ARIC Manuscript Proposal # 1662

PC Reviewed: 6/8/10	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title:
Pleiotropy and pathway analyses of genetic variants associated with both type 2 diabetes and colorectal, breast, and prostate cancers

b. Abbreviated Title (Length 26 characters):
Pleiotropy and pathway

2. Writing Group:
Writing group members: Laura A. Raynor, James S. Pankow, Laura Rasmussen-Torvik, Weihong Tang, Anna Prizment, David Couper

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

F**irst author**: **Laura A. Raynor**
Address: Division of Epidemiology and Community Health
University of Minnesota
1300 S. Second Street, Suite 300
Minneapolis, MN 55454-1015

Phone: 612-626-7921 Fax: 612- 624-0315
E-mail: rayno007@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**
Address: Division of Epidemiology and Community Health
University of Minnesota
1300 S. Second Street, Suite 300
Minneapolis, MN 55454-1015

Phone: 612-624 2883 Fax: 612-624-0315
E-mail: panko001@umn.edu

3. Timeline:

June 2010: Proposal submitted
July-September 2010: Analysis
October-November 2010: Writing of the paper
December 2010-: Revisions by co-authors
January 2011: Submission to journals

4. Rationale:

Type 2 diabetes is an epidemic in the United States, affecting around 7% of the population and leading to significant increases in morbidity, mortality, and long-term healthcare costs (Rahman et al. 2008; Kolberg et al. 2009). There is evidence that type 2 diabetes is a risk factor for several cancers including colon, prostate, and breast cancers (Xue and Michels 2007; Folsom et al. 2009). Pathophysiological explanations have been generated to explain the connections between these diseases and several genes associated with type 2 diabetes risk have also been found to be associated with cancer outcomes such as prostate and colorectal cancer (Severi et al. 2007; Folsom et al. 2008; Pal et al. 2009; Meyer et al. 2010). However, to date there have been no extensive searches of GWA datasets for evidence of shared genetic regions between type 2 diabetes and cancer outcomes.

References

1. Folsom AR et al. Variation in TCF7L2 and increased risk of colon cancer: the Atherosclerosis Risk in Communities (ARIC) Study. Diabetes Care. 2008;31(5):905-9.
2. Kolberg JA et al. Development of a type 2 diabetes risk model from a panel of serum biomarkers from the Inter99 cohort. Diabetes Care. 2009; 32(7):1207-12
3. Meyer TE et al. Diabetes genes and prostate cancer in the Atherosclerosis Risk in Communities study. Cancer Epidemiology Biomarkers and Prevention. 2010;19(2):558-65.
4. Pal et al. Common variants in 8q24 are associated with risk for prostate cancer and tumor aggressiveness in men of European ancestry. The Prostate; 2009:69:1548-56.
5. Rahman M et al. A simple risk score identifies individuals at high risk of developing Type 2 diabetes: a prospective cohort study. Family Practice. 2008; 25(3):191-6.
6. Severi G et al. The common variant rs1447295 on chromosome 8q24 and prostate cancer risk: results from an Australian population-based case-control study. Cancer Epidemiology Biomarkers and Prevention. 2007;16(3):610-2.
7. Xue F and Michels KB. Diabetes, metabolic syndrome, and breast cancer: a review of the current evidence. American Journal of Clinical Nutrition. 2007; 86(3): s823-35.

5. Main Hypothesis/Study Questions:

We will use the GWA data in ARIC to look for genes that are associated with both incident type 2 diabetes and colorectal cancer and if numbers permit, prostate and breast cancer outcomes. We will do so by summarizing variation across nominally

associated loci into quantitative risk scores and determining if these scores are associated with type 2 diabetes and cancer outcomes. Furthermore, we will conduct a pathway analysis to determine if biochemical pathways are shared between the type 2 diabetes and cancers.

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).

Analyses will be restricted to Caucasian participants. Individuals with a history of cancer or prevalent diabetes at the baseline examination will be excluded from analysis in addition, to individuals with missing risk factor data. All remaining participants will be followed through 2006, which is the most recent data available on incident cancer outcomes. Incident type 2 diabetes is defined a history of physician-diagnosed diabetes, current use of anti-diabetes mediation, having a fasting serum glucose ≥126 mg/dL, or having a non-fasting glucose measure ≥200 mg/dL. Incident cancer outcomes in this cohort have been ascertained by linkage to cancer registries.

To analyze events for the GWA, we will use Cox proportional hazard models to calculate hazard ratios and corresponding 95% confidence intervals using SAS v.9.2 (Cary, NC) and assuming an additive genetic model. Cox models will be adjusted for age, sex, and field site. Following the GWA, we will randomly divide the ARIC dataset into two sub-sets of equal size, the training and the testing sets. To create a risk score in the training dataset, we will reduce the number of SNPs available for analysis by filtering on minor allele frequency, genotyping rate, and linkage disequilibrium independent of their association with type 2 diabetes. We will use the training set to obtain sets of alleles that are significantly associated with type 2 diabetes at increasingly liberal thresholds (P_T<0.01, 0.2, 0.3, 0.4, and 0.5) in Cox regression. For each individual in the testing set we will calculate the number of score alleles they have, weighted by the log hazard ratio for each allele, from the training dataset. To assess whether the aggregate scores reflect diabetes risk, we will test for higher mean scores in cases compared to controls (Purcell et al. 2009). We will also examine diabetes risk by quantiles of aggregate scores to examine the pattern and strength of association and determine the degree to which these patterns differ according to the threshold used to select SNPs.

In order to determine if genes are shared between diabetes and cancers, we will test whether the derived diabetes risk score is associated with colorectal, prostate, and/or breast cancer outcomes for any P_T thresholds. If there is an association between the diabetes risk score and any cancer, we will examine the risk by quantiles of aggregate scores to examine the pattern and strength of association and determine the degree to which these patterns differ according to the threshold used to select SNPs

Subsequently, we will use the SNPs that were filtered by minor allele frequency, genotyping rate, and linkage disequilibrium and look for pathways that overlap between disease outcomes. We will enter a list of these SNPs and their corresponding p-values into the Gene Set-based Analysis of Polymorphism (GeSBAP), which will return significant functional categories, the associated p-values, and the genes included in each category that are connected with both cancer and diabetes outcomes.

GeSBAP selects SNPs that map into genes or their neighborhoods (±5 kb). The SNP with the highest association to the trait studied is taken as the proxy of the gene. All the genes are mapped to their corresponding functional categories, which we will select a priori, and ranked accordingly to their proxy polymorphisms. A gene set analysis (GSA) test is used to check for functional categories showing significant associations to the trait studied. We will compare the pathways between diabetes and cancer outcomes to identify those that overlap.

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 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 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)?

1. ARIC proposal 1444- Adiponectin (ADIPOQ) and adiponectin receptor (ADIPOR1, and ADIPOR2) SNPs and the incidence of cancer: Atherosclerosis Risk in Communities(ARIC) study
2. ARIC proposal 1227- TCF7L2 variants and cancer
3. ARIC proposal 1280- Interactions between diabetes, diabetes genes, and the androgen receptor gene on risk of prostate cancer.

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 GWA for loci influencing incident CHD and other HLB phenotypes (NHLBI RFA for large scale genotyping) (STAMPEED) (GEI)
2007.02 The National Heart Lung and Blood Institute's Candidate Gene Association Resource (CARE): Phase I (CARE)

___ 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/

12. **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.**