ARIC Manuscript Proposal #2164

PC Reviewed: 7/9/13	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: **A Bayesian Dimension Reduction Approach for Detection of Multi-locus Interaction in Case-Control Studies**

b. Abbreviated Title (Length 26 characters):
Bayesian modeling of multi-locus interaction

2. Writing Group:
Writing group members:

Debashree Ray, Xiang Li, Wei Pan, James S. Pankow, Saonli Basu

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

First author:
Debashree Ray
Division of Biostatistics,
University Of Minnesota
A 460 Mayo Building, MMC 303,
420 Delaware St SE
Minneapolis, MN 55455
Fax: 612-626-0660
Email: rayxx267@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: panko001@umn.edu

3. Timeline:

4. Rationale:

Several genome-wide association studies (GWAS) have been published on various complex diseases. Although new loci are found to be associated with these diseases, only very little of the genetic risk for these diseases is typically explained. The major limitation of single-locus association analysis in a GWAS is that most genome-wide tests require a high significance threshold, which makes it difficult to detect variants with small effect sizes. Also it ignores interaction among SNPs. Each locus alone may have little or no effect on the risk of the complex disease but collectively may increase the risk of the disease substantially (gene-gene interaction). There is evidence that diseases often arise as a result of complicated interactions among SNPs (Gabriel et al., 2002; Merryweather-Clarke et al., 2003). Here we attempt to develop a Bayesian multi-locus modeling approach to improve detection of loci by modeling interaction.

5. Main Hypothesis/Study Questions:

Multi-locus association analysis can improve power to detect loci associated with a disease by jointly modeling the SNP effects within a gene and by reducing the burden of multiple hypotheses testing in GWAS. The overall objective of this proposal is to develop methods and software for studying association between multiple SNPs and a disease. We intend to jointly model multiple SNPs and study their effect on type-2 diabetes. Our goal is to improve power of detection of different groups of interacting SNPs by using dimension-reduction techniques and by modeling interaction. We intend to perform our analysis using data on type 2 diabetes prevalence in Caucasian population of ARIC cohort and compare our findings with the single SNP and gene-based association findings for the disease.

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:

Diabetes status at baseline (prevalence), demographic factors (age, gender, race/ethnicity);; genotypes already generated using Affymetrix 6.0 array for ongoing GWAS projects.

Design and analysis:

We propose to investigate the marginal as well as joint effects of groups of SNPs on type 2 diabetes in the Atherosclerosis Risk in Communities (ARIC) Study. We will also identify genes related to type 2 diabetes and use our model as well as existing approaches to compare the effects of these genes.

We intend to use a Bayesian partitioning model, which clusters SNPs according to their direction of association, models higher order interactions using a flexible scoring scheme, and uses posterior marginal probabilities to detect association between the SNP-set and the disease. For any SNP-set, only three parameters are needed to model multi-locus interaction, which is considerably less than any other competitive parametric method. We intend to use extensive simulation studies and type 2 diabetes data on Atherosclerosis Risk in Communities (ARIC) study to illustrate our model and compare our approach with several other multi-locus approaches.

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)?

1470: Genome-wide association study of prevalent type 2 diabetes in the ARIC Study

The current proposal differs from the ms 1470 in that it proposes Bayesian modeling of multiple SNPs in a gene-based association study.

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* _____)

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

2006.03 (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 central.

References:

1. Gabriel, S. B., Salomon, R., Pelet, A., Angrist, M., et al. (2002). Segregation at three loci explains familial and population risk in hirschsprung disease. Nature Genetics 31, 89-93.

2. Merryweather-Clarke, A. T., Cadet, E., Bomford, A., Capron, D., et al. (2003). Digenic inheritance of mutations in HAMP and HFE results in different types of haemochromatosis. Hum. Mol. Genet. 12, 2241-2247.