

UCSF

UC San Francisco Previously Published Works

Title

Germline Genetic Variants and Lung Cancer Survival in African Americans

Permalink

<https://escholarship.org/uc/item/6x97n3hv>

Journal

Cancer Epidemiology Biomarkers & Prevention, 26(8)

ISSN

1055-9965

Authors

Jones, Carissa C
Bush, William S
Crawford, Dana C
et al.

Publication Date

2017-08-01

DOI

10.1158/1055-9965.epi-16-0998

Peer reviewed



Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2017 August ; 26(8): 1288–1295. doi:
10.1158/1055-9965.EPI-16-0998.

Germline genetic variants and lung cancer survival in African Americans

Carissa C. Jones^{a,b}, William S. Bush^{c,*}, Dana C. Crawford^{c,*}, Angela S. Wenzlaff^d, Ann G. Schwartz^d, John K. Wiencke^e, Margaret R. Wrensch^{e,f,g}, William J. Blot^{h,i}, Stephen J. Chanock^j, Eric L. Grogan^{k,a}, and Melinda C. Aldrich^{a,b,h}

^aDepartment of Thoracic Surgery, Vanderbilt University Medical School, Nashville, TN

^bVanderbilt Genetics Institute, Vanderbilt University Medical School, Nashville, TN

^cDepartment of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, TN

^dKarmanos Cancer Institute, Wayne State University, Detroit, MI

^eDepartment of Neurological Surgery, University of California, San Francisco, San Francisco, CA

^fDepartment of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA

^gInstitute of Human Genetics, University of California San Francisco, San Francisco, CA

^hDivision of Epidemiology, Vanderbilt University Medical School, Nashville, TN

ⁱInternational Epidemiology Institute, Rockville, MD

^jDivision of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD

^kTennessee Valley Health System Veterans Affairs, Nashville, TN

Abstract

Background—African Americans (AAs) have the highest lung cancer mortality in the U.S. Genome-wide association studies of germline variants influencing lung cancer survival have not yet been conducted in AAs. We examined five previously reported GWAS catalog variants and explored additional genome-wide associations among AA lung cancer cases.

Methods—Incident non-small cell lung cancer cases (N=286) in the Southern Community Cohort Study were genotyped on the Illumina HumanExome BeadChip. We used Cox proportional hazards models to estimate hazard ratios (HR) and 95% confidence intervals (CI) for overall mortality. Two independent AA studies (N=316 and N=298) were used for replication.

Results—One previously reported variant, rs1878022 on 12q23.3, was significantly associated with mortality (HR=0.70, 95% CI: 0.54–0.92). Replication findings were in the same direction,

CORRESPONDENCE: Melinda C. Aldrich, 609 Oxford House, 1313 21st Avenue South, Nashville, TN 37232-4682, Phone: 615-875-2627, Fax: 615-936-7003, melinda.aldrich@vanderbilt.edu.

*Present Address: Institute for Computational Biology, Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH

although attenuated (HR=0.87 and HR=0.94). Meta-analysis had a HR of 0.83 (95% CI: 0.71–0.97). Analysis of common variants identified an association between chromosome 6q21.33 and mortality (HR=0.46, 95% CI: 0.33–0.66).

Conclusion—We identified an association between rs1878022 in *CMKLR1* and lung cancer survival. However, our results in AAs have a different direction of effect compared to a prior study in European Americans, suggesting a different genetic architecture or presence of gene-environment interactions. We also identified variants on chromosome 6 within the gene-rich HLA region, which has been previously implicated in lung cancer risk and survival.

Impact—We find evidence that inherited genetic risk factors influence lung cancer survival in AAs. Replication in additional populations is necessary to confirm potential genetic differences in lung cancer survival across populations.

Keywords

Non-small cell lung cancer; genetics; survival; African American

INTRODUCTION

Lung cancer (LC) is the leading cause of cancer-related death in the United States (1). Notably African Americans have poorer LC survival compared to whites, namely a 14% 5-year survival in African Americans compared to 17% in whites (2). This poor survival can largely be attributed to presentation at a later stage in African Americans, which is associated with reduced LC survival (3). In the clinical setting, an individual's treatment plan is tailored to the stage of LC diagnosis and can impact LC survival (4). Other factors associated with poor LC survival include low socioeconomic status (SES), smoking, older age, male sex, site of origin, tumor grade, histologic subtype and somatic mutation profile (5, 6).

Tumor mutation status is frequently incorporated into genetic medicine to guide therapeutic decisions (7, 8). However, germline variation could also influence treatment response (9, 10). Genetic association studies have successfully identified germline variants that contribute to LC survival, but genome-wide association studies (GWAS) of LC survival have been conducted only in populations of European or Asian descent (11–18). To improve the precision of approaches based on population genetic history, genetic studies in diverse racial/ethnic populations need to be conducted. Overall, African descent populations have shorter segments of linkage disequilibrium compared to European descent populations, thus examining genetic associations of disease in African American populations may allow for improved fine mapping of causal variants with potential therapeutic benefits applicable across racial/ethnic groups.

Using a prospective cohort of African Americans, we sought to validate LC survival variants previously reported in the NHGRI-EBI GWAS catalog (19), accounting for known factors influencing survival. We also sought to identify novel and potentially functional genome-wide germline variants associated with LC survival in African Americans. Though the present study is underpowered to detect genome-wide associations, this is the first study to

our knowledge to examine genome-wide germline genetic associations of LC survival in African Americans.

MATERIALS AND METHODS

Study Population

The Southern Community Cohort Study (SCCS) is a large prospective cohort study designed to examine racial disparities in cancer. Adults between the ages of 40–79 years were enrolled primarily at community health centers throughout a 12-state region across the Southeastern United States (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia) between March 2002 and September 2009. Approximately 90% of enrolled individuals agreed to donate a biologic specimen. Nearly two-thirds of participants self-identified as African American. Additional details regarding the study design and recruitment have been previously published (20, 21). The SCCS was approved by Institutional Review Boards at Vanderbilt University and Meharry Medical College. Written informed consent was obtained from all participants.

Case Identification and Mortality Assessment

Incident African American LC cases occurring in the SCCS through September 2012 were identified via linkage with the 12 state cancer registries throughout the study catchment area. Epidemiologic data such as demographic characteristics, self-reported medical history, and tobacco use, were ascertained at study enrollment by trained personnel via in-person computer-assisted interviews at community health centers or by questionnaire mailed to a random subset of the general population in the 12-state region (20, 21). Histology, stage at diagnosis, and clinical treatment were collected from state cancer registry data. LC stage was determined using the American Joint Committee on Cancers staging system and Surveillance, Epidemiology, and End Results (SEER) summary stage. Treatment information was identified from cancer registries and summarized as a 5-level design variable: no treatment, chemotherapy only, radiation only, surgery only, or multi-modality (any combination of the above therapy regimens). Participants were followed for all-cause mortality through 2013, which was assessed via linkage with the Social Security Administration and the National Death Index. Survival time (years) was defined as the time between date of diagnosis and date of death, loss to follow-up, or end of follow-up period for the study.

Genotyping and Quality Control

Individuals with an available blood or buccal cell biospecimen from which germline DNA could be extracted as of September 2012 were genotyped on the Illumina HumanExome BeadChip v1.1 (San Diego, CA). The HumanExome array contains >240,000 variants, including a panel of >3,000 Ancestry Informative Markers (AIMs) for distinguishing African and European ancestries, and >4,700 disease-associated tag markers identified from GWAS. A detailed description of the quality control (QC) process is presented in Supplemental Figure 1. Briefly, QC procedures removed individuals having mismatched genetic sex, <98% genotyping efficiency and related individuals. Autosomal variants were

excluded that had Mendelian errors, a call rate <98%, or a minor allele frequency (MAF) < 5%. QC filters were applied using PLINK and R (version 3.0.3) (22). A total of 14 samples were used to verify genotype reproducibility, with a genotype concordance of >99.99%.

Admixture Estimation

Global ancestry was estimated using the ADMIXTURE software (23). The CEU (Utah residents with ancestry from northern and western Europe) and YRI (Yoruba from Ibadan, Nigeria) HapMap populations were included as ancestral reference populations to inform the admixture estimation process. After merging with the HapMap reference populations and pruning based on linkage disequilibrium ($r^2 > 0.4$), 1,137 AIMs remained among the 2,587 AIMs in the final QC dataset for estimation of genetic admixture. Supervised ($k=2$) admixture analysis was performed to estimate the percent African (YRI) and European (CEU) ancestry for each African American individual.

Statistical Analysis

Trans-ethnic Replication of NHGRI-EBI GWAS Catalog Lung Cancer Survival Variants—Variants previously associated with LC survival and genotyped on the HumanExome array were identified from the NHGRI-EBI GWAS catalog (rs1878022, rs1209950, rs9981861, rs1656402, and rs716274, Supplemental Table 1) (19). For each additively coded variant, we ran a Cox proportional hazards model, controlling for age at diagnosis, sex, disease stage, treatment, and percent African ancestry. Statistical significance was determined using Bonferroni correction based on the number of *a priori* variants tested (five variants, $\alpha=0.05/5=0.01$). We also examined whether smoking pack-years, education status, and self-reported COPD status modified our genetic associations. All statistical analyses were conducted using the R survival package (survival, version 2.37–7).

Discovery of New Common Variants Associated with Lung Cancer Survival—Cox proportional hazards models were used to estimate hazard ratios assessing the association between common (MAF $\geq 5\%$) variants and LC survival. Genotypes were coded additively and models were adjusted for age at diagnosis, sex, disease stage, treatment, and percent African ancestry. We further adjusted our models to assess the impact of smoking, education, and self-reported COPD status on our genetic associations. Functional predictions were conducted using PolyPhen-2 and SIFT gathered from the ENSEMBL website (24, 25). A type I error rate of $\alpha=0.05/28,041$ variants = 1.78×10^{-6} was set for inference of statistical ‘significance’.

Replication Populations—Two independent African American LC study populations were available for replication: 1) the Karmanos Cancer Institute (KCI) at Wayne State University (WSU) and 2) University of California San Francisco (UCSF). Non-small cell lung cancer (NSCLC) cases from both study populations (N=316 and N=298, respectively) were previously genotyped on the Illumina HumanHap 1M Duo array (San Diego, CA) and filtered based on standard QC measures. Four variants meeting our p-value threshold for replication that were not directly genotyped were imputed with IMPUTE2 using a cosmopolitan reference population from 1000 Genomes (phase 1, b37, June 2014 release), with phasing performed using SHAPEIT (26). KCI/WSU African American LC cases were

included from three WSU studies: the Family Health Study (FHS study III), the Women's Epidemiology of Lung Disease (WELD) Study, and the EXHALE (Exploring Health, Ancestry and Lung Epidemiology) Study, and have been previously described (27). Cases were ascertained using rapid case ascertainment through the population-based Metropolitan Detroit Cancer Surveillance System, an NCI-funded SEER registry. Stage, treatment, and vital status were ascertained from the Detroit SEER registry. African Americans participating in the UCSF LC study were identified using rapid case ascertainment methods from September 1998 to March 2003 (28). Specifically, cancer histology and stage were determined using ICD-O codes abstracted from SEER data from the California Cancer Registry (CCR). The CCR also provided data for whether surgery, radiation, and chemotherapy were given to the patients, their last known vital status, and date of death or date of last contact. Stage and treatment information for both replication populations were summarized using the same algorithm as implemented within the SCCS. Percent African ancestry was estimated using genome-wide data and a supervised analysis ($k=2$) implemented with ADMIXTURE (23). A fixed effects meta-analysis of the discovery (SCCS) and replication (KCI/WSU and UCSF) cohorts was performed using METAL (29).

RESULTS

We identified a total of 336 incident African American LC cases in the SCCS cohort with an available biospecimen. After implementation of QC procedures, 286 African Americans with NSCLC were included in our study (Supplemental Figure 1). The mean follow-up time was 1.3 years (min=0.003, max=8.6), and 87% of individuals were deceased at the end of follow-up. LC cases were diagnosed at a mean age of 60 years and were primarily (60%) male (Table 1). Ninety-four percent of individuals were recruited from community health centers resulting in a primarily low SES population with 68% having an annual household income less than \$15,000, and 47% having less than a high school education. Fifty-two percent of African American LC cases in the SCCS were diagnosed with stage IV disease (Table 1, Supplemental Figure 2). Median African ancestry was 86% (Supplemental Figure 3A) among SCCS African American NSCLC cases. African American NSCLC cases in the KCI/WSU ($N=316$) and UCSF ($N=298$) studies had lower African ancestry (83% and 82%, respectively; Supplemental Figures 3B and 3C), higher SES and fewer stage IV diagnoses (26% KCI/WSU and 29% UCSF; Supplemental Table 2 and Supplemental Figure 2) than African Americans in the SCCS. The mean follow-up time for the KCI/WSU and UCSF populations was 4.0 years (min=0.3, max=13.7) and 2.8 years (min=0.2, max=9.0), respectively. Our final sample size for complete case analysis was $N=275$ in the SCCS, $N=312$ in KCI/WSU, and $N=284$ in UCSF.

NHGRI-EBI GWAS Catalog Variants

Of the variants previously associated with LC survival in the GWAS catalog, five were genotyped on the HumanExome array and passed QC (14, 16, 18). Variants rs1209950, rs9981861, rs1656402, and rs716274 were previously identified in East Asian populations, while rs1878022 was previously reported in a European descent population (Supplemental Table 1). Variants had similar allele frequencies as those observed in the 1000 Genomes Americans of African Ancestry in the Southwestern US (ASW) population (Supplemental

Table 3). We observed that the C allele for rs1878022 was significantly associated with lower mortality after correction for multiple testing (Hazard Ratio, HR = 0.70, 95% Confidence Interval (CI): 0.54–0.92), adjusting for age, sex, stage, treatment, and African ancestry (Table 2 and Supplemental Table 3). Though hazard ratios were in the same direction as previously reported for the remaining variants, confidence intervals were wide and results were not significant. Inclusion of smoking pack years, education, or COPD status in our models did not appreciably alter (<10% change) our genetic associations (Supplemental Table 4). Associations of rs1878022 in the KCI/WSU and UCSF study populations had hazard ratios in the same direction as the SCCS, although not statistically significant (KCI/WSU: HR = 0.87, 95% CI: 0.66–1.14; UCSF: HR = 0.94, 95% CI: 0.72–1.24, Table 2 and Supplemental Table 3). A fixed effects meta-analysis of the three populations revealed a summary hazard ratio of 0.83 (95% CI: 0.71–0.97, $p = 0.02$, $I^2 = 0.30$, Figure 1). Stratification of the SCCS cohort by stage revealed a significant association among African Americans with stage IV LC (HR = 0.62, 95% CI: 0.44–0.88, $p = 0.008$, Supplemental Table 5). Pooled analysis of stage IV individuals from KCI/WSU and UCSF resulted in a HR of 1.20 (95% CI: 0.85–1.69, Supplemental Table 5) and a stage-stratified meta-analysis of the three populations (N=308) found an attenuated and non-significant association among stage IV individuals (HR = 0.83, 95% CI: 0.65–1.07).

ExomeChip Common Variant Associations

We examined common variants associated with LC survival among African Americans. After QC, 28,041 common variants remained for analysis. While no variant met our threshold for statistical significance, thirteen variants had p -values $< 1.0 \times 10^{-4}$ (Table 3 and Figure 2). Of the 13 variants with the smallest p -values, seven were associated with greater all-cause mortality and six were associated with reduced all-cause mortality in additive models with increasing copies of the minor allele. All variants had similar allele frequencies to those observed in the 1000 Genomes ASW reference population (Supplemental Table 6). Four of the 13 variants (rs2072633, rs1505229, rs7502216, rs537160) were located within gene introns and an additional five (rs1133358, rs8176785, rs7626962, rs35761244, rs1639122) were exonic. A variant in the *SUN5* gene on chromosome 20q11, rs1133358, was the most strongly associated with mortality (HR = 0.61, 95% CI: 0.49–0.76) and predicted by SIFT to be deleterious to protein function, though with low confidence (Table 3 and Table 4). A peak on chromosome 6 revealed three variants (rs605203, rs2072633, rs537160, $r^2 < 0.5$) associated with reduced mortality (HR=0.46, 95% CI: 0.33–0.66; HR=0.66, 95% CI: 0.54–0.81; HR=0.61, 95% CI: 0.48–0.78, respectively). Variant rs7626962 was also predicted to have deleterious effects by SIFT (Table 4). Adding smoking pack-years, education status, and self-reported COPD status to our statistical models did not substantially alter the observed hazard ratios (Supplemental Table 7). We further examined these 13 variants in our KCI/WSU and UCSF replication populations (Table 3 and Supplemental Table 6). Within the KCI/WSU study population, the G allele at rs7302017 was associated with increased mortality (HR: 1.29, 95% CI: 1.01–1.63), while the G allele at rs605203 was associated with reduced mortality in the UCSF study population (HR: 0.55, 95% CI: 0.38–0.78). Both of these associations have the same magnitude and direction of effect as the discovery cohort.

DISCUSSION

We sought to validate LC survival variants previously identified in Asian or European populations in an incident study of African Americans. We additionally sought to identify novel common variants associated with survival in African Americans. Our evaluation of five previously identified LC survival GWAS variants identified one variant in the *CMKLR1* gene significantly associated with reduced mortality in African Americans with NSCLC. The remaining four variants were not significantly associated. The observed reduced mortality for rs1878022 is in contrast to a prior study by Wu and colleagues conducted in European Americans in which they reported that the minor allele of rs1878022 was associated with increased mortality (18). All individuals in the study by Wu *et al.* were white ever-smokers with stage III or IV NSCLC who received platinum-based chemotherapy and no surgery; whereas, in the present study of African Americans no restriction was made on stage, treatment, or smoking status for study inclusion. Of note, the frequency of the C allele in the 1000 Genomes populations differs between European and African populations (35% versus 13%). The observed allele frequency of 14% in the present study of SCCS African Americans is similar to the reported frequency (18%) in the 1000 Genomes African American (ASW) population and in the UCSF and WSU LC cases (17.4% and 18.8%, respectively). While the observed population-specific effects of rs1878022 may simply be the result of a spurious association, it could also result from differences in linkage disequilibrium (LD) patterns between Africans and Europeans as a reflection of distinct backbone haplotypes that could have different causal variants. Examination of LD structure surrounding the *CMKLR1* gene revealed differences between the YRI and CEU populations, with the CEU having larger blocks of strong LD (Supplemental Figure 4). Using LDlink (30) and data from an African (YRI) population, we find five variants are in strong LD with rs1878022 ($r^2 > 0.8$) and an additional eight variants in moderate LD ($r^2 > 0.3$ and $r^2 < 0.8$) while six variants are in moderate LD in whites (CEU) and none in strong LD. RegulomeDB (31) reveals two of the variants in moderate LD in the YRI (rs4964244 and rs4964245, $r^2 = 0.44$ and 0.38 , respectively) and one variant in the CEU (rs4964242, $r^2 = 0.55$) have high regulatory potential and are likely to affect binding. Variant rs4964242 is not in LD with rs4964244 or rs4964245 in either the CEU or YRI population ($r^2 < 0.1$), indicating that different regulatory variants could account for the observed population-specific effects. It is possible that rs1878022 is tagging a different causal variant in the YRI population than in the CEU population and thus it remains necessary to fine map the *CMKLR1* gene to determine the causal variant influencing LC survival in African Americans compared to European Americans. Given the high prevalence of rare variants, targeted nature of the HumanExome array and lack of an imputation backbone, we were unable to examine LD surrounding rs1878022 within the SCCS.

Variant rs1878022 is located within the second intron within the 5'-untranslated region of the chemokine-like receptor 1 (*CMKLR1*) gene on chromosome 12q23.3. *CMKLR1* encodes a seven transmembrane G-protein coupled receptor that has been associated with adiposity, glucose intolerance, and inflammation (32, 33). It has been shown to play a role in the immune response to cigarette smoke in murine models of chronic obstructive pulmonary disease, an established risk factor for LC risk (34). Furthermore, in response to binding of its

ligand, chemerin, *CMKLR1* has been shown to activate MAPK, ERK1/2, and Akt signaling cascades involved in cell cycle regulation (35). Examination of ENCODE data at rs1878022 identified a small peak of H3K27Ac and H3K4me1 in normal human lung fibroblasts (Supplemental Figure 5). Epigenetic markers found surrounding rs1878022 are commonly found at active regulatory elements and suggest that this variant might play a role in the regulation of *CMKLR1* activity.

While our analysis of common variants did not identify any significant variants, several variants had promising p-values for associations with mortality, including multiple protein coding variants. Three variants on chromosome 6p21.33 in weak LD (rs605203, rs2072633, and rs537160, $r^2 < 0.5$) were associated with a reduction in mortality. All three variants were non-protein coding. Variant rs605203, which had a similar hazard ratio in the UCSF population, is located 70bp upstream of the *SLC44A4* gene and downstream of *EHMT2*. Variants rs2072633 and rs537160 are located within the 29th and 19th intron of the complement factor B (*CFB*) gene, respectively. Chromosome 6p21.33 is part of the gene-rich human leukocyte antigen (HLA) region. The 6p22.1–p21.31 chromosome region has been previously associated with LC risk in both European and African Americans, though these results have inconsistently replicated (36–40). Furthermore, rs4324798 located at 6p22.1 was previously associated with increased survival in a European descent population of never-smoking small cell LC cases (41, 42). LC susceptibility within this region has been largely attributed to the *BAT3* and *MSH5* genes, which play an important role in DNA damage response and potentially response to cancer therapeutics (43–45). Our findings support an association of the 6p22.1–p21.33 region with overall survival in African American NSCLC cases.

We identified variant rs1639122 on chromosome 12, located within the chromodomain helicase DNA binding protein 4 (*CHD4*). A change from the A to C allele results in a missense mutation that changes the amino acid from glutamate to aspartate. *CHD4* is involved in nucleosome remodeling and transcriptional silencing as part of the nucleosome remodeling and deacetylation (NuRD) complex (46, 47). Additionally, *CHD4* plays a role in DNA damage response and cell cycle progression (48, 49). Somatic mutations in *CHD4* were identified in endometrial tumors and a germline protein coding variant was associated with increased risk of overall cancer, malignant lymphoma, rectal cancer, and LC, providing further evidence for a potential role of this gene in LC survival (50–52). While the majority of these common variants failed to replicate in the KCI/WSU and UCSF populations, the totality of evidence from previously published associations with LC risk and biological plausibility suggest our findings should be further investigated in additional African American LC populations.

To our knowledge, this study is the first to examine genome-wide genetic variants associated with all-cause mortality among African American NSCLC cases. While we acknowledge our sample size is small, these analyses serve as a starting point for further investigation. With our limited sample size of African American NSCLC cases we have reduced statistical power to detect true associations and thus replication in a larger population of African American NSCLC cases is necessary. However, we had *a priori* evidence to assess variants previously associated with LC survival, thus reducing the burden of multiple testing typical

of genetic association studies. Importantly, assessing trans-ethnic replication of the NHGRI-EBI GWAS Catalog LC survival variants provides genetic information about LC desperately needed among diverse racial/ethnic populations. Moreover, meta-analysis of the discovery and replication cohorts provides additional support for this association. Our findings provide greater evidence of an association between the *CMKLR1* gene and survival among LC cases.

It is important to note that the present analyses are limited by the design of the genotyping array used in the discovery population. The HumanExome array was designed to capture biologically-relevant coding variation. As a result, variants are not evenly distributed across the genome and are predominantly rare in frequency. This array also includes >4,500 variants identified in genome-wide association studies. While there were more than 20 variants associated with LC survival in the NHGRI-EBI GWAS catalog at the time of our data extraction, only five were present on the array for examination in the current study. The unique design of the HumanExome array precluded us from analyzing variants in LD with the five previously reported variants examined here. Specifically, examination of LD patterns surrounding the five variants in the relevant 1000 Genomes reference population revealed four variants in LD (± 100 kb, $r^2 > 0.6$, CEU reference population) with rs1878022, the only variant previously identified in a European descent population, and 57 variants in LD (± 100 kb, $r^2 > 0.6$, Han Chinese in Beijing, China (CHB) reference population) with one of the four variants previously identified in Asian populations. None of these variants in LD were present on the HumanExome array, preventing inclusion in the statistical analyses. Future analyses should include previously reported variants not present here and perform a thorough examination of LD structure surrounding these variants.

While the KCI/WSU and UCSF replication populations were similar, we note key differences between the discovery (SCCS) and replication study populations likely due to differences in the composition of the study base from which the LC cases arose. The discovery and the replication LC cases were all sampled using cancer registries within primary study bases; however, the study base of the SCCS is defined by its prospective cohort design of primarily individuals seeking medical care at community health centers across the Southeastern United States and KCI/WSU and UCSF are defined by geographic regions encompassing the Detroit and the San Francisco Bay Area metropolitan areas, respectively. These different geographic samplings likely contributed to important differences across the three studies in education level, current smoking prevalence, and stage distribution. Specifically, a greater percentage of SCCS LC cases were diagnosed at later stages of disease compared to KCI/WSU and UCSF cases (Supplemental Figure 2). This difference in stage distribution between discovery and replication studies could be attributed to differences in cohort vs. case-control ascertainment since case-control designs often miss the sickest individuals. The observed difference in results across the discovery and replication studies among stage IV LC cases for the association between rs1878022 and survival may be due to unique gene-environment interactions or simply a matter of chance.

In summary, this study identified several variants associated with survival in African Americans NSCLC cases and controls, a high-risk and underrepresented population in LC genetics research. By examining variants previously associated with LC survival, we

observed that rs1878022 is significantly associated with survival in African Americans. However, the direction of effect observed is in contrast to a previous study in Europeans, suggesting rs1878022 may have populations-specific effects on LC survival due to possible differing causal variants between European descent populations and African Americans due to distinct LD substructure (18). Additionally, we identified several potential novel variants, both protein coding and non-coding, associated with survival in African Americans, including a region on chromosome 6p21.33 that has been previously associated with LC risk and small cell LC survival. Future studies fine-mapping *CMKLR1* and the 6p21–22 region and conducting functional studies could lead to potential therapeutic interventions to improve LC survival, especially in African Americans.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

SOURCE OF FUNDING:

This work was supported by a Department of Defense Early Investigator Synergistic Idea Award granted to M.C. Aldrich (W81XWH-12-1-0547) and E.L. Grogan (W81XWH-12-1-0544). M.C. Aldrich was supported by a NIH/National Cancer Institute K07 CA172294. C.C. Jones was supported by the Training Program on Genetic Variation and Human Phenotypes, a NIH/NIGMS training grant (4T32GM080178-10, PI: N.J. Cox). E.L. Grogan was supported by a Veterans Affairs Career Development Award 10-024. Studies at the Karmanos Cancer Institute at Wayne State University were supported by NIH grants/contracts R01CA060691, R01CA87895, and P30CA22453, and a Department of Health and Human Services contract HHSN261201000028C to A.G. Schwartz. The funding source had no role in study design; the collection, analysis and interpretation of data; in the writing of the report; or in the design to submit the article for publication.

Data on SCCS cancer cases used in this publication were provided by the Alabama Statewide Cancer Registry; Kentucky Cancer Registry, Lexington, KY; Tennessee Department of Health, Office of Cancer Surveillance; Florida Cancer Data System; North Carolina Central Cancer Registry, North Carolina Division of Public Health; Georgia Comprehensive Cancer Registry; Louisiana Tumor Registry; Mississippi Cancer Registry; South Carolina Central Cancer Registry; Virginia Department of Health, Virginia Cancer Registry; Arkansas Department of Health, Cancer Registry, 4815 W. Markham, Little Rock, AR 72205. The Arkansas Central Cancer Registry is fully funded by a grant from National Program of Cancer Registries, Centers for Disease Control and Prevention (CDC). Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry which participates in the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC). The contents of this publication are solely the responsibility of the authors and do not represent the official views of the CDC or the Mississippi Cancer Registry.

References

1. American Cancer Society. Cancer Facts & Figures 2015. Atlanta: American Cancer Society; 2015.
2. Howlader, NNA.Krapcho, M.Garshell, J.Miller, D.Altekruse, SF.Kosary, CL.Yu, M.Ruhl, J.Tatalovich, Z.Mariotto, A.Lewis, DR.Chen, HS.Feuer, EJ., Cronin, KA., editors. SEER Cancer Statistics Review, 1975–2012. National Cancer Institute; Bethesda, MD: Apr. 2015 http://seer.cancer.gov/csr/1975_2012/, based on November 2014 SEER data submission, posted to the SEER web site
3. Efirid JT, Landrine H, Shiue KY, O'Neal WT, Podder T, Rosenman JG, et al. Race, insurance type, and stage of presentation among lung cancer patients. SpringerPlus. 2014; 3:710. [PubMed: 25674451]
4. National Comprehensive Cancer Network. National Comprehensive Cancer Network (NCCN) clinical practice guidelines in oncology: non-small cell lung cancer. V42014.
5. Islami F, Ward EM, Jacobs EJ, Ma J, Goding Sauer A, Lortet-Tieulent J, et al. Potentially preventable premature lung cancer deaths in the USA if overall population rates were reduced to

- those of educated whites in lower-risk states. *Cancer Causes & Control*. 2015; 26:409–18. [PubMed: 25555993]
6. Travis, William DEB, Konrad Muller-Hermelink, H., Harris, Curtis C., editors. *Pathology and Genetics of Tumors of the Lung, Pleura, Thymus and Heart*. Lyon: IARC Pres; 2004. World Health Organization Classification of Tumors.
 7. Pao W, Chmielecki J. Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer. *Nature Reviews Cancer*. 2010; 10:760–74. [PubMed: 20966921]
 8. Pao W, Miller VA, Politi KA, Riely GJ, Somwar R, Zakowski MF, et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Medicine*. 2005; 2:e73. [PubMed: 15737014]
 9. Bell DW, Gore I, Okimoto RA, Godin-Heymann N, Sordella R, Mulloy R, et al. Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR. *Nature Genetics*. 2005; 37:1315–6. [PubMed: 16258541]
 10. Gregorc V, Hidalgo M, Spreafico A, Cusatis G, Ludovini V, Ingersoll RG, et al. Germline polymorphisms in EGFR and survival in patients with lung cancer receiving gefitinib. *Clinical Pharmacology and Therapeutics*. 2008; 83:477–84. [PubMed: 17713473]
 11. Han JY, Lee YS, Shin ES, Hwang JA, Nam S, Hong SH, et al. A genome-wide association study of survival in small-cell lung cancer patients treated with irinotecan plus cisplatin chemotherapy. *Pharmacogenomics J*. 2014; 14:20–7. [PubMed: 23478653]
 12. Hu L, Wu C, Zhao X, Heist R, Su L, Zhao Y, et al. Genome-wide association study of prognosis in advanced non-small cell lung cancer patients receiving platinum-based chemotherapy. *Clinical Cancer Research*. 2012; 18:5507–14. [PubMed: 22872573]
 13. Lee Y, Yoon KA, Joo J, Lee D, Bae K, Han JY, et al. Prognostic implications of genetic variants in advanced non-small cell lung cancer: a genome-wide association study. *Carcinogenesis*. 2013; 34:307–13. [PubMed: 23144319]
 14. Sato Y, Yamamoto N, Kunitoh H, Ohe Y, Minami H, Laird NM, et al. Genome-wide association study on overall survival of advanced non-small cell lung cancer patients treated with carboplatin and paclitaxel. *Journal of Thoracic Oncology*. 2011; 6:132–8. [PubMed: 21079520]
 15. Tan XL, Moyer AM, Fridley BL, Schaid DJ, Niu N, Batzler AJ, et al. Genetic variation predicting cisplatin cytotoxicity associated with overall survival in lung cancer patients receiving platinum-based chemotherapy. *Clinical Cancer Research*. 2011; 17:5801–11. [PubMed: 21775533]
 16. Wu C, Xu B, Yuan P, Miao X, Liu Y, Guan Y, et al. Genome-wide interrogation identifies YAP1 variants associated with survival of small-cell lung cancer patients. *Cancer Research*. 2010; 70:9721–9. [PubMed: 21118971]
 17. Wu X, Wang L, Ye Y, Aakre JA, Pu X, Chang GC, et al. Genome-wide association study of genetic predictors of overall survival for non-small cell lung cancer in never smokers. *Cancer Research*. 2013; 73:4028–38. [PubMed: 23704207]
 18. Wu X, Ye Y, Rosell R, Amos CI, Stewart DJ, Hildebrandt MA, et al. Genome-wide association study of survival in non-small cell lung cancer patients receiving platinum-based chemotherapy. *Journal of the National Cancer Institute*. 2011; 103:817–25. [PubMed: 21483023]
 19. Burdett, T., (EBI) HPN, Hastings, E., (EBI) Hindorff, LA., (NHGRI), Junkins, HA., (NHGRI), Klemm, AK., (NHGRI), MacArthur, J., (EBI), Manolio, TA., (NHGRI), Morales, J., (EBI), Parkinson, H., (EBI), Welter, D., (EBI). [Accessed February 21, 2014] The NHGRI-EBI Catalog of published genome-wide association studies. Available from: www.ebi.ac.uk/gwas
 20. Signorello LB, Hargreaves MK, Steinwandel MD, Zheng W, Cai Q, Schlundt DG, et al. Southern community cohort study: establishing a cohort to investigate health disparities. *Journal of the National Medical Association*. 2005; 97:972–9. [PubMed: 16080667]
 21. Signorello LB, Hargreaves MK, Blot WJ. The Southern Community Cohort Study: investigating health disparities. *Journal of Health Care for the Poor and Underserved*. 2010; 21:26–37.
 22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*. 2007; 81:559–75. [PubMed: 17701901]
 23. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. 2009; 19:1655–64. [PubMed: 19648217]

24. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*. 2009; 4:1073–81. [PubMed: 19561590]
25. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nature Methods*. 2010; 7:248–9. [PubMed: 20354512]
26. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*. 2009; 5:e1000529. [PubMed: 19543373]
27. Schwartz AG, Cote ML, Wenzlaff AS, Land S, Amos CI. Racial differences in the association between SNPs on 15q25.1, smoking behavior, and risk of non-small cell lung cancer. *Journal of Thoracic Oncology*. 2009; 4:1195–201. [PubMed: 19641473]
28. 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; 113:141–7. [PubMed: 15386428]
29. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26:2190–1. [PubMed: 20616382]
30. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015; 31:3555–7. [PubMed: 26139635]
31. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research*. 2012; 22:1790–7. [PubMed: 22955989]
32. Ernst MC, Haidl ID, Zuniga LA, Dranse HJ, Rourke JL, Zabel BA, et al. Disruption of the chemokine-like receptor-1 (CMKLR1) gene is associated with reduced adiposity and glucose intolerance. *Endocrinology*. 2012; 153:672–82. [PubMed: 22186410]
33. Rama D, Esendagli G, Guc D. Expression of chemokine-like receptor 1 (CMKLR1) on J744A.1 macrophages co-cultured with fibroblast and/or tumor cells: modeling the influence of microenvironment. *Cellular Immunology*. 2011; 271:134–40. [PubMed: 21752353]
34. Demoor T, Bracke KR, Dupont LL, Plantinga M, Bondue B, Roy MO, et al. The role of ChemR23 in the induction and resolution of cigarette smoke-induced inflammation. *Journal of Immunology*. 2011; 186:5457–67.
35. Kaur J, Adya R, Tan BK, Chen J, Randeve HS. Identification of chemerin receptor (ChemR23) in human endothelial cells: chemerin-induced endothelial angiogenesis. *Biochemical and Biophysical Research Communications*. 2010; 391:1762–8. [PubMed: 20044979]
36. Bae EY, Lee SY, Kang BK, Lee EJ, Choi YY, Kang HG, et al. Replication of results of genome-wide association studies on lung cancer susceptibility loci in a Korean population. *Respirology*. 2012; 17:699–706. [PubMed: 22404340]
37. Timofeeva MN, Hung RJ, Rafnar T, Christiani DC, Field JK, Bickeboller H, et al. Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Human Molecular Genetics*. 2012; 21:4980–95. [PubMed: 22899653]
38. Walsh KM, Gorlov IP, Hansen HM, Wu X, Spitz MR, Zhang H, et al. Fine-mapping of the 5p15.33, 6p22.1–p21.31, and 15q25.1 regions identifies functional and histology-specific lung cancer susceptibility loci in African-Americans. *Cancer Epidemiol Biomarkers Prev*. 2013; 22:251–60. [PubMed: 23221128]
39. Zhang M, Hu L, Shen H, Dong J, Shu Y, Xu L, et al. Candidate variants at 6p21.33 and 6p22.1 and risk of non-small cell lung cancer in a Chinese population. *International Journal of Molecular Epidemiology and Genetics*. 2010; 1:11–8. [PubMed: 21537448]
40. Truong T, Hung RJ, Amos CI, Wu X, Bickeboller H, Rosenberger A, et al. Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. *Journal of the National Cancer Institute*. 2010; 102:959–71. [PubMed: 20548021]
41. Xun WW, Brennan P, Tjonneland A, Vogel U, Overvad K, Kaaks R, et al. Single-nucleotide polymorphisms (5p15.33, 15q25.1, 6p22.1, 6q27 and 7p15.3) and lung cancer survival in the

- European Prospective Investigation into Cancer and Nutrition (EPIC). *Mutagenesis*. 2011; 26:657–66. [PubMed: 21750227]
42. Yang P, Li Y, Jiang R, Cunningham JM, Li Y, Zhang F, et al. A rigorous and comprehensive validation: common genetic variations and lung cancer. *Cancer Epidemiol Biomarkers Prev*. 2010; 19:240–4. [PubMed: 20056643]
 43. Sasaki T, Gan EC, Wakeham A, Kornbluth S, Mak TW, Okada H. HLA-B-associated transcript 3 (Bat3)/Scythe is essential for p300-mediated acetylation of p53. *Genes & Development*. 2007; 21:848–61. [PubMed: 17403783]
 44. Tompkins JD, Wu X, Chu YL, Her C. Evidence for a direct involvement of hMSH5 in promoting ionizing radiation induced apoptosis. *Experimental Cell Research*. 2009; 315:2420–32. [PubMed: 19442657]
 45. Wu X, Xu Y, Feng K, Tompkins JD, Her C. MutS homologue hMSH5: recombinational DSB repair and non-synonymous polymorphic variants. *PloS One*. 2013; 8:e73284. [PubMed: 24023853]
 46. Tong JK, Hassig CA, Schnitzler GR, Kingston RE, Schreiber SL. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature*. 1998; 395:917–21. [PubMed: 9804427]
 47. Xue Y, Wong J, Moreno GT, Young MK, Cote J, Wang W. NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Molecular Cell*. 1998; 2:851–61. [PubMed: 9885572]
 48. Polo SE, Kaidi A, Baskcomb L, Galanty Y, Jackson SP. Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. *The EMBO Journal*. 2010; 29:3130–9. [PubMed: 20693977]
 49. Smeenk G, Wiegant WW, Vrolijk H, Solari AP, Pastink A, van Attikum H. The NuRD chromatin-remodeling complex regulates signaling and repair of DNA damage. *The Journal of Cell Biology*. 2010; 190:741–9. [PubMed: 20805320]
 50. Le Gallo M, O’Hara AJ, Rudd ML, Urlick ME, Hansen NF, O’Neil NJ, et al. Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. *Nature Genetics*. 2012; 44:1310–5. [PubMed: 23104009]
 51. Yamada M, Sato N, Ikeda S, Arai T, Sawabe M, Mori S, et al. Association of the chromodomain helicase DNA-binding protein 4 (CHD4) missense variation p.D140E with cancer: potential interaction with smoking. *Genes, Chromosomes & Cancer*. 2015; 54:122–8. [PubMed: 25407497]
 52. Zhao S, Choi M, Overton JD, Bellone S, Roque DM, Cocco E, et al. Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:2916–21. [PubMed: 23359684]

Study	HR
SCCS	0.70
WSU	0.87
UCSF	0.94
Summary	0.83

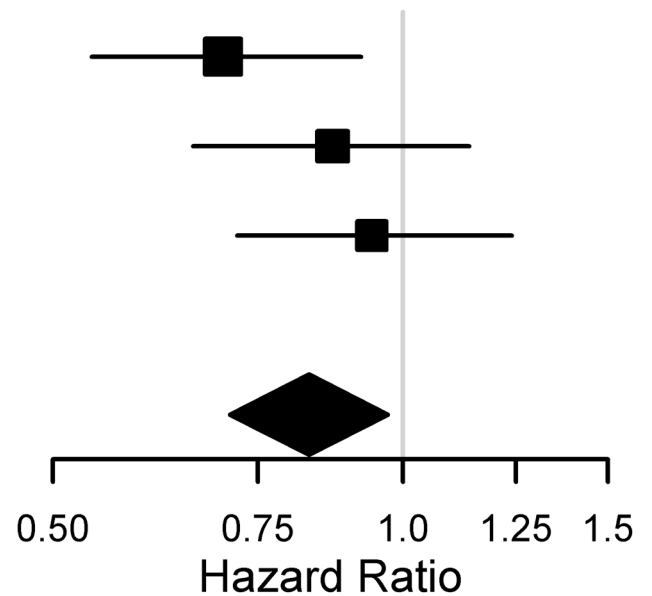
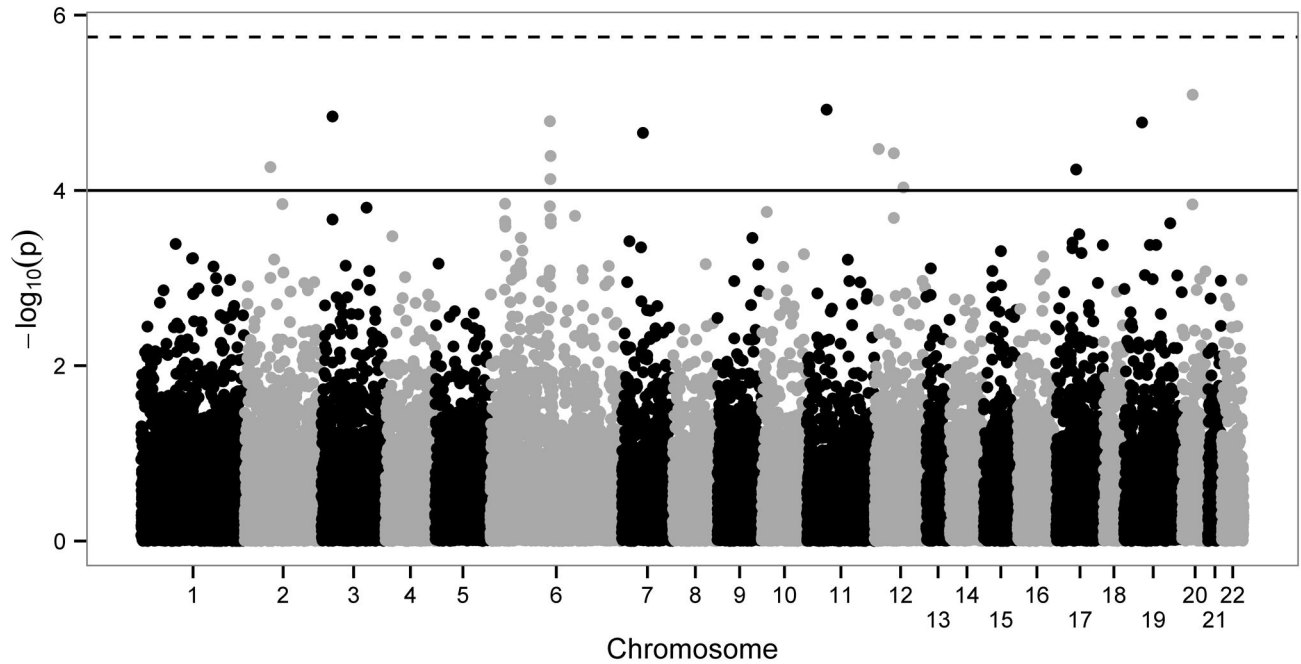


FIGURE 1.

Meta-analysis of Cox proportional hazard results for rs1878022 in the discovery (SCCS, N=275) and replication (KCI/WSU and UCSF, N=312 and 284, respectively) African American studies of non-small cell lung cancer.

$I^2 = 0.30$; measure of heterogeneity

**FIGURE 2.**

Multivariable Cox proportional hazards results for common SNPs assayed on the Illumina HumanExome BeadChip v1.1 and lung cancer survival among African Americans with NSCLC in the SCCS (N=275).

Models are adjusted for age, sex, treatment, stage, and percent African ancestry.

Solid black line represents our suggestive significance threshold (p value $< 1.0 \times 10^{-4}$).

Dashed black line represents a Bonferroni corrected significance threshold (p value $< 1.78 \times 10^{-6}$)

TABLE 1

Descriptive characteristics of African Americans with incident non-small cell lung cancer participating in the Southern Community Cohort Study (N=286).

	N (%)
Sex	
Male	171 (59.8)
Female	115 (40.2)
Vital status	
Alive	38 (13.3)
Dead	248 (86.7)
Median African ancestry, %	85.6
Lung cancer stage at diagnosis	
I	44 (15.7)
II/III	90 (32.0)
IV	147 (52.3)
Unknown	5
Treatment	
No treatment	75 (26.8)
Surgery only	31 (11.1)
Chemotherapy only	51 (18.2)
Radiation only	36 (12.9)
Multi-modality	87 (31.1)
Unknown	6
Histology	
Adenocarcinoma	113 (39.5)
Non-small cell lung cancer-NOS ^a	78 (27.3)
Squamous	72 (25.2)
Other NSCLC	22 (7.7)
Multiple histologies	1 (0.3)
Mean age at diagnosis, yr (SD)	59.6 (9.1)
Mean duration of disease among those who died, yr (SD)	0.88 (1.1)
Mean duration of disease among those alive at last follow-up, yr (SD)	4.1 (1.7)
Highest education level, yr	
<12	134 (47.2)
12	150 (52.8)
Unknown	2
Smoking status at cohort entry	
Current	206 (72.8)
Former	59 (20.8)
Never	18 (6.4)
Unknown	3
Mean cigarettes per day (SD)	15.7 (13.1)

	N (%)
Mean smoking pack-years (SD)	37.5 (30.7)
First-degree relative with lung cancer	
Yes	21 (9.3)
No	205 (90.7)
Unknown	60

^aNOS = not otherwise specified

SD = standard deviation

Yr = years

Multivariable Cox proportional hazards results for African Americans with incident non-small cell lung cancer and the five SNPs previously associated with lung cancer survival in the NHGRI GWAS catalog assayed by the Illumina HumanExome BeadChip v1.1.

TABLE 2

SNP	Chr.	bp Position ^a	Gene ^b	SCCS (N=275)			KCI/WSU (N=312)			UCSF (N=284)		
				HR	95% CI	p value	HR	95% CI	p value	HR	95% CI	p value
rs1878022	12q23.3	108699032	<i>CMKLR1</i>	0.70	(0.54–0.92)	9.8x10 ⁻³	0.87	(0.66–1.14)	0.30	0.94	(0.71–1.24)	0.66
rs1209950	21q22.2	40173528	// <i>ETS2</i>	1.16	(0.90–1.50)	0.24	0.93	(0.70–1.25)	0.65	0.91	(0.68–1.22)	0.53
rs9981861	21q22.2	41415044	<i>DSCAM</i>	1.08	(0.88–1.33)	0.46	1.13	(0.91–1.39)	0.27	1.15	(0.91–1.45)	0.24
rs1656402	2q37.1	233426526	<i>EIF4E2</i>	0.93	(0.76–1.14)	0.48	0.90	(0.72–1.11)	0.30	1.11	(0.86–1.43)	0.41
rs716274 ^c	11q22.3	103418158	<i>DYNC2H1</i> //	1.02	(0.84–1.23)	0.86	0.97	(0.80–1.17)	0.72	1.21	(0.97–1.50)	0.08

Models are adjusted for age, sex, treatment, stage, and percent African ancestry.

^a dbSNP build 137/GRCH37.p5

^b For variants outside of gene boundaries,

// denotes the location of the variant relative to the closest gene.

^c Imputed in KCI/WSU and UCSF populations. Imputation score=0.99.

HR = hazard ratio, CI = confidence interval

TABLE 3

Multivariable Cox proportional hazards results for African Americans with incident non-small cell lung cancer and common variants assayed by the Illumina HumanExome BeadChip v1.1.1. Results are shown for associations with a p value $< 1.0 \times 10^{-4}$.

SNP	Chr	bp Position ^d	Gene ^b	SCCS (N=275)			KCI/WSU (N=312)			USCF (N=284)		
				HR	95% CI	p value	HR	95% CI	p value	HR	95% CI	p value
rs1133358	20q11.21	31590686	<i>SUN5</i>	0.61	(0.49–0.76)	8.12×10^{-06}	1.01	(0.82–1.24)	0.94	1.08	(0.87–1.35)	0.49
rs8176785	11p15.1	20805286	<i>NELL1</i>	1.65	(1.32–2.07)	1.20×10^{-05}	0.98	(0.79–1.21)	0.82	1.11	(0.87–1.42)	0.39
rs7626962	3p22.2	38620907	<i>SCN5A</i>	1.95	(1.44–2.64)	1.44×10^{-05}	1.05	(0.75–1.46)	0.78	0.80	(0.53–1.22)	0.30
rs605203	6p21.33	31847012	<i>SLC44A4 // EHMT2</i>	0.46	(0.33–0.66)	1.63×10^{-05}	1.13	(0.82–1.57)	0.43	0.55	(0.38–0.79)	0.001
rs35761244 ^c	19p13.2	13873698	<i>CCDC130</i>	1.66	(1.32–2.09)	1.68×10^{-05}	0.98	(0.70–1.36)	0.90	0.94	(0.68–1.32)	0.73
rs6959964 ^c	7q11.22	68905738	<i>// AUTS2</i>	1.56	(1.27–1.91)	2.21×10^{-05}	0.84	(0.67–1.04)	0.10	0.88	(0.71–1.10)	0.27
rs1639122	12p13.31	6711147	<i>CHD4</i>	1.80	(1.36–2.38)	3.38×10^{-05}	0.93	(0.74–1.19)	0.58	1.05	(0.78–1.40)	0.76
rs7138803	12q13.12	50247468	<i>BCDIN3D // FAIM2</i>	1.70	(1.32–2.18)	3.78×10^{-05}	0.96	(0.74–1.24)	0.74	0.90	(0.69–1.19)	0.48
rs2072633	6p21.33	31919578	<i>CFB</i>	0.66	(0.54–0.81)	4.07×10^{-05}	1.00	(0.81–1.23)	0.98	0.91	(0.73–1.13)	0.40
rs1505229	2p12	77589901	<i>LRRTM4</i>	0.67	(0.56–0.82)	5.44×10^{-05}	0.96	(0.78–1.19)	0.72	1.04	(0.84–1.30)	0.72
rs7502216	17q12	36612948	<i>ARHGAP23</i>	0.67	(0.55–0.81)	5.78×10^{-05}	0.85	(0.69–1.04)	0.11	1.10	(0.89–1.36)	0.40
rs537160	6p21.33	31916400	<i>CFB</i>	0.61	(0.48–0.78)	7.44×10^{-05}	0.98	(0.74–1.29)	0.87	0.81	(0.61–1.06)	0.13
rs7302017 ^c	12q14.1	63004583	<i>MIRLET7 // PPM1H</i>	1.53	(1.24–1.89)	9.28×10^{-05}	1.29	(1.01–1.63)	0.04	0.87	(0.67–1.13)	0.30

Models are adjusted for age, sex, treatment, stage, and percent African ancestry.

^a dbSNP build 137/GRCH37.p5

^b For variants outside of gene boundaries,

// denotes the location of the variant relative to the closest gene(s).

^c Imputed in KCI/WSU and USCF populations. Imputation score=0.90 for rs35761244, =0.98 for rs6959964, and =0.99 for rs7302017.

HR = hazard ratio, CI = confidence interval

TABLE 4

Functional predictions (PolyPhen-2, SIFT) for protein coding variants with a p value $< 1.0 \times 10^{-4}$ from the common variant analysis among African Americans with incident non-small cell lung cancer participating in the Southern Community Cohort Study (N=275). Amino acid changes and functional predictions (PolyPhen-2, SIFT) provided for protein coding variants.

SNP	Gene ^a	Amino Acid Change	PolyPhen-2 Prediction	SIFT Prediction
rs1133358	<i>SUN5</i>	E > D	Benign	Deleterious
rs8176785	<i>NELL1</i>	R > Q	Benign	Tolerated
rs7626962	<i>SCN5A</i>	S > Y	Benign	Deleterious
rs35761244	<i>CCDC130</i>	C > S	Benign	Tolerated
rs1639122	<i>CHD4</i>	E > D	Benign	Tolerated