ARIC Manuscript Proposal #4008

PC Reviewed: 2/8/22	**Status: _____**	**Priorty: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: DNA methylation markers of immune response and cancer risk and mortality

b. Abbreviated Title (Length 26 characters): Immunity and cancer risk

2. Writing Group:

Dominique S. Michaud, Naisi Zhao, Jiayun Lu, Devin Koestler, Karl Kelsey, Anna Prizment, Eric Boerwinkle, Jan Bressler, Elizabeth A. Platz

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

First author: Dominique S. Michaud
Address: Department of Public Health & Community Medicine
Tufts University School of Medicine
136 Harrison Avenue
Boston, MA 02111
(203) 918-5528 (cell phone)
E-mail: Dominique.Michaud@tufts.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: Department of Epidemiology, Rm E6132
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe St.
Baltimore, MD 21230

Phone: 410-614-9674 Fax: 410-614-2632
E-mail: eplatz1@jhu.edu

3. Timeline: Manuscript drafted by January 2023 (assuming TopMed methylation data is available 9 months prior to that).

4. Rationale:

There is overwhelming evidence that the immune response plays a role in cancer development (1). Many epidemiological studies have focused on the role of systemic inflammation on cancer risk and progression. Elevated CRP levels (2-6), white blood cell count (7-9), neutrophil-to-lymphocyte ratios (6, 10, 11), and serum levels of pro-inflammatory cytokines (12) have been associated with higher risk of cancer or decreased cancer survival. Yet, due to the difficulty in measuring immune cell subtypes in peripheral white blood cells (WBC) on a population level, few studies have been able to examine the role of immune cell subtypes involved in the adaptive response in cancer incidence. Given that cancer itself will cause changes in immune status, retrospective analyses cannot address the etiological question of how the immune response may impact risk of developing cancer. Prospective studies that have measured WBC counts, including the ARIC study, have been limited to measuring innate cells (i.e., neutrophils, eosinophils, monocytes and basophils) and all lymphocytes combined. However, there is increasing evidence suggesting that immune cells such as T-regulatory cell (Tregs) play a key role in cancer development (13). In addition, growing data support the hypothesis that naïve CD8 T cells play a key role in ageing and immunity and may also impact cancer development (14, 15). New studies are needed to provide insight into the role of these immune cells in risk of developing cancer. To address this gap, we propose to use DNA methylation to estimate 12 immune cell subsets using a newly expanded reference-based deconvolution methodology (16) and examine how these immune cell proportions relate to cancer risk and mortality in the ARIC study.

5. Main Hypothesis/Study Questions:

The main hypothesis is that the innate and adaptive immune response can impact risk and mortality and that they do so in a similar manner across all cancer types. For major cancer types (i.e., colorectal, breast, lung cancers) we propose that known risk factors, especially, race, age, smoking, and obesity, may modulate these associations.

Using data from the Atherosclerosis Risk in Communities (ARIC) Study we propose to conduct prospective analyses for these aims/hypotheses:

1) Total neutrophil counts and elevated neutrophil proportions are positively associated with risk of total cancer and cancer mortality, especially within a 5-year period before cancer diagnosis. Note: This analysis will be different from Proposal #1825 (which used directly measured WBC differentials from 3 field centers) as it will include epigenetic predicted neutrophil proportions (and estimate counts) and mutually adjust for adaptive immune cells (CD8, CD4, B cells and NK cells). Racial differences will be evaluated for this aim.

2) Elevated T-regulatory cells (Tregs), counts and proportions, are positively associated with risk of total cancer incidence and mortality. Analyses will be stratified on smoking and obesity.
3) Elevated CD8 naïve cells (as a ratio of total CD8 cells) are inversely associated risk of total cancer incidence and mortality. Analyses will be stratified on smoking and obesity.
4) Neutrophil-to-lymphocyte ratio (NLR) is positively associated with risk of cancer mortality.

While these are our main hypotheses, we will also examine associations of CD4 cells (naïve and memory), B cells (naïve and memory), NK cells, and the innate immune cells derived from the deconvolution of white blood cells with cancer incidence and mortality.

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 design.

Analytic population: All participants with DNA methylation data from Visit 3.
Exclusions: Individuals with prevalent cancer at Visit 3 will be excluded.

Exposure: DNA methylation data obtained with the Illumina MethylationEPIC arrays (TOPMed) from Visit 3.

Outcomes: The outcomes will be: (1) total cancer incidence, (2) total cancer mortality, (3) cancer-specific incidence for lung, breast, and colorectal cancers, and (4) smoking, obesity, and hormone-associated cancer incidence. Follow-up time will be calculated from date of Visit 3 to date of cancer diagnosis, date of death, or end of follow-up (December 31, 2015) for the incidence analyses. For mortality analyses, censoring will be either date of death or end of December 2018.

Other variables: Age, race, derived BMI at Visit 3, current smoking status and packyears smoked by Visit; alcohol drinking at Visit 3 (never, former, or current drinker), diabetes status at Visit 3 (diagnosed: MD diagnosis, medications; undiagnosed: fasting glucose ≥126 mg/dL at any visit and/or glycated hemoglobin ≥6.5% at Visit 2; at risk for diabetes: fasting glucose of 100 to <126 mg/dL at visit 4; if not fasting, prior visit concentration will carried forward); ever use of hormone replacement therapy (women only; Visits 1 and 3); lifecourse SES calculated using data from ancillary study at Visit 4 as done previously in ARIC (17); US Census tract data on neighborhood income for the year 2000 (18). We will also include a pack-year variable estimated from the DNA methylation markers (19).

Estimation of immune cell composition.
Peripheral blood leukocyte subtypes proportions, including myeloid lineage sub-types [neutrophils (Neu), eosinophils (Eos), basophils (Bas), and monocytes (Mono)] and lymphoid lineage subtypes [B lymphocytes naïve (Bnv), B lymphocytes memory (Bmem), T helper lymphocytes naïve

(CD4nv), T helper lymphocytes memory (CD4mem), T regulatory cells (Treg), T cytotoxic lymphocytes naïve (CD8nv), T cytotoxic lymphocytes memory (CD8mem), and natural killer lymphocytes (NK)], will be estimated using a newly expanded reference-based deconvolution library EPIC IDOL-Ext (16). The expanded reference library adds new immune cell types to the current function *estimateCellCounts2* (available in the FlowSorted.Blood.EPIC Bioconductor package (20)), which is based on previously published reference-based cell mixture deconvolution algorithm developed by our team (Kelsey & Koestler) (21). This library uses the IDOL methodology (21) to optimize the currently available six-cell reference library in order to deconvolve the proportions of 12 leukocyte subtypes in peripheral blood. This EPIC IDOL-Ext library (Bioconductor package FlowSorted.BloodExtended.EPIC) was validated using flow cytometry gold standard data and substantiated by including publicly available data from >100,000 samples (16). Up to 56 different immune cell ratios (e.g., Neutrophil-to-Lymphocyte ratio) can be estimated using the new libraries.

Methylation-Derived Neutrophil Lymphocyte Ratio (mdNLR): We will use a DNA methylation-derived NLR (mdNLR) index to predict the common clinical NLR parameter using a previously described approach (22). This index is based on normal isolated leukocyte reference DNA methylation libraries and established reference-based cell-mixture deconvolution algorithms (22, 23).

Estimation of immune cell **counts**: Using the Visit 2 total WBC count we will estimate immune specific cell counts by multiplying the immune cell proportions (estimated from DNA methylation) to the total WBC count. Given that these two measures were obtained at different time points (Visits 2 and 3), the assumption will be that the WBC count do not change substantially over time. To check this assumption, we will regress WBC counts from Visit 2 on WBC from Visit 1 – individuals with large residual values will be removed from the analysis. We will also remove participants who started (or stopped) taking corticosteroids, or who changed smoking status, between Visits 2 and 3. Only 3 field centers will be included in these analyses (those with cell count data).

Data analysis:

All statistical analyses will be performed in R (version 3.5.1). We will use Cox proportional hazard models to examine the association between DNA methylation-based immunity measures and cancer risk and mortality. Immune cell proportions, counts and ratios will be included in models as continuous variables. In addition, each regression model will be adjusted for age, sex, education (<high school, high school, >high school), field center*race (black from suburban Minneapolis, Forsyth County or Washington County; white from Forsyth County or Washington County; black from Jackson), smoking (current, former, never; packyears smoked [continuous]), BMI (continuous), diabetes status (diagnosed, undiagnosed, at risk for diabetes, none), alcohol drinking (never, former, or current drinker), lifecourse SES, US Census tract data on neighborhood income, batch effect, and a previously described methylation-predicted pack-years smoked (19). The proportional hazards assumption was checked by conducting global tests of correlating the set of scaled Schoenfeld residuals with time for each covariate.

Associations will be additional examined by strata of smoking status, median age, and race. Interactions will be tested if associations differ substantially by strata. In addition, sensitivity analyses will be conducted to evaluate if associations differ for cancer cases that were diagnosed less than 5 years after baseline (i.e., blood draw) compared to those diagnosed more than 5 years after baseline (to rule out reverse causation). Sensitivity analyses will also be conducted to examine total cancers excluding non-solid cancers.

As a sensitivity analysis for analyses measuring immune cell counts, we will remove participants who had highly inconsistent total counts between Visits 1 and 2.

Methodologic challenges:

The main goal of the proposed study is to examine the relationship between the immune response and cancer incidence and mortality. We will use epigenetic markers of immune cell types to estimate proportions of subsets of immune cells – these will be measured at Visit 3 and may not necessarily reflect cell proportions over a long period of time. Other limitations include a lack of detailed data on cancer screening; however, we do have information on frequency of doctor's visit, which we will adjust for in multivariate models.

Data needed for analysis:

Table 1: Summary of ARIC variables for proposed analysis

Variable	**Time of Collection**	**Analytic Plan**
Total WBC count & WBC subtype (neutrophil, lymphocyte, monocyte, eosinophil, basophiles) count	Visit 1 & Visit 2	Exposure of interest; Exclude participants whose WBC (total & subtype) counts change from Visit 1 to Visit 2 in sensitivity analyses (see analytic plan for more details).
Overall and site-specific cancer incidence	Through 12/31/2015	Outcome 1
Overall and site-specific cancer-specific mortality	Through 12/31/2015	Outcome 2
Sex	Visit 1	Covariate
Age	Visit 3	Covariate
Education level	Visit 1	Covariate
Socioeconomic status	Visit 1	Covariate
Study center	Visit 1	Covariate

BMI	Visit 3	Covariate
Medication use: immunomodulating medications, corticosteroids, NSAIDS/aspirin, antibiotics, antihistamines	Visits 1-3	Sensitivity analyses will be performed to account for corticosteroids, NSAIDs, antibiotics and antihistamine use. We will also adjust for NSAID use during follow-up by treating NSAIDs use as a time-varying covariate.
Medical history: cardiovascular disease, cancer, asthma, diabetes	Visit 3	Cancer cases at baseline will be excluded; Sensitivity analyses will also be performed excluding cases of CVD, diabetes and asthma
Reproductive History: Menopausal status, age at menopause, pregnancy, exogenous hormone use	Visit 3	Covariate – for breast cancer analyses
Alcohol use	Visit 3	Covariate
Smoking (Never, Former, Current & pack-years)	Visit 3	Covariate
Physical Activity	Visit 3	Covariate
Screening Questions: Frequency of doctor visits	Visit 3	Covariate

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

No overlap

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)?

MS#1896 and MS #1825. Dr. Prizment will be included in this analysis (and is on both these proposals).

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* 1996.0; 2011.07 Enhancing ARIC Infrastructure to Yield a New Cancer Epidemiology Cohort; 1995.04 Cancer Study)

TOPMed DNA methylation data

_X__ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* 1998.02, 2004.05)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

- Lifecourse SES data generated as part of 1998.02 (Life courses SES, social context, and CVD) – Dr. Heiss
- Census tract income data generated as part of 2004.05 (Socioeconomic characteristics of place of residence: impact on rates and trends in nonfatal and fatal CHD in the ARIC Surveillance Communities) – Drs. Heiss and Rose

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.

References

1. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. Cell. 2010;140(6):883-99.
2. Van Hemelrijck M, Holmberg L, Garmo H, Hammar N, Walldius G, Binda E, et al. Association between levels of C-reactive protein and leukocytes and cancer: three repeated measurements in the Swedish AMORIS study. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2011;20(3):428-37.
3. Trichopoulos D, Psaltopoulou T, Orfanos P, Trichopoulou A, Boffetta P. Plasma C-reactive protein and risk of cancer: a prospective study from Greece. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2006;15(2):381-4.
4. Gunter MJ, Stolzenberg-Solomon R, Cross AJ, Leitzmann MF, Weinstein S, Wood RJ, et al. A prospective study of serum C-reactive protein and colorectal cancer risk in men. Cancer research. 2006;66(4):2483-7.
5. Prizment AE, Folsom AR, Dreyfus J, Anderson KE, Visvanathan K, Joshu CE, et al. Plasma C-reactive protein, genetic risk score, and risk of common cancers in the Atherosclerosis Risk in Communities study. Cancer causes & control : CCC. 2013;24(12):2077-87.
6. Sollie S, Michaud DS, Sarker D, Karagiannis SN, Josephs DH, Hammar N, et al. Chronic inflammation markers are associated with risk of pancreatic cancer in the Swedish AMORIS cohort study. BMC cancer. 2019;19(1):858.
7. Erlinger TP, Muntner P, Helzlsouer KJ. WBC count and the risk of cancer mortality in a national sample of U.S. adults: results from the Second National Health and Nutrition Examination Survey mortality study. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2004;13(6):1052-6.
8. Shankar A, Wang JJ, Rochtchina E, Yu MC, Kefford R, Mitchell P. Association between circulating white blood cell count and cancer mortality: a population-based cohort study. Archives of internal medicine. 2006;166(2):188-94.
9. Margolis KL, Rodabough RJ, Thomson CA, Lopez AM, McTiernan A. Prospective study of leukocyte count as a predictor of incident breast, colorectal, endometrial, and lung cancer and mortality in postmenopausal women. Archives of internal medicine. 2007;167(17):1837-44.
10. Templeton AJ, McNamara MG, Seruga B, Vera-Badillo FE, Aneja P, Ocana A, et al. Prognostic role of neutrophil-to-lymphocyte ratio in solid tumors: a systematic review and meta-analysis. Journal of the National Cancer Institute. 2014;106(6):dju124.
11. Guthrie GJ, Charles KA, Roxburgh CS, Horgan PG, McMillan DC, Clarke SJ. The systemic inflammation-based neutrophil-lymphocyte ratio: experience in patients with cancer. Critical reviews in oncology/hematology. 2013;88(1):218-30.

12. Il'yasova D, Colbert LH, Harris TB, Newman AB, Bauer DC, Satterfield S, et al. Circulating levels of inflammatory markers and cancer risk in the health aging and body composition cohort. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2005;14(10):2413-8.
13. Najafi M, Farhood B, Mortezaee K. Contribution of regulatory T cells to cancer: A review. Journal of cellular physiology. 2019;234(6):7983-93.
14. Pawelec G. Immunosenenescence: role of cytomegalovirus. Exp Gerontol. 2014;54:1-5.
15. Saavedra D, Garcia B, Lage A. T Cell Subpopulations in Healthy Elderly and Lung Cancer Patients: Insights from Cuban Studies. Frontiers in immunology. 2017;8:146.
16. Salas LA, Zhang Z, Koestler DC, Butler RA, Hansen HM, Molinaro AM, et al. Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling bioRxiv. 2021.
17. Shoham DA, Vupputuri S, Diez Roux AV, Kaufman JS, Coresh J, Kshirsagar AV, et al. Kidney disease in life-course socioeconomic context: the Atherosclerosis Risk in Communities (ARIC) Study. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2007;49(2):217-26.
18. Foraker RE, Patel MD, Whitsel EA, Suchindran CM, Heiss G, Rose KM. Neighborhood socioeconomic disparities and 1-year case fatality after incident myocardial infarction: the Atherosclerosis Risk in Communities (ARIC) Community Surveillance (1992-2002). Am Heart J. 2013;165(1):102-7.
19. Sugden K, Hannon EJ, Arseneault L, Belsky DW, Broadbent JM, Corcoran DL, et al. Establishing a generalized polyepigenetic biomarker for tobacco smoking. Transl Psychiatry. 2019;9(1):92.
20. Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biology 2018;19(64).
21. Koestler DC, Jones MJ, Usset J, Christensen BC, Butler RA, Kobor MS, et al. Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL). BMC bioinformatics. 2016;17:120.
22. Koestler DC, Usset J, Christensen BC, Marsit CJ, Karagas MR, Kelsey KT, et al. DNA Methylation-Derived Neutrophil-to-Lymphocyte Ratio: An Epigenetic Tool to Explore Cancer Inflammation and Outcomes. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2017;26(3):328-38.
23. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012;13:86.