# ARIC Manuscript Proposal #2668

**PC Reviewed: 11/10/15**  **Status: <u>A</u>**  **Priority: <u>2</u>**
**SC Reviewed: _____**  **Status: _____**  **Priority: ____**

**1a. Full Title:** Genome-wide Analysis of SNP-by-Omega-3 Fatty Acid Interaction on Pulmonary Function

**b. Abbreviated Title:** SNP-by-Omega 3 Fatty Acid Interaction on Pulmonary Function

**2. Writing Group:** Lyn Steffen, Ruixue Hou, Saroja Voruganti, Geetha Chittoor, Dana Hancock, Patricia Cassano, Misa Graff

ARIC Authors: Kari North, Lyn Steffen, other interested ARIC Authors are welcome.

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. KEN

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

Kari North
Associate Professor
Department of Epidemiology
Carolina Center for Genome Sciences
UNC Gillings School of Global Public Health
Chapel Hill, NC
919-966-2148 (voice)
919-966-9800 (Fax)

**3. Timeline (The following are estimates.  Please see project phases and timeline below for more details):**
 Individual cohort statistical analyses:   Fall 2015
 Consortium meta-analyses: Spring 2016
 Manuscript preparation:   Fall 2016
 Manuscript submission:   Spring 2017

**4. Rationale:** Pulmonary function, as measured by spirometry, is used clinically to diagnose and follow the progression of chronic obstructive pulmonary disease (COPD), the third leading cause of death in the United States. Pulmonary function is detrimentally affected by cigarette smoke, which causes oxidative stress and inflammation in the lung and airways and often leads to airflow obstruction and eventually COPD. However, dietary nutrients that attenuate inflammatory responses, specifically omega-3 fatty acids ($\omega$3 FAs), fiber, and vitamin D are among the few lifestyle factors associated with slowing pulmonary function decline and lowering risk of COPD, especially for people with a history of smoking. Furthermore a health dietary pattern has been identified as a means to reduce inflammation, and the 2015 Dietary Guidelines

Advisory Committee identified healthy diet patterns for chronic disease reduction, based on extensive systematic reviews of diet patterns and chronic disease risk. The two most commonly used measures of pulmonary function are forced expiratory volume in 1 second (FEV1) and its ratio to forced vital capacity (FEV1/FVC). Over 40% of the variability in these parameters is attributable to genetic factors, but few studies have considered gene by nutrient interactions given limitations in power in single studies.
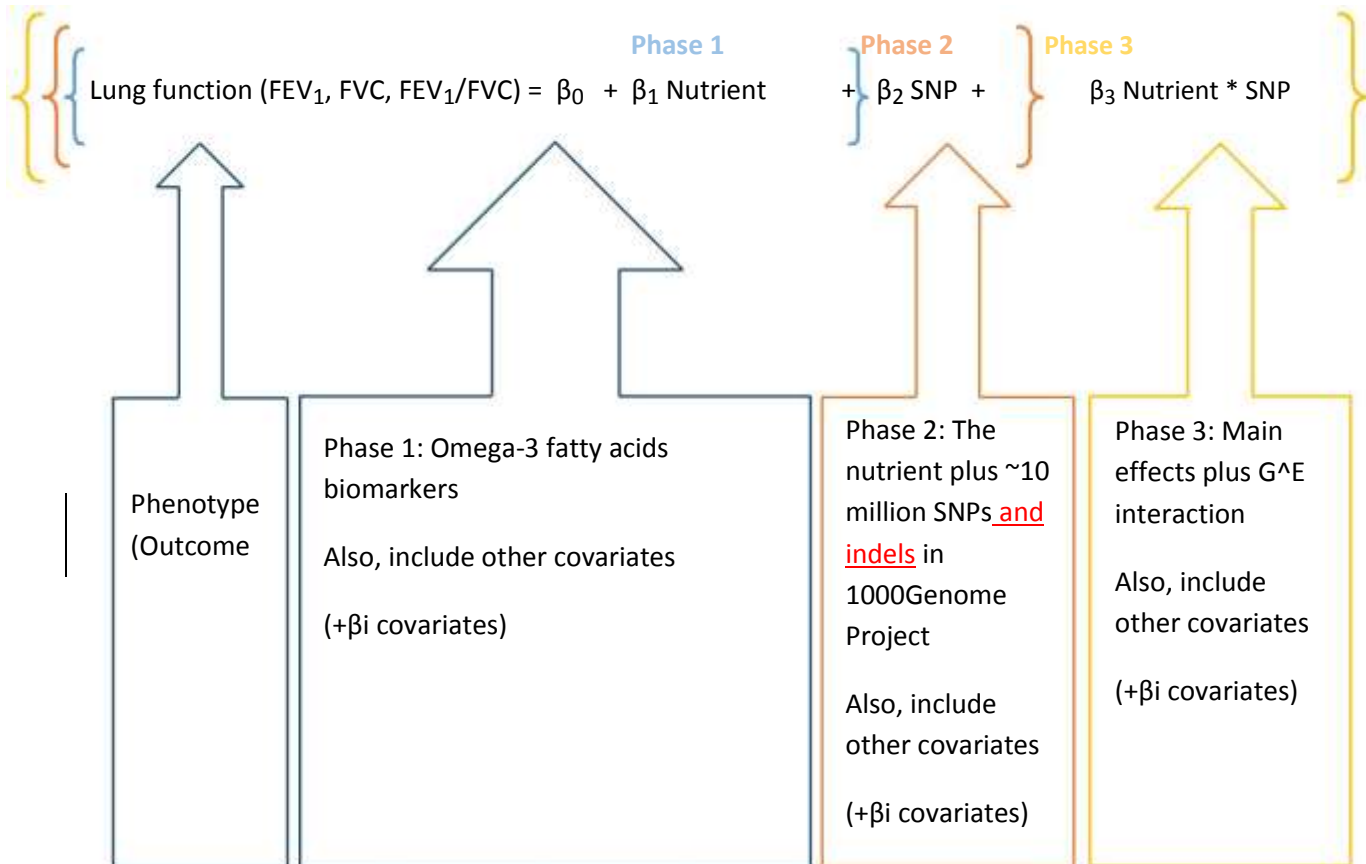
Our hypotheses focus on three specific nutrients, $\omega 3$ FAs), vitamin D and fiber, in addition to healthy dietary pattern. These nutrients, which are positively associated with pulmonary function in past studies, are hypothesized to play a role through anti-inflammatory mechanisms (Litonjua and ebrary Inc. 2012, Fonseca Wald, van den Borst et al. 2014). Dietary intake is a complex behavior with many contributing factors. Genetic variation affects nutritional status through variation in taste, variation in enzymes governing the absorption, transport and metabolism of nutrients, and variation in regulatory regions given the role of some nutrients as transcription factors.

This proposal specifically focuses on the relationship between $\omega 3$ FAs, both as derived from food frequency questionnaires and as a measured biomarker, and three pulmonary function measures (FEV1, FVC, and FEV1/FVC).

## 5. Main Hypotheses/Study Questions:

SNP-by-$\omega 3$ FA interaction across the genome will identify novel gene regions that influence the relation of $\omega 3$ FAs with pulmonary function and will also help characterize known gene regions.

## 6. Design and Analysis:

**Phase 1**     **Phase 2**     **Phase 3**

Lung function (FEV$_1$, FVC, FEV$_1$/FVC) = $\beta_0$ + $\beta_1$ Nutrient + $\beta_2$ SNP + $\beta_3$ Nutrient * SNP

Phenotype (Outcome

Phase 1: Omega-3 fatty acids biomarkers

Also, include other covariates

(+βi covariates)

Phase 2: The nutrient plus ~10 million SNPs and indels in 1000Genome Project

Also, include other covariates

(+βi covariates)

Phase 3: Main effects plus G^E interaction

Also, include other covariates

(+βi covariates)

- Other covariates include age, $age^2$, height, $height^2$, weight, gender, study site, smoking status (never/former/current), and pack-years.

- Stratification: all the models should be stratified by race/ethnicity (Caucasian or African American)


Phase 1 should be conducted on all participants with available pulmonary function and genetic data, as well as omega-3 fatty acids biomarkers and covariate data.

Phase 2 should be conducted on the model of phase 1 plus ~10 million SNPs and indels in 1000Genome Project (MAF > 0.01).

Phase 3 should be the main effects plus G*E interaction

### *Exclusion and Inclusion Criteria for Participants*

1 Pulmonary function test (PFT) data

Participants who have $FEV_1$, FVC, and $FEV_1$/FVC data will be included. Usually, $N_{FVC}$ should be equal to $N_{FEV1}$, and also equal to $N_{FEV1/FVC}$ (where N = number of participants).

In addition, for quality control, the standards for spirometry data are the same as in all previous CHARGE GWAS meta-analyses. Thus, $FEV_1$ and FVC measurements meeting the ATS/ERS criteria for acceptability will be included.

2 Omega-3 fatty acid biomarkers in units of % of total plasma fatty acids

Optimally, the timing of blood collection for the omega-3 fatty acid biomarker assay should be concurrent with the PFT data. If not, the timing difference, in units of days, between omega-3 fatty acid (ω-3 FAs) biomarker and PFT measurements will be computed. In this pipeline, we will use four ω-3 FA biomarkers measured in plasma and expressed as %total fatty acids; the four biomarkers are α-linolenic acid (ALA), eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA), and docosapentaenoic acid (DPA). All of these ω-3 FA biomarkers were measured in plasma in the ARIC study.

3 Cigarette smoking

Three smoking variables should be adjusted in all models: former smoker yes/no (1/0 coding), current smoker yes/no (1/0 coding), and pack-years (unit of cigarette pack-years; provide confirming details of calculations). Participants who do not have smoking data (never/ former/ current smoker and pack-years) should be excluded from all analyses.

1) <u>Smoking pack-years</u>, calculated cumulatively up to the time point corresponding to the PFTs. If missing at the time of PFTs, the previous closest time point (including the study baseline) should be used to calculate the pack-years.
2) <u>Smoking status</u> (never/former/current) should be included as two dummy variables: current smoking (yes/no) and former smoking (yes/no), at the time point corresponding to the PFTs being used for this study. If missing, the smoking status at the previous closest time point should be used.

4 Height

Individuals without baseline height data should be excluded. Height should be concurrent with the PFTs.

If not, the height data from the previous closest examination, including the study baseline, should be used. Height units should be **meters, m**.

5 Weight

Individuals without weight data can still be included in the analysis of $FEV_1$ and $FEV_1/FVC$, but they would be excluded in the analysis of FVC. In terms of FVC analysis, weight should be concurrent with the PFTs. If not, the weight from the previous closest examination, including baseline, should be used, as done for height. Weight units should be in kilograms, **kg**.

6 Age

Individuals without age data should be excluded from all analyses. Age should be concurrent with the $FEV_1$ measurement. Age should be in units of **years** of age.

7 Gender

Individuals without gender data should be excluded from all analyses.

8 Center

For cohorts with one study site, this covariate is omitted.

**PHASE 1- Main Effect of Omega-3 Fatty Acid Biomarkers on Pulmonary Function**

*Statistical Analysis*

Preliminary Modeling. We will construct a baseline model below to compute residuals and examine the corresponding residual diagnostic plot. Using the studentized residuals, we will apply a filter of ±3 or ±4 to identify outliers. We will examine outliers to confirm the presence of aberrant data, and edit data set accordingly. With the edited data set, we will run the same model again and compare the residual diagnostic plots for improved residual distribution.

Model Setup

*1 Baseline models without the nutrient term*

Standard ordinary least squares regression models, without including any of the nutrient-related variables, will be run to facilitate interpretation of results. This is the baseline model without considering nutrition:

1. $\mathbf{FEV_1} = \alpha + \beta_1$ current smoking $+ \beta_2$ former smoking $+ \beta_3$ pack-years $+ \beta_4$ age $+ \beta_5$ age$^2 + \beta_6$ sex $+ \beta_7$ height $+ \beta_8$ height$^2 + \beta_9$ center
2. $\mathbf{FVC} = \alpha + \beta_1$ current smoking $+ \beta_2$ former smoking $+ \beta_3$ pack-years $+ \beta_4$ age $+ \beta_5$ age$^2 + \beta_6$ sex $+ \beta_7$ height $+ \beta_8$ height$^2 + \beta_9$ center $+ \beta_{10}$ weight
3. $\mathbf{FEV_1/FVC} = \alpha + \beta_1$ current smoking $+ \beta_2$ former smoking $+ \beta_3$ pack-years $+ \beta_4$ age $+ \beta_5$ age$^2 + \beta_6$ sex $+ \beta_7$ height $+ \beta_8$ height$^2 + \beta_9$ center

*2 Extended models including omega-3 fatty acid biomarkers to test associations with PFT outcomes: note each model will be run separately for each biomarker, thus models 4a, 4b,*

*4c, and 4d include ALA alone, EPA alone, DHA alone and DPA alone, respectively.*

4. **FEV$_1$** = α + β$_1$ current smoking + β$_2$ former smoking + β$_3$ pack-years + β$_4$ age + β$_5$ age$^2$ + β$_6$ sex + β$_7$ height + β$_8$ height$^2$ + β$_9$ center + β$_{10}$ [**ALA or EPA or DHA or DPA**] (biological marker)

5. **FVC** = α + β$_1$ current smoking + β$_2$ former smoking + β$_3$ pack-years + β$_4$ age + β$_5$ age$^2$ + β$_6$ sex + β$_7$ height + β$_8$ height$^2$ + β$_9$ center + β$_{10}$ weight + β$_{11}$ [**ALA or EPA or DHA or DPA**] (biological marker)

6. **FEV$_1$/FVC** = α + β$_1$ current smoking + β$_2$ former smoking + β$_3$ pack-years + β$_4$ age + β$_5$ age$^2$ + β$_6$ sex + β$_7$ height + β$_8$ height$^2$ + β$_9$ center + β$_{10}$ [**ALA or EPA or DHA or DPA**] (biological marker)

*Below is for <u>dietary</u> omega-3 fatty acids, which will be different from omega-3 fatty acids <u>biomarkers</u> model, especially for the covariate "total caloric intake".*

**PHASE 2- MAIN GENETIC EFFECTS**

*We will next use the FEV1, FVC, and FEV1/FVC standardized residuals to run linear regression models that test genome-wide SNP and SNP-by-ω3 FA associations with FEV1 and, in parallel, with FVC and with FEV1/FVC. The stringent adjustment for smoking variables is needed, given that smoking is potentially a strong confounder correlated with SNPs, ω3 FAs, and pulmonary function. Of note, there may be some changes made to the model specified below, depending on the results of the nutrient effects on PFTs in phase 1 above.*

**Equation in all participants: FEV1 (or FVC or FEV1/FVC) = α + β1 SNP + β2 ω3 FAs + β3 SNP x ω3 FAs + β4 smoking status (current/former/never) + β5 pack-years + β6 SNP x pack-years + β7 total caloric intake**

**PHASE 3- INTERACTION MODEL**

*In addition to analyzing all participants, we will run the genome-wide analyses in current smokers only. Because of the strong biologic priors for ω3 FAs combating the harmful effect of cigarette smoking, we expect that stronger genetic associations may be seen in current smokers. Again, there may be some changes made to the model below, depending on the results of the nutrient effects on PFTs in phase I below.*

**Equation in current smokers only: FEV1 (or FVC or FEV1/FVC) = α + β1 SNP + β2 ω3 FAs + β3 SNP x ω3 FAs + β4 pack-years+ β5 SNP x pack-years + β6 total caloric intake**

*In general, we will declare genome-wide statistical significance in the full sample or in current smokers only, using a Bonferroni-corrected threshold. The exact p-value threshold will be determined to account for correlated outcomes and exposures; the most conservative threshold set at $P<8.3x10^{-9}$, based on one million independent tests across the genome for each of the 6 combinations of nutrients and phenotypes being studied (3 PFT measurements and 2 ω3 FAs [dietary and biomarker]).*

**References:**

Fonseca Wald, E. L., B. van den Borst, H. R. Gosker and A. M. Schols (2014). "Dietary fibre and fatty acids in chronic obstructive pulmonary disease risk and progression: a systematic review." Respirology 19(2): 176---184.

Litonjua, A. A. and ebrary Inc. (2012). Vitamin D and the lung mechanisms and disease associations.Respiratory medicine. New York, Humana Press/Springer,: 1 online resource.

Scragg, R., M. Sowers and C. Bell (2007). "Serum 25---hydroxyvitamin D, ethnicity, and blood pressure in the Third National Health and Nutrition Examination Survey." Am J Hypertens 20(7): 713---719.

**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)?**

GWAS of pulmonary function. Kari North was an Author on this work and is a member of the CHARGE pulmonary function group.

**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\* 2006.03 & 2007.02)**
**___ 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/

**12a. 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.**

**12b. The NIH instituted a Public Access Policy in April, 2008** which ensures that the public has access to the published results of NIH funded research.  It is **your responsibility to upload manuscripts to PUBMED Central** whenever the journal does not and be in compliance with this policy.  Four files about the public access policy  from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.