ARIC Manuscript Proposal # 1357

PC Reviewed: 04/08/08 **Status: A** **Priority: 2**
SC Reviewed: _____ **Status: __** **Priority: _**

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

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

2. Writing Group: Stephanie J. London (NIEHS), Dana B. Hancock (Postdoc, NIEHS), Grace Chiu (NIEHS contractor), Matthew B. Schabath (UTH), Jennifer.A.Nettleton (UTH), Lyn Steffen (U of MN), 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. _SJL____ **[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

Corresponding/senior author (if different from first author correspondence will be sent to both the first author & the corresponding author):
Name: Stephanie J. London

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

3. Timeline:

Statistical Analysis:	May 2008 to November 2008
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 know 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). Mounting evidence also suggests a role for diet in influencing pulmonary function. In the ARIC cohort, we have found pulmonary function and the presence of chronic obstructive pulmonary disease (COPD) as defined using pulmonary function test criteria, in cross-sectional analyses, to be influenced by dietary fiber (Kan et al. 2008) in both smokers and nonsmokers. Dietary fiber was protective for COPD and reduced lung function after taking into account intakes of antioxidant (vitamins C, E and carotenoids) and other micronutrients (vitamin D) which have been associated with better lung function in previous studies that did not examine fiber. In our ARIC analyses, higher fiber intake was consistently positively related to lung function based on FEV1, FVC, FEV1/FVC ratio as well as COPD defined by spirometry. These findings based on spirometry agree with our results on chronic bronchitis from a cohort of adults in Singapore (Butler et al. 2004) in which fiber intake was protective and appeared to explain inverse associations with intakes of antioxidant micronutrients. In our ARIC analyses, with fiber variables in the model, Vitamin D remained inversely related to the FEV1 and FVC, although with a weaker and less statistically significant association than fiber – vitamin C, vitamin E and carotenoids were no longer beneficially associated with lung function. Previous analyses in ARIC have shown intake of omega-3 fatty acids to be related to chronic obstructive pulmonary disease (Shahar et al. 1994). We previously described a gene by diet interaction in lung cancer which has been well replicated (London et al. 2000). We have described a genetic interaction involving antioxidant supplementation in relation to lung function in childhood asthma (Romieu et al. 2004) but there are few data on gene by diet interaction in relation to lung function.

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 nutrients related to COPD and PFTs in ARIC. Our primary interest is in dietary fiber, given our recent findings for main effects. However, we will also consider vitamin D which was independently inversely associated with lung function and COPD in our fiber analyses as well as omega-3 fatty acids which have previously been associated with COPD in ARIC. Because most previous studies of diet and COPD did not examine fiber but reported beneficial association with antioxidant micronutrients (vitamins C and E and carotenoids) that are correlated with fiber intake we will also need to examine these dietary factors in our analyses. The goal is both to examine hits in the main effects analysis as interacting factors with diet but also to use the whole genome analysis to identify SNP associations that may work only by interaction with the dietary factors.

5. Main Hypothesis/Study Questions:

1. Genetic variants interact with several exposure that are related to lung function in the ARIC cohort To identify the variants that interact with these factors, 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). The exposures of interest have been show by us or others to have main effects for pulmonary function in ARIC. While our primary interest is on dietary fiber, the focus of our recent paper, we will also examine vitamin D which was inversely related to some of the lung function and COPD parameters in models with fiber. Based on earlier ARIC findings, we will also examine interactions with omega-3 fatty acids.
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 within strata of exposure to the factors of interest. 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 do analyses based on the percent predicteds (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 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.

Self-reported doctor diagnosis of emphysema or chronic bronchitis is not regarded as a reliable index of COPD in the general population. In our analysis of ARIC data, 32% of subjects reporting MD diagnosis of emphysema did not have COPD by pulmonary function test criteria. Presumably these individuals have asthma or chronic respiratory symptoms due to other pathologies including cardiac disease or perhaps have some variability in their lung function close to the cutoff for the definition of COPD. Chronic bronchitis is defined clinically and epidemiologically as cough with phlegm production for at least three months out of the year for two years in a row. As expected the proportion of individuals with chronic bronchitis by these standard criteria who did not have COPD by pulmonary function test criteria was higher at 62%. The presence of chronic bronchitis by symptoms with normal pulmonary function tests is classified in the GOLD criteria as stage-0 COPD. 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 manner by symptoms.

Other variables of interest: In addition to dietary, pulmonary function and respiratory symptoms/disease variables, we will need to examine age, height, weight, gender, center, race, smoking (status, duration, amount), occupation, 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 the dietary and environmental factors of interest and includes testing of gene by environment interactions. PLINK includes permutation tests to adjust for multiple comparisons.

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 diet interaction in both racial groups would give credence to a given result.

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

___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)? None at this time.

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.

13. Literature Cited

Butler LM, Koh WP, Lee HP, Yu MC, London SJ. Dietary fiber and reduced cough with phlegm: a cohort study in Singapore. Am J Respir Crit Care Med. 2004 Aug 1;170(3):279-87. Epub 2004 Apr 29.

DeMeo DL, Silverman EK.Genetics of chronic obstructive pulmonary disease. Semin Respir Crit Care Med. 2003 Apr;24(2):151-60.

Kan H, Stevens J, Heiss G, Rose KM, London SJ.
Dietary fiber, lung function, and chronic obstructive pulmonary disease in the atherosclerosis risk in communities study. Am J Epidemiol. 2008 Mar 1;167(5):570-8. Epub 2007 Dec 5.

London SJ, Yuan JM, Chung FL, Gao YT, Coetzee GA, Ross RK, Yu MC. Isothiocyanates, glutathione S-transferase M1 and T1 polymorphisms, and lung-cancer risk: a prospective study of men in Shanghai, China. Lancet. 2000 Aug 26;356(9231):724-9.

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.

Romieu I, Sienra-Monge JJ, Ramírez-Aguilar M, Moreno-Macías H, Reyes-Ruiz NI,
Estela del Río-Navarro B, Hernández-Avila M, London SJ.
Genetic polymorphism of GSTM1 and antioxidant supplementation influence lung function in relation to ozone exposure in asthmatic children in Mexico City. Thorax. 2004 Jan;59(1):8-10.

Shahar E, Folsom AR, Melnick SL, Tockman MS, Comstock GW, Gennaro V, Higgins MW, Sorlie PD, Ko WJ, Szklo M. Dietary n-3 polyunsaturated fatty acids and smoking-related chronic obstructive pulmonary disease. Atherosclerosis Risk in Communities Study Investigators. N Engl J Med. 1994 Jul 28;331(4):228-33.

Vestbo J. COPD in the ECRHS. Thorax. 2004 Feb;59(2):89-90.

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

Wilk JB, DeStefano AL, Joost O, Myers RH, Cupples LA, Slater K, Atwood LD,
Heard-Costa NL, Herbert A, O'Connor GT, Gottlieb DJ. Linkage and association with pulmonary function measures on chromosome 6q27 in the Framingham Heart Study.
Hum Mol Genet. 2003 Nov 1;12(21):2745-51. Epub 2003 Sep 9.

Wilk JB, Walter RE, Laramie JM, Gottlieb DJ, O'Connor GT.
Framingham Heart Study genome-wide association: results for pulmonary function measures.BMC Med Genet. 2007 Sep 19;8 Suppl 1:S8.