ARIC Manuscript Proposal #2496

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

1.a. Full Title: USAT: A Unified Score-based Association Test for Multiple Phenotype-Genotype Analysis

b. Abbreviated Title (Length 26 characters): USAT for Multiple Phenotypes

2. Writing Group:
Writing group members:

Debashree Ray, 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:

Genome-wide Association Studies (GWASs) for complex diseases often collect data on multiple correlated endo-phenotypes. The standard approach to analyze these phenotypes is to perform single-trait analyses separately and report the findings for each individual trait. van der Sluis et al. (2013) demonstrated several alternative models which would benefit from a joint analysis. Blair et al. (2013) illustrated the comorbidity between Mendelian disorders and different complex disorders, which indicates that there may be common genetic variants affecting several of these complex traits. Thus, multivariate analysis of these correlated phenotypes can improve the power to detect genetic variants. Several methods are developed to model correlated phenotypes, but they are computationally very intensive, especially when we are dealing with GWAS data (Kent Jr, 2009). Multivariate analysis of variance (MANOVA) can perform such association analysis at a GWAS level, but the behavior of MANOVA under different trait models has not been carefully investigated. Here we investigate the performance of MANOVA under various alternative trait models and propose a new method, USAT, that can analyze the joint association of a genetic variant with multiple traits.

5. Main Hypothesis/Study Questions:

Analyzing multiple disease-related phenotypes could potentially increase power to detect association of SNPs/genes with a disease. Moreover this joint analysis could reveal some pleiotropic genes involved in the biological development of the disease. The overall objective of this proposal is to study the performance of MANOVA in detecting association under different multiple trait models. We also intend to develop a novel method and software for studying association between genes and multiple correlated phenotypes. We intend to jointly model multiple disease (type 2 diabetes)-related phenotypes to study pleiotropic effect as well as to increase the power to detect association of SNPs with the disease. We intend to perform our analysis using the Caucasian population of ARIC cohort and compare our findings with the single SNP 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-related quantitative traits (fasting glucose, fasting insulin, 2-hr glucose) at visit 4, demographic factors (age, gender, race/ethnicity); other diabetes risk factors (BMI, waist circumference) at visit 4; genotypes already generated using Affymetrix 6.0 array for ongoing GWAS projects. The diabetes status at visit 4 will be used to screen out diabetic individuals.

Design and analysis:

We propose to investigate the marginal effects of SNPs on type 2 diabetes related quantitative traits in the Atherosclerosis Risk in Communities (ARIC) Study. We will perform single SNP analysis with multiple correlated type 2 diabetes-related phenotypes.

The traditional methods for multivariate modeling impose significant computational challenges. Here we compare five approaches such as minimum p-value method (minP), multivariate analysis of variance (MANOVA), Fisher's combination of p-values, Trait based association test (TATES), Sum of Squared Score (SSU) test through extensive simulation studies. We propose computationally efficient ways of applying these methods to study the effects of single genetic variant on multiple traits and show under what scenarios we gain power by modeling these traits simultaneously.

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 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.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)?

1409: Genome-wide association study of diabetes-related quantitative traits in ARIC whites
1468: Genetic variants are associated cross-sectionally and longitudinally with multiple measures of fasting glucose
1470: Genome-wide association study of prevalent type 2 diabetes in the ARIC Study

The objective of all of these related manuscripts is to conduct a GWAS for individual diabetes-related traits; none propose multivariate modeling of correlated traits.

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)* _____ _____ _____)

*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 :

Blair, D., Lyttle, C., Mortensen, J., et al. (2013). A nondegenerate code of deleterious variants in mendelian loci contributes to complex disease risk. Cell 155, 70-80.

Kent Jr, JW (2009), Analysis of Multiple Phenotypes, Genet Epidemiology 2009, 33 (Suppl 1): S33-S39.

van der Sluis, S., Posthuma, D., and Dolan, C. (2013). TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. PLoS Genet 9(1), e1003235.