ARIC Manuscript Proposal #2763

PC Reviewed: 6/7/16           Status: A           Priority: 2
SC Reviewed: _______           Status: _____          Priority: _____

1.a. Full Title: A pathway-based analysis of germline variants in DNA-repair genes and prostate cancer mortality in the Atherosclerosis Risk in Communities (ARIC) Cancer Study

b. Abbreviated Title (Length 26 characters): DNA-repair SNPs and prostate cancer

2. Writing Group:
   Writing group members: Danyelle Winchester, Elizabeth A. Platz, Corinne Joshu, Jianfeng Xu, Nilanjan Chatterjee, Mara Vitolins, and other interesting ARIC investigators

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

First author: Danyelle A. Winchester
Address: 615 N. Wolfe Street
         Office E6133
         Baltimore, MD, 21205
         Phone:               Fax: (410) 614- 2632
         E-mail: dwinche2@jhu.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: Elizabeth A. Platz
Address: 615 N. Wolfe Street
         Office E6132
         Baltimore, MD 21205
         Phone: (410) 614-9674      Fax: (410) 614- 2632
         E-mail: eplatz1@jhu.edu

3. Timeline:
Data analysis – 8 months
First draft of manuscript – 8 months
4. **Rationale:**
In the United States, the incidence and mortality rates of prostate cancer have declined over the past 10 years (1). However, in 2016 it is estimated that over 26,000 men will die from prostate cancer (1). The accumulation of DNA lesions and reduced DNA repair activity in cancer cells has been reported as a contributing factor of prostate cancer (2–4). Several studies have previously reported that carriers of rare germline mutations in DNA-repair genes *ATM, BRCA1, BRCA2, BRIPL, CHEK2* and *NBN1* are at greater risk of developing prostate cancer, and increased risk of advanced disease, metastatic spread and poorer survival outcome (5–11).

The DNA repair pathways are a set of mechanisms that detects and corrects genomic damages that occur across the genome. There are four main pathways involved in DNA repair, which includes base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR) and double strand break repair (DSBR); and are essential to maintaining genetic stability in normal and cancer cells (12). Variations and mutations in genes involved in DNA repair can cause genetic instability, ultimately leading cancer development and a more aggressive disease through the alteration of DNA repair pathways (13).

There are several studies that have examined the association between sequence variants in genes involved in DNA repair and prostate cancer (14–19). A previous study has implicated single nucleotide polymorphisms (SNPs) in *ERCCI* and *ATM* in the development of higher-grade prostate cancer (20). Also, a SNP in *MLH1* was reported to be associated with overall prostate cancer risk, aggressive disease, and prostate cancer recurrence, but not prostate cancer–specific mortality (21). A single variant in *NBSI* was found to be associated with a higher risk of advanced prostate cancer, which suggests a possible role in prostate cancer progression as well (22). Recently, a study reported that a gene set in the DNA damage and repair (DDR) pathways were significantly associated with biochemical recurrence-free, metastasis-free, and overall survival in prostate cancer cases who underwent prostatectomy. This study also reported that seven DDR pathways were associated with prognosis (23).

Despite these research advances, there remains limited information about SNPs in genes involved in DDR pathways and prostate cancer mortality and case fatality. Thus, we propose to analyze data from a GWAS performed in ARIC to determine whether variants in genes involved the DDR pathways are associated with prostate cancer mortality and case fatality. The previous studies referenced above have been limited to white populations. The ARIC study provides an opportunity to further examine whether variants in genes involved in the DDR pathways are associated with lethal prostate cancer in black men as well.

In addition, DNA lesions accumulate due to damage caused by reactive oxygen species (ROS), which can be produced by certain lifestyle factors including smoking and obesity (24). Therefore, we will conduct stratified analyses to determine whether such modifiable factors that are also associated with fatal prostate cancer – smoking and obesity – modify the association between the SNPs and prostate cancer mortality and case-fatality.

5. **Main Hypothesis/Study Questions:**
1. To determine whether SNPs in genes involved in the DNA damage and repair DDR pathways are associated with prostate cancer incidence, especially disease with a lethal phenotype, and
prostate cancer mortality in a cohort of white and black men without the diagnosis of prostate cancer at baseline.

2. To determine whether SNPs in genes involved in the DDR pathways are associated with death from prostate cancer in men from the cohort of white and black men with the diagnosis of prostate cancer.

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).

Study design
We will use the BioCarta pathway database to identify genes involved in the DNA damage and repair (DDR) pathways pathway, and then identify the SNPs in those genes that are present on the Affymetrix Genomewide SNP Array 6.0, which was used in ARIC. Using those SNPs, we will conduct two prospective cohort analyses to determine if SNPs in genes involved the DDR are associated with risk of prostate cancer (incidence) and mortality in men without the diagnosis at baseline (cohort 1) and risk of death from prostate cancer in men with the diagnosis (cohort 2).

Inclusion/exclusion
All ARIC participants with adequate genotyping data available will be included in this study. We will exclude all participants who did not provide informed consent to DNA usage for future studies and/or to studies of other chronic diseases like cancer.

Cohort 1: men only, no reported cancer diagnosis of any site at baseline

Cohort 2: men only, no cancer reported cancer diagnosis of any site at baseline, had a prostate cancer diagnosis during follow-up, and has a date of prostate cancer diagnosis (needed because date of diagnosis becomes the start of follow-up).

Exposure:
Already genotyped (Affymetrix Genomewide SNP Array 6.0) and imputed SNPs in genes involved in the DNA repair pathways.

Outcome:

Prostate cancer incidence
We will use the ARIC prostate cancer case file 1987-2012, which includes total incidence as well as incidence of prostate cancer with a lethal phenotype. Prostate cancer incidence data was obtained from cancer registries, hospital records, and death certificates. A subanalysis will be conducted excluding men whose prostate cancer diagnosis was only ascertained by death certificate.

Prostate cancer mortality and case-fatality
Event endpoints will include prostate cancer mortality among all men and prostate cancer case fatality among men diagnosed with prostate cancer. Only deaths from prostate cancer as the
underlying cause will be considered. Separate analyses will be conducted for each endpoint. Deaths from prostate cancer (ICD-10 code C61) during the follow-up period have been identified from death certificates and the National Death Index through 2012.

**Statistical analysis**

To determine whether SNPs in genes involved in the DNA damage and repair (DDR) pathways are associated with prostate cancer incidence, we will use Cox proportional hazards regression to estimate hazard ratios (HR) and 95% confidence intervals (CI) assuming an additive model or a co-dominant model adjusting for race*field center and using age as the time scale. At this time, modifiable risk factors for prostate cancer incidence are uncertain. To determine if the association between DDR SNPs and prostate cancer differs by race, we will repeat the analysis stratified by race and will test for statistical interaction between the SNPs and race using the likelihood ratio test.

To determine whether SNPs in genes involved in the DNA damage and repair (DDR) pathways are associated with prostate cancer mortality, we will perform Cox proportional hazards models to estimate HRs and 95% CI of prostate cancer mortality assuming an additive model or a co-dominant model adjusting for race and using age as the time scale. While it is unlikely that risk factors for fatal prostate cancer confound the genes-prostate cancer association, we will nevertheless additionally for smoking, body mass index, and height. We will determine whether the number of black men is large enough for stratified analysis, and if so, will perform the analysis stratified as per above. If no, in a subanalysis, we will restrict to white men.

To determine whether SNPs in genes involved in the DNA damage and repair (DDR) pathways are associated with prostate cancer case fatality, we will perform Cox proportional hazards regression to estimate HRs and 95% CI of prostate cancer specific death among men with a diagnosis of the disease assuming an additive model or a co-dominant model adjusting for age and race and beginning follow-up at the date of diagnosis (time scale is time since diagnosis). Where data are available, we will additionally adjust for stage and grade at diagnosis to be able to assess the association between these SNPs and outcome beyond the influence of these SNPs on the development of more aggressive disease and/or stratify by stage (localized or advanced) or grade (Gleason <7, 7+) of disease. We will additionally adjust for risk factors for the prostate cancer death – smoking, body mass index, and height. If the sample size for black men is large enough for stratified analysis, will perform the analysis stratified as per above. If no, in a subanalysis, we will restrict to white men.

After performing the above analyses, we will develop genetic risk scores (GRS) using the top-hits approach, in which we will select those SNPs that were associated with incidence, mortality, or case fatality in this cohort. To generate a GRS, we will sum across the associated SNPs the number of risk alleles for each SNP. We will consider whether weighting each risk allele by its association in other studies is appropriate for this cohort. We will not use the RRs from this study as the weights to avoid over fitting the data. To determine whether the derived GRSs are associated with prostate cancer outcomes above, we will use Cox proportional hazards models to estimate HRs and 95% CIs. We will model the GRS in several ways, including as a count and in quantiles based on distribution in cohort 1.
For each of the above outcomes, we will also conduct analyses stratified by the modifiable factors smoking and obesity, and also for the non-modifiable factors age and height. The presence of a statistical interaction for each SNP or GRS and modifiable and non-modifiable factors will be evaluated using the likelihood ratio test.

Finally, we will consider whether a pathway-based approach to the analysis is feasible (25).

All statistical analyses will be performed using R programing software.

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 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)?

- Proposal# 1045 (first author: Kari E. North): Gene-by-smoking interaction. subclinical atherosclerosis, incident CHD and stroke in ARIC (AS#2002.06)
- Proposal #1238 (first author: Kari E. North): DNA-damage pathway and genetic susceptibility to type 2 diabetes mellitus and insulin resistance states
- Proposal #2157 (first author: Stephanie London): Pathway analysis based on meta-analysis of genome wide association studies of FEV₁ and FEV₁/FVC
- Proposal #2157 (first author: Jason Wu): Genome-wide association study (GWAS) of plasma fatty acid biomarkers in the de novo lipogenesis pathway: CHARGE fatty acid 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* 2011.07, 1995.04)
____ 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.csccl.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.csccl.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


21. Langeberg WJ, Kwon EM, Koopmeiners JS, Ostrander EA, Stanford JL. Population-based study of the association of variants in mismatch repair genes with prostate cancer risk and


