Title
Socioeconomic status and lung cancer: unraveling the contribution of genetic admixture.

Journal
American Journal of Public Health, 103(10)

ISSN
0090-0036

Authors
Aldrich, Melinda C
Selvin, Steve
Wrensch, Margaret R
et al.

Publication Date
2013-10-01

DOI
10.2105/ajph.2013.301370

Peer reviewed

Published in final edited form as:
Am J Public Health. 2013 October ; 103(10): e73–e80. doi:10.2105/AJPH.2013.301370.

Socioeconomic Status and Lung Cancer: Unraveling the Contribution of Genetic Admixture

Melinda C. Aldrich, PhD, MPH, **Steve Selvin, PhD**, **Margaret R. Wrensch, PhD, MPH**, **Jennette D. Sison, MPH**, **Helen M. Hansen, BA**, **Charles P. Quesenberry Jr., PhD**, **Michael F. Seldin, MD, PhD**, **Lisa F. Barcellos, PhD**, **Patricia A. Buffler, PhD, MPH**, and **John K. Wiencke, PhD**

Department of Thoracic Surgery and Division of Epidemiology, Vanderbilt University (MCA); Department of Medicine and Department of Neurological Surgery (MW, JS, HMH, JKW), University of California, San Francisco, CA; Division of Epidemiology (LFB, PAB) and Division of Biostatistics (SS), School of Public Health, University of California, Berkeley, CA; Division of Research, Kaiser Permanente, Oakland, CA (CPQ); Departments of Biological Chemistry and Medicine, University of California, Davis, CA (MFS)

Abstract

Objectives—We examined the relationship between genetic ancestry, socioeconomic status and lung cancer among African Americans and Latinos.

Methods—We evaluated socioeconomic status and genetic ancestry in a Northern California lung cancer case-control study (1998-2003) of African Americans and Latinos. Lung cancer cases and controls were frequency matched on age, sex and race/ethnicity. Case-control differences in individual admixture proportions were assessed using two-sample t-tests, and analysis of covariance. Logistic regression models examined associations between genetic ancestry, socioeconomic characteristics and lung cancer.

Results—Decreased Amerindian ancestry was associated with higher education among Latino controls and greater African ancestry was associated with decreased education among African

Correspondence should be sent to Melinda C. Aldrich, PhD, MPH, Department of Thoracic Surgery and Division of Epidemiology, Institute for Medicine and Public Health, Vanderbilt University, 609 Oxford House, 1313 21st Avenue South, Nashville, TN. (melinda.aldrich@vanderbilt.edu), Phone: 615-322-3386; Fax: 615-936-7003;.

About the Authors: At the time of the study, Melinda Aldrich was with the Division of Epidemiology, University of California, Berkeley, School of Public Health, Berkeley, CA. Steve Selvin is with the Division of Biostatistics, University of California, Berkeley, School of Public Health, Berkeley, CA. Margaret Wrensch, Helen Hansen and John Wiencke are with the Department of Neurological Surgery, University of California, San Francisco, CA. At the time of the study, Jennette Sison was with the Department of Neurological Surgery, University of California, San Francisco, CA. Charles Quesenberry is with the Division of Research, Kaiser Permanente, Oakland, CA. Michael Seldin is with the Departments of Biological Chemistry and Medicine, University of California, Davis, CA. Lisa Barcellos and Patricia Buffler are with the Division of Epidemiology and School of Public Health, University of California, Berkeley, CA.

Contributors: M.C. Aldrich conceptualized the analysis, conducted the statistical analysis, interpreted the data, and drafted and finalized the manuscript. S. Selvin guided the statistical analysis, interpretation of data, and provided substantive editorial input. M. R. Wrensch and J. K. Wiencke conceptualized the study, provided oversight for the study, and provided substantial editorial input. J. Sison provided administrative data support and contributed to the acquisition of the data. H. Hansen contributed to the acquisition of the data. C. P. Quesenberry provided statistical expertise and study conceptualization. M. Seldin contributed to acquisition of the data. L. F. Barcellos and P.A. Buffler provided interpretation of the data. All authors reviewed and revised the manuscript for important intellectual content and approved the final version of the manuscript.

Human Participant Protection: The study was approved by the University of California Committee for the Protection of Human Subjects. All participants provided written informed consent.

lung cancer cases. Education was associated with lung cancer in both Latinos and African Americans, independently of smoking, ancestry, age and sex. Genetic ancestry was not associated with lung cancer in African Americans.

Conclusions—Findings suggest socioeconomic factors may have a greater impact than genetic ancestry for lung cancer among African Americans. The genetic heterogeneity and recent dynamic migration and acculturation of Latinos complicate recruitment, thus epidemiologic analyses and findings should be interpreted cautiously.

Introduction

Associations between socioeconomic status and cancer incidence or mortality and accompanying racial differences are common findings across cancers and populations.[1-9] An inverse association between socioeconomic measures and lung cancer incidence and mortality is a consistent observation among populations,[7, 10-18] especially among men, although for lung cancer mortality in the U.S. this is a reversal of the pattern from earlier decades.[19] Socioeconomic measurements are also known to vary across diverse populations.[20] In the United States, African Americans and Latinos have, on average, lower education, larger household sizes, lower income and are frequently unmarried compared to whites.[21-24] Smoking is more prevalent among persons characterized by low socioeconomic factors such as low education, low income and working class occupations.[20, 25-27] Studies examining the relationship between socioeconomic status and lung cancer, or cancer in general, have used surveys and registries with large sample sizes, thereby increasing the precision of effect estimates.[7, 11, 12] However, these studies are constrained by the lack of data on important risk factors for lung cancer [11] or link aggregate socioeconomic exposure data to individual-level disease status.[6, 7, 11] Ascribing attributes of a group to an individual may not be appropriate and can result in inaccurate inferences, especially if the exposure, socioeconomic status, is misclassified.[28, 29]

Despite known disparities in lung cancer incidence[30] and consistently observed associations between socioeconomic status and both lung cancer and race/ethnicity, few studies have examined this interrelationship which is thought to result from a complex interplay of environmental, social, economic and genetic factors. Using incident cancer registry data, Krieger *et al*. observed an inverse relationship between lung cancer incidence and socioeconomic deprivation among African Americans but an increase in incidence with economic prosperity among Latinos.[31] A study examining lung cancer among Latinos found that incidence increased as income increased and the percent of Latinos residing in the census tract decreased.[32] Many studies examining socioeconomic differences in lung cancer risk suggest the increased risk cannot be fully explained by smoking, occupational or dietary exposures,[13, 15, 16, 33, 34] whereas others found controlling for several measures such as smoking,[35] dietary fat and perceived health removed associations with socioeconomic status.[17] Some studies examining racial/ethnic differences in lung cancer found ethnic differences disappeared after adjusting for socioeconomic status.[6, 7, 11] Together these findings highlight the complexities of understanding the relationship between socioeconomic status, lung cancer and race/ethnicity.

Self-reported race/ethnicity represents a combination of several factors which are genetic, social, economic and environmental.[36] Moreover, due to the ancestral heterogeneity of Latinos and African Americans, self-reported race/ethnicity does not provide precise genetic information. Recent advances in statistical tools and identification of genetic markers informative for ancestry have enabled the genetic heterogeneity of populations to be described and applied to epidemiologic studies. Genetic ancestry associations are a useful tool to suggest that a genetic component contributes to disease disparities and admixture mapping is implemented to identify genetic factors contributing to disease.[37, 38] Importantly, genetic ancestry may be associated with socioeconomic factors.[39-43] For example, Sanchez *et al*. revealed Amerindian ancestry was greater in individuals with fewer years of education.[42] Complex associations between socioeconomic status, ancestry and lung cancer require examination to disentangle their contributions to lung cancer. We examined the relationship between socioeconomic status, genetic ancestry and lung cancer in a case-control study conducted in African Americans and Latinos.

Methods

Study participants

Persons with newly diagnosed primary lung cancer residing in the five San Francisco Bay Area counties were identified from September 1998 through March 2003 by the Northern California Cancer Center (NCCC), a Surveillance, Epidemiology, and End Results cancer registry, using rapid case ascertainment methods. One hospital in the catchment area, Summit Medical Center, was not a participant in the NCCC rapid case ascertainment program; therefore, lung cancer cases diagnosed at this hospital were ascertained independently using methods comparable to those implemented by NCCC. A total of 368 cases (255 African Americans and 113 Latinos) are included in this analysis.

Population and community-based methods were used to recruit potential controls from 1) random-digit dialing (RDD), 2) Health Care Financing Administration (HCFA) records for persons aged 65 or older, and 3) community-based sources, such as churches, health fairs, and senior centers. Cases and controls were eligible for participation if they 1) self-identified as African American or Latino (using U.S. Census categories), 2) were 21 years or older, and 3) resided within the five counties of Alameda, Contra Costa, Santa Clara, San Francisco, or San Mateo. For each case, approximately twice as many eligible controls were recruited having the same age (+/- 10 years), sex, and self-identified race/ethnicity. A total of 579 controls (280 African Americans and 299 Latinos) are included in the analysis.

Eligible cases and controls were invited to participate in an in-person interview and to donate a biologic (blood or buccal swab) sample. Extensive details of case and control recruitment, including participation frequencies, are summarized elsewhere.[44, 45] Briefly, response rates for lung cancer cases that completed the full questionnaire and provided a biologic sample were 72% for African Americans and 68% for Latinos. Response rates for eligible controls were 62% for RDD controls, 21% for HCFA controls and 81% for community-based controls, with an overall participation of 55% among all eligible controls.[45] The Northern California Lung Cancer Study was developed to understand racial/ethnic differences in lung cancer susceptibility between two populations with extreme

differences in lung cancer risk. The study was approved by the University of California Committee for the Protection of Human Subjects. Written, informed consent was obtained from all participating subjects.

Interview Data Collection

Epidemiologic data were collected during in-person interviews using a structured questionnaire. Information was collected on household income, number of persons living in the household, highest degree obtained, number of years of education, marital status, and smoking history. Total household income was measured as an ordinal variable from the previous year for controls and year before diagnosis for cases. Number of people living in the household was ascertained one year before diagnosis for the cases and at time of interview for the controls. Education was collected as a categorical variable indicating highest degree obtained: less than high school, high school, technical/trade school, community college, college, and graduate school. For analyses, highest degree obtained was dichotomized into high school degree or less and post-secondary education. Total number of education years was collected and coded as reported. Current marital status was collected as a categorical variable: single, married/partner, separated, divorced, and widowed and coded as a nominal variable. Smoking level was coded as a three-level ordinal variable, capturing the amount and duration of smoking (non-smoker, < 30 pack-years, and ≥ 30 pack-years).

Specimen Processing

At the time of interview, either a blood or buccal specimen was collected. Specimens were transported to the Molecular Epidemiology Laboratory at the University of California, San Francisco within 48 hours of collection. Samples were processed for long-term storage until ready for future genotyping. When samples from all study participants were collected, DNA was isolated by automated phenol chloroform extraction using the Autogen 3000 (Autogen, Inc., Holliston, MA). DNA concentration was measured by fluorescence measurement (PicoGreen, Invitrogen Corporation, Carlsbad, CA) and normalized to 30-100ng/ul, for a total concentration of 150-500ng.

Biologic samples yielding insufficient quantities of DNA (blood, n = 2 and buccal, n = 4) underwent whole genome amplification as previously described.[46] Genotypes from genomic and whole genome amplified (WGA) DNA were called using separate clustering analyses. Genotype call rates (GenCall ≥ 0.25) averaged 99.41% and 99.16% for genomic and WGA samples, respectively. Genotype reproducibility was verified with duplicates of unamplified DNA and WGA/genomic DNA pairs. Unamplified duplicates (n = 31) had a mean reproducibility of 99.99%. WGA/genomic DNA pairs amplified from blood (n = 18 pairs) and buccal specimens (n = 28 pairs) exhibited a mean genotype reproducibility of 99.39% and 98.49%, respectively.[47]

Genotyping of Ancestry Informative Markers

Samples were genotyped at the UC Davis Genome Center using the Illumina Bead Station 500G Golden Gate™ genotyping platform and a custom SNP panel.[46] A panel of 184 autosomal biallelic ancestry informative genetic markers (AIMs) distinguishing the continental ancestral populations comprising Latinos and African Americans was genotyped

to determine the genetic ancestry of each participant as described previously.[48] Subjects were removed from statistical analyses if they self-reported more than one race/ethnicity (n = 44) or failed genotyping on the Illumina panel (n = 5) for a final sample size for analysis of 947 subjects who self-described as African American or Latino.

DNA collected from the ancestral populations, specifically Europeans (San Francisco Bay Area, US, N = 47), West Africans (Bantu and Nilo Saharan speakers, Nigeria, N = 46), and Amerindians (Mayans, Guatemala, N = 46) was also included on the genotyping platform. The mean difference ("delta") in allele frequencies between the parental populations ranged from 0.43 to 0.49.

Statistical Analysis

Subjects included in this analysis are those which self-identified as African American (N = 535) or Latino (N = 412). Individual genetic ancestry (percent European, Amerindian, and African ancestry) of African American and Latino participants was determined using 184 AIMs and a maximum likelihood method written with R statistical software (http://cran.r-project.org/) and following estimation methods described by Chakraborty *et al*.[49, 50] and Hanis *et al*.[51] Briefly, we assumed a three population model to estimate the contribution of the parental populations to the admixed population. Given the expected genotype probabilities, maximum likelihood estimates were obtained for the proportion of alleles the admixed individual received from the parental populations and summed over all loci. This method provides similar ancestry estimates to STRUCTURE, a commonly used Bayesian algorithm.[48, 52] Ancestral population allele frequencies were included to improve genetic ancestry estimates.

Analyses were conducted separately for African Americans and Latinos, as these groups are genetically and epidemiologically different. Differences in mean genetic ancestry between cases and controls by socioeconomic status were assessed using a two-sample t-test. Analysis of covariance was used to identify associations between individual admixture estimates and socioeconomic status variables and smoking, adjusting for the variables age and sex. Logistic regression models were used to estimate odds ratios assessing the relationship between lung cancer and socioeconomic status variables, adjusting for admixture, age and sex. Tests assessing the influence of admixture were not adjusted for smoking since adjusting for a collider (a variable directly influenced by two or more other variables) in a causal pathway, such as smoking (Supplemental Figure 1), can induce biased effect estimates in causal models and can change associations between its causes.[53, 54] A two-sided significance probability of 0.05 was used to infer non-random influences.

Results

African American and Latino lung cancer cases and controls did not differ according to the frequency strata-matched variables age and sex (Table 1). African American and Latino lung cancer cases and controls significantly differed by individual and household-level socioeconomic indicators. African American controls were more educated and had fewer persons living in the household compared to cases (Table 1). Latinos controls had significantly lower household income compared to Latino cases (Table 1). Among both

African Americans and Latinos, lung cancer cases smoked significantly more than controls. More Latino cases were born in the U.S. compared to Latino controls. Latino cases had significantly greater European ancestry and lower Amerindian ancestry compared to Latino controls; whereas, African American cases and controls did not differ by genetic ancestry (Table 1).

For several socioeconomic strata two-sample t-tests showed significant differences in African, European and Amerindian ancestry between Latino cases and controls (Supplemental Table 1). For all differences reaching statistical significance among Latinos, mean European genetic ancestry was significantly increased and Amerindian ancestry was decreased in cases compared to controls (Supplemental Table 1). Among African Americans, ancestry differences were small with no consistent pattern and significant for only two socioeconomic strata (Supplemental Table 2).

Analyses indicated that Latino controls with a post-secondary education had decreased Amerindian ancestry compared to controls with a high-school degree or less (36% versus 40%, respectively), adjusting for age and sex ($P = 0.03$, Table 2). African American cases without a post-secondary education had increased African ancestry compared to cases with a post-secondary education, adjusting for age and sex ($P < 0.01$, Table 3). Among both Latino and African American cases, increased smoking amount was associated with increased European ancestry ($P = 0.05$ and $P = 0.04$, respectively, data not shown).

Controlling for genetic admixture, smoking, age and sex (Table 4), having a post-secondary education was associated with significantly decreased risk of lung cancer compared to having a high school education or less (odds ratio (OR) = 0.41, 95% confidence interval (CI): 0.23 – 0.75). In Latinos, no evidence of an association was observed between household income, household size or marital status and lung cancer, adjusting for admixture, smoking, age and sex. Among Latinos, genetic admixture was significantly associated with lung cancer in all models, controlling for socioeconomic status, age and sex ($P < 0.01$) (Table 4). Smoking was significantly associated with lung cancer in all models, adjusting for genetic admixture, age, sex and each socioeconomic status variable ($P < 0.01$).

In African Americans, lung cancer was significantly associated with fewer education years (OR = 0.92, 95% CI: 0.86 – 0.98) and larger household size (OR = 1.31, 95% CI: 1.13 – 1.53), controlling for admixture, smoking, age and sex (Table 4). Income and marital status showed no evidence of an association with lung cancer in African Americans, controlling for admixture, smoking, age and sex. Genetic admixture was not associated with lung cancer among African Americans. Smoking was significantly associated with an increased risk of lung cancer in African Americans, controlling for genetic admixture, each socioeconomic status variable, age and sex ($P < 0.01$).

Discussion

Results from this analysis reveal a complex relationship between lung cancer, socioeconomic status and genetic ancestry. The genetic ancestry differences observed between Latinos and African Americans reflect their unique origins and gene flow that has

occurred over time.[36] In Latinos, European and Amerindian genetic ancestry differed between cases and controls when stratified by socioeconomic variables. Greater covariance between Amerindian ancestry and income and marital status for Latino cases versus controls suggests that income and marital status vary with ancestral population among cases but not among controls, although not statistically significant after multiple testing correction. Since cases were more likely to be US born, covariance between ancestry and socioeconomic factors may reflect acculturation within the Latino population.[21] Regression analyses consistently indicated an association between genetic ancestry and lung cancer among Latinos, adjusted for socioeconomic indicators, self-reported race, age and sex. Together these results suggest genetic factors may play an important role in lung cancer incidence among Latinos, although this finding should be interpreted with caution due to known differences in the sampling schemes and study base between cases and controls, resulting in an observed increased European ancestry among controls as previously published.[44, 48] A significant association between post-secondary education and lung cancer in Latinos remained after adjusting for genetic ancestry and smoking, suggesting increased education is associated with reduced lung cancer risk not explained by genetic ancestry or smoking. Krieger *et al*. observed the incidence of lung cancer increased with economic prosperity in Latinos in the San Francisco Bay Area and suggested that this could be due to smoking or occupational carcinogens.[31] Similarly, an increased association was observed between household income and lung cancer among Latinos, although non-significant. This association may be influenced by access to medical care since poorer Latinos are more likely to be recent immigrants with no health insurance.[55] Unlike the Krieger *et al*. study,[31] this lung cancer study benefited from available individual-level smoking data and was able to control for smoking in analyses, allowing an assessment of the influence of household income adjusted for smoking. Importantly, greater European ancestry was associated with increased smoking highlighting that ancestry may serve as a biomarker for targeted smoking cessation programs.[56] Moreover, if European ancestry is a proxy for acculturation this link may support the complex relationship between smoking, high socioeconomic status and lung cancer among Latinos.

Associations among self-reported African Americans differed from Latinos. Case and control European, African and Amerindian genetic ancestry differed little and regression analyses revealed no important associations. This does not rule out individual susceptibility based on a specific loci among African Americans as we and others have demonstrated.[57] However, both education and household size were associated with incident lung cancer in African Americans. A greater number of education years was associated with a decreased risk of lung cancer and a larger household size was associated with a increased risk of lung cancer, indicating the importance of social factors, independent of smoking or genetic ancestry.

It is possible the inconsistent associations between African Americans and Latinos are due to different genetic ancestry compositions resulting from their differing historical and societal contexts.[36, 58] Socioeconomic status consists of a constellation of environmental, occupational, behavioral and lifestyle determinants of lung cancer.[3, 59] Socioeconomic variables, such as income and household size, are dynamic and change over the lifetime of

an individual. For a chronic disease such as lung cancer, changing socioeconomic conditions over time may be relevant. However, for both African Americans and Latinos, education was inversely associated with lung cancer in this study. This variable may be more informative than others since it is relevant to each participant, is stable over time unlike occupation or income and is accurately assessed.[17] Future lung cancer studies would benefit from careful attention and a more detailed assessment of socioeconomic status, including measures of census block-group which may be relevant to racial/ethnic populations,[60] in an attempt to elucidate the association with lung cancer.

Community-based controls may be an imperfect counterfactual for the lung cancer cases potentially limiting our inferences. However, comparisons of our study controls with individuals in the target study base, demonstrate education status and household size are comparable.[45] The large percentage of the foreign-born Latinos in this study suggests future investigations should examine selection bias and acculturation. Ascertainment of cases may vary by socioeconomic status and aggressive lung cancers may be missed, however using the NCCC rapid case ascertainment we sought to identify all incident lung cancers, irrespective of socioeconomic status or lung cancer stage. An important limitation of this study is its modest sample size, limiting the statistical power. Replication of our findings is warranted in independent studies with larger sample sizes of African Americans and diverse Latino populations, which can greatly vary in their ancestry,[61] to confirm our observed associations and overcome the multiple testing present in our analyses. Exclusion of 44 multiethnic subjects may have resulted in a selection bias, however, including only those identifying as African American or Latino may better control for unmeasured factors.

Estimation of admixture makes several important assumptions yet it is difficult to know whether the assumptions hold. [51, 52] Measurement error of admixture proportions can either result in biased effect estimates or residual confounding.[48] Genetic ancestry associations are suggestive of a genetic contribution to disease; however, associations may be confounded by non-genetic factors (environmental or social) correlated with genetic ancestry and associated with disease. Genetic ancestry may also be a proxy for acculturation since it correlates with country-of-origin and birth site.[58, 62] Thus associations with genetic ancestry can be reflective of either genetic or non-genetic factors, accentuating the need for comprehensive multifaceted research.[36, 39] By incorporating several socioeconomic variables and smoking, we were able to more thoroughly control for varied socioeconomic influences.

A strength of this study is the measurement of the genetic heterogeneity of self-reported African Americans and Latinos. Reports of racial/ethnic differences in lung cancer have used self-reported ethnicity. Although this measure is a useful summary of unmeasured genetic, environmental and social factors, it does not provide information about genetic heterogeneity. Using genetic admixture in statistical models provides a concise way to examine genetic background that may be associated with particular genetic risk factors. Thus, associations between socioeconomic variables and lung cancer were controlled not only for self-reported ethnicity through matching, but also for genetic background through analyses. The number of ancestry informative markers used is a strength of this study and are sufficiently informative for estimation of ancestry since accurate estimates of ancestry proportions can be achieved with as few as 24-30 markers.[63, 64] This analysis also benefited

from explicitly stating its causal graph and using this as a roadmap to make decisions about which factors to control for as possible confounders. Our findings are similar to investigations of other outcomes emphasizing the important role of socioeconomic status (e.g. education) that may better explain disease risk relative to genetic ancestry.[41, 43, 65, 66]

In summary, African Americans and Latinos experience not only heterogeneous social and environmental exposures but are themselves genetically heterogeneous. Our findings reinforce the need to understand non-smoking environmental risk factors and genetic heritage in lung cancer. Although racial/ethnic differences in associations between socioeconomic status and lung cancer will be difficult to disentangle, future studies will benefit from a multifactorial and transdisciplinary framework to better understand the etiology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants from National Institute of Environmental Health Sciences (R01ES06717, 2R01ES09137-06); National Institute of Arthritis and Musculoskeletal and Skin Diseases (R01AR050267); and National Institute of Diabetes and Digestive and Kidney Diseases (R01K071185).

The authors thank the Northern California Cancer Center and Summit Medical Center for their assistance with case ascertainment.

References

1. Ward E, Jemal A, Cokkinides V, Singh GK, Cardinez C, Ghafoor A, et al. Cancer disparities by race/ethnicity and socioeconomic status. CA Cancer J Clin. 2004; 54(2):78–93. [PubMed: 15061598]
2. Social inequalities and cancer. IARC Sci Publ. 1997; (138):1–15.
3. Marmot MG, Kogevinas M, Elston MA. Social/economic status and disease. Annu Rev Public Health. 1987; 8:111–35. [PubMed: 3555518]
4. Singh GK, Miller BA, Hankey BF, Feuer EJ, Pickle LW. Changing area socioeconomic patterns in U.S. cancer mortality, 1950-1998: Part I--All cancers among men. J Natl Cancer Inst. 2002; 94(12): 904–15. [PubMed: 12072544]
5. Schwartz KL, Crossley-May H, Vigneau FD, Brown K, Banerjee M. Race, socioeconomic status and stage at diagnosis for five common malignancies. Cancer Causes Control. 2003; 14(8):761–6. [PubMed: 14674740]
6. McWhorter WP, Schatzkin AG, Horm JW, Brown CC. Contribution of socioeconomic status to black/white differences in cancer incidence. Cancer. 1989; 63(5):982–7. [PubMed: 2914303]
7. Baquet CR, Horm JW, Gibbs T, Greenwald P. Socioeconomic factors and cancer incidence among blacks and whites. J Natl Cancer Inst. 1991; 83(8):551–7. [PubMed: 2005640]
8. Gorey KM, Vena JE. The association of near poverty status with cancer incidence among black and white adults. J Community Health. 1995; 20(4):359–66. [PubMed: 7593741]
9. Banks J, Marmot M, Oldfield Z, Smith JP. Disease and disadvantage in the United States and in England. JAMA. 2006; 295(17):2037–45. [PubMed: 16670412]
10. Brown SM, Selvin S, Winkelstein W Jr. The association of economic status with the occurrence of lung cancer. Cancer. 1975; 36(5):1903–11. [PubMed: 1192375]
11. Devesa SS, Diamond EL. Socioeconomic and racial differences in lung cancer incidence. Am J Epidemiol. 1983; 118(6):818–31. [PubMed: 6650483]

12. Shaw C, Blakely T, Sarfati D, Fawcett J, Hill S. Varying evolution of the New Zealand lung cancer epidemic by ethnicity and socioeconomic position (1981-1999). N Z Med J. 2005; 118(1213):U1411. [PubMed: 15843840]
13. Mao Y, Hu J, Ugnat AM, Semenciw R, Fincham S. Socioeconomic status and lung cancer risk in Canada. Int J Epidemiol. 2001; 30(4):809–17. [PubMed: 11511609]
14. Smith GD, Leon D, Shipley MJ, Rose G. Socioeconomic differentials in cancer among men. Int J Epidemiol. 1991; 20(2):339–45. [PubMed: 1917232]
15. Ekberg-Aronsson M, Nilsson PM, Nilsson JA, Pehrsson K, Lofdahl CG. Socio-economic status and lung cancer risk including histologic subtyping--a longitudinal study. Lung Cancer. 2006; 51(1):21–9. [PubMed: 16337709]
16. Hein HO, Suadicani P, Gyntelberg F. Lung cancer risk and social class. The Copenhagen Male Study--17-year follow up. Dan Med Bull. 1992; 39(2):173–6. [PubMed: 1611922]
17. Braaten T, Weiderpass E, Kumle M, Lund E. Explaining the socioeconomic variation in cancer risk in the Norwegian Women and Cancer Study. Cancer Epidemiol Biomarkers Prev. 2005; 14(11 Pt 1):2591–7. [PubMed: 16284383]
18. Albano JD, Ward E, Jemal A, Anderson R, Cokkinides VE, Murray T, et al. Cancer mortality in the United States by education level and race. J Natl Cancer Inst. 2007; 99(18):1384–94. [PubMed: 17848670]
19. Singh GK, Miller BA, Hankey BF. Changing area socioeconomic patterns in U.S. cancer mortality, 1950-1998: Part II--Lung and colorectal cancers. J Natl Cancer Inst. 2002; 94(12):916–25. [PubMed: 12072545]
20. Barbeau EM, Krieger N, Soobader MJ. Working class matters: socioeconomic disadvantage, race/ethnicity, gender, and smoking in NHIS 2000. Am J Public Health. 2004; 94(2):269–78. [PubMed: 14759942]
21. Ramirez, RR.; de la Cruz, GP. The Hispanic Population in the United States: March 2002. Washington DC: U.S. Census Bureau; 2003.
22. McKinnon, J. The Black Population in the United States: March 2002. Washington DC: U.S. Census Bureau; 2003.
23. Bauman KJ, Graf NL. Educational attainment: Census 2000 Brief: US Department of Commerce, US Census Bureau 2000. 2003
24. Kreider RM, Simmons T. Marital status: Census 2000 Brief: US Department of Commerce, US Census Bureau; 2003. Oct.2003
25. Cokkinides V, Bandi P, McMahon C, Jemal A, Glynn T, Ward E. Tobacco control in the United States--recent progress and opportunities. CA Cancer J Clin. 2009; 59(6):352–65. [PubMed: 19897839]
26. Gadgeel SM, Kalemkerian GP. Racial differences in lung cancer. Cancer Metastasis Rev. 2003; 22(1):39–46. [PubMed: 12716035]
27. Centers for Disease Control and Prevention. Tobacco use among U.S. racial/ethnic minority groups--African Americans, American Indians and Alaska Natives, Asian Americans and Pacific Islanders, Hispanics. A Report of the Surgeon General. Executive summary. MMWR Recomm Rep. 1998; 47(RR-18):v–xv. 1–16. [PubMed: 9784089]
28. Brenner H, Savitz DA, Jockel KH, Greenland S. Effects of nondifferential exposure misclassification in ecologic studies. Am J Epidemiol. 1992; 135(1):85–95. [PubMed: 1736664]
29. Morgenstern H. Uses of ecologic analysis in epidemiologic research. Am J Public Health. 1982; 72(12):1336–44. [PubMed: 7137430]
30. Howlader, N.; Noone, AM.; Krapcho, M.; Neyman, N.; Aminou, R.; Altekruse, SF., et al. SEER Cancer Statistics Review, 1975-2009. Bethesda, MD: National Cancer Institute; 2012.
31. Krieger N, Quesenberry C Jr, Peng T, Horn-Ross P, Stewart S, Brown S, et al. Social class, race/ethnicity, and incidence of breast, cervix, colon, lung, and prostate cancer among Asian, Black, Hispanic, and White residents of the San Francisco Bay Area, 1988-92 (United States). Cancer Causes Control. 1999; 10(6):525–37. [PubMed: 10616822]
32. Eschbach K, Mahnken JD, Goodwin JS. Neighborhood composition and incidence of cancer among Hispanics in the United States. Cancer. 2005; 103(5):1036–44. [PubMed: 15672387]

33. van Loon AJ, Goldbohm RA, Kant IJ, Swaen GM, Kremer AM, van den Brandt PA. Socioeconomic status and lung cancer incidence in men in The Netherlands: is there a role for occupational exposure? J Epidemiol Community Health. 1997; 51(1):24–9. [PubMed: 9135784]
34. van Loon AJ, Goldbohm RA, van den Brandt PA. Lung cancer: is there an association with socioeconomic status in The Netherlands? J Epidemiol Community Health. 1995; 49(1):65–9. [PubMed: 7707009]
35. Stellman SD, Chen Y, Muscat JE, Djordjevic MV, Richie JP Jr, Lazarus P, et al. Lung cancer risk in white and black Americans. Ann Epidemiol. 2003; 13(4):294–302. [PubMed: 12684197]
36. Risch N, Burchard E, Ziv E, Tang H. Categorization of humans in biomedical research: genes, race and disease. Genome Biol. 2002; 3(7) comment 2007.
37. Cheng CY, Kao WH, Patterson N, Tandon A, Haiman CA, Harris TB, et al. Admixture mapping of 15,280 African Americans identifies obesity susceptibility loci on chromosomes 5 and X. PLoS Genet. 2009; 5(5):e1000490. [PubMed: 19461885]
38. Nalls MA, Wilson JG, Patterson NJ, Tandon A, Zmuda JM, Huntsman S, et al. Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. Am J Hum Genet. 2008; 82(1):81–7. [PubMed: 18179887]
39. Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, Karter AJ, et al. The importance of race and ethnic background in biomedical research and clinical practice. N Engl J Med. 2003; 348(12): 1170–5. [PubMed: 12646676]
40. Choudhry S, Burchard EG, Borrell LN, Tang H, Gomez I, Naqvi M, et al. Ancestry-Environment Interactions and Asthma Risk among Puerto Ricans. 2006:1088–1093.
41. Florez JC, Price AL, Campbell D, Riba L, Parra MV, Yu F, et al. Strong association of socioeconomic status with genetic ancestry in Latinos: implications for admixture studies of type 2 diabetes. Diabetologia. 2009; 52(8):1528–36. [PubMed: 19526211]
42. Sanchez E, Rasmussen A, Riba L, Acevedo-Vasquez E, Kelly JA, Langefeld CD, et al. Impact of genetic ancestry and sociodemographic status on the clinical expression of systemic lupus erythematosus in American Indian-European populations. Arthritis Rheum. 2012; 64(11):3687–94. [PubMed: 22886787]
43. Campbell DD, Parra MV, Duque C, Gallego N, Franco L, Tandon A, et al. Amerind ancestry, socioeconomic status and the genetics of type 2 diabetes in a Colombian population. PLoS ONE. 2012; 7(4):e33570. [PubMed: 22529894]
44. Wrensch MR, Miike R, Sison JD, Kelsey KT, Liu M, McMillan A, et al. CYP1A1 variants and smoking-related lung cancer in San Francisco Bay area Latinos and African Americans. International journal of cancer. 2005; 113(1):141–7.
45. Cabral DN, Napoles-Springer AM, Miike R, McMillan A, Sison JD, Wrensch MR, et al. Population- and Community-based Recruitment of African Americans and Latinos: The San Francisco Bay Area Lung Cancer Study. 2003:272–279.
46. Hansen HM, Wiemels JL, Wrensch M, Wiencke JK. DNA quantification of whole genome amplified samples for genotyping on a multiplexed bead array platform. Cancer Epidemiol Biomarkers Prev. 2007; 16(8):1686–90. [PubMed: 17684147]
47. Aldrich MC, Selvin S, Hansen HM, Barcellos LF, Wrensch MR, Sison JD, et al. CYP1A1/2 haplotypes and lung cancer and assessment of confounding by population stratification. Cancer Res. 2009; 69(6):2340–8. [PubMed: 19276377]
48. Aldrich MC, Selvin S, Hansen HM, Barcellos LF, Wrensch MR, Sison JD, et al. Comparison of statistical methods for estimating genetic admixture in a lung cancer study of African Americans and Latinos. Am J Epidemiol. 2008; 168(9):1035–46. [PubMed: 18791191]
49. Chakraborty R, Weiss KM. Frequencies of complex diseases in hybrid populations. 1986:489–503.
50. Chakraborty R. Gene admixture in human populations: Models and predictions. 1986:1–43.
51. Hanis CL, Chakraborty R, Ferrell RE, Schull WJ. Individual admixture estimates: disease associations and individual risk of diabetes and gallbladder disease among Mexican-Americans in Starr County, Texas. Am J Phys Anthropol. 1986; 70(4):433–41. [PubMed: 3766713]
52. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155(2):945–59. [PubMed: 10835412]

53. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology. 2003; 14(3):300–6. [PubMed: 12859030]
54. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999; 10(1):37–48. [PubMed: 9888278]
55. Dockterman D, Velasco G. Statistical Portrait of Hispanics in the United States 2008. 2010
56. Aldrich, MC.; Kritchevsky, SB.; Meibohm, B.; Leak, TS.; Harris, T.; Kumar, R., et al. African ancestry-tobacco interaction accelerates lung function decline among African Americans; American Society of Human Genetics Meeting; 2009; Honolulu, HI. 2009.
57. Hansen HM, Xiao Y, Rice T, Bracci PM, Wrensch MR, Sison JD, et al. Fine mapping of chromosome 15q25.1 lung cancer susceptibility in African-Americans. Hum Mol Genet. 2010; 19(18):3652–61. [PubMed: 20587604]
58. Burchard EG, Borrell LN, Choudhry S, Naqvi M, Tsai HJ, Rodriguez-Santana JR, et al. Latino populations: a unique opportunity for the study of race, genetics, and social environment in epidemiological research. Am J Public Health. 2005:2161–2168. [PubMed: 16257940]
59. Alberg AJ, Samet JM. Epidemiology of Lung Cancer. 2003:21S–49.
60. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. Annu Rev Public Health. 1997; 18:341–78. [PubMed: 9143723]
61. Salari K, Choudhry S, Tang H, Naqvi M, Lind D, Avila PC, et al. Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. Genet Epidemiol. 2005; 29(1):76–86. [PubMed: 15918156]
62. Yaeger R, Avila-Bront A, Abdul K, Nolan PC, Grann VR, Birchette MG, et al. Comparing genetic ancestry and self-described race in african americans born in the United States and in Africa. Cancer Epidemiol Biomarkers Prev. 2008; 17(6):1329–38. [PubMed: 18559547]
63. Ruiz-Narvaez EA, Rosenberg L, Wise LA, Reich D, Palmer JR. Validation of a small set of ancestral informative markers for control of population admixture in African Americans. American journal of epidemiology. 2011; 173(5):587–92. [PubMed: 21262910]
64. Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, et al. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Hum Mutat. 2009; 30(1):69–78. [PubMed: 18683858]
65. Non AL, Gravlee CC, Mulligan CJ. Education, genetic ancestry, and blood pressure in African Americans and Whites. American Journal of Public Health. 2012; 102(8):1559–65. [PubMed: 22698014]
66. Klimentidis YC, Dulin-Keita A, Casazza K, Willig AL, Allison DB, Fernandez JR. Genetic admixture, social-behavioural factors and body composition are associated with blood pressure differently by racial-ethnic group among children. Journal of human hypertension. 2012; 26(2): 98–107. [PubMed: 21248781]

Table 1

Socioeconomic, genetic ancestry, and smoking characteristics of Northern California Lung Cancer Study participants, 1998-2003

	African Americans (N=535)			Latinos (N=412)		
	Controls (N=280)	Cases (N=255)		Controls (N=299)	Cases (N=113)	
Characteristic	**n (%)**	**n (%)**	***p*[b]**	**n (%)**	**n (%)**	***p*[b]**
Age[a], years						
<45	14 (5.0)	9 (3.5)		11 (3.7)	5 (4.4)	
45-54	71 (25.4)	43 (16.9)		50 (16.7)	17 (15.0)	
55-64	78 (27.9)	90 (35.3)		52 (17.4)	29 (25.7)	
65-74	73 (26.1)	70 (27.5)		105 (35.1)	31 (27.4)	
≥ 75	44 (15.7)	43 (16.9)		81 (27.1)	31 (27.4)	
Mean (s.e.)	61.8 (0.7)	63.5 (0.7)		66.3 (0.7)	65.8 (1.1)	
Min	30	30		28	36	
Max	87	92		95	90	
Sex[a]						
Male	120 (42.9)	120 (47.1)		146 (48.8)	59 (52.2)	
Female	160 (57.1)	135 (52.9)		153 (51.2)	54 (47.8)	
Foreign born						
Yes	3 (1.1)	6 (2.4)		196 (65.6)	52 (46.0)	
No	277 (98.9)	249 (97.7)		103 (34.5)	61 (54.0)	
Household income ($)						
< 19999	158 (57.3)	136 (59.1)		140 (53.4)	35 (38.9)	
≥ 20000	118 (42.8)	94 (40.9)		122 (27.1)	55 (61.1)	
Unknown	4	25		37	23	
% Genetic ancestry (s.e.)						
European	20.09 (0.8)	18.98 (0.9)	0.35	53.99 (0.98)	60.24 (1.83)	<0.01
African	77.42 (0.8)	78.0 (0.9)	0.62	7.71 (0.43)	6.99 (0.76)	0.39
Amerindian	2.49 (0.2)	3.01 (0.23)	0.08	38.30 (0.95)	32.78 (1.74)	<0.01
Mean number of people in household[c] (s.e.)	2.1 (0.07)	2.5 (0.08)		2.6 (0.08)	2.8 (0.13)	
Mean education[d], years (s.e.)	13.6 (0.19)	12.3 (0.21)		11.0 (0.3)	10.1 (0.4)	

	African Americans (N=535)			Latinos (N=412)		
	Controls (N=280)	Cases (N=255)		Controls (N=299)	Cases (N=113)	
Characteristic	**n (%)**	**n (%)**	***p*[b]**	**n (%)**	**n (%)**	***p*[b]**
Marital status						
Single	41 (14.6)	27 (11.3)		34 (11.4)	9 (8.0)	
Married/partner	66 (23.6)	71 (29.7)		140 (47.0)	48 (42.9)	
Separated	23 (8.2)	20 (8.4)		15 (5.0)	3 (2.7)	
Divorced	95 (33.9)	61 (25.5)		50 (16.8)	20 (17.9)	
Widowed	55 (19.6)	60 (25.1)		59 (19.8)	32 (28.6)	
Unknown		16		1	1	
Smoking level						
Non-smoker	90 (32.1)	15 (6.0)		158 (53.0)	32 (28.3)	
< 30 pack-years	134 (47.9)	105 (41.7)		108 (36.2)	31 (27.4)	
≥30 pack-years	56 (20.0)	132 (52.4)		32 (10.7)	50 (44.3)	
Unknown		3		1		

[a] Frequency matched variable.

[b] *P* values are from the two-sample t-test.

[c] Subjects indicating group housing (n=2) or homeless (n=2) are not included.

[d] Number of subjects with unreported total years of education for African Americans is n=16 and for Latinos is n=5.

Table 2

Association between Amerindian ancestry and socioeconomic status indicators and smoking pack-years for Latinos in the Northern California Lung Cancer Study, 1998-2003

Model	Variable	df	F	*P*	F	*P*
			Controls		Cases	
1	Income	9	0.65	0.78	1.69	0.11
2	Post-secondary education	1	4.61	0.03	3.45	0.07
3	Household size	5	2.10	0.07	1.81	0.12
4	Marital status	4	0.21	0.93	1.13	0.35
5	Smoking level	2	0.10	0.91	3.62	0.03

F-tests are derived from analysis of covariance models, adjusted for age and sex.

Table 3

Association between African ancestry and socioeconomic status indicators and smoking pack-years for African Americans in the Northern California Lung Cancer Study, 1998-2003

Model	Variable	df	F	*P*	F	*P*
			Controls		Cases	
1	Income	9	1.21	0.29	1.37	0.20
2	Post-secondary education	1	2.98	0.09	14.70	< 0.01
3	Household size	5	0.74	0.57	0.87	0.50
4	Marital status	4	0.59	0.67	0.75	0.56
5	Smoking level	2	0.25	0.78	3.60	0.03

F-tests are derived from analysis of covariance models, adjusted for age and sex.

Table 4

Association between lung cancer, admixture, socioeconomic, and smoking variables among Latinos and African Americans participating in the Northern California Lung Cancer Study, 1998-2003

		Latinos		African Americans	
	Variable	**OR**	**95% CI**	**OR**	**95% CI**
Model 1[a]	Household income	1.09	(0.98-1.21)	1.04	(0.96-1.11)
Model 2[b]	African ancestry	1.00	(0.96-1.02)	1.01	(0.99-1.02)
	Amerindian ancestry	0.99	(0.97-1.00)	1.05	(0.99-1.11)
		Global admixture association $P < 0.01$		Global admixture association $P = 0.26$	
Model 3[a]	Post-secondary education	0.41	(0.23-0.75)	0.72	(0.49-1.05)
Model 4[b]	African ancestry	0.98	(0.95-1.02)	1.01	(0.99-1.02)
	Amerindian ancestry	0.99	(0.97-1.00)	1.05	(0.99-1.11)
		Global admixture association $P < 0.01$		Global admixture association $P = 0.28$	
Model 5[a]	Education years	0.96	(0.90-1.01)	0.92	(0.86-0.98)
Model 6[b]	African ancestry	0.98	(0.95-1.02)	1.00	(0.99-1.02)
	Amerindian ancestry	0.99	(0.97-1.00)	1.05	(0.99-1.11)
		Global admixture association $P < 0.01$		Global admixture association $P = 0.20$	
Model 7[a]	Household size[c]	1.17	(0.99-1.39)	1.31	(1.13-1.53)
Model 8[b]	African ancestry	0.99	(0.96-1.02)	1.01	(0.99-1.02)
	Amerindian ancestry	0.99	(0.97-1.00)	1.05	(0.99-1.11)
		Global admixture association $P < 0.01$		Global admixture association $P = 0.11$	
Model 9[a]	Ever married	1.74	(0.72-4.19)	1.24	(0.70-2.21)
Model 10[b]	African ancestry	0.99	(0.96-1.02)	1.01	(0.99-1.02)
	Amerindian ancestry	0.99	(0.98-1.00)	1.05	(0.99-1.11)
		Global admixture association $P < 0.01$		Global admixture association $P = 0.17$	

[a] Odds ratios are derived from logistic regression model, adjusting for age, sex, admixture, and smoking.

[b] Odds ratios are derived from logistic regression model, adjusting for age, sex and socioeconomic variable. The impact of admixture is assessed using a likelihood ratio test comparing the full model to a restricted model without the two admixture variables.

[c] Subjects reporting group housing are not included in analysis.