1.a. Full Title: Genome-wide association study of epithelial ovarian cancer (EOC) loci in women of African Ancestry

b. Abbreviated Title (Length 26 characters): GWAS of ovarian cancer in AA

2. Writing Group:
   Writing group members:
   Ani Manichaikul, University of Virginia
   Lauren Cole Peres, University of Virginia
   Joellen M. Schildkraut, University of Virginia
   Kala Visvanathan, Johns Hopkins Bloomberg School of Public Health
   Additional authors from the African American Cancer Epidemiology Study (AACES) and the Ovarian Cancer Association Consortium (OCAC)

   We also welcome other ARIC study investigators with interest in ovarian cancer as co-authors

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

First author: Ani Manichaikul
   Address: West Complex, Rm 6115
   Center for Public Health Genomics
   University of Virginia School of Medicine
   Charlottesville, VA 22903

   Phone: 434-982-1612    Fax: 434-982-1815
   E-mail: amanicha@virginia.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: Kala Visvanathan
   Address: Johns Hopkins Bloomberg School of Public Health
   615 N. Wolfe Street
   Room E6142
   Baltimore, Maryland 21205
   Phone: 410-614-2632
   E-mail: kvisvan1@jhu.edu
3. **Timeline:** Following approval of the manuscript proposal, we will complete case-control genome-wide association analysis of ARIC African American women combined with existing samples from the Ovarian Cancer Association Consortium by July 2017. Once the GWAS analyses are complete, we will perform functional validation analysis of selected resulting candidate genes. We will target submission of the manuscript by February 2018.

4. **Rationale:** Epithelial ovarian cancer (EOC) is a rare but deadly disease for which there is notably poorer survival in women of African Ancestry (AA) compared to women of European Ancestry (EA). Previous EA-based genome-wide association studies (GWAS) have identified 30 common, low penetrant EOC susceptibility alleles. Using the custom-designed 533,631 SNP Illumina OncoArray and imputation to ~12 million genetic variants in the 1000 Genomes Phase 3, we conducted a GWAS in 755 AA EOC cases, including 537 high-grade serous ovarian cancer (HGSOC) cases, and 1,235 AA controls from the Ovarian Cancer Association Consortium (OCAC). We identified three novel susceptibility loci with suggestive evidence of association with EOC, based on a threshold of $P < 1 \times 10^{-6}$. We propose to expand our analysis by including additional cases and controls from ARIC, MESA and other population-based cohorts with African American women.

5. **Main Hypothesis/Study Questions:** We anticipate our expanded case-control analysis of EOC incorporating African American women from ARIC will lead to improved power and identification of additional genetic loci for EOC at improved statistical significance compared to the OCAC analysis completed to date.

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).**

The proposed effort will implement a case-control design incorporating EOC cases from OCAC together with additional EOC cases from ARIC based on pathology confirmation, controls from OCAC, and controls identified from ARIC, MESA and other population-based cohorts. Controls selected from ARIC will be limited to African American women having at least one ovary at the baseline examination. We will further implement the selection criteria as described below.

We recognize there are likely differences in underlying composition of ancestry represented among OCAC cases and controls compared to controls identified from ARIC and MESA due to differences in recruitment sites. Notably, OCAC includes cases recruited from both Europe and the United States. In addition, we expect there could be differences in ancestral composition among African Americans within the United States. With this in mind, we are particularly interested in including the African American women from ARIC as they include representation from North Carolina, which is also heavily represented in our existing EOC cases. Specifically, the included EOC cases have significant representation from the North Carolina Ovarian Cancer Study (NCOCS, 14% of EOC cases in our study) and the African American Cancer...
In order to obtain good matching between cases and controls, we propose to select controls from among the eligible ARIC African American women as follows. First, principal components (PCs) of ancestry will be computed using the FASTPOP software for the OCAC African Ancestry women using a set of SNPs shared by both the OncoArray (source of genotypes for OCAC) and the Affymetrix 6.0 array genotypes from ARIC and MESA. We will then project the computed PCs from OCAC to the ARIC and MESA women to obtain a shared set of PCs across the cases and potential controls. Self-reported African American women will be filtered on inferred African ancestry > 50%. Finally, we will implement a selection algorithm whereby each case will be matched with ~3 controls of comparable ancestry (within a window of ~0.5 standard deviations for the first two PCs of ancestry). The target ratio of controls to cases will be determined depending on the number of eligible controls identified from ARIC, MESA and potentially other sources of controls. The proposed strategy for identification of controls will ensure a similar distribution of ancestry among the OCAC cases and external controls.

Upon completion of the selection of controls for analysis, genetic data from OCAC, ARIC and MESA will be merged for imputation to the 1000 Genomes Phase 3 reference panel. Genome-wide association analysis will be performed by logistic regression with adjustment for two principal components of ancestry using a score test to account for genotype uncertainty as implemented in SNPTESTv2.5.2. For genotyped SNPs, we will include results only for those SNPs with Hardy-Weinberg Equilibrium p-value > 1x10^-5 and heterozygosity count (HC) > 30, where HC is defined as N x MAF x (1-MAF) for each SNP, N represents the sample size (either the number of cases or the number of controls), and MAF represents the SNP minor allele frequency. For imputed SNPs, we include those SNPs with imputation R-squared > 0.5 and effective heterozygosity count (effHC) > 30, where effHC is defined as the imputation R-squared x HC. Note that quality control filters will be applied separately for cases and controls to select SNPs carried forward for genetic association analysis, such that a minimum HC (or effective HC) of 30 will be observed among each of the case and control groups.

**Summary of eligibility criteria:** For this effort, we will request GWAS data on all self-reported African American women in ARIC with the following exclusions:
- individuals dropped in genotype quality control
- individuals with less than 50% African ancestry
- women who have undergone bilateral oophorectomy
- women with age > 45 years at the time of recruitment (NOTE: this cutoff should include all of the ARIC participants)

After selecting the eligible cases and controls, we will permanently destroy the data for the potential controls not sampled.

We recognize that our eligibility criteria may limit the number of controls that may be included from ARIC and MESA. Accordingly, we are in the process of pursuing additional sources of African American female controls for this study from other cohorts including the Women’s Health Initiative (WHI) and the Black Women’s Health Study (BWHS).
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)?

Eric Boerwinkle was responsible for the GWAS effort.

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, 2009.17, 2010.01) [These are the cancer ones; they help determined who is eligible to be a control, and if you do want the cases, certainly they contributed to case ascertainment and adjudication.]

_____ 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.csecc.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.