1.a. Full Title: Association between genetic polymorphisms affecting eosinophil count and colorectal cancer (CRC) risk in the ARIC study.

b. Abbreviated Title (Length 26 characters): Eosinophil SNPs and CRC

2. Writing Group:
Writing group members: Anna Prizment, Heather Nelson, Corinne E Joshu, Kala Visvanathan, Aaron R Folsom, other interested investigators

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

First author:
Address: Division of Epidemiology and Community Health
1300 S. Second St., Suite 300
Minneapolis, MN 55454

Phone: 626-8862 Fax: 612-624-0315
E-mail: prizm@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: Aaron Folsom
Address: Division of Epidemiology and Community Health
1300 S. Second St., Suite 300
Minneapolis, MN 55454

Phone: 626-8862 Fax: 612-624-0315
E-mail: folso001@umn.edu

3. Timeline:
Analyses will begin after the ARIC Committee approves the proposal. We anticipate the manuscript will be completed within 2 years.

4. Rationale:
Recently, we reported an inverse association between eosinophil count and the risk of colorectal cancer in the ARIC study (1). Eosinophils are granulocytic white blood cells
involved in the immune response and capable of killing pathogens and tumor cells (2, 3). Eosinophil counts are increased in those with allergy and are correlated with severity of asthma and atopic dermatitis (4, 5).

The inverse association of eosinophil count with colorectal cancer is in accord with murine studies supporting a protective role of eosinophils in carcinogenesis (6-9). A potential protective role of eosinophils was also observed in the development of colorectal cancer in human studies. Several studies reported better prognostic indicators of colorectal carcinoma among those with increased blood and tissue eosinophil counts (10-16). Recently, a Dutch nested case-control study detected and inverse association of peripheral blood eosinophil count with CRC mortality, especially among men and smokers (17).

However, our study of eosinophil counts and colorectal cancer incidence in ARIC had an important limitation -- eosinophil counts were measured at three different ARIC centers resulting in different means and ranges of absolute eosinophil counts. In addition, since it was an observational study, we could not exclude residual confounding and reverse causality.

Now we suggest examining the risk of colorectal cancer in relation to five single nucleotide polymorphisms (SNPs) associated with eosinophil count in a discovery and replication GWAS by Gudbjartsson et al published in Nature Genetics in 2009: rs1420101 at IL1RL1 gene, rs12619285 at IKZF2 gene, rs4857855 at GATA2, rs4143832 at IL5 gene, and rs3184504 at SH2B3 (18).

The advantages of using genetic variants associated with trait instead of the trait itself (using the principal of Mendelian randomization) are the following: genetic variants are not associated with known and unknown confounders; they can be measured very accurately; and they are not susceptible to reverse causality (19). Therefore, the proxy use of SNPs related to eosinophil count instead of the count itself may be used to further examine an association between eosinophil count and colorectal cancer risk and may help to establish causality.

All SNPs from the above-mentioned GWAS (18) study have been genotyped in ARIC. We will create a weighted eosinophil genetic risk score (GRS) as a sum of risk alleles for each person using parameter estimates from that GWAS study and examine associations of this GRS with eosinophil count and colorectal cancer risk among Caucasians. Using the eosinophil GRS instead of individual SNPs may help to increase power, avoid multiple comparisons, and lessen the pleiotropic effect of SNPs.

In an additional analysis, we will try to replicate associations between eosinophil-related SNPs and asthma observed in the above-mentioned GWAS study by Gudbjartsson et al [2009] (18). They reported that rs1420101 was associated with asthma in ten different populations (7,996 cases and 44,890 controls). They also reported that SNPs rs2416257 at WDR36, rs3939286 at IL33 and rs9494145 at MYB that showed suggestive associations with eosinophil counts were also associated with asthma.

5. **Hypothesis/Study Questions:**

Our main hypothesis: a GRS based on the variants related to eosinophil count (eosinophil GRS) is inversely associated with the colorectal cancer risk in the ARIC study.

- Examine association of eosinophil GRS with colorectal cancer risk.
Secondary aims:
- Examine association of eosinophil GRS with blood eosinophil count.
- Examine association of eosinophil GRS with asthma and assess reproducibility of associations between individual eosinophil-related SNPs and asthma reported in the GWAS study (18).

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: Prospective cohort study

Inclusion/Exclusion: inclusion: all Caucasians that participated at Visit 1 that were free of cancer; exclusion: those who did not provide consent to participate in cancer studies, those who did not provide consent for genetic analyses, and those with missing genetic information.

Independent variables: Five SNPs associated with eosinophil count in the GWAS study (18). An additive genetic model will be used with SNPs coded as 0, 1, or 2, where 2 will designate a risk allele.

Dependent Variable: Colorectal cancer incidence (N=205) for Caucasians are available through 2006. In an additional cross-sectional analysis, we will test association of eosinophil-related SNPs with blood eosinophil count and asthma. Participants will be considered as having asthma if they self-reported asthma at any of the 4 visits.

Other variables of interest: Because our main analysis is genetic, confounding is not expected, but we will consider confounding by major risk factors such as age, sex, center, smoking, BMI, pack-years, education, aspirin use, hormone therapy use, and diabetes measured at visit 1.

Since eosinophil count is related to smoking and sex, interactions with sex and smoking at Visit 1 will be considered if we find a significant association between eosinophil GRS and colorectal cancer risk.

Analysis plan:
Main analysis: We will use a proportional hazards model to estimate the risk of colorectal cancer in relation to eosinophil GRS in unadjusted and adjusted models. The standardized eosinophil GRS will be computed for each subject by multiplying the number of alleles associated with higher eosinophil count by the coefficient from the GWAS studies and taking the sum over the SNPs. To make the genetic risk score easier to interpret, we will rescale the score so it ranges from 0 to 10 (10 is the maximum number of risk alleles in one person as there are five SNPs under study). The GRS will be presented as a continuous variable.
If we find a significant association between GRS and colorectal cancer, we will use interaction terms in the proportional hazards models to determine if these associations vary by smoking at baseline.

**Additional analyses:**
- We will utilize linear regression to explore the association of individual SNPs related to eosinophil count and eosinophil GRS with the blood eosinophil count. Log transformation of the count will be used because of skewed distribution.
- Associations of eosinophil GRS and individual eosinophil-related SNPs with asthma will be examined using logistic regression.

The following models will be used to analyze the associations:
Model 1: adjusted for age, sex, and ARIC study site
Model 2: Model 1 additionally adjusted for BMI, smoking status, pack-years of smoking, education level, aspirin use, hormone therapy, diabetes.

**Power calculation for main analysis:** Assuming incidence rate of 0.002 for CRC and annual loss to follow-up of 0.5% among ~8,700 Caucasians in ARIC, we will have 80% power to detect RR=0.71 (two-sided α=0.05) for eosinophil GRS, dichotomized at median [Power program, Kaiser Permanente].

**Methodological limitations:** potentially low power and pleiotropic effects of the genetic variants. However, using GRS instead of individual SNPs may increase power and will help to avoid multiple comparisons as well as account for a substantive amount of variation in circulating eosinophil count. We hypothesize that GRS will give more information than a single measurement of eosinophil counts because it provides information about cumulative lifetime exposure to eosinophils (provided that those with a high GRS will have a slightly higher eosinophil count over life time). The GRS may also account for pleiotropic effects of individual SNPs.

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”?
8.c. If yes, is the author aware that the participants with RES_DNA = ‘not for profit’ restriction must be excluded if the data are used by a for profit group?  
_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)?

Prizment AE, Anderson KE, Visvanathan K, Folsom AR “Inverse association of eosinophil count with colorectal cancer incidence: Atherosclerosis Risk in Communities study”. The lead authors are the same as in this proposal

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* __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.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.

References
17. Taghizadeh N, Vonk JM, Boezen HM. Peripheral blood eosinophil counts and risk of colorectal cancer mortality in a large general population-based cohort study. 2011.