

5/9/2024

DNA Methylation-Derived Immune Cell Proportions and Cancer Risk, Including Lung Cancer, in Black Participants

Authors: Christopher S. Semancik¹, Naisi Zhao¹, Devin C. Koestler^{2,3}, Eric Boerwinkle^{4,5}, Jan Bressler⁴, Rachel J. Buchsbaum⁶, Karl T. Kelsey^{7,8}, Elizabeth A. Platz^{9,10}, Dominique S. Michaud¹

Author Affiliations:

¹Department of Public Health & Community Medicine, Tufts University School of Medicine, Tufts University, Boston, MA, USA

²The University of Kansas Cancer Center, Kansas City, KS, USA

³Department of Biostatistics & Data Science, University of Kansas Medical Center, Kansas City, KS, USA

⁴Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA

⁵Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA

⁶Division of Hematology/Oncology, Tufts Medical Center, Boston, MA, USA

⁷Department of Epidemiology, Brown University, Providence, RI, USA

⁸Department of Pathology and Laboratory Medicine, Brown University, Providence, RI, USA

⁹Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

¹⁰The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, USA

Running Title: Immune Cell Profiles and Cancer Risk in Black Participants

Corresponding Author:

Dominique S. Michaud, ScD

Tufts University School of Medicine

136 Harrison Avenue

Boston, MA 02111

Tel: 617-636-0482

Email: Dominique.Michaud@tufts.edu

Conflicts of Interest: Dr. Karl Kelsey is a founder and scientific advisor to Cellintec, which had no role in this research. Otherwise, the authors have no conflicts of interest to report.

Manuscript Notes:

Word Count: 3,745 words

Total Number of Tables/Figures: 6 tables (Main Text); 9 tables, 4 figures (Supplementary Materials)

Keywords: DNA methylation, immune cell profiles, cancer risk, immunology, epidemiology

Section: Epidemiology and Prevention

List of Abbreviations (alphabetical order):

95% CI: 95% confidence interval

ARIC: Atherosclerosis Risk in Communities (cohort study)

CD8+ cells: cytotoxic T lymphocytes

CpG: CpG oligodeoxynucleotide sites

DMR: differentially methylated region
DNA: deoxyribonucleic acid
HR: hazard ratio
MD: Maryland
mdNLR: methylation-derived neutrophil-to-lymphocyte ratio
MN: Minnesota
MS: Mississippi
NC: North Carolina
SD: standard deviation
TME: tumor microenvironment
Tregs: T regulatory cells
WHI: Women's Health Initiative (cohort study)

Abstract

Prior cohort studies assessing cancer risk based on immune cell subtype profiles have predominantly focused on White populations. This limitation obscures vital insights into how cancer risk varies across race. Immune cell subtype proportions were estimated using deconvolution based on leukocyte DNA methylation markers from blood samples collected at baseline on participants without cancer in the Atherosclerosis Risk in Communities (ARIC) Study. Over a mean of 17.5 years of follow-up, 668 incident cancers were diagnosed in 2,467 Black participants. Cox proportional hazards regression was used to examine immune cell subtype proportions and overall cancer incidence and site-specific incidence (lung, breast, and prostate cancers). Higher T regulatory cell proportions were associated with statistically significantly higher lung cancer risk (hazard ratio = 1.22, 95% confidence interval = 1.06-1.41 per percent increase). Increased memory B cell proportions were associated with significantly higher risk of prostate cancer (1.17, 1.04-1.33) and all cancers (1.13, 1.05-1.22). Increased CD8+ naïve cell proportions were associated with significantly lower risk of all cancers in participants ≥ 55 years (0.91, 0.83-0.98). Other immune cell subtypes did not display statistically significant associations with cancer risk. These results in Black participants align closely with prior findings in largely White populations. Findings from this study could help identify those at high cancer risk and outline risk stratifying to target patients for cancer screening, prevention, and other interventions. Further studies should assess these relationships in other cancer types, better elucidate the interplay of B cells in cancer risk, and identify biomarkers for personalized risk stratification.

Introduction

The immune system plays a key role in protecting against cancer¹. Studies using animal models and cancer patients have shown the immune system's ability to recognize and eliminate tumor cells through immunosurveillance, which involves both innate and adaptive immune response^{2,4}. However, tumor cells can evade immunosurveillance responses by suppressing the immune system^{5,6}. Within tumors, higher proportions of cytotoxic T lymphocytes (CD8+ cells) have been associated with more favorable cancer outcomes, whereas higher proportions of T regulatory cells (Tregs) have been associated with immunosuppression, accelerated cancer development, and decreased survival⁷⁻¹⁰. Intertumoral accumulation of Tregs has been consistently associated with greater tumor aggressiveness in patients with various cancer types¹¹⁻¹⁴.

Despite these insights, the role of peripheral blood immune cell profiles in the pre-cancerous state and their influence on subsequent cancer risk remains unclear. Further, cohort studies cannot systematically use flow cytometry to identify immune cell type profiles in peripheral blood at regular, consistent intervals because cohorts seldom have whole blood stored under appropriate conditions. To date, few observational studies have examined the relationship between immune cells measured in pre-diagnostic blood and cancer risk.

Recent advances in high-dimensional arrays enable measurement of DNA methylation at 450,000 to 850,000 CpG oligodeoxynucleotide sites (CpGs) throughout the genome, allowing precise estimates of immune cell proportions from frozen blood samples¹⁵⁻¹⁸. Pre-diagnostic blood collection is essential to assess DNA methylation states and immune cell proportions because the cancers themselves may alter these profiles¹⁹. Located throughout the genome, differentially methylated regions (DMRs) can distinctly identify the lineage of differentiated immune cell subtypes²⁰. These unique differentially methylated CpGs can be used to identify immune cell lineages, and their proportions can be estimated using a statistical method called "deconvolution"²⁰. Resulting immune cell

proportions can be employed to assess relationships between immune cell profiles and cancer risk²⁰.

Since DNA methylation analysis can be done using DNA from archived blood, immune profiles can now be assessed using the resources of large epidemiologic studies that have banked specimens.

For this analysis, we utilized the Atherosclerosis Risk in Communities (ARIC) study to investigate the risk of lung cancer, breast cancer, prostate cancer, and all cancers pooled together (excluding hematological cancers) in relation to DNA methylation-derived relative proportions of peripheral blood immune cell types in Black participants. This represents an important opportunity to analyze this relationship in an understudied population, as Black individuals are known to have lower average neutrophil counts than White individuals^{21,22}.

Materials and Methods

Study Population

Participants were members of the ARIC study (RRID: SCR_021769), a prospective cohort study of cardiovascular disease risk that enrolled 15,792 people between 1987 and 1989 from four different communities in the United States (Jackson, MS; Washington County, MD; suburban Minneapolis, MN; and Forsyth County, NC)^{23,24}. Participants underwent a baseline clinical examination (Visit 1; 1987-1989), which included an in-home interview and clinical examination, assessing medical and lifestyle factors²⁵. Participants returned for follow-up clinical examinations in 1990-1992 (Visit 2), 1993-1995 (Visit 3), 1996-1998 (Visit 4), 2011-2013 (Visit 5), and so on until Visit 10. Blood specimens were banked at each visit, and participants were followed by annual telephone calls until 2011 with semi-annual contact thereafter. The ARIC study protocol was approved by institutional review boards at each site and participants gave informed consent.

For this analysis, we included Black participants (initial total of n = 2,520) from the Jackson, MS community (n = 2,287) and the Forsyth County, NC community (n = 233) who previously had

methylation profiling performed, fully consented to cancer and genetic research, and had no cancer history before blood collection.

DNA Methylation Profiling and Prediction of Immune Cell Proportions

At Visit 2 or Visit 3, DNA methylation levels were measured^{26,27}. Then, using lineage CpG markers for immune subsets, immune cell deconvolution was conducted to estimate proportions for 12 leukocyte subtypes: neutrophils, eosinophils, basophils, monocytes, memory and naïve B cells, CD4+ and CD8+ naïve and memory cells, natural killer cells, and Tregs²⁶. Deconvolution has been validated against flow cytometry^{15,26,28-30}. Implausibly low immune cell proportion values were assigned the limit of detection, as described by Bell-Glenn and colleagues³¹. Values for the methylation-derived neutrophil-to-lymphocyte ratio (mdNLR) were calculated by dividing neutrophil proportions by lymphocyte proportions, and are represented as a ratio. The R function EstDimRMT was utilized to determine the number of principal components with non-zero eigenvalues needed to correct batch effects³².

Cancer Ascertainment

Based on cancer registry data and follow-up, cancer cases were ascertained for all cancer types, except for non-melanoma skin cancer. The outcome of interest, cancer incidence of any type, was positive if a participant developed any type of cancer over the follow-up period. The only outcome of interest was first incidences of cancer, not cancer recurrences. For some analyses, the outcome was restricted to the most common types of cancer, including lung cancer, breast cancer, and prostate cancers. In the case of breast cancer individually, cases in premenopausal women were excluded due to small numbers, meaning all included cases of breast cancer individually occurred in postmenopausal women. The premenopausal breast cancer cases, however, were included in all cancers combined. Besides lung, breast, and prostate cancers, other individual cancers were not analyzed due to the low

numbers of cases associated with other types of cancer, thus resulting in low and reduced statistical power if these cancers were to be analyzed individually.

Incident cancers were ascertained from baseline (either Visit 2 or Visit 3) until the end of 2015 through linkage with state cancer registries in Minnesota, North Carolina, Maryland, and Mississippi. From baseline through 2015, we ascertained 721 primary cancer cases and 345 cancer deaths in Black participants. These cases occurred over a mean of 17.5 years of follow-up. For analyses, all hematological cancers ($n = 53$) were removed, as hematological cancers originate in the progenitors that give rise to immune cells and may result in spurious immune profiles. This left 2,467 participants and 668 cancers for analysis.

Covariate Assessment

Risk factors associated with cancer include: age, sex, BMI, cigarette smoking (self-reported smoking status, self-reported pack-years, and methylation-derived pack-years), postmenopausal hormone use, and alcohol consumption (self-reported drinking status). Data on cigarette smoking and drinking status (current, former, never) and cigarette smoking cumulative dose (pack-years) were collected at each visit during follow-up, and we used the corresponding data from the visit of blood draw. Postmenopausal hormone use at Visit 2 was used.

Estimation of Peripheral Blood Leukocyte Composition

Briefly, the Illumina HumanMethylation450 BeadChip array was used for genome-wide DNA methylation profiling in 2,853 Black participants at 483,525 CpG sites. Genomic DNA was extracted from peripheral blood leukocyte samples using the Gentra Puregene Blood Kit (Qiagen), and bisulfite conversion of 1 μg genomic DNA was performed using the EZ-96 DNA Methylation Kit (Deep Well Format) (Zymo Research) ²⁷.

As the main exposures of interest, peripheral blood leukocyte subtype proportions were estimated, including myeloid lineage subtypes (neutrophils, eosinophils, basophils, and monocytes) and lymphoid lineage subtypes (B lymphocytes naïve, B lymphocytes memory, T helper lymphocytes naïve (CD4+ naïve cells), T helper lymphocytes memory (CD4+ memory cells), T regulatory cells (Tregs), T cytotoxic lymphocytes naïve (CD8+ naïve cells), T cytotoxic lymphocytes memory (CD8+ memory cells), and natural killer lymphocytes). This estimation was done using a newly expanded reference-based deconvolution library EPIC IDOL-Ext.²⁶ This library used the IDOL methodology to optimize the currently available six-cell reference library, to deconvolve the proportions of 12 leukocyte subtypes in peripheral blood^{33,34}. This EPIC IDOL-Ext library was validated using gold standard flow cytometry data and substantiated by including publicly available data from more than 100,000 samples²⁶.

Measurement of Complete Blood Count and Total Leukocyte Count

In addition to using the deconvolution technique to estimate peripheral blood leukocyte composition for each participant, complete blood count (CBC) was measured on archived blood of participants from Visit 3. For most participants, CBC was measured without differential analysis; 30 participants did not have CBC (n = 2,437 participants with CBC data).

Statistical Analysis

To estimate the association of methylation-derived immune cell proportions with cancer incidence, we used Cox proportional hazards regression to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) of total cancer, lung cancer, breast cancer, and prostate cancer, adjusting for known risk factors described in the covariate assessment. Participants contributed time at risk from blood draw for profiling at Visit 2 (89.1% of participants) or Visit 3 (10.9% of participants) until cancer diagnosis of any site, death, or administrative censoring at the end of 2015, whichever came first.

Specifically, the covariates included: age (continuous), sex, BMI (continuous), cigarette smoking status, cigarette smoking dose (continuous), postmenopausal hormone use, methylation-derived pack-years (continuous), methylation-derived neutrophil-to-lymphocyte ratio (mdNLR) (continuous), drinking status (for breast cancer only), and batch effect (continuous).

Analyses were performed using R Project for Statistical Computing (v4.0.2; R Core Team 2020 [RRID: SCR_001905]). Statistical tests were two-sided, and a P-value of less than 0.05 was considered statistically significant.

Results

In this population, being older, a smoker (current smoker or higher pack-years), not using menopausal hormones (among women), having less education, or having a lower mdNLR was more common with increasing proportion of Tregs (**Table 1**). Being younger, female, a menopausal hormone user (among women), having more education, or having a lower mdNLR was more common with increasing CD8+ naïve cell proportions (**Table 2**).

Looking at cancer risks overall based on all 12 immune cell subtypes, many findings were not statistically significant (**Table 3**). However, three cell subtypes especially demonstrated statistically significant results when analyzed as continuous variables: Tregs, CD8+ naïve cells, and memory B cells. Further, total leukocyte counts (derived from complete blood count) exhibited positive statistical significance.

Assessing Treg proportions, each one-percent increase was associated with a 6% increased risk of all cancers (HR: 1.06, 95% CI = 1.00 to 1.11) (**Table 3**). For lung cancer, a one-percent increase of Treg proportion was associated with a 22% increased risk of lung cancer (HR: 1.22, 95% CI = 1.06 to 1.41) (**Table 3**). Treg proportions were associated with increased risk of breast cancer and prostate cancer but were not statistically significant (**Table 3**). In a sensitivity analysis, we further adjusted for education

level and diabetes status at baseline, and the associations in all cancers and lung cancer were not appreciably changed (**Supplementary Table 1**). Treg associations did not vary substantially when stratified by age (55 years old or younger versus more than 55 years old) or by sex (**Tables 4-6; Supplementary Tables 7-9**). Mutually adjusting for measured total leukocyte count and methylation-derived Treg proportions did not attenuate findings for either measure (**Supplementary Table 6**).

Analyzing CD8+ naïve cell proportions, a one-percent increase was associated with a 4% decreased risk of all cancers (HR: 0.96, 95% CI = 0.91 to 1.01) (**Table 3**). Each one-percent increase in CD8+ naïve cell proportion was associated with a 15% decreased risk of lung cancer (HR: 0.85, 95% CI = 0.71 to 1.01; **Table 3**). CD8+ naïve cell proportion was not associated with breast cancer or prostate cancer risk (**Table 3**). In a sensitivity analysis, we further adjusted for education level and diabetes status at baseline, and associations in all cancers and lung cancer were not appreciably changed (**Supplementary Table 1**). However, for age-stratified analyses, results were statistically significant in all cancers for individuals over 55 years old (HR: 0.91, 95% CI = 0.83 to 0.98), but null for individuals 55 years old or younger, and cases were similarly distributed by age group (**Table 4**). Age-stratified results were directionally similar in individuals over 55 years old, but not statistically significant, in the individual cancers assessed (**Tables 5-6**).

Assessing memory B cell proportions, each one-percent increase was associated with a 13% increased risk of all cancers (HR: 1.13, 95% CI = 1.05 to 1.22), a 17% increased risk of prostate cancer (HR: 1.17, 95% CI = 1.04 to 1.33), and a 20% increased risk of lung cancer (HR: 1.20, 95% CI = 0.99 to 1.46) (**Table 3**). Since prostate cancer affects some men minimally and others quite aggressively, an analysis of only prostate cancer cases without a lethal phenotype exhibited a similarly statistically significant memory B cell association (HR: 1.19, 95% CI = 1.05, 1.34). Increasing memory B cell proportion was suggestive of reduced breast cancer risk, although this relationship was not statistically significant. Associations were similar in all cancers, prostate cancer, and lung cancer after further

adjusting for education level and diabetes status (**Supplementary Table 1**). However, for age-stratified analyses, results were statistically significant in all cancers (HR: 1.16, 95% CI = 1.05 to 1.27) and prostate cancer (HR: 1.25, 95% CI = 1.10 to 1.43) for individuals 55 years old or younger, but null for individuals over 55 years old (**Tables 4, 6**). Age-stratified results were not statistically significant in breast cancer, although the relationships were suggestively protective in both age groups (**Table 5**). The associations in men were statistically significant for all cancers (HR: 1.16, 95% CI = 1.05 to 1.28) but not women, and in lung cancer for men (HR: 1.34, 95% CI = 1.04 to 1.71) but not women; this relationship may be partly influenced by the direct contribution of prostate cancer cases in men only (**Supplementary Tables 7-8**).

Although the associations for Tregs, CD8+ naïve cells, and memory B cells for all cancers and lung cancer followed a linear dose-response (**Supplementary Figures 1-2**), spline plots showed that Treg and CD8+ naïve cell associations for breast and prostate cancers follow a non-linear pattern (**Supplementary Figures 3-4**). Therefore, we conducted quantile-based analyses for all cancers and individual cancers to address nuance in relationships modeled in previously discussed continuous models. In quartile analysis for all cancers, a 31% elevated risk of all cancers was observed for the highest quartile of Treg proportion, relative to the lowest quartile (HR: 1.31, 95% CI = 1.04 to 1.65) (**Supplementary Table 3**). Furthermore, a 47% increased risk of all cancers was observed for the highest compared with the lowest quartile of memory B cell proportion (HR: 1.47, 95% CI = 1.13 to 1.91) (**Supplementary Table 3**).

Using tertiles for the lung cancer analysis (due to a smaller number of cases), there was a dose-response, but not statistically significant relationship for Treg and memory B cell proportion, while for CD8+ naïve cell proportion, a statistically significant relationship was observed for the highest compared with the lowest tertile (HR: 0.47, 95% CI = 0.25 to 0.87) (**Supplementary Table 4**). For white blood cell count, a 167% elevated risk of lung cancer was observed for the highest compared with the lowest tertile of white blood cell count (HR: 2.67, 95% CI = 1.43 to 4.97) (**Supplementary Table 4**). For breast

cancer, relationships were generally less dose-response oriented (**Supplementary Table 5**). Only white blood cell count followed a somewhat dose-response oriented relationship for breast cancer. For prostate cancer, associations were similarly non-linear (**Supplementary Table 6**).

Discussion

Our study uniquely highlights that in a Black population, higher methylation-derived peripheral blood Treg proportion was associated with elevated risk of lung cancer, even when adjusting for smoking status and pack-years. Furthermore, for Treg proportion, a non-significant elevated risk of prostate cancer was noted. Additionally, increased memory B cell proportion was associated with elevated risk of all cancers and prostate cancer. Conversely, CD8+ naïve cell proportion was associated with a decreased risk of developing lung cancer and all cancers, an association that was statistically significant for individuals over 55 years old, suggesting that adaptive immunity in older age could be critical to combatting primary tumorigenesis. A positive association between total white blood cell count (directly measured) and lung cancer risk was also observed, which was independent of relationships measured in methylation-derived immune cell subtypes. Our findings extend prior findings which examined immune cell profiles and risk of major cancer types to a new population, using a cutting-edge algorithm for predicting immune cell subtype proportions.

Our work is among the first to investigate these associations in a Black cohort. Nonetheless, our findings merit comparison to other existing studies. Several studies have addressed the association between total white blood cell count and cancer risk. Studies conducted in the Women's Health Initiative (WHI) and UK Biobank cohorts have reported that elevated white blood cell counts are associated with statistically significant increases in risk of invasive breast cancer in postmenopausal women, as well as endometrial cancer and lung cancer in a general cohort of individuals between 40 and 69 years old^{35,36}. Fewer studies have examined immune cell subtypes. Generally, higher proportions of

Tregs relative to total leukocytes or other immune cells have been associated with higher risks of lung, colorectal, breast, and pancreatic cancers¹⁶⁻¹⁸. On the other hand, higher relative proportions of CD8+ cells have been inversely associated with risk of lung cancer and breast cancer^{16,18}. Using a different cohort (the CLUE study), we previously detected a statistically significant increase in risk of non-small cell lung cancer for an increase of one standard deviation in mdNLR³⁷. Additionally, a 2020 paper by Kresovich et al. reported that increased B cell proportions are associated with higher breast cancer risk, and that increased monocyte proportions are associated with lower breast cancer risk among premenopausal women³⁸. In contrast, we did not observe any statistically significant relationships between four measures of immune cell proportions (mdNLR, CD4+/CD8+ total cells, B cells/lymphocytes, and T cells/lymphocytes) and risk of pancreatic cancer in a separate study¹⁹.

In addition to observational studies, investigations into cancer cells, hosts, and microenvironments have postulated mechanisms through which cancerous cells are eliminated efficiently by CD8+ cells, while Tregs weaken cellular immune response by impeding the activation of T effector cells, preventing cancer cells from being destroyed and promoting tumor growth^{2,3,39}. This induces cellular and molecular networks, which induce an immunosuppressive environment that favors tumor growth⁴⁰⁻⁴². Increased ratios of Tregs to CD8+ cells have been shown to be an indicator of this immune evasion and tumor growth within the tumor microenvironment (TME)⁴³⁻⁴⁵. It is still largely unknown how Tregs and B cells may interplay to initiate or accelerate tumorigenesis. On the other hand, CD8+ naïve cells are preferential immune cell types for targeting and preventing carcinogenesis⁴⁶. During carcinogenesis, CD8+ cells encounter dysfunction and exhaustion due to immune-related tolerance and immunosuppression within the TME, which favors adaptive immune resistance⁴⁶. Upon their activation, CD8+ cells infiltrate to the core of the tumor's invading site and kill cancer cells⁴⁶. By killing malignant cells upon recognition of specific antigenic peptides by T-cell receptors, CD8+ cells play a central protective role in cancer immunity, unlike Tregs⁴⁷.

There are many key strengths of this analysis. First, the ARIC study is a prospective cohort with a large number of Black participants²³. Prior studies utilizing cohorts such as EPIC, WHI, UK Biobank, the Sister Study, and CLUE II have focused on White populations^{16-18,48}. To our knowledge, no existing study has evaluated the association between methylation-derived immune cell subtypes and cancer risk in a Black study population, extending prior findings into a historically under-researched population. Furthermore, many prior studies have focused only on major immune cell types, and did not study immune cell subtypes such as Tregs, CD8+ naïve cells, and memory B cells. This analysis utilizes the innovative deconvolution algorithm to predict proportions of 12 distinct immune cell subtypes^{19,39,46}. Unlike previously used methods, deconvolution allows for immune cell proportions to be estimated and measured from archived blood (peripheral blood leukocytes), allowing for a prospective study design that examines the role of systemic immune response in cancer risk^{18,20}.

One key limitation of this study is the relatively small number of individual cancer cases, such as lung cancer (n = 84 cases). Despite this, statistically significant associations observed indicate robust findings. However, larger sample sizes are necessary for a more detailed analysis of specific cancer types, such as colorectal cancer (n = 67 cases). We cannot rule out measurement error (including batch-to-batch variation in array data), residual confounding, and reverse causation (including undetected cancer incidence increasing immune cell proportions during short follow-up periods)^{18,20}. While cohort studies inherently carry risks of such biases, we have accounted for them through comprehensive multivariable models, sensitivity analyses (**Supplementary Tables 1, 7**), time-lag analyses (**Supplementary Table 2**), adjustment for batch effect by accounting for heterogeneity of measurements at CpG sites, and a biological measure of smoking pack-years. Taken together, accounting for these biases allows us to conclude that undetected, developing, or incident cancers were likely not driving increased or decreased cell proportions and that residual confounding or measurement error minimally contributed to the observed results. Using the Bonferroni correction for multiple comparisons (16 tests;

adjusted significance level of $P = 0.05/16 = 0.003$), some findings are no longer statistically significant (Tregs in lung cancer and B memory cells in prostate cancer), so these results should be interpreted with caution as they could represent chance findings. However, our findings are largely consistent with prior studies, so results are likely not due to chance alone. Finally, because only postmenopausal women were included in breast cancer analysis, we could not directly replicate prior B cell findings in premenopausal women³⁸.

There are multiple areas for future study, chief among which is the need to identify biomarkers for personalized cancer risk stratification. This will permit better individualized decisions regarding preventive interventions, such as designing guidelines, laying out early screening regimens, and informing patients on lifestyle choices. Future research should strive to validate and refine deconvolution in Black individuals. Further studies should examine the impact of B cells on premenopausal breast cancer risk. Finally, larger studies in Black populations could confirm our findings and allow for assessment of other individual cancers.

In summary, this study shows that in a Black population, higher methylation-derived proportion of Tregs is associated with increased risk of lung cancer, higher memory B cell proportion is associated with increased risk of all cancers and prostate cancer, and higher CD8+ naïve cell proportion is associated with decreased risk of lung cancer and all cancers, particularly at older age. Our study underscores the complex interplay between various immune cell types and cancer risk in a Black population. These insights contribute significantly to our understanding of cancer immunology and highlight the need for further research in diverse populations to enhance cancer prevention and treatment strategies.

Acknowledgements:

The authors thank the staff and participants of the ARIC study for their important contributions. Cancer data were provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP007114, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

Author Contributions:

The CRediT contribution roles that each author played in the development of this manuscript are detailed next to their initials below.

CSS: Data curation; formal analysis; investigation; methodology; validation; visualization; writing – original draft; writing – review & editing

NZ: Data curation; methodology; supervision; writing – review & editing

DCK: Software; supervision; writing – review & editing

EB: Funding acquisition; resources; writing – review & editing

JB: Writing – review & editing

RJB: Funding acquisition; writing – review & editing

KTK: Conceptualization; methodology; writing – review & editing

EAP: Conceptualization; funding acquisition; methodology; project administration; resources; supervision; writing – review & editing

DSM: Conceptualization; funding acquisition; investigation; methodology; project administration; supervision; validation; visualization; writing – original draft; writing – review & editing

The work reported in the paper has been performed by the authors, unless clearly specified in the text.

Funding:

This work was supported by the Jennifer M. Bjercke Scholar Award Fund for Breast Cancer Research at Tufts Medical Center (awarded to Dr. Dominique Michaud), which provided support for the analysis and writing of this manuscript. The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (75N92022D00001, 75N92022D00002, 75N92022D00003, 75N92022D00004, 75N92022D00005). Funding was also supported by 5RC2HL102419 and R01NS087541. Studies on cancer in ARIC are also supported by the National Cancer Institute (U01 CA164975). The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the US Government.

References:

1. Gonzalez H, Hagerling C, Werb Z. Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes Dev.* Oct 1 2018;32(19-20):1267-1284. doi:10.1101/gad.314617.118
2. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science.* Mar 25 2011;331(6024):1565-70. doi:10.1126/science.1203486
3. Finn OJ. Immuno-oncology: understanding the function and dysfunction of the immune system in cancer. *Ann Oncol.* Sep 2012;23 Suppl 8(Suppl 8):viii6-9. doi:10.1093/annonc/mds256
4. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature.* Jan 18 2017;541(7637):321-330. doi:10.1038/nature21349
5. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* Mar 4 2011;144(5):646-74. doi:10.1016/j.cell.2011.02.013
6. Cavallo F, De Giovanni C, Nanni P, Forni G, Lollini PL. 2011: the immune hallmarks of cancer. *Cancer Immunol Immunother.* Mar 2011;60(3):319-26. doi:10.1007/s00262-010-0968-0
7. Karakhanova S, Ryschich E, Mosl B, Harig S, Jager D, Schmidt J, et al.. Prognostic and predictive value of immunological parameters for chemoradioimmunotherapy in patients with pancreatic adenocarcinoma. *Br J Cancer.* Mar 17 2015;112(6):1027-36. doi:10.1038/bjc.2015.72
8. Shang B, Liu Y, Jiang SJ, Liu Y. Prognostic value of tumor-infiltrating FoxP3+ regulatory T cells in cancers: a systematic review and meta-analysis. *Sci Rep.* Oct 14 2015;5:15179. doi:10.1038/srep15179
9. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science.* Sep 29 2006;313(5795):1960-4. doi:10.1126/science.1129139
10. Ryschich E, Notzel T, Hinz U, Autschbach F, Ferguson J, Simon I, et al. Control of T-cell-mediated immune response by HLA class I in human pancreatic carcinoma. *Clin Cancer Res.* Jan 15 2005;11(2 Pt 1):498-504.
11. Bates GJ, Fox SB, Han C, Leek RD, Garcia JF, Harris AL, et al. Quantification of regulatory T cells enables the identification of high-risk breast cancer patients and those at risk of late relapse. *J Clin Oncol.* Dec 1 2006;24(34):5373-80. doi:10.1200/JCO.2006.05.9584
12. Curiel TJ, Coukos G, Zou L, Alvarez X, Cheng P, Mottram P, et al. Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. *Nat Med.* Sep 2004;10(9):942-9. doi:10.1038/nm1093
13. Flammiger A, Weisbach L, Huland H, Tennstedt P, Simon R, Minner S, et al. High tissue density of FOXP3+ T cells is associated with clinical outcome in prostate cancer. *Eur J Cancer.* Apr 2013;49(6):1273-9. doi:10.1016/j.ejca.2012.11.035
14. Suzuki H, Chikazawa N, Tasaka T, Wada J, Yamasaki A, Kitaura Y, et al. Intratumoral CD8(+)/T/FOXP3 (+) cell ratio is a predictive marker for survival in patients with colorectal cancer. *Cancer Immunol Immunother.* May 2010;59(5):653-61. doi:10.1007/s00262-009-0781-9
15. 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.* May 8 2012;13:86. doi:10.1186/1471-2105-13-86
16. Le Cornet C, Schildknecht K, Rossello Chornet A, Fortner RT, Gonzalez Maldonado S, Katzke VA, et al. Circulating Immune Cell Composition and Cancer Risk: A Prospective Study Using Epigenetic Cell Count Measures. *Cancer Res.* May 1 2020;80(9):1885-1892. doi:10.1158/0008-5472.CAN-19-3178
17. Barth SD, Schulze JJ, Kuhn T, Raschke E, Husing A, Johnson T, et al. Treg-Mediated Immune Tolerance and the Risk of Solid Cancers: Findings From EPIC-Heidelberg. *J Natl Cancer Inst.* Nov 2015;107(11)doi:10.1093/jnci/djv224
18. Katzke VA, Le Cornet C, Mahfouz R, Brauer B, Johnson T, Canzian F, et al. Are Circulating Immune Cells a Determinant of Pancreatic Cancer Risk? A Prospective Study Using Epigenetic Cell Count

- Measures. *Cancer Epidemiol Biomarkers Prev*. Dec 2021;30(12):2179-2187. doi:10.1158/1055-9965.EPI-21-0169
19. Michaud DS, Ruan M, Koestler DC, Alonso L, Molina-Montes E, Pei D, et al. DNA Methylation-Derived Immune Cell Profiles, CpG Markers of Inflammation, and Pancreatic Cancer Risk. *Cancer Epidemiol Biomarkers Prev*. Aug 2020;29(8):1577-1585. doi:10.1158/1055-9965.EPI-20-0378
20. Michaud DS, Kelsey KT. DNA Methylation in Peripheral Blood: Providing Novel Biomarkers of Exposure and Immunity to Examine Cancer Risk. *Cancer Epidemiol Biomarkers Prev*. Dec 2021;30(12):2176-2178. doi:10.1158/1055-9965.EPI-21-0866
21. Zhang W, Xu H, Qiao R, Zhong B, Zhang X, Gu J, et al. ARIC: accurate and robust inference of cell type proportions from bulk gene expression or DNA methylation data. *Brief Bioinform*. Jan 17 2022;23(1)doi:10.1093/bib/bbab362
22. Atallah-Yunes SA, Ready A, Newburger PE. Benign ethnic neutropenia. *Blood Rev*. Sep 2019;37:100586. doi:10.1016/j.blre.2019.06.003
23. Wright JD, Folsom AR, Coresh J, Sharrett AR, Couper D, Wagenknecht LE, et al. The ARIC (Atherosclerosis Risk In Communities) Study: JACC Focus Seminar 3/8. *J Am Coll Cardiol*. Jun 15 2021;77(23):2939-2959. doi:10.1016/j.jacc.2021.04.035
24. Jackson R, Chambless LE, Yang K, Byrne T, Watson R, Folsom A, et al. Differences between respondents and nonrespondents in a multicenter community-based study vary by gender ethnicity. The Atherosclerosis Risk in Communities (ARIC) Study Investigators. *J Clin Epidemiol*. Dec 1996;49(12):1441-46. doi:10.1016/0895-4356(95)00047-x
25. Joshi CE, Barber JR, Coresh J, Couper DJ, Mosley TH, Vitolins MZ, et al. Enhancing the Infrastructure of the Atherosclerosis Risk in Communities (ARIC) Study for Cancer Epidemiology Research: ARIC Cancer. *Cancer Epidemiol Biomarkers Prev*. Mar 2018;27(3):295-305. doi:10.1158/1055-9965.EPI-17-0696
26. 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. *Nat Commun*. Feb 9 2022;13(1):761. doi:10.1038/s41467-021-27864-7
27. Trame CB, Chang Y, Axelrod HL, Eberhardt RY, Coggill P, Punta M, et al. New mini- zincin structures provide a minimal scaffold for members of this metallopeptidase superfamily. *BMC Bioinformatics*. Jan 3 2014;15:1. doi:10.1186/1471-2105-15-1
28. Koestler DC, Christensen B, Karagas MR, Marsit CJ, Langevin SM, Kelsey KT, et al. Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a validation analysis. *Epigenetics*. Aug 2013;8(8):816-26. doi:10.4161/epi.25430
29. Titus AJ, Gallimore RM, Salas LA, Christensen BC. Cell-type deconvolution from DNA methylation: a review of recent applications. *Hum Mol Genet*. Oct 1 2017;26(R2):R216-R224. doi:10.1093/hmg/ddx275
30. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One*. 2012;7(7):e41361. doi:10.1371/journal.pone.0041361
31. Bell-Glenn S, Salas LA, Molinaro AM, Butler RA, Christensen BC, Kelsey KT, et al. Calculating detection limits and uncertainty of reference-based deconvolution of whole-blood DNA methylation data. *Epigenomics*. Apr 2023;15(7):435-451. doi:10.2217/epi-2023-0006
32. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. Oct 2010;11(10):733-9. doi:10.1038/nrg2825
33. 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*. Mar 8 2016;17:120. doi:10.1186/s12859-016-0943-7

34. 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 Biol.* May 29 2018;19(1):64. doi:10.1186/s13059-018-1448-7
35. Margolis KL, Rodabough RJ, Thomson CA, Lopez AM, McTiernan A, Women's Health Initiative Research G. Prospective study of leukocyte count as a predictor of incident breast, colorectal, endometrial, and lung cancer and mortality in postmenopausal women. *Arch Intern Med.* Sep 24 2007;167(17):1837-44. doi:10.1001/archinte.167.17.1837
36. Wong JYY, Bassig BA, Loftfield E, Hu W, Freedman ND, Ji BT, et al. White Blood Cell Count and Risk of Incident Lung Cancer in the UK Biobank. *JNCI Cancer Spectr.* Apr 2020;4(2):pkz102. doi:10.1093/jncics/pkz102
37. Zhao N, Ruan M, Koestler DC, Lu J, Salas LA, Kelsey KT, et al. Methylation-derived inflammatory measures and lung cancer risk and survival. *Clin Epigenetics.* Dec 16 2021;13(1):222. doi:10.1186/s13148-021-01214-2
38. Kresovich JK, O'Brien KM, Xu Z, Weinberg CR, Sandler DP, Taylor JA. Prediagnostic Immune Cell Profiles and Breast Cancer. *JAMA Netw Open.* Jan 3 2020;3(1):e1919536. doi:10.1001/jamanetworkopen.2019.19536
39. Saleh R, Elkord E. FoxP3(+) T regulatory cells in cancer: Prognostic biomarkers and therapeutic targets. *Cancer Lett.* Oct 10 2020;490:174-185. doi:10.1016/j.canlet.2020.07.022
40. Takeuchi Y, Nishikawa H. Roles of regulatory T cells in cancer immunity. *Int Immunol.* Aug 2016;28(8):401-9. doi:10.1093/intimm/dxw025
41. Shimizu J, Yamazaki S, Sakaguchi S. Induction of tumor immunity by removing CD25+CD4+ T cells: a common basis between tumor immunity and autoimmunity. *J Immunol.* Nov 15 1999;163(10):5211-8.
42. Shitara K, Nishikawa H. Regulatory T cells: a potential target in cancer immunotherapy. *Ann N Y Acad Sci.* Apr 2018;1417(1):104-115. doi:10.1111/nyas.13625
43. Jagger A, Shimojima Y, Goronzy JJ, Weyand CM. Regulatory T cells and the immune aging process: a mini-review. *Gerontology.* 2014;60(2):130-7. doi:10.1159/000355303
44. Yoon HH, Orrock JM, Foster NR, Sargent DJ, Smyrk TC, Sinicrope FA. Prognostic impact of FoxP3+ regulatory T cells in relation to CD8+ T lymphocyte density in human colon carcinomas. *PLoS One.* 2012;7(8):e42274. doi:10.1371/journal.pone.0042274
45. deLeeuw RJ, Kost SE, Kakal JA, Nelson BH. The prognostic value of FoxP3+ tumor-infiltrating lymphocytes in cancer: a critical review of the literature. *Clin Cancer Res.* Jun 1 2012;18(11):3022-9. doi:10.1158/1078-0432.CCR-11-3216
46. Farhood B, Najafi M, Mortezaee K. CD8(+) cytotoxic T lymphocytes in cancer immunotherapy: A review. *J Cell Physiol.* Jun 2019;234(6):8509-8521. doi:10.1002/jcp.27782
47. Durgeau A, Virk Y, Corgnac S, Mami-Chouaib F. Recent Advances in Targeting CD8 T-Cell Immunity for More Effective Cancer Immunotherapy. *Front Immunol.* 2018;9:14. doi:10.3389/fimmu.2018.00014
48. Xu Z, Sandler DP, Taylor JA. Blood DNA Methylation and Breast Cancer: A Prospective Case-Cohort Analysis in the Sister Study. *J Natl Cancer Inst.* Jan 1 2020;112(1):87-94. doi:10.1093/jnci/djz065

Table 1: Baseline Characteristics for the Study Population, Characterized by Treg Proportion, Black Participants in ARIC

Characteristic	Quartile of Treg Composition				Overall (n = 2467)
	Quartile 1 (n = 845)	Quartile 2 (n = 391)	Quartile 3 (n = 614)	Quartile 4 (n = 617)	
Age					
Mean (SD)	55.8 (5.74)	56.3 (5.86)	56.4 (5.80)	57.4 (5.73)	56.4 (5.80)
Median	55.0	56.0	56.0	57.0	56.0
[Min, Max]	[47.0, 71.0]	[47.0, 69.0]	[47.0, 69.0]	[47.0, 70.0]	[47.0, 71.0]
BMI					
Mean (SD)	30.0 (5.76)	30.0 (6.36)	30.6 (6.67)	29.8 (6.31)	30.1 (6.25)
Median	29.2	28.9	29.5	29.2	29.2
[Min, Max]	[17.7, 53.8]	[15.1, 62.4]	[16.1, 57.9]	[14.7, 54.8]	[14.7, 62.4]
Sex (%)					
Female	529 (62.6%)	239 (61.1%)	399 (65.0%)	401 (65.0%)	1568 (63.6%)
Male	316 (37.4%)	152 (38.9%)	215 (35.0%)	216 (35.0%)	899 (36.4%)
Smoking Status (%)					
Never	487 (57.6%)	183 (46.8%)	291 (47.3%)	265 (42.9%)	1226 (49.7%)
Former smoker	193 (22.8%)	98 (25.1%)	171 (27.9%)	164 (26.6%)	626 (25.4%)
Current smoker	165 (19.6%)	110 (28.1%)	152 (24.8%)	188 (30.5%)	615 (24.9%)
Smoking Pack-Years, Current and Former Smokers					
Mean (SD)	21.0 (23.4)	20.9 (21.6)	19.78 (18.7)	25.5 (25.2)	21.9 (22.6)
Median	14.6	15.7	16.0	19.9	16.3
[Min, Max]	[0, 199]	[0, 180]	[0, 138]	[0, 189]	[0, 199]
Methylation-Derived Pack-Years					
Mean (SD)	19.1 (34.5)	21.5 (34.9)	19.3 (35.6)	22.5 (37.1)	20.4 (35.5)
Median	0	0	0	0	0
[Min, Max]	[0, 302]	[0, 183]	[0, 239]	[0, 202]	[0, 302]
Menopausal Hormone Therapy Use, Women Only (%)					
Never	305 (57.7%)	142 (59.4%)	233 (58.4%)	274 (68.3%)	954 (60.8%)
Former user	17 (3.2%)	8 (3.3%)	13 (3.3%)	6 (1.5%)	44 (2.8%)
Current user	118 (22.3%)	47 (19.7%)	78 (19.5%)	48 (12.0%)	291 (18.6%)
Unknown	89 (16.8%)	42 (17.6%)	75 (18.8%)	73 (18.2%)	279 (17.8%)
Education Level (%)					
Basic Education (less than completed high school)	269 (31.8%)	160 (40.9%)	252 (41.0%)	290 (47.0%)	971 (39.4%)
Intermediate Education (high school or equivalent)	245 (29.0%)	109 (27.9%)	182 (29.6%)	168 (27.2%)	704 (28.5%)
Advanced Education (at least some college)	331 (39.2%)	119 (30.4%)	176 (28.7%)	158 (25.6%)	784 (31.8%)
Missing	0 (0%)	3 (0.8%)	4 (0.7%)	1 (0.2%)	8 (0.3%)
Drinking Status (%)					
Never	316 (37.4%)	134 (34.3%)	217 (35.3%)	200 (32.4%)	867 (35.1%)
Former drinker	244 (28.9%)	124 (31.7%)	202 (32.9%)	211 (34.2%)	781 (31.7%)
Current drinker	285 (33.7%)	133 (34.0%)	195 (31.8%)	206 (33.4%)	819 (33.2%)
Methylation-Derived Neutrophil-to-Lymphocyte Ratio					
Mean (SD)	3.22 (6.43)	1.99 (2.24)	1.67 (1.31)	1.46 (0.967)	2.21 (4.02)
Median	1.36	1.37	1.34	1.17	1.32
[Min, Max]	[0.223, 46.3]	[0.308, 33.1]	[0.176, 12.2]	[0.263, 11.5]	[0.176, 46.3]

Table 2: Baseline Characteristics for the Study Population, Characterized by CD8+ Naïve Cell Proportion, Black Participants in ARIC

Characteristic	Quartile of CD8+ Naïve Cell Composition				Overall (n = 2467)
	Quartile 1 (n = 1111)	Quartile 2 (n = 452)	Quartile 3 (n = 452)	Quartile 4 (n = 452)	
Age					
Mean (SD)	57.7 (5.91)	56.2 (5.62)	55.6 (5.44)	54.3 (5.28)	56.4 (5.80)
Median	58.0	56.0	55.0	53.0	56.0
[Min, Max]	[47.0, 71.0]	[47.0, 70.0]	[47.0, 70.0]	[47.0, 69.0]	[47.0, 71.0]
BMI					
Mean (SD)	30.2 (6.28)	30.2 (6.67)	29.7 (5.84)	30.2 (6.03)	30.1 (6.23)
Median	29.3	29.4	28.9	29.2	29.3
[Min, Max]	[14.7, 54.9]	[15.1, 62.4]	[15.8, 53.8]	[16.1, 52.8]	[14.7, 62.4]
Sex (%)					
Female	638 (57.4%)	267 (59.1%)	321 (71.0%)	342 (75.7%)	1568 (63.6%)
Male	473 (42.6%)	185 (40.9%)	131 (29.0%)	110 (24.3%)	899 (36.4%)
Smoking Status (%)					
Never	572 (51.5%)	219 (48.6%)	222 (49.1%)	213 (47.1%)	1226 (49.7%)
Former smoker	277 (24.9%)	116 (25.7%)	117 (25.9%)	116 (25.7%)	626 (25.4%)
Current smoker	262 (23.6%)	117 (25.9%)	113 (25.0%)	123 (27.2%)	615 (24.9%)
Smoking Pack-Years, Current and Former Smokers					
Mean (SD)	23.1 (22.6)	22.3 (22.6)	22.0 (26.8)	18.9 (17.5)	21.9 (22.6)
Median	18.6	16.9	15.2	14.0	16.3
[Min, Max]	[0,199]	[0, 160]	[0, 189]	[0, 82.0]	[0, 199]
Methylation-Derived Pack-Years					
Mean (SD)	20.8 (36.0)	19.8 (35.1)	18.8 (33.7)	21.5 (36.6)	20.4 (35.5)
Median	0	0	0	0	0
[Min, Max]	[0, 302]	[0, 188]	[0, 167]	[0, 239]	[0, 302]
Menopausal Hormone Therapy Use, Women Only (%)					
Never	401 (62.8%)	162 (60.7%)	188 (58.6%)	203 (59.4%)	954 (60.8%)
Former user	12 (1.9%)	10 (3.7%)	10 (3.1%)	12 (3.5%)	44 (2.8%)
Current user	111 (17.4%)	50 (18.7%)	57 (17.8%)	73 (21.3%)	291 (18.6%)
Unknown	114 (17.9%)	45 (16.9%)	66 (20.5%)	54 (15.8%)	279 (17.8%)
Education Level (%)					
Basic Education (less than completed high school)	492 (44.3%)	168 (37.2%)	179 (39.6%)	132 (29.2%)	971 (39.4%)
Intermediate Education (high school or equivalent)	304 (27.4%)	137 (30.3%)	135 (29.9%)	128 (28.3%)	704 (28.5%)
Advanced Education (at least some college)	309 (27.8%)	146 (32.3%)	138 (30.5%)	191 (42.3%)	784 (31.8%)
Missing	6 (0.5%)	1 (0.2%)	0 (0%)	1 (0.2%)	8 (0.3%)
Drinking Status (%)					
Never	366 (32.9%)	153 (33.8%)	170 (37.6%)	178 (39.4%)	867 (35.1%)
Former drinker	376 (33.8%)	153 (37.6%)	135 (29.9%)	118 (26.1%)	781 (31.7%)
Current drinker	369 (33.3%)	147 (32.5%)	147 (32.5%)	156 (34.5%)	819 (33.2%)
Methylation-Derived Neutrophil-to-Lymphocyte Ratio					
Mean (SD)	2.61 (5.27)	2.44 (3.58)	1.91 (2.25)	1.38 (1.11)	2.21 (4.02)
Median	1.42	1.51	1.30	0.984	1.32
[Min, Max]	[0.263, 46.3]	[0.176, 37.5]	[0.230, 26.2]	[0.223, 9.92]	[0.176, 46.3]

Table 3: Hazard Ratios (HRs) for Immune Cell Proportions and Cancer Risk, Black Participants in ARIC
 HR^b (95% CI) per 1 percent increase in methylation-derived immune cell proportion or 1 unit increase in ratios or white blood cell count

Methylation-Derived Immune Cell Type or Measure	All Cancer ^a (668 cases)	Lung Cancer (84 cases)	Postmenopausal Breast Cancer ^c (114 cases)	Prostate Cancer ^d (173 cases)
CD4+	0.99 (0.98, 1.01)	1.00 (0.96, 1.03)	1.00 (0.97, 1.02)	1.00 (0.98, 1.03)
CD4+ memory	0.99 (0.98, 1.01)	0.99 (0.95, 1.04)	1.01 (0.98, 1.05)	1.00 (0.97, 1.03)
CD4+ naïve	1.00 (0.98, 1.02)	1.00 (0.95, 1.06)	0.97 (0.93, 1.01)	1.01 (0.97, 1.06)
CD4+ naïve-to-memory ratio	1.02 (0.98, 1.05)	0.96 (0.85, 1.09)	1.00 (0.93, 1.08)	0.98 (0.90, 1.06)
CD8+	1.00 (0.99, 1.01)	1.02 (0.99, 1.06)	1.02 (0.98, 1.05)	1.01 (0.99, 1.03)
CD8+ memory	1.00 (0.99, 1.02)	1.03 (1.00, 1.06)	1.02 (0.99, 1.05)	1.01 (0.99, 1.03)
CD8+ naïve	0.96 (0.91, 1.01)	0.85 (0.71, 1.01)	0.96 (0.86, 1.07)	1.00 (0.89, 1.12)
CD8+ naïve-to-memory ratio	0.98 (0.87, 1.09)	0.87 (0.59, 1.29)	0.96 (0.75, 1.24)	0.84 (0.62, 1.13)
CD4+-to-CD8+ ratio	0.97 (0.91, 1.04)	0.98 (0.84, 1.15)	0.89 (0.75, 1.07)	0.94 (0.82, 1.07)
Treg	1.06 (1.00, 1.12)	1.22 (1.06, 1.41)	1.04 (0.92, 1.18)	1.08 (0.96, 1.22)
B cell	0.99 (0.97, 1.02)	0.98 (0.91, 1.06)	1.04 (0.98, 1.10)	0.97 (0.92, 1.03)
B cell memory	1.13 (1.05, 1.22)	1.20 (0.99, 1.46)	0.95 (0.74, 1.24)	1.17 (1.04, 1.33)
B cell naïve	0.98 (0.95, 1.01)	0.96 (0.88, 1.04)	1.04 (0.98, 1.10)	0.94 (0.89, 1.00)
B cell naïve-to-memory ratio	0.99 (0.98, 1.00)	1.06 (0.94, 1.00)	1.01 (0.99, 1.04)	0.98 (0.96, 1.00)
NLR	1.00 (0.98, 1.02)	0.95 (0.87, 1.03)	0.99 (0.94, 1.04)	1.01 (0.98, 1.04)
Lymphocyte to Monocyte Ratio	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.03 (0.95, 1.12)
White blood cell count (from CBC differential)	1.03 (0.98, 1.08) (n = 659)	1.19 (1.06, 1.33) (n = 83)	1.05 (0.94, 1.18) (n = 112)	0.99 (0.91, 1.08) (n = 171)

^a Excluding hematologic cancers. ^b Multivariable models were adjusted for age, sex, BMI, self-reported smoking status, self-reported smoking pack-years, postmenopausal hormone use (creating an 'NA' category for males), methylation-derived smoking pack-years, mdNLR (in all models except mdNLR and lymphocyte to monocyte ratio), and batch effect. ^c The breast cancer model also adjusted for self-reported drinking status and did not adjust for sex. ^d The prostate cancer model did not adjust for sex or postmenopausal hormone use (men only).

Table 4: HRs for Age-Stratified Model in All Cancers^a, Black Participants in ARIC

HR^b (95% CI) per 1 percent increase in methylation-derived immune cell proportion or 1 unit increase in ratios or white blood cell count

Methylation-Derived Immune Cell Type or Measure	Age Group		p-value for interaction
	55 and Under All Cancer HR (95% CI) (294 cases)	Over 55 All Cancer HR (95% CI) (374 cases)	
CD4+	1.00 (0.98, 1.02)	0.99 (0.98, 1.01)	0.31
CD4+ memory	0.99 (0.96, 1.01)	1.00 (0.98, 1.02)	0.58
CD4+ naïve	1.02 (1.00, 1.05)	0.99 (0.96, 1.01)	0.27
CD4+ naïve-to-memory ratio	1.03 (0.99, 1.08)	1.00 (0.95, 1.06)	0.83
CD8+	1.00 (0.98, 1.02)	1.01 (0.99, 1.02)	0.90
CD8+ memory	1.00 (0.98, 1.02)	1.01 (0.99, 1.02)	0.73
CD8+ naïve	0.99 (0.93, 1.07)	0.91 (0.83, 0.98)	0.13
CD8+ naïve-to-memory ratio	1.01 (0.88, 1.16)	0.93 (0.77, 1.12)	0.54
CD4+-to-CD8+ ratio	1.04 (0.95, 1.05)	0.93 (0.86, 1.02)	0.11
Treg	1.06 (0.97, 1.16)	1.05 (0.98, 1.12)	0.89
B cell	0.98 (0.94, 1.02)	1.00 (0.97, 1.04)	0.91
B cell memory	1.16 (1.05, 1.27)	1.10 (0.99, 1.23)	0.25
B cell naïve	0.96 (0.92, 1.01)	0.99 (0.96, 1.03)	0.51
B cell naïve-to-memory ratio	0.99 (0.97, 1.00)	0.99 (0.98, 1.01)	0.59
NLR	0.99 (0.96, 1.02)	1.00 (0.98, 1.03)	0.62
Lymphocyte to Monocyte Ratio	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	0.98
White blood cell count (from CBC differential)	1.04 (0.98, 1.11) (n = 290)	1.02 (0.96, 1.08) (n = 369)	0.86

^a Excluding hematologic cancers. ^b Multivariable models were adjusted for sex, BMI, self-reported smoking status, self-reported smoking pack-years, methylation-derived smoking pack-years, postmenopausal hormone use, mdNLR, and batch effect.

Table 5: HRs for Age-Stratified Model in Postmenopausal Breast Cancer, Black Participants in ARIC
 HR^a (95% CI) per 1 percent increase in methylation-derived immune cell proportion or 1 unit increase in ratios or white blood cell count

Methylation-Derived Immune Cell Type or Measure	Age Group		p-value for interaction
	55 and Under Breast Cancer HR (95% CI) (49 cases)	Over 55 Breast Cancer HR (95% CI) (65 cases)	
CD4+	0.99 (0.94, 1.04)	1.00 (0.96, 1.03)	0.67
CD4+ memory	1.00 (0.95, 1.06)	1.02 (0.97, 1.06)	0.65
CD4+ naïve	0.98 (0.92, 1.05)	0.96 (0.90, 1.02)	0.76
CD4+ naïve-to-memory ratio	1.04 (0.95, 1.13)	0.95 (0.83, 1.09)	0.64
CD8+	1.03 (0.98, 1.08)	1.01 (0.97, 1.05)	0.19
CD8+ memory	1.02 (0.98, 1.07)	1.01 (0.98, 1.05)	0.23
CD8+ naïve	1.00 (0.85, 1.17)	0.90 (0.75, 1.07)	0.68
CD8+ naïve-to-memory ratio	1.09 (0.82, 1.45)	0.79 (0.48, 1.28)	0.40
CD4+-to-CD8+ ratio	0.91 (0.68, 1.22)	0.89 (0.71, 1.11)	0.83
Treg	1.09 (0.90, 1.32)	1.02 (0.87, 1.19)	0.60
B cell	0.98 (0.89, 1.09)	1.06 (0.99, 1.14)	0.93
B cell memory	0.92 (0.64, 1.30)	0.95 (0.64, 1.42)	0.54
B cell naïve	0.99 (0.90, 1.10)	1.07 (1.00, 1.15)	0.82
B cell naïve-to-memory ratio	1.00 (0.96, 1.03)	1.03 (1.00, 1.05)	0.46
NLR	0.88 (0.71, 1.09)	1.02 (0.97, 1.07)	0.37
Lymphocyte to Monocyte Ratio	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	0.67
White blood cell count (from CBC differential)	1.09 (0.92, 1.30) (n = 48)	1.02 (0.87, 1.20) (n = 64)	0.85

^a Multivariable models were adjusted for BMI, self-reported smoking status, self-reported smoking pack-years, methylation-derived smoking pack-years, self-reported drinking status, postmenopausal hormone use, mdNLR, and batch effect.

Table 6: HRs for Age-Stratified Model in Prostate Cancer, Black Participants in ARIC

HR^a (95% CI) per 1 percent increase in methylation-derived immune cell proportion or 1 unit increase in ratios or white blood cell count

Methylation-Derived Immune Cell Type or Measure	Age Group		p-value for interaction
	55 and Under Prostate Cancer HR (95% CI) (79 cases)	Over 55 Prostate Cancer HR (95% CI) (94 cases)	
CD4+	1.01 (0.97, 1.05)	1.00 (0.97, 1.03)	0.06
CD4+ memory	0.99 (0.94, 1.03)	1.01 (0.97, 1.05)	0.26
CD4+ naïve	1.05 (0.98, 1.12)	0.99 (0.93, 1.05)	0.12
CD4+ naïve-to-memory ratio	1.00 (0.88, 1.14)	0.96 (0.85, 1.08)	0.83
CD8+	1.00 (0.97, 1.01)	1.01 (0.98, 1.04)	0.38
CD8+ memory	1.00 (0.96, 1.03)	1.01 (0.99, 1.04)	0.68
CD8+ naïve	1.07 (0.95, 1.01)	0.81 (0.64, 1.04)	0.02
CD8+ naïve-to-memory ratio	0.94 (0.68, 1.29)	0.61 (0.32, 1.17)	0.28
CD4+ to-CD8+ ratio	1.02 (0.86, 1.20)	0.86 (0.71, 1.06)	0.22
Treg	1.08 (0.89, 1.09)	1.07 (0.91, 1.25)	0.37
B cell	1.00 (0.92, 1.09)	0.94 (0.87, 1.02)	0.04
B cell memory	1.25 (1.10, 1.43)	0.98 (0.77, 1.26)	0.09
B cell naïve	0.94 (0.85, 1.03)	0.94 (0.86, 1.02)	0.36
B cell naïve-to-memory ratio	0.97 (0.94, 1.01)	0.99 (0.96, 1.02)	0.88
NLR	0.96 (0.88, 1.05)	1.03 (0.99, 1.06)	0.10
Lymphocyte to Monocyte Ratio	1.11 (0.94, 1.31)	1.00 (0.99, 1.01)	0.04
White blood cell count (from CBC differential)	1.01 (0.88, 1.15) (n = 77)	0.98 (0.87, 1.10) (n = 94)	0.22

^a Multivariable models represent hazard ratios for each one-percent increase in cell subtype composition in prostate cancer; models were adjusted for BMI, self-reported smoking status, self-reported smoking pack-years, methylation-derived smoking pack-years, mdNLR, and batch effect.