1.a. Full Title: Multiethnic GWAS of Prostate Cancer

b. Abbreviated Title (Length 26 characters): prostate cancer GWAS

2. Writing Group:

The writing group has not yet been determined. Those leading the multiethnic GWAS meta-analysis in the PRACTICAL Consortium include:

Chris Haiman¹, David Conti¹, Ros Eeles² and Zsofia Kote-Jarai²
¹Department of Preventive Medicine, University of Southern California, Los Angeles, CA
²Institute for Cancer Research, London UK

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

First author: Chris Haiman
Address: Department of Preventive Medicine
University of Southern California
1440 Biggy St
Los Angeles, CA 90033

Phone: 323 442 7755 Fax: 323 442 7749
E-mail: haiman@usc.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).

Elizabeth A. Platz, ScD, MPH
Professor and Martin D. Abeloff, MD Scholar in Cancer Prevention
Deputy Chair
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe St., Room E6132
Baltimore, MD 21205
P 410.614.9674
F 410.614.2632
3. **Timeline:** ~ 3 years

4. **Rationale:**

There has been substantial progress in the discovery of susceptibility loci for prostate cancer (PCa), with genome-wide association studies (GWAS) revealing ~270 loci that explain ~50% of familial risk. Greater than 90% of these loci were found in studies conducted in men of European ancestry. Men of African ancestry are at greater risk of developing and dying from PCa compared to men from other populations, and there is growing evidence to support a genetic contribution to PCa risk disparities. In 2008, we established the African Ancestry Prostate Cancer Consortium (AAPC), which today includes 10,368 cases and 10,986 controls of African ancestry from 37 studies throughout the U.S., Africa, and Caribbean. In AAPC, multiple novel PCa risk variants have been discovered with sizable relative risks (1.5-2.2) that are only found in populations of African descent. We have also led GWAS and fine-mapping analyses in populations of African, European, Latino and Asian ancestry which have highlighted the importance and power of multiethnic genetic studies for discovery and fine-mapping. The majority of GWAS variants are thought to alter risk for PCa by disrupting regulatory elements, which in turn alter transcript levels of their target genes. To examine this hypothesis, we also introduced gene expression imputation as means to integrate regulation of gene expression (eQTL) with GWAS signals to identify novel susceptibility genes. Although transcriptome-wide association scans (TWAS) have proven extremely powerful for a wide-range of diseases, including PCa in men of European ancestry, it has yet to be applied/validated in non-European studies of PCa.

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

Thus, we propose to (1) substantially augment the sample size of the AAPC Consortium as well as the size of GWAS in Latino and Asian men for increased power for discovery of risk alleles for overall PCa and aggressive disease, (2) measure and correlate transcriptomic profiles in prostate tissue with genetic variation to identify novel susceptibility genes for PCa in men of African, Asian and Latino ancestry and, (3) apply innovative approaches for developing and evaluating a polygenic risk score to quantify the contribution of germline variation to ethnic differences in PCa incidence, using all available data on individuals of European, African, Asian, and Latino ancestry. Our overarching goals are to identify and quantify the contribution of genetic risk factors to the greater risk of overall and aggressive PCa in men of African ancestry, and to develop a polygenic risk score for PCa that is effective across diverse racial and ethnic populations.

The current multiethnic GWAS of PCa in the PRACTICAL Consortium includes ~115,000 cases and ~120,000 controls. Over the next 4-5 years we plan to double the size of this effort through bring together existing studies with GWAS data, such as ARIC, as well as new recruitment and genotyping.
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).

We propose to use the Affymetrix GWAS data in ARIC which is part of the PAGE Consortium and available to PAGE investigators (C. Haiman, co-PI). We plan to merge the genetic data with ARIC prostate cancer phenotype information (PMID: 29263187) for case-control analyses. These studies will include 596 cases and 5,058 controls of European ancestry and 277 cases and 1,548 controls of African ancestry (2015 ARIC cancer case file).

We propose to impute variants using the Michigan imputation server (https://imputationserver.sph.umich.edu/index.html) and the Haplotype Reference Consortium (HRC) panel, which over the next 2-3 years will include >40,000 African ancestry, >25,000 Latino and >20,000 Asian whole-genome sequences from projects such as the NHLBI Trans-omics for Precision Medicine (TOPMed) and NHRGI Center for Common Disease Genomics (CCDG) programs as well as >3,500 sequences from sub-African populations in the African Genome Variation Project, H3Africa and the 1000 Genomes Project.

The association analysis of single common and less-common variants (MAF>0.5%) will be based on a logistic regression framework with appropriate adjustment variables such as age, sub-study (for AAPC) and population structure, estimated from unlinked markers using principal component analysis paying particular attention to potential estimation of subgroups within Africa. Any residual confounding will be accounted for using a structured covariance matrix within a linear mixed modeling framework. We will be analyzing case-control comparisons using logistic regression, estimating the log additive odds ratio for any genotyped or imputed variant (estimated to be ~20-25 million variants). The resulting study and race-specific effect estimates will be meta-analyzed with the results from other studies in the Consortium using a fixed-effects approach and corresponding p-values will be used to determine significance at an \( \alpha \)-level of 5x10\(^{-8}\).

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)?

# 1724 Pleiotropic Effects of Cancer Risk Variants on Prostate Cancer Risk (Schumacher, Haiman et al.)

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

___  A. primarily the result of an ancillary study (list number*)
    2011.07 Enhancing ARIC Infrastructure to Yield a New Cancer Epidemiology Cohort
    1995.04 Cancer Study)
___  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.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __x__ No.