1. **Full Title**: Meta-analysis of 1000 Genomes imputed variants and pulmonary function in the CHARGE Consortium

   **Abbreviated Title (Length 26 characters)**: 1000G and PFTs: CHARGE

2. **Writing Group**: Annah Wyss, Stephanie London, Alanna Morrison, Kari North, Misa Graff, Nora Franceschini, Sarah Reese. Since this is a consortium paper a set number of authors per cohort will also be included (subject to change).

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

**First author**: Annah Wyss, PhD  
Address: NIEHS, PO Box 12233, MD A3-05. RTP NC 27709  
Phone: 919-541-0420  
Fax:  
E-mail: annah.wyss@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): Stephanie London  
Address: NIEHS, PO Box 12233, MD A3-05. RTP NC 27709  
Phone: 919-541-5772  
Fax:  
E-mail: london2@niehs.nih.gov


4. **Rationale**:

Spirometric measures of pulmonary function, such as forced expiratory volume in one second (FEV1), forced vital capacity (FVC) and their ratio (FEV1/FVC) are heritable traits influenced by both environment and genes. Recent genome-wide association studies (GWAS) in subjects of European ethnicity have identified at least 27 loci associated with FEV1 and/or FEV1/FVC (Hancock et al. Nat Genet 2010, Soler-Artigas et al. Nat Genet 2011, Hancock et al. Plos Genet 2012). We have also found novel signals for FVC in subjects of European ethnicity that replicate in other ethnic groups (Loth et al. Nat Genet...
Unfortunately, these signals are responsible for a minimal amount of the trait variation: for FEV1, the GWAS loci explain ~1.5% of inter-individual variance (adjusting for age, gender and height) and for FEV1/FVC the implicated loci explain 3.2% of the variation in this trait. The extent to which rare variants contribute to variation in these traits within the population is unknown. We seek to test the hypothesis that rare coding variation in conjunction with common variants contributes to inter-individual variability in these three spirometric phenotypes (FEV1, FVC, and FEV1/FVC).

The availability of 1000 Genomes imputation in ARIC and other cohorts enables evaluation of rare variants as well as variants that were not captured by HapMap imputation. This dataset includes about 40 million SNPs. Our partner consortium, SpiroMeta, has revisited meta-analysis of these same traits with 1000 Genomes imputation and identified novel signals – both relatively common SNPs not well captured by the previous HapMap imputation and rare variants. They have both identified novel loci and additional independent hits in our previously identified loci. These variants are not the same as those that we have identified in analysis of exome chip data.

Here we propose analyses of ARIC samples with both 1000 Genomes imputation and spirometry phenotype data (FEV1, FVC, and FEV1/FVC) with the goal of discovering new genes harboring rare coding variants associated with pulmonary function. The ARIC results will be combined with other studies participating in the CHARge Pulmonary Working Group using meta-analysis. The analysis will include European and non-European ethnic groups. Depending on the time course of SpiroMeta’s 1000 Genomes work, we will either publish our findings separately, if we have novel ones not identified by SpiroMeta, or combine the data with SpiroMeta as we have done in previous publications (Hancock et al. Nat Genet 2010, Soler-Artigas et al. Nat Genet 2011, Hancock et al. Plos Genet 2012, Loth et al. Nat Genet in press).

5. Main Hypothesis/Study Questions:

1. Are rare or common variants not covered in the earlier HapMap imputed GWAS dataset related to pulmonary function - FEV1/FVC, FEV1 and FVC?
2. Do these variants implicate novel loci?
3. Are these variants in previously implicated loci and do they explain our previous GWAS hits or are they independent?

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).

Detailed CHARge Pulmonary Working Group 1000 Genomes analysis plan is attached.

7.a. Will the data be used for non-CVD analysis in this manuscript?
     _X_ Yes  ____ No
7.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 ICTDER 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.csc.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)?

Our analysis plan is modeled on the GIANT 1000 Genomes adiposity group analysis plan which one of our co-authors, Kari North, kindly provided to us. In addition Alanna Morrison, is involved in 1000G analysis of lipid traits and other phenotypes.

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.csc.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.csc.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.