ARIC Manuscript Proposal #2639

PC Reviewed: 10/13/15  Status: A  Priority: 2
SC Reviewed: _________  Status: _____  Priority: ____

1.a. Full Title:
Genome-wide Association Study Using an Additive Model

b. Abbreviated Title (Length 26 characters):
GWAS Using Additive Model

2. Writing Group:
Writing group members:

Abhijit Mandal, James S. Pankow and Saonli Basu

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

First author:
Abhijit Mandal
Division of Biostatistics
University of Minnesota
A 460 Mayo Building, MMC 303
420 Delaware St SE
Minneapolis, MN 55455
Phone: 612-392-1005
Fax: 612-626-0660
Email: mandala@umn.edu

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: James S. Pankow
Division of Epidemiology & Community Health
1300 S 2nd St, Suite 300
Minneapolis, MN 55454
Phone (Work): 624-2883
Or: Barb Lux, 626-9092, b-lux@umn.edu
Email: pank0001@umn.edu
3. **Planned Timeline:** Complete ARIC analysis by October, 2015. Submit manuscript by November, 2015.

4. **Rationale:**

Genome-wide association studies (GWASs) have identified several important genetic variants associated with complex diseases and traits. However, in most instances these variants explain very little of the disease or trait heritability. The typical single locus association analysis in a GWAS fails to detect variants with small effect sizes, and also unable to capture the higher order interactions among these variants. Multilocus association analysis provides a powerful alternative by jointly modeling the variants within a gene. It also reduces the burden of multiple hypothesis testing in a GWAS.

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

There are several methods for gene-based analysis in the literature. Typically the regression based methods assume that the effect of single nucleotide polymorphisms (SNPs) and covariates are linearly associated with the trait. However, these tests may give poor performance if the linear model fails to hold properly. To overcome this problem we have developed a nonparametric test based on the additive model. The test is very simple and computationally efficient to conduct at the GWAS level. First the advantage of this method will be shown using simulated data. To demonstrate the usefulness of this test and compare to other methods, we wish to provide a real-life example that includes a continuous outcome variable and several continuous non-genetic predictors. One outcome that we think will fit well within these parameters is lung function. We propose to analyze associations between GWAS SNPs and lung function in Caucasians from the Atherosclerosis Risk in Communities (ARIC). The primary lung function variable is FEV1/FVC, which is the ratio between the forced expiratory volume in 1 second (FEV1) and the forced vital capacity (FVC). We will assess whether there are any nonlinear effects of SNPs, and covariates like cigarette-years of smoking, height, weight and age on the ratio FEV1/FVC. Finally, we will compare our method with the classical regression analysis, and some popular meta analysis methods like GATES, VEGAS and the min-P test.

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).**

Variables/measurements from the ARIC main study database to be analyzed:

Lung function quantitative traits (FEV1/FVC) measured at visit 1, demographic factors (age, gender); other risk factors (cigarette-years of smoking, height, weight); genotypes already generated using Affymetrix 6.0 array for ongoing GWAS projects.
Design and analysis:

We want to investigate the joint effect of all SNPs in a gene on FEV1/FVC using ARIC data. We will adjust the test by taking related factors, like cigarette-years of smoking, age, gender, height, weight etc, as covariates.

The traditional methods in regression parametrically model the data, however they show poor performance if a suitable model is difficult to find. To overcome this problem we will build a nonparametric additive model, and propose a computationally efficient method to test the association of a gene with the disease. We will check if there is any nonlinear effect of SNPs and covariates on FEV1/FVC.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes _x__ 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? _ ____ 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.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)?

1360: Genome-Wide Association Study of Chronic Obstructive Pulmonary Disease (COPD) Phenotypes and Lung Function Parameters: The Atherosclerosis Risk in Communities Study.

2131: Meta-analysis of exome chip variants and pulmonary function in the CHARGE consortium.

2157: Pathway analysis based on meta-analysis of genome wide association studies of FEV1 and FEV1/FVC.

2397: Meta-analysis of 1000 Genomes imputed variants and pulmonary function in the CHARGE Consortium.

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)
   ____   B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* __________ __________ __________)

   (Stampeed and Geneva genotype funding in Caucasians)

*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.
References:


