Cancer Causes & Control (2019) 30:1259–1268
https://doi.org/10.1007/s10552-019-01212-z

ORIGINAL PAPER

Relationship between West African ancestry with lung cancer risk and survival in African Americans

Khadijah A. Mitchell[1] · Ebony Shah[2] · Elise D. Bowman[1] · Adriana Zingone[1] · Noah Nichols[1] · Sharon R. Pine[3] · Rick A. Kittles[4] · Bríd M. Ryan[1,5]

Received: 31 December 2018 / Accepted: 30 July 2019 / Published online: 29 August 2019

Abstract

Purpose African Americans, especially men, have a higher incidence of lung cancer compared with all other racial and ethnic groups in the US. Self-reported race is frequently used in genomic research studies to capture an individual's race or ethnicity. However, it is clear from studies of genetic admixture that human genetic variation does not segregate into the same biologically discrete categories as socially defined categories of race. Previous studies have suggested that the degree of West African ancestry among African Americans can contribute to cancer risk in this population, though few studies have addressed this question in lung cancer.

Methods Using a genetic ancestry panel of 100 SNPs, we estimated West African, European, and Native American ancestry in 1,407 self-described African Americans and 2,413 European Americans.

Results We found that increasing West African ancestry was associated with increased risk of lung cancer among African American men ($OR_{Q5\ vs\ Q1} = 2.55$ (1.45–4.48), $p = 0.001$), while no association was observed in African American women ($OR_{Q5\ vs\ Q1} = 0.90$ (0.51–1.59), $p = 0.56$). This relationship diminished following adjustment for income and education.

Conclusions Genetic ancestry is not a major contributor to lung cancer risk or survival disparities.

Keywords Lung cancer · Cancer disparities · Risk · Survival · Genetic ancestry · African American

Abbreviations

OR	Odds ratio
CI	Confidence interval
AIMS	Ancestry-informative markers
SNP	Single-nucleotide polymorphism
HR	Hazard ratio

Background

Lung cancer is the second most common cancer in the United States and the leading cause of cancer-related deaths [1]. First noted in 1972 [2], racial differences in lung cancer incidence exist especially among African American men who have the highest incidence of lung cancer among all racial and ethnic groups in the country [3]. In addition, lung cancer survival varies considerably by self-reported race [4–6]. Several factors, including biological, environmental [7, 8], social, economic, and cultural causes, including access to care [5, 9] could explain the disparities.

African Americans are an admixed population [10] with a high degree of genetic heterogeneity due to African and

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10552-019-01212-z) contains supplementary material, which is available to authorized users.

✉ Bríd M. Ryan
Brid.Ryan@nih.gov

[1] Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, USA

[2] Indiana University School of Medicine, Indianapolis, IN 46202, USA

[3] Rutgers Cancer Institute of New Jersey, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, New Brunswick, NJ 08854, USA

[4] Division of Health Equities, Department of Populations Sciences, City of Hope Comprehensive Cancer Center, Duarte, CA 91010, USA

[5] Laboratory of Human Carcinogenesis, Center for Cancer Research, NCI, Building 37, Room 3060C, Bethesda, MD 20892, USA

non-African heritage. This genetic variation predominantly reflects Western and West central African (~71%), European (~13%), and other African (~8%) populations [10]. However, the contribution of African and European ancestry differs greatly at the individual level in African Americans from the same geographic area in the US. Also, the proportion of African ancestry can vary as much as tenfold in self-reported African American and European American populations [11, 12]. Populations in the United States also have a proportion of Native American ancestry but its contribution to health and disease remains understudied. Recent investigations found that the proportion of Native American ancestry is associated with risk of acute lymphoblastic leukemia in Hispanic/Latino children [13] and breast cancer among Hispanic/Latinos [14] in admixed populations and that the proportion of African ancestry is related to the efficacy of the smoking cessation drug, Naltrexone [15]. Genetic ancestry is associated with disease risk for other cancers with a known health disparity in African American populations, such as breast and prostate cancer. Excessive African ancestry at 2q37.1 and 3p24.2 is associated with a higher incidence of triple-negative breast cancer (TNBC), an aggressive breast cancer subtype [16]. African-enriched genetic variation found in promoter regions affecting transcription factor binding sites have been significantly associated with prostate cancer risk after adjusting for age and West African ancestry [17]. These studies suggest breast and prostate cancer susceptibility alleles only present in African ancestry populations are associated with disease.

Self-reported race is often used in biomedical research to capture an individual's race or ethnicity. However, admixture studies show that genetic variation does not segregate into the same biologically discrete categories as socially defined categories of race [12]. Ancestry-informative markers (AIM) are collections of DNA polymorphisms with significant frequency differences between populations across continents and can generate a more accurate measure of genetic ancestral background than self-reported race. AIMs allow us to study the relationship between genetic ancestry and cancer phenotypes, including lung cancer risk and survival. Given that the main lung cancer disparity is among men, we aimed to determine whether genetic ancestry was associated with lung cancer risk and whether this relationship differed by sex.

Materials and methods

Ethics

All participants provided informed consent for the collection of biospecimens and demographic information in accordance with the Declaration of Helsinki. The study was approved by the institutional review board of the National Cancer Institute and the University of Maryland (IRB Number: OH98-C-N027).

The NCI-MD case–control study

Cases were recruited from seven different hospitals in the Baltimore metropolitan area. The recruitment of self-reported African American and European American participants in this study occurred between 1998 and 2016, as described previously [18]. Exclusion criteria for eligibility included; being more than 24 months after initial cancer diagnosis, non-US resident, non-English speaking, residing in an institution such as a prison, nursing home or shelter, severely ill, or unable to give informed consent. Approximately 16% of cases identified were not eligible. Population controls were contacted via Maryland Department of Motor Vehicles records. Hospital controls, which were not collected past 2007, were recruited through University of Maryland outpatient clinics. The demographics of cases and controls are summarized in Table 1.

Global genetic ancestry

Global genetic ancestry analysis was performed as previously described [19]. Briefly, DNA from blood was genotyped for 100 AIMs using the Sequenom MassARRAY iPLEX platform. The AIMs panel consisted of carefully selected autosomal markers previously identified and validated for estimating continental ancestry information in admixed populations (Supplementary Table 1) [20–25]. Individual single-nucleotide polymorphism (SNP) genotype calls were generated using Sequenom TYPER software. A genotype concordance rate of 99.5% was observed for all markers. Genotyping call rates exceeded 96% for all individuals included in the analyses. Individual admixture estimates for each study participant were calculated using a model-based clustering method as implemented in the program STRUCTURE v2.3 [26]. As we were unsure about the ancestries of our samples, we used the admixture model to determine which estimation of K (number of sub populations) is the best fit for the data. We set K from 2 to 5 and ran 100 iterations. We determined that $K=3$ had the best fit and used the $K=3$ estimates for our analyses. There was no evidence of any hidden relationships among the participants.

GWAS

Genome wide association study (GWAS) data previously collected for African American lung cancer patients [27] were mined for SNPs in *PDE4D*. Main effect summary statistics were extracted following stratification by sex. These data are deposited at dbGAP (accession number phs001210.

Table 1 Demographics of the study population

	African Americans $n=1,407$			European Americans $n=2,413$		
	Control $n=949$	Case $n=458$	p^a	Control $n=1,363$	Case $n=1,050$	p^b
Demographics						
Age (median, IQR)	65 (59–70)	63 (57–71)	< 0.0001	66 (60–73)	67 (59–74)	< 0.0001
Sex						
Male	631 (66.5%)	242 (52.8%)	< 0.0001	797 (58.5%)	560 (53.3%)	0.86
Female	318 (33.5%)	216 (47.2%)		566 (41.5%)	490 (46.7%)	
Smoking status						
Never	334 (35.2%)	31 (6.8%)	0.004	489 (35.9%)	96 (9.1%)	< 0.0001
Former	406 (42.7%)	166 (36.2%)		648 (47.5%)	478 (45.5%)	
Current	209 (22.1%)	261 (57.1%)		226 (16.6%)	476 (45.4%)	
Missing	0	0		0	0	
Pack-years (median, IQR)	7.4 (0–25.1)	31 (16.0–49.1)	< 0.0001	11 (0–37.5)	42 (23.5–62.5)	< 0.0001
Histology						
Adenocarcinoma		205 (45.3%)			518 (48.9%)	
Squamous		124 (27.4%)			268 (25.8%)	
Large Cell		5 (1.1%)			16 (1.5%)	
NSCLC		70 15.5%)			138 (13.3%)	
BAC		10 (2.2%)			34 (3.3%)	
Adenosquamous		7 (1.6%)			21 (2.0%)	
Other		6 (1.3%)			11 (1.1%)	
Small cell		6 (1.3%)			11 (1.1%)	
Missing		25 (4.3%)			33 (3.0%)	
Stage						
I		139 (36.5%)			336 (36.5%)	
II		52 (14.3%)			132 (14.3%)	
III		91 (21.9%)			202 (21.9%)	
IV		93 (22.6%)			208 (22.6%)	
Missing		25 (6.3%)			43 (4.7%)	

P^a denotes differences between AA controls and EA controls and P^b refers to differences between AA cases and EA cases. Student's t test was used for continuous variables and a χ^2 test was used for categorical variables

v1.p1). The program LDlink [28] was used to assess linkage between alleles.

Statistical analysis

All statistical analyses, including χ^2 tests on categorical variables, student's t test on continuous variables, logistic regression, likelihood-ratio tests, and univariable and multivariable Cox regression analyses were conducted using STATA version 14 (StataCorp LP, College Station, TX). Baseline characteristics of the population were compared across self-reported racial/ethnic groups using χ^2 tests on categorical variables or Student's t test on continuous variables. Former cigarette smokers were defined as having quit one or more years prior to being interviewed to control for behavior changes in smoking and never smokers were defined as having smoked < 100 cigarettes in their lifetime. Summary measures of genetic ancestry were assessed based on the individual genetic ancestry estimates within self-reported racial/ethnic groups. We used logistic regression to estimate odds ratios (OR) and 95% confidence intervals (CI) for the relationship between specific AIM SNPs and lung cancer risk without adjustment for false discovery. GWAS data were used to follow up interesting genes. P values for interaction were determined using a likelihood-ratio test comparing a full model including an ordinal multiplicative interaction term to a reduced model without an interaction term. Hazard ratios (HR) were estimated using univariable and multivariable Cox regression analysis, controlling for age, sex, smoking status, pack-years of smoking, stage, histology, education, and income. Lung cancer-specific survival was assessed, where a lung cancer death was noted if lung

cancer was listed as the primary, secondary, or tertiary cause of death. Results were considered significant if $p < 0.05$ and the proportional hazards assumption was met. All statistical analyses were conducted using STATA version 14 (StataCorp LP, College Station, TX). Given the range of West African ancestry in our dataset and our goal to determine whether or not there was a relationship between ancestry and lung cancer risk, we chose quintiles instead of median or tertile for two main reasons. First, we anticipated that any relationship between ancestry with risk would be monotonic. Second, median and tertile categories were too narrow. By choosing quintiles, we were able to capture a broader spread of the data and capture individuals with low and high ancestry at the extremes. Categorical descriptions of ancestry were based on the quintiles of ancestry in each individual population and for the risk analysis, based on the quintiles of the control population.

TCGA Data

The Cancer Genome Atlas (TCGA) is supported by the National Cancer Institute and the National Human Genome Research Institute (https://cancergenome.nih.gov). We downloaded participant sequence data—that were processed by the Illumina GA Sequencer RNA-seq version 2 pipeline (Mapsplice alignment algorithm and the RNA-Seq by Expectation–Maximization (RSEM) algorithm)—to generate expression values for *PDE4D,* which were subsequently analyzed with stratification by sex and adjustment for age, smoking status, and pack-years of smoking.

Results

Demographics

We performed genetic admixture analysis on a total of 3,820 individuals. Of these, 1,407 were African Americans (controls = 949, cases = 458), and 2,413 (controls = 1,363, cases = 1,050) were European American (by self-report) (Table 1). The majority of cases in both populations were either former or current smokers. When comparing African American cases with controls, African American cases had fewer never (6.8% versus 35.2%) and former smokers (36.2% versus 42.7%), and more current smokers (57.1% versus 22.1%). Similarly, European American case and control comparisons revealed European American cases also had fewer never (9.1% versus 35.9%) and former smokers (45.5% versus 47.5%), and more current smokers (45.4% versus 16.6%). However, African Americans with lung cancer had a lower median smoking pack-year history than European American lung cancer cases, which is consistent with previous studies [29].

West African and European genetic admixture

Consistent with previous genetic studies and the history of the Transatlantic Slave Trade, self-reported African Americans had a broader distribution of genetic admixture than European Americans (Fig. 1a, b). Among African Americans, the median West African ancestry was 77% (range 0% to 98%) (Table 2). Among European Americans, the median West African ancestry was 3% (range 0% to 90%) (Table 2). African Americans had a median European ancestry of 22% (range 2% to 99%) (Table 2). European Americans had a median of 95% European ancestry (range 9% to 100%) (Table 2). Interestingly, 14 (1%) self-reported African Americans had less than 20% West African ancestry, while 91 (6.5%) self-reported African Americans had less than 50% West African ancestry. Equally notable, there were 10 (0.4%) self-reported European Americans with greater than 70% West African ancestry. It is possible that these individuals were misclassified.

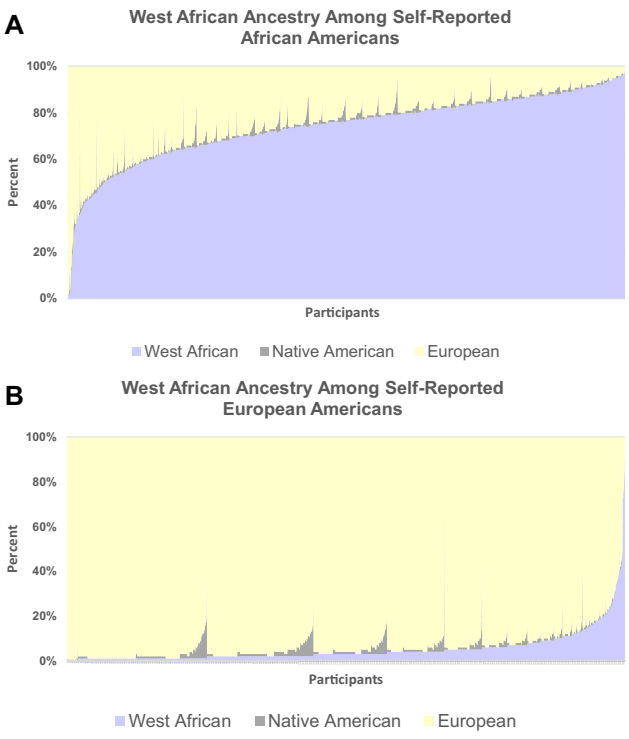

Fig. 1 Proportion of West African, European, and Native American ancestry among self-reported African Americans (**a**) and European Americans (**b**)

Table 2 Distribution of individual West African and European genetic ancestry in cases and controls by self-reported race

	West African Ancestry		European Ancestry	
	African Americans (%)	European Americans (%)	African Americans (%)	European Americans (%)
Min	0	0	2	9
Max	98	90	99	100
Mean	74	5	25	93
Median	77	3	22	95
20th percentile	64	1	13	90
40th percentile	74	2	19	94
60th percentile	80	4	25	96
80th percentile	86	6	34	98

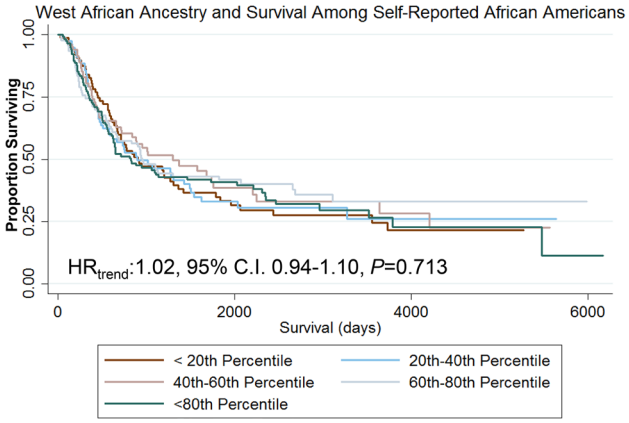

Fig. 2 Relationship between increasing proportions of genetic ancestry and lung cancer-specific survival in African Americans. Cox models adjusted for age, sex smoking, tumor stage, and tumor histology

Native American ancestry

The proportion of Native American ancestry in both African Americans and European Americans was low. The median Native American ancestry in both populations was 1%, while one European American possessed 75% Native American ancestry. Sixty-four European Americans possessed 10% or greater Native American ancestry. The maximum proportion of Native American ancestry in African Americans was 50%, while 39 had 10% or greater Native American ancestry.

Relationship between self-reported race and genetic ancestry with lung cancer survival

We first assessed whether increasing West African ancestry was associated with poor prognosis in African Americans, using quintiles of West African ancestry in African Americans as cut-points in a threshold-finding phase. As shown in Fig. 2, increasing West African ancestry was not associated with an increased risk of death (HR_{trend} 1.02, 95% CI 0.94–1.10, $p=0.71$) after adjustment for age, sex, smoking status, pack-years of smoking, stage, and histology. The relationship with survival did not change after adjustment for education and income (HR_{trend} 0.97, 95% CI 0.90–1.06, $p=0.51$). Together, these data suggest that societal, cultural, and socioeconomic factors are more relevant to survival than ancestry in African Americans. There was no difference in the relationship between genetic ancestry and survival by sex (data not shown).

Recent work on acute lymphocytic leukemia has suggested that Native American ancestry can contribute to cancer survival among European Americans [13]. To our knowledge, the contribution, if any, of Native American ancestry to lung cancer outcomes in either African Americans or European Americans has not been tested. We therefore compared outcomes between individuals with increasing proportions of Native American ancestry, using 10% as a cut-point. As shown in Supplementary Fig. 1, Native American ancestry could not be observed to contribute to survival in African Americans (HR 1.02, 95% CI 0.45–2.30, $p=0.959$) or European Americans (HR 0.87, 95% CI 0.56–1.39, $p=0.581$).

Proportions of West African Ancestry in African American cases and controls

Recent work has suggested a genetic contribution to lung cancer susceptibility and, potentially, to lung cancer racial disparities [7, 27, 30, 31]. Among African Americans, the median proportion of West African ancestry was 78% in cases and 76% in controls (Kruskal–Wallis test: $p=0.012$), suggesting that African Americans with lung cancer present with a higher proportion of West African ancestry. As the main disparity in lung cancer incidence is observed among men, we stratified this analysis by sex. Interestingly, there was no difference in median West African ancestry between cases and controls among women (78% and 79%, respectively) ($p=0.286$). However, African American men had significantly more West African ancestry in cases as compared with controls (median in cases, 79%, compared with median of 75% in controls) (Kruskal–Wallis test: $p=0.0001$) (Figs. 2, 3a and b).

We examined the association between the proportion of West African ancestry with risk of lung cancer among African Americans. As shown in Table 3, increasing West

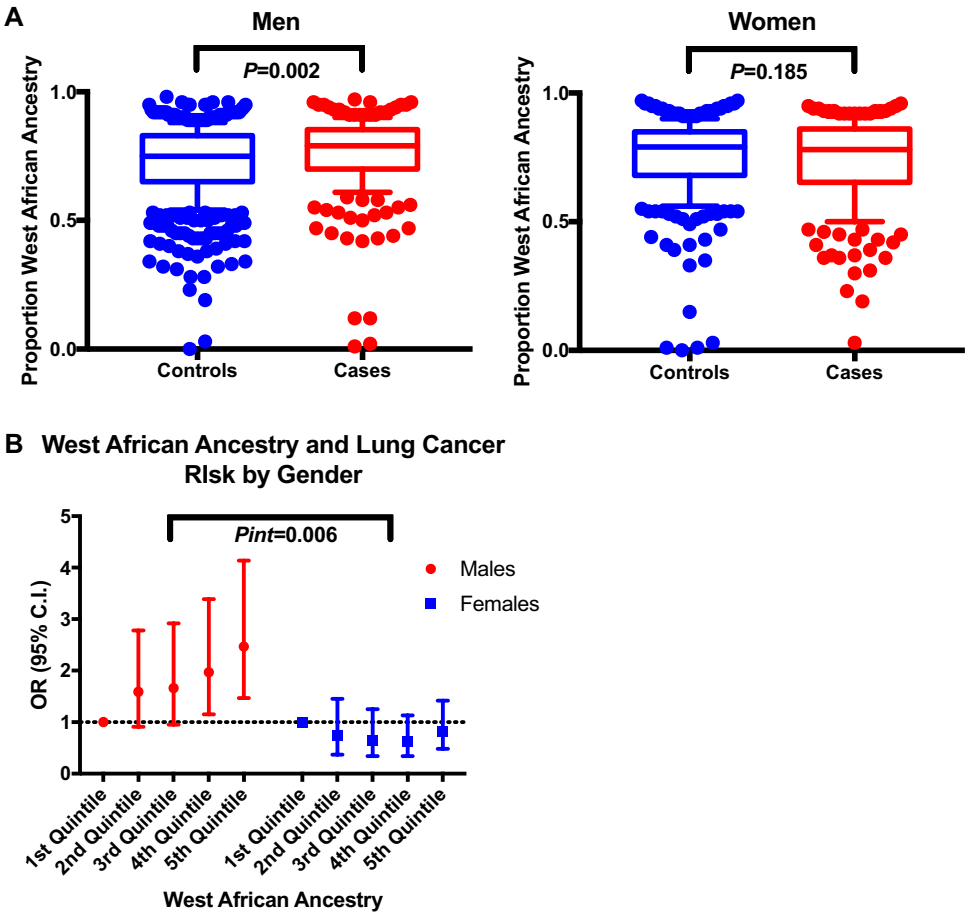

Fig. 3 Proportion of West African ancestry in African American cases and controls, stratified by sex (**a**). Evidence for effect modification of the relationship between West African ancestry and lung cancer risk by sex (**b**)

African ancestry was associated with an increased risk of lung cancer ($OR_{Q5\ vs\ Q1}$ 1.56, 95% CI 1.05–2.31, $p=0.03$). Adjustment for non-SES and SES factors included age, sex, smoking status, pack-years of smoking, education, and income (Table 3). After adjusting for sex, the association between West African ancestry was only significant in men (Males: $OR_{Q5\ vs\ Q1}$ 2.55, 95% CI 1.45–4.48, $p=0.001$) (Females: $OR_{Q5\ vs\ Q1}$ 0.90, 95% CI 0.51–1.59, $p=0.56$) (Table 3). An interaction test for West African ancestry and sex was significant ($p=0.006$) (Table 3). Our analysis also showed preliminary evidence that this relationship is stronger for lung squamous cell carcinoma (Supplementary Fig. 2). However, further adjustment of the model for crude measures of SES, such as income and education, almost nullified the result (OR_{trend} 1.11, 95% CI 0.97–1.26, $p=0.13$) (Table 3). We also stratified our analysis by smoking status (Table 4). Overall, there was no clear trend observed for current, never, and former smokers (Supplementary Table 2).

We subsequently examined the association between individual AIMs with lung cancer risk, stratified by sex. As shown in Supplementary Table 1, one of the interesting SNPs associated with risk only in African American men was rs10059859 in *PDE4D* (OR_{trend} 1.47, 95% CI 1.15–1.87, $p=0.002$), but not African American women (OR_{trend} 1.06, 95% CI 0.80–1.40, $p=0.70$). This SNP was also not associated with risk among European American men (OR_{trend} 1.19, 95% CI 0.91–1.55, $p=0.198$). We performed a stratification of the lung cancer GWAS in African Americans [27] by sex and searched for *PDE4D* SNPs in linkage with rs10059859 in both men and women. Of the 722 *PDE4D* SNPs on the array, 41 were associated with lung cancer risk in men, while just seven were significant in women (Supplementary Table 3). The most significant SNP, rs17380508, was associated with an increased risk of lung cancer in African American men (OR 5.32, 95% CI 2.19–12.93, $p<0.0001$), but not African American women (OR_{trend} 1.13, 95% CI 0.44–2.88, $p=0.67$). Using LDlink, this SNP was in moderate linkage with rs10059859. From the AIMs analysis, there were several other SNPs that demonstrated sex-specific risk associations, several of which could be investigated in future studies.

Proportions of Native American ancestry in African American cases and controls

No significant relationship between increasing Native American ancestry with risk of lung cancer could be observed among African Americans (Supplementary Table 4). We

Table 3 Association between West African ancestry with risk of lung cancer in African Americans

	Control (*n*)	Case (*n*)	OR	95% CI	p^a	OR	95% CI	p^b
All								
< 20th percentile	206	88	Referent	Referent		Referent	Referent	
≥ 20th and < 40th percentile	183	79	1.19	0.79–1.79	0.40	1.03	0.66–1.62	0.90
≥ 40th and < 60th percentile	193	84	1.04	0.69–1.55	0.86	0.82	0.52–1.28	0.37
≥ 60th and < 80th percentile	188	92	1.20	0.81–1.78	0.37	0.88	0.57–1.37	0.58
> 80th percentile	182	115	1.58	1.08–2.31	**0.02**	1.06	0.69–1.62	0.79
Trend					0.03			0.99
Men								
< 20th percentile	148	36	Referent	Referent		Referent	Referent	
≥ 20th and < 40th Percentile	138	47	1.66	0.97–2.86	0.07	1.50	0.83–2.74	0.18
≥ 40th and < 60th Percentile	133	48	1.61	0.93–2.79	0.09	1.23	0.68–2.23	0.50
≥ 60th and < 80th percentile	108	51	2.04	1.19–3.52	**0.01**	1.47	0.80–2.68	0.21
> 80th percentile	104	60	2.53	1.48–4.31	**0.001**	1.68	0.93–3.02	0.08
Trend					0.001			0.13
Women								
< 20th percentile	58	52	Referent	Referent		Referent	Referent	
≥ 20th and < 40th percentile	45	32	0.77	0.39–1.49	0.43	0.66	0.32–1.37	0.26
≥ 40th and < 60th percentile	60	36	0.62	0.33–1.17	0.14	0.52	0.26–1.06	0.07
≥ 60th and < 80th percentile	77	41	0.60	0.33–1.10	0.10	0.51	0.26–1.00	0.05
> 80th percentile	78	55	0.90	0.51–1.59	0.72	0.65	0.33–1.24	0.19
Trend					0.56			0.16

Bold determines statistical significance

OR odds ratio, *CI* confidence interval

[a]Adjusted for age, sex, smoking status, and pack-years of smoking

[b]Adjusted for age, sex, smoking status, pack-years of smoking, education, and income

stratified this analysis by sex, but no significant relationship was observed (Supplementary Table 4).

Discussion

The proportion of West African ancestry among African American male cases was higher than male controls, with no differences observed among African American female cases and controls. Indeed, we observed a relationship between increasing West African ancestry with lung cancer risk. A previous study did not find a relationship between West African ancestry with lung cancer risk, but this analysis was not stratified by sex to our knowledge [25]. Discordant results between this and previous studies could also be explained by geographic differences in the populations studied as the degree of African ancestry in African Americans has been shown to vary by both state and region [11]—our study sampled participants on the East Coast (Baltimore, MD), while a previous study included participants from the West Coast (San Francisco, California). Further, income and education levels could also vary based on geography. It is also possible that different environmental factors, such as smoking, could mediate the relationship between genetic factors and lung cancer risk in African Americans. Our analysis indicated a linear relationship between genetic ancestry and lung cancer risk, suggesting that the result is capturing some aspect of the shared environment rather than a distinct genetic event. In support of this, when we adjust the risk analysis for income and education, the relationship is considerably attenuated. However, why this phenotype is restricted to men is still unclear. While we conducted a stratified analysis by smoking status, additional studies with greater representation of African American male never smokers is needed to draw any clear conclusions. Further, histological specific analyses in the context of smoking stratification could also be informative.

There was no significant association between increased West African ancestry and lung cancer survival, confirming previous studies that suggest genetic ancestry does not adversely contribute to lung cancer outcomes and that social determinants are stronger contributors to lung cancer survival [9, 32]. This analysis supports studies that have examined lung cancer outcomes within equal access to care settings [4, 9, 33] and data showing differences in outcomes diminished when patients were given the same access to

Table 4 Association between West African ancestry with risk of lung cancer in African Americans stratified by smoking status

	Control (*n*)	Case (*n*)	OR	95% CI	p^a	OR	95% CI	p^b
Current								
< 20th percentile	40	34	Referent	Referent		Referent	Referent	
≥ 20th and < 40th percentile	35	26	0.99	0.48–2.05	0.65	0.84	0.39–1.81	0.65
≥ 40th and < 60th percentile	40	39	1.20	0.61–2.34	0.87	0.94	0.46–1.94	0.87
≥ 60th and < 80th percentile	31	51	1.97	1.01–3.85	0.18	1.65	0.79–3.45	0.18
> 80th percentile	39	46	1.56	0.80–3.02	0.78	1.11	0.54–2.27	0.78
Trend					0.04			0.31
Former								
< 20th percentile	100	49	Referent	Referent		Referent	Referent	
≥ 20th and < 40th percentile	82	48	1.44	0.84–2.47	0.19	1.17	0.64–2.14	0.61
≥ 40th and < 60th percentile	81	41	1.04	0.59–1.82	0..89	0.77	0.41–1.45	0.42
≥ 60th and < 80th percentile	88	35	0.81	0.46–1.42	0.46	0.51	0.27–0.98	0.04
> 80th percentile	79	58	1.62	0.96–2.79	0.07	1.06	0.58–1.94	0.84
Trend					0.40			0.41
Never								
< 20th percentile	66	5	Referent	Referent		Referent	Referent	
≥ 20th and < 40th percentile	66	5	1.05	0.29–3.83	0.95	1.13	0.30–4.97	0.87
≥ 40th and < 60th percentile	72	4	0.75	0.19–2.95	0.68	0.92	0.21–4.00	0.92
≥ 60th and < 80th percentile	66	6	0.99	0.28–3.48	0.99	0.85	0.20–3.57	0.82
> 80th percentile	64	11	1.80	0.58–5.61	0.31	1.31	0.36–4.81	0.68
Trend					0.30			0.77

[a]Adjusted for age, sex, smoking status, and pack-years of smoking

[b]Adjusted for age, sex, smoking status, pack-years of smoking, education, and income

OR odds ratio, *CI* confidence interval

clinical care [33]. It was interesting to note that before adjustment for SES factors such as education and income, genetic ancestry was associated with risk but not survival. One possibility for this observation is that dimensions of SES and access to care that affect lung cancer survival were already captured in the data—as in many other studies, we observe that AAs have shorter lung cancer survival because the disease is diagnosed at a later stage or access to expert medical care is lacking. In lung cancer risk however, given that there are no (major) population differences in terms of *access* to a lung cancer diagnosis (LDCT screening was not introduced for the majority of the study period), it is possible that confounding by SES and factors related to the social environment persisted in the risk model.

Previous studies have described Native American genetic ancestry as a modifier of cancer outcomes [34], drug response [13], and several other conditions [14, 35–37]. Indeed, specific subtypes of Native American ancestry were associated with a risk of lung cancer and mortality among admixed Chileans [38]. In our study, we found no relationship between Native American ancestry with lung cancer risk or survival. However, to have a sufficient power of 80% to detect the observed OR = 0.69 a sample 3 times of size would be necessary. Thus, our ability to detect associations between Native American ancestry and lung cancer among these populations was limited.

As done previously [19], we examined the association of individual AIMs markers with lung cancer risk to determine whether any of these loci are markers or regions of interest for lung cancer risk. We identified a SNP in *PDE4D* that was associated with lung cancer risk in African American men, but not women. Of note, we also tested the relationship between genetic ancestry with lung cancer risk using estimates of genetic ancestry that did not include the *PDE4D* locus, and, as shown in Supplementary Table 5, increasing West African ancestry remained associated with lung cancer risk in men. Using available data from the recently compiled GWAS of lung cancer in African Americans [27], we found further evidence for a sex-specific association between *PDE4D* and lung cancer in men. Interestingly, a previous GWAS of esophageal squamous cell carcinoma identified *PDE4D* as a risk locus, where the effect was also mainly observed among men [39], while other GWAS studies have also linked *PDE4D* with immune response [40, 41], lung spirometry [42, 43], asthma [44], and cancer [45, 46].

While not common, there are a few examples of genes/SNPs linked with sexual dimorphism in epidemiological studies [47–49], while other studies demonstrate sex differences in the relationship between West African ancestry with other traits [50]. One possible mechanism to explain such dimorphic findings is hormone control as *PDE4D* is regulated by several hormones, including testosterone. Specifically, androgen stimulation switches the promoter start site and modulates isoform production [51]. In lung cancer, we found evidence that expression of *PDE4D* could be regulated by male hormones using TCGA data (Supplementary Fig. 2). Thus, as smoking is associated with both higher testosterone levels and PDE4D expression [52, 53], our data suggest that *PDE4D* responsiveness to hormone stimulation could be one mechanism that explains the epidemiological dimorphism seen in our study. Of note, Hispanic patients with a high percentage of Native American ancestry were more likely to carry a variant of *PDE4B*, a close relative of *PDE4D*, that was strongly associated with leukemia relapse [54]. While our data suggest that this SNP could be related to sex disparities, it remains to be determined whether this SNP contributes to race/ethnicity disparities.

In summary, our data show that West African ancestry is not associated with survival among African Americans. Our data indicate that increasing West African ancestry is associated with increased risk of lung cancer among men. Given that the relationship is linear, and diminishes following adjustment for crude measures of SES, we postulate that the analysis suggests socioeconomic or aspects of the shared environment were captured in this analysis. Future studies that include sociological measures of the shared environment, beyond education status and income, are needed to disentangle this observation. We further report a second finding from our exploratory analysis of individual AIMs SNPs where we found the *PDE4D* gene associated with lung cancer risk among African American men. These findings require further follow-up and investigation. There were several other SNPs that demonstrated sex-specific risk associations, several of which could be investigated in future studies. Our genetic ancestry analysis of a large cohort of lung cancer patients in the Baltimore region of Maryland demonstrated that considerable admixture exists at the individual and population level in self-reported African Americans, and to a lesser degree among European Americans. Such ancestry estimates can be used to control for genetic heterogeneity and population substructure.

Funding This work was supported by the intramural program of the National Cancer Institute and does not have an intramural grant number.

Compliance with ethical standards

Conflict of interest There are no conflicts of interest from any author.

References

1. Siegel RL, Miller KD, Jemal A (2016) Cancer statistics, 2016. CA Cancer J Clin 66:7–30
2. Burbank F, Fraumeni JF (1972) U.S. cancer mortality: nonwhite predominance. J Natl Cancer Inst 49:649–659
3. DeSantis CE, Siegel RL, Sauer AG et al (2016) Cancer statistics for African Americans, 2016: progress and opportunities in reducing racial disparities. CA Cancer J Clin 66:290–308
4. Lathan CS, Neville BA, Earle CC (2006) The effect of race on invasive staging and surgery in non-small-cell lung cancer. J Clin Oncol 24:413–418
5. Bach PB, Cramer LD, Warren JL, Begg CB (1999) Racial differences in the treatment of early-stage lung cancer. N Engl J Med 341:1198–1205
6. Bach PB, Schrag D, Brawley OW, Galaznik A, Yakren S, Begg CB (2002) Survival of blacks and whites after a cancer diagnosis. JAMA 287:2106–2113
7. Haiman CA, Stram DO, Wilkens LR et al (2006) Ethnic and racial differences in the smoking-related risk of lung cancer. N Engl J Med 354:333–342
8. Blot WJ, Cohen SS, Aldrich M, McLaughlin JK, Hargreaves MK, Signorello LB (2011) Lung cancer risk among smokers of menthol cigarettes. J Natl Cancer Inst 103:810–816
9. Lathan CS (2015) Lung cancer care: the impact of facilities and area measures. Transl Lung Cancer Res 4:385–391
10. Tishkoff SA, Reed FA, Friedlaender FR et al (2009) The genetic structure and history of Africans and African Americans. Science 324:1035–1044
11. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL (2015) The genetic ancestry of African Americans, Latinos, and European Americans across the United States. Am J Hum Genet 96:37–53
12. Mersha TB, Abebe T (2015) Self-reported race/ethnicity in the age of genomic research: its potential impact on understanding health disparities. Hum Genomics 9:1
13. Yang JJ, Cheng C, Devidas M et al (2011) Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. Nat Genet 43:237–241
14. Fejerman L, Hu D, Huntsman S et al (2013) Genetic ancestry and risk of mortality among U.S Latinas with breast cancer. Cancer Res 73:7243–7253
15. Bress A, Kittles R, Wing C, Hooker SE, King A (2015) Genetic ancestry as an effect modifier of naltrexone in smoking cessation among African Americans: an analysis of a randomized controlled trial. Pharmacogenet Genomics. 25:305–312
16. Ruiz-Narvaez EA, Sucheston-Campbell L, Bensen JT et al (2016) Admixture mapping of African-American women in the AMBER consortium identifies new loci for breast cancer and estrogen-receptor subtypes. Front Genet 7:170
17. Ricks-Santi LJ, Apprey V, Mason T et al (2012) Identification of genetic risk associated with prostate cancer using ancestry informative markers. Prostate Cancer Prostatic Dis 15:359–364
18. Zheng YL, Loffredo CA, Yu Z et al (2003) Bleomycin-induced chromosome breaks as a risk marker for lung cancer: a case–control study with population and hospital controls. Carcinogenesis 24:269–274

19. Al-Alem U, Rauscher G, Shah E et al (2014) Association of genetic ancestry with breast cancer in ethnically diverse women from Chicago. PLoS ONE 9:e112916
20. Kosoy R, Nassir R, Tian C et al (2009) Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Hum Mutat 30:69–78
21. Nassir R, Kosoy R, Tian C et al (2009) An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels. BMC Genet 10:39
22. Collins-Schramm HE, Kittles RA, Operario DJ et al (2002) Markers that discriminate between European and African ancestry show limited variation within Africa. Hum Genet 111:566–569
23. Yang N, Li H, Criswell LA et al (2005) Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. Hum Genet 118:382–392
24. Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF (2006) A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. Am J Hum Genet 79:640–649
25. Aldrich MC, Selvin S, Hansen HM et al (2008) Comparison of statistical methods for estimating genetic admixture in a lung cancer study of African Americans and Latinos. Am J Epidemiol 168:1035–1046
26. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587
27. Zanetti KA, Wang Z, Aldrich M et al (2016) Genome-wide association study confirms lung cancer susceptibility loci on chromosomes 5p15 and 15q25 in an African-American population. Lung Cancer 98:33–42
28. Machiela MJ, Chanock SJ (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31:3555–3557
29. Holford TR, Levy DT, Meza R (2016) Comparison of smoking history patterns Among African American and white cohorts in the United States born 1890 to 1990. Nicotine Tob Res 18(Suppl 1):S16–29
30. Schwartz AG, Wenzlaff AS, Bock CH et al (2011) Admixture mapping of lung cancer in 1812 African-Americans. Carcinogenesis 32:312–317
31. Coté ML, Kardia SL, Wenzlaff AS, Ruckdeschel JC, Schwartz AG (2005) Risk of lung cancer among white and black relatives of individuals with early-onset lung cancer. JAMA 293:3036–3042
32. Jones CC, Mercaldo SF, Blume JD et al (2018) Racial disparities in lung cancer survival: the contribution of stage, treatment, and ancestry. J Thorac Oncol. 13:1464–1473
33. Zheng L, Enewold L, Zahm SH et al (2012) Lung cancer survival among black and white patients in an equal access health system. Cancer Epidemiol Biomark 21:1841–1847
34. Rocconi RP, Lankes HA, Brady WE et al (2016) The role of racial genetic admixture with endometrial cancer outcomes: an NRG Oncology/Gynecologic Oncology Group study. Gynecol Oncol 140:264–269
35. Kosoy R, Qi L, Nassir R et al (2012) Relationship between hypertension and admixture in post-menopausal African American and Hispanic American women. J Hum Hypertens 26:365–373
36. Ruiz-Narváez EA, Bare L, Arellano A, Catanese J, Campos H (2010) West African and Amerindian ancestry and risk of myocardial infarction and metabolic syndrome in the Central Valley population of Costa Rica. Hum Genet 127:629–638
37. Slattery ML, Lundgreen A, Stern MC et al (2014) The influence of genetic ancestry and ethnicity on breast cancer survival associated with genetic variation in the TGF-β-signaling pathway: the Breast Cancer Health Disparities Study. Cancer Causes Control 25:293–307
38. Lorenzo Bermejo J, Boekstegers F, Gonzalez Silos R et al (2017) Subtypes of Native American ancestry and leading causes of death: Mapuche ancestry-specific associations with gallbladder cancer risk in Chile. PLoS Genet 13:e1006756
39. Wu C, Hu Z, He Z et al (2011) Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. Nat Genet 43:679–684
40. Kennedy RB, Ovsyannikova IG, Pankratz VS, Haralambieva IH, Vierkant RA, Poland GA (2012) Genome-wide analysis of polymorphisms associated with cytokine responses in smallpox vaccine recipients. Hum Genet 131:1403–1421
41. Low SK, Chung S, Takahashi A et al (2013) Genome-wide association study of chemotherapeutic agent-induced severe neutropenia/leucopenia for patients in Biobank Japan. Cancer Sci 104:1074–1082
42. Lutz SM, Cho MH, Young K et al (2015) A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. BMC Genet 16:138
43. de Jong K, Vonk JM, Timens W et al (2015) Genome-wide interaction study of gene-by-occupational exposure and effects on FEV1 levels. J Allergy Clin Immunol 136:1664–1672.e14.
44. Himes BE, Hunninghake GM, Baurley JW et al (2009) Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. Am J Hum Genet 84:581–593
45. Michailidou K, Hall P, Gonzalez-Neira A et al (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 45:353–361, 361e1–36e1-2
46. Wu C, Wang Z, Song X et al (2014) Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations. Nat Genet 46:1001–1006
47. Robles AI, Ryan BM (2016) KRT81 miR-SNP rs3660 is associated with risk and survival of NSCLC. Ann Oncol 27:360–361
48. Dorak MT, Karpuzoglu E (2012) Gender differences in cancer susceptibility: an inadequately addressed issue. Front Genet 3:268
49. Liu LY, Schaub MA, Sirota M, Butte AJ (2012) Sex differences in disease risk from reported genome-wide association study findings. Hum Genet 131:353–364
50. Klimentidis YC, Arora A, Zhou J, Kittles R, Allison DB (2016) The Genetic contribution of West-African Ancestry to protection against central obesity in African-American men but not women: Results from the ARIC and MESA studies. Front Genet 7:89
51. Rajan P, Dalgliesh C, Carling PJ et al (2011) Identification of novel androgen-regulated pathways and mRNA isoforms through genome-wide exon-specific profiling of the LNCaP transcriptome. PLoS ONE 6:e29088
52. Zhao J, Leung JY, Lin SL, Schooling CM (2016) Cigarette smoking and testosterone in men and women: a systematic review and meta-analysis of observational studies. Prev Med 85:1–10
53. Zuo H, Han B, Poppinga WJ et al (2018) Cigarette smoke upregulates PDE3 and PDE4 to decrease cAMP in airway cells. Br J Pharmacol 175:2988–3006
54. Yang JJ, Cheng C, Meenakshi D et al (2012) Genome-wide association study identifies germline polymorphisms associated with relapse of childhood acute lymphoblastic leukemia. Blood 120(20):4197–4204