predisposition and environmental factors in previous studies. For each of these factors of primary interest, we will perform analyses based on the percent predicted values (which remove the well-established effects of age, height, sex, and ethnicity). Analyses will be adjusted for other exposures related to these outcomes in ARIC including smoking, physical activity, occupation, BMI and education. For the percent predicted values, we will do separate analyses with visits 1 and 2 as an internal first level replication. We would assume that true interactions should show up with data from both visits.

The other published whole genome association study of pulmonary function was in the Framingham study (Wilk et al. 2007). In that study, in addition to calculating the percent predicted they examined the mean of two measures from closely timed visits. Thus, we will also calculate the mean of the actual parameters for ARIC visits 1 and 2 which were three years apart. Vollmer has stated that due to measurement error in pulmonary function test measurements, it is not possible to study longitudinal changes with only two measures of less than 4 years apart (Vollmer 1993). Taking the mean value of the parameters from the two visits which are only three years apart, we should reduce the standard error and will also give use results comparable to the Framingham analysis. Analyses of mean values will be adjusted for appropriate transformation of factors influence pulmonary function including age, height, ethnicity, gender, BMI, smoking status (never, past, current) and pack-years. We will also repeat analyses of interaction within strata of never versus ever smoking status to examine whether interactions are present in both strata or not.

2. Qualitative traits

Standard criteria for diagnosis of chronic obstructive pulmonary disease (COPD) are based on pulmonary function measured after administration of bronchodilator because improvement after bronchodilator is more consistent with asthma than COPD. However, bronchodilator administration is difficult in large epidemiologic studies and pre-bronchodilator pulmonary function values have been regarded as an acceptable for classifying COPD in large studies (Vestbo 2004). Indeed, previous analyses in the ARIC study, including ours, have found classic associations using pre-bronchodilator pulmonary function to classify COPD. Applying the GOLD criteria to the prebronchodilator pulmonary function tests, we will classify COPD as FEV1/FVC <0.70 & FEV1 < 80% predicted.

Although self-reported doctor diagnosis of emphysema or chronic bronchitis is not regarded as a reliable index of COPD in the general population to have some consistency with the main effects analysis being proposed by Schabath in which we are also participating, we will do analyses compared to a “normal” group defined by individuals without COPD by pulmonary function test criteria, without MD diagnosis of emphysema or chronic bronchitis and without chronic bronchitis defined in the standard epidemiologic manner by subject report of symptoms (cough with phlegm on most days for three months out of year for two or more years in a row).

**Other variables of interest:** In addition to the traffic variables, pulmonary function and respiratory symptoms/disease variables, we will need to examine age, height, weight, gender, center, race, smoking (status, duration, amount), environmental tobacco smoke, background air pollution, occupation, dietary factors related to lung function in ARIC (fiber, vitamin D, omega-3 fatty acids, vitamin C, carotenoids and vitamin E), and physical activity variables that we have adjusted for in previous ARIC analyses of these outcomes.
Summary of data analysis:

We presume that the dataset we receive will already have passed standard quality control checks for genome-wide association studies. However, we will verify that the dataset we receive meets these. This includes checks for bad chromosomal mapping of SNPs, excess missingness, low mean allele frequency, excess homozygosity, Hardy Weinberg Equilibrium, etc. These quality control checks will be done using PLINK (Purcell et al, 2007) which was developed specifically for GWAS analyses and is being used in ARIC for GWAS analyses of other outcomes. We are using PLINK in another study and have gained significant experience with its implementation.

Analyses of both the quantitative (pulmonary function) and qualitative traits (COPD classification) will be done in PLINK. PLINK includes both logistic and linear regression and allows adjustment for covariates. PLINK allows for analyses stratified by environmental variables and includes testing of gene by environment interactions. PLINK includes permutation tests to adjust for multiple comparisons.

Stratified analyses will be important in addition to formal testing of interactions. For example if the group of subject living in close proximity to major roads is more susceptible there may be more power to find genome-wide association in this subgroup than to formally test for an interaction with traffic.

All analyses will be performed within broad racial strata of Caucasians and African-Americans. Within these strata we will check for population stratification using PLINK and adjust for it as necessary. The finding of the same gene by traffic interaction or stratum-specific association in both racial groups would give credence to a given result.

Replication of results:

We are aware of another study with pulmonary function and measures of traffic related air pollution in a large population study. This is the SAPALDIA study in Switzerland. While they are not doing a whole genome association study, we will discuss collaboration to replicate our top ranked findings both for interaction and stratum specific results. We should be able to do the genotyping with intramural funds to Dr. London at NIEHS. Dr. Nicole Probst-Hensch who leads the genetic part of the SALPADIA study has been receptive to collaborations with us.

Any anticipated methodological limitations or challenges: The analysis of GWAS data is just beginning in ARIC. We expect to learn from issues that come up in analyses that will be done before we complete ours. We recognize that the field of GWAS is rapidly changing. However, PLINK is widely regarded as a state of the art software package for this purpose and new functionality is constantly being added. We have members of our writing team who well versed in emerging issues in GWAS analysis techniques and familiar with PLINK.

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 ICTDER02 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 ICTDER02 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 ICTDER02 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
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)? None at this time.

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* 2003.03 )
Traffic-related air pollution and pulmonary function and symptomatic respiratory outcomes (AirPoll) 2003.03 -- in this project Dr. London added assessment of traffic exposure and baseline air pollution.

___  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.

13. Literature Cited


Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559-75. Epub 2007 Jul 25.


Vollmer WM. Reconciling cross-sectional with longitudinal observations on annual decline. Occup Med. 1993;8:339-51


Wilk JB, Walter RE, Laramie JM, Gottlieb DJ, O'Connor GT.