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

Kyle M. Walsh1,2, Ivan P. Gorlov5, Helen M. Hansen2, Xifeng Wu6, Margaret R. Spitz5, Huifeng Zhang7, Emily Y. Lu8, Angela S. Wenzlaff10, Jennette D. Sison2, Chongjuan Wei6, Stacy M. Lloyd8, Wei Chen7, Marsha L. Frazier6, Michael F. Seldin4, Laura J. Bierut11, Paige M. Bracci3, Margaret R. Wrensch2, Ann G. Schwartz10, John K. Wiencke2, and Christopher I. Amos7

Abstract

Background: Genome-wide association studies of European and East Asian populations have identified lung cancer susceptibility loci on chromosomes 5p15.33, 6p22.1-p21.31, and 15q25.1. We investigated whether these regions contain lung cancer susceptibility loci in African-Americans and refined previous association signals by using the reduced linkage disequilibrium observed in African-Americans.

Methods: 1,308 African-American cases and 1,241 African-American controls from 3 centers were genotyped for 780 single-nucleotide polymorphisms (SNP) spanning 3 regions, and additional SNP imputation was carried out. Associations between polymorphisms and lung cancer risk were estimated using logistic regression, stratified by tumor histology where appropriate.

Results: The strongest associations were observed on 15q25.1 in/near CHRNA5, including a missense substitution [rs16969968: OR, 1.57; 95% confidence interval (CI), 1.25–1.97; P, 1.1 × 10\(^{-5}\)] and variants in the 5'-UTR. Associations on 6p22.1-p21.31 were histology specific and included a missense variant in BAT2 associated with squamous cell carcinoma (rs2736158: OR, 0.64; 95% CI, 0.48–0.85; P, 1.82 × 10\(^{-5}\)). Associations on 5p15.33 were detected near TERT, the strongest of which was rs2735940 (OR, 0.82; 95% CI, 0.73–0.93; P, 1.1 × 10\(^{-5}\)). This association was stronger among cases with adenocarcinoma (OR, 0.75; 95% CI, 0.65–0.86; P, 8.1 × 10\(^{-5}\)).


Impact: Results implicate the BAT2, TERT, and CHRNA5 genes in the pathogenesis of specific lung cancer histologies. Cancer Epidemiol Biomarkers Prev; 22(2); 251–60. ©2012 AACR.
consistently been associated with both lung cancer and lung cancer histology. Specifically, rs2736100 in TERT has been shown to be significantly more common in cases with adenocarcinoma than with other lung cancer histologies (3). There is also evidence from GWAS that 5p15.33 is a susceptibility locus for multiple cancer sites, including: the pancreas (11), bladder (12), prostate (13), and brain (14).

The 6p22.1-p21.31 locus is part of the HLA region and is highly polymorphic. Association studies of lung cancer susceptibility have found inconsistent associations in this region, with several GWAS identifying significant associations (1–3, 6) and several follow-up replication studies failing to do so (10, 15, 16). Notably, multiple risk SNPs identified in European populations are nonpolymorphic in East Asian populations (10). Because previously associated SNPs in the HLA region may represent European-specific susceptibility loci, studies conducted in other ethnic groups should assay additional SNPs in this region to capture genetic variation not seen in Europeans.

The 15q25.1 region associated with lung cancer contains 6 genes, 3 of which (CHRNA5, CHRNA3, and CHNB4) are nicotinic receptor subunit genes. SNPs in these genes have been associated with smoking behavior, with risk alleles conferring a propensity to smoke with greater frequency and intensity (17–19). This would suggest an indirect effect of genetic variation at 15q25.1 on lung cancer risk, where variants influence cancer risk through a causal pathway in which smoking is an obligate intermediate. However, the effects of SNPs in this region on lung cancer risk appear more complex than this. Multiple studies have found that SNPs at 15q25.1 also influence lung cancer risk among never smokers (6, 20). Functional analyses have further shown that 1 of the 6 genes in this region, proteasome 20S subunit isoform 1 (PSMA4), plays a role in cancer cell proliferation and apoptosis (21). Additional functional studies identified a lung cancer risk SNP on 15q25.1 that increases IREB2 expression (22). Identifying the causal variants at 15q25.1 has been challenging because SNPs in the region are in high linkage disequilibrium (LD) among Europeans (23). Fine mapping of the associated region in African-Americans, where LD is reduced, provides an opportunity to refine the location of lung cancer risk alleles by exploiting the shorter haplotype blocks (24, 25).

To better understand the role of genetic variation in these 3 regions, we conducted a multicenter case-control study of African-Americans to identify genetic variants associated with an increased risk of lung cancer. Samples were genotyped on an Illumina Golden Gate custom panel containing SNPs selected for fine-mapping of previous GWAS hits on chromosomes 5, 6, and 15. The goals of the study were to replicate SNP associations previously detected in European populations, identify novel or African-American-specific lung cancer susceptibility SNPs in these regions, and refine previous association signals by exploiting the reduced LD observed in African-Americans.

### Materials and Methods

#### Study population

A multicenter case-control study was designed to include African-American lung cancer cases and controls from 3 collaborating institutions: The University of California, San Francisco (UCSF; 447 cases and 453 controls), Wayne State University (WSU; 459 cases and 460 controls), and the MD Anderson Cancer Center (MDA; 479 cases and 376 controls). All participating institutions received Institutional Review Board approval, and appropriate written informed consent was obtained from human subjects. All study participants reported being of African-American ethnicity.

UCSF cases and controls were enrolled as part of The Northern California Lung Cancer Study, which has been described in detail previously (26). Cases and controls older than 18 years of age were identified during 2 collection periods, spanning September 1998 to March 2003 and July 2005 to March 2008. Cases were Northern California residents presenting with previously untreated, histologically confirmed lung cancer. Cases in the first accrual period were identified primarily through the Northern California Cancer Center (NCCC) rapid case ascertainment program and the Alta Bates/Summit Hospital. Cases in the second accrual period were identified through both the NCCC and the Kaiser Permanente Medical Care Program (KPMCP). Control participants ascertained in the first accrual period were recruited through sources: random-digit dialing, Health Care Financing Administration records, and community-based recruitment. Controls in the second accrual period were recruited through the KPMCP.

Wayne State University cases were identified through the population-based Metropolitan Detroit Cancer Surveillance System, an NCI-funded SEER registry, as part of the EXHALÉ study (27). Rapid case ascertainment was used to identify histologically-confirmed cases within several months of diagnosis. African-Americans diagnosed with a first primary lung cancer from November 1, 2005 through June 30, 2010 were recruited for the study. Controls were recruited through community-based methods and were frequency matched on race, sex, and 5-year age group.

MDA cases were recruited from The University of Texas M. D. Anderson Cancer Center and the Michael E. DeBakey VA Medical Center in Houston. All cases with newly diagnosed, histopathologically confirmed lung cancer were eligible. Case-exclusion criteria were previous chemotherapy or radiotherapy or recent blood transfusion. African-American controls were recruited from Houston-area community centers and the Kelsey-Seybold Foundation, Houston’s largest multispecialty physician group practice. Controls were matched to the cases on age (±5 years), sex, and African-American ethnicity.

#### SNP selection

Three regions found in previous GWAS to be associated with an increased risk for lung cancer were selected for
fine-mapping in African-American subjects. The custom SNP panel included 120 ancestry-informative markers (AIM) for the calculation of percentage of sub-Saharan African ancestry and 760 SNPs selected to carry out fine-mapping of the 5p15.33, 6p22.1-p21.31, and 15q25.1 regions. This includes 138 SNPs on 5p15.33, 356 SNPs on 6p22.1-p21.31, and 266 SNPs on 15q25.1. Markers used in analyses were selected on the basis of: known functional effect on activity of nicotinic acetylcholine receptors, previously associated in East Asian or European populations, allele frequency greater than 0.01 in African populations, position across the region, predicted effect on function, $r^2$ value with respect to other markers less than 0.70 as determined by SNP browser version 4.0 (28), inclusion in 1 of 3 previous studies of African-American lung cancer susceptibility (24, 25, 29), or discovery by targeted sequencing (30).

**Genotyping**

UCSF samples were genotyped at the University of California, San Francisco Genome Center using an Illumina custom panel of 1,536 SNPs. Unamplified genomic DNA samples extracted from whole blood ($n = 750$) were genotyped together with whole-genome-amplified (WGA) blood or buccal DNA samples ($n = 150$), prepared as previously described (31). Genotypes for unamplified DNA and WGA DNA samples were clustered separately. A GenCall genotype quality threshold of 0.25 was used. Genotype reproducibility was verified with 12 duplicate samples with average concordance of 99.97% (range, 99.68–100%). Cepheid Trios were genotyped to assess the accuracy of assigned genotype clusters. The average heritability for 9 Parent–Parent–Child trios was 99.88% (99.19–100%). All cluster plots were visually inspected.

Wayne State University samples were genotyped at the Applied Genomics Technology Center (AGTC) at Wayne State University using the same Illumina Golden Gate Custom panel of 1,536 SNPs. All WSU samples were unamplified genomic DNA, extracted from whole blood. Genotype reproducibility was verified with 30 duplicate samples, each with more than 99% concordance. Ten CEPH controls were genotyped and checked for concordance with published HapMap SNP genotypes at loci overlapping with those assayed by the Illumina custom panel. MDA samples were genotyped at the MD Anderson Cancer Center using the same Illumina Golden Gate Custom panel as the other sites. Genotyping was carried out on unamplified genomic DNA derived from peripheral whole blood. Genotype reproducibility was verified with 7 duplicate samples, with concordance ranging from 99.93% to 100%.

For all study sites, samples with genotyping call-rates less than 95% were excluded from analysis. SNPs with genotyping call rates less than 95% in more than 1 study site were excluded from analyses. To exclude poorly genotyped SNPs, any SNP with a Hardy–Weinberg Equilibrium (HWE) $P$ value less than 1.0 x $10^{-4}$ in controls, stratified by site, was removed from analysis. All SNP quality control was carried out using Plink v1.07 (32).

**Calculation of % sub-Saharan African ancestry**

The genetic structure of African-American subjects was evaluated using Structure v2.3.1 to estimate percentage of membership in 3 distinct founder populations: sub-Saharan African, European, and East Asian (33), where East Asian population ancestry was also used as a surrogate for Amerindian descent. Founder population allele

---

**Table 1. Demographic, smoking, and tumor histology characteristics of African-American lung cancer cases and controls appearing in the association analyses**

<table>
<thead>
<tr>
<th>Sample size</th>
<th>UCSF ($n = 879$)</th>
<th>MDA ($n = 839$)</th>
<th>WSU ($n = 831$)</th>
<th>Total ($n = 2549$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>Control</td>
<td>Case</td>
<td>Control</td>
<td>Case</td>
</tr>
<tr>
<td>Sample size</td>
<td>432</td>
<td>447</td>
<td>473</td>
<td>366</td>
</tr>
<tr>
<td>Mean ± SE</td>
<td>64.36 ± 0.53</td>
<td>64.17 ± 0.54</td>
<td>64.20 ± 0.47</td>
<td>55.41 ± 0.59</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% African Ancestry</td>
<td>0.76 ± 0.007</td>
<td>0.76 ± 0.006</td>
<td>0.76 ± 0.007</td>
<td>0.76 ± 0.007</td>
</tr>
<tr>
<td>% European Ancestry</td>
<td>0.19 ± 0.006</td>
<td>0.19 ± 0.007</td>
<td>0.18 ± 0.006</td>
<td>0.18 ± 0.006</td>
</tr>
<tr>
<td>Pack-years smoked*</td>
<td>33.39 ± 1.15</td>
<td>27.39 ± 1.74</td>
<td>39.92 ± 1.37</td>
<td>24.66 ± 1.32</td>
</tr>
<tr>
<td>n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>196 (45.37)</td>
<td>204 (46.64)</td>
<td>251 (63.07)</td>
<td>154 (42.08)</td>
</tr>
<tr>
<td>Ever smoker</td>
<td>403 (93.29)</td>
<td>289 (64.65)</td>
<td>397 (83.93)</td>
<td>270 (73.77)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>181 (41.90)</td>
<td>134 (29.98)</td>
<td>246 (62.01)</td>
<td>143 (39.07)</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>176 (40.74)</td>
<td>NA</td>
<td>215 (45.45)</td>
<td>NA</td>
</tr>
<tr>
<td>Squamous cell</td>
<td>107 (24.77)</td>
<td>NA</td>
<td>131 (27.78)</td>
<td>NA</td>
</tr>
<tr>
<td>Small cell</td>
<td>34 (7.87)</td>
<td>NA</td>
<td>24 (5.07)</td>
<td>NA</td>
</tr>
<tr>
<td>Other/not specified</td>
<td>115 (26.62)</td>
<td>NA</td>
<td>103 (21.78)</td>
<td>NA</td>
</tr>
</tbody>
</table>

*Among ever smokers

Abbreviations: UCSF, University of California San Francisco; MDA, M.D. Anderson Cancer Center; WSU, Wayne State University.
frequencies were defined using SNP data from 102 unlinked \( (r^2 < 0.20) \) ancestry informative markers (AIM), genotyped in 502 unrelated HapMap individuals (167 Yoruban Africans, 165 Europeans, 84 Chinese, and 86 Japanese; ref. 34). These same AIMs were genotyped in our study participants for use with the Structure program.

**Collection of covariates**

Cases and controls recruited at all study sites completed interviews conducted by trained interviewers. Data on sex, age, and smoking behaviors were collected as covariates for this study. "Never smokers" were defined as those who had smoked less than 100 cigarettes in their lifetimes; "former smokers" were those who had quit smoking more than 1 year before diagnosis (cases) or interview (controls); "current smokers" included those who had quit smoking within the past 12 months. Pack-years for former and current smokers were calculated as the years smoked times the average number of cigarettes per day divided by 20.

Cancer histology was determined using ICD-O codes abstracted from Surveillance Epidemiology and End Results (SEER) data from the California Cancer Registry (UCSF cases) or Detroit Cancer Registry (WSU cases). For MDA cases, histology was determined by extraction from medical records. The following ICD-O groupings were made: adenocarcinoma (ICD-O: 8140, 8230, 8250–8255, 8260, 8310, 8333, 8470, 8480, 8481, 8490, and 8550), squamous cell carcinoma (8052, 8070–8073, 8083, and 8084), and small cell carcinoma (8041–8045).

**SNP imputation and association analyses**

To refine association peaks from case-control analyses and identify additional risk variants, we carried out SNP imputation in the 5p15.33, 6p22.1-p21.31, and 15q25.1 regions using the Impute2 v2.1.2 software and its standard Markov chain Monte Carlo algorithm using the default settings for targeted imputation (35). All 1,000 Genomes Phase I interim release haplotypes were provided as the reference haplotype panel (36). Using a cosmopolitan set of reference haplotypes is currently recommended for imputation, and is especially critical for imputation of recently admixed populations, such as African-Americans (37).

SNPs with imputation quality (info) scores less than 0.70 or posterior probabilities less than 0.90 were excluded to remove poorly imputed SNPs. Any SNP with a minor allele frequency of less than 1% in case subjects was excluded from association tests. All association statistics, both for imputed and for directly genotyped SNPs, were calculated using logistic regression in SNPTEST v2 assuming a log-additive model. A missing-data likelihood score-test was applied to the imputed variants, as this produces standard errors which account for the additional uncertainty inherent in the analysis of imputed genotypes. The effect of individual SNPs on lung cancer risk was calculated while adjusting for sex, age, % African ancestry, % European ancestry, and number of pack-years smoked. Adjustment for current smoking did not affect associations and this covariate was not included in the final models.
Because the 15q25.1 region contains nicotinic receptor gene variants that have been associated with smoking behavior and, therefore, SNPs in this region may influence lung cancer risk by modifying smoking behavior, these SNP associations were also calculated without adjustment for pack-years. All SNP associations were assessed in the full case-control sample, and also stratified by histology (adenocarcinoma vs. controls, squamous cell carcinoma vs. controls, and small-cell lung cancer vs. controls). All associations are for an allelic additive model, adjusted for the indicated covariates, where odds ratios (OR) are for each additional copy of the minor allele.

**Haplotype analyses**

Haplotype analyses were carried out using Haplovew v4.2 to identify haplotype blocks and calculate $r^2$ values (38). Associations between the identified haplotypes and lung cancer risk were calculated in Plink while controlling for appropriate covariates.

**Results**

After excluding samples with call rates of less than 95%, 2,549 African-American participants remained for analysis (1,308 cases and 1,241 controls). Compared with controls, cases were more likely to be male, older, smoke, and have smoked a greater number of pack-years (Table 1). Cases and controls had similar African ancestry and similar Caucasian ancestry (Table 1; Supplementary Fig. S1). After excluding monomorphic SNPs and SNPs that did not meet call-rate or HWE thresholds, 660 genotyped SNPs remained for analysis. This included 111 SNPs on 5p15.33, 320 SNPs on 6p22.1-p21.31, and 242 SNPs on 15q25.1. An additional 103,928 SNPs were imputed in the 3 regions, of which 38.5% were excluded for having a MAF less than 1% in cases, 20.7% were excluded for imputation quality (info) scores less than 0.70, and 4.6% were excluded for posterior probabilities less than 0.90. This left 37,660 imputed SNPs for analysis.
Table 2. Top 10 directly genotyped SNP associations from the case-control analysis of lung cancer in African-Americans

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position</th>
<th>Gene</th>
<th>Allele</th>
<th>MAF</th>
<th>OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2036527</td>
<td>Chr15:78851615</td>
<td>6.3 kb upstream of CHRNA5</td>
<td>C/T</td>
<td>0.2102</td>
<td>1.34 (1.17–1.54)</td>
<td>2.0 × 10^{-5}</td>
</tr>
<tr>
<td>rs17486278</td>
<td>Chr15:78867482</td>
<td>Intron 1 of CHRNA5</td>
<td>A/C</td>
<td>0.2802</td>
<td>1.31 (1.15–1.48)</td>
<td>2.7 × 10^{-5}</td>
</tr>
<tr>
<td>rs16996968</td>
<td>Chr15:78882925</td>
<td>Exon 5 of CHRNA5 (missense Asp → Asn)</td>
<td>G/A</td>
<td>0.06008</td>
<td>1.57 (1.25–1.97)</td>
<td>1.1 × 10^{-4}</td>
</tr>
<tr>
<td>rs7180002</td>
<td>Chr15:78873993</td>
<td>Intron 2 of CHRNA5</td>
<td>A/T</td>
<td>0.102</td>
<td>1.40 (1.17–1.68)</td>
<td>2.1 × 10^{-4}</td>
</tr>
<tr>
<td>rs951266</td>
<td>Chr15:78878541</td>
<td>Intron 2 of CHRNA5</td>
<td>C/T</td>
<td>0.1031</td>
<td>1.39 (1.16–1.66)</td>
<td>2.8 × 10^{-4}</td>
</tr>
<tr>
<td>rs17486195</td>
<td>Chr15:78865197</td>
<td>Intron 1 of CHRNA5</td>
<td>A/G</td>
<td>0.114</td>
<td>1.36 (1.14–1.61)</td>
<td>5.4 × 10^{-4}</td>
</tr>
<tr>
<td>rs4243084</td>
<td>Chr15:78911672</td>
<td>Intron 1 of CHRNA3</td>
<td>C/G</td>
<td>0.1537</td>
<td>1.28 (1.11–1.48)</td>
<td>8.5 × 10^{-4}</td>
</tr>
<tr>
<td>rs2735940</td>
<td>Chr15:1296486</td>
<td>1.3 kb upstream of TERT</td>
<td>T/C</td>
<td>0.479</td>
<td>0.82 (0.73–0.93)</td>
<td>1.1 × 10^{-3}</td>
</tr>
<tr>
<td>rs4635969</td>
<td>Chr15:1308552</td>
<td>13.3 kb upstream of TERT</td>
<td>C/T</td>
<td>0.3328</td>
<td>0.81 (0.72–0.92)</td>
<td>1.2 × 10^{-3}</td>
</tr>
<tr>
<td>rs17405217</td>
<td>Chr15:78731149</td>
<td>Intron 1 of IREB2</td>
<td>C/T</td>
<td>0.05842</td>
<td>1.43 (1.14–1.81)</td>
<td>2.5 × 10^{-3}</td>
</tr>
</tbody>
</table>

*ordered by P value; |  
*bminor allele listed second; |  
*MAF in controls only; |  
*dOR for each additional copy of the minor allele, estimated in a logistic regression model adjusted for: age, sex, study site, % sub-Saharan African ancestry, % European ancestry, and number of pack-years smoked; |  
*eP value in an allelic additive logistic regression model, adjusted for: age, sex, study site, % sub-Saharan African ancestry, % European ancestry, and number of pack-years smoked. |  

Figs. 1 and 2 show results of the association analyses for the 5p15.33, 6p22.1-p21.31, and 15q25.1 regions, respectively, with triangles identifying SNPs previously determined to be associated with lung cancer risk according to the NHGRI Catalog of Published Genome-Wide Association Studies. Supplementary Table S1 provides association results for all the imputed SNPs included in Figs. 1 and 2 that had P values less than 0.01. Histology-specific association results are also provided.

Four previously identified lung cancer risk SNPs on 5p15.33 were associated with lung cancer in this African-American sample (P < 0.05, effect size in same direction as previous reports), but were not the most statistically significant associations observed in the region. The most significantly associated genotyped SNP in the region was rs2735940 (OR, 2.31; 95% CI, 1.45–3.69; P = 0.00045) and rs111819086 in an intron of AGPUD (OR, 1.92; 95% CI, 1.31–2.82; P,
Table 3. Top 3 directly genotyped SNP associations in each lung cancer histologic subgroup

<table>
<thead>
<tr>
<th>Histologic subtype</th>
<th>SNP</th>
<th>Position</th>
<th>Gene</th>
<th>Allele</th>
<th>OR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenocarcinoma (n = 538)</td>
<td>rs2735940 Chr5:1296486 1.3 kb upstream of TERT</td>
<td>T/C</td>
<td>0.75</td>
<td>8.10</td>
<td>0.75</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>rs7725218 Chr5:1282414 Intronic 2 of TERT</td>
<td>G/A</td>
<td>1.33</td>
<td>2.00</td>
<td>1.13</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>rs2736100 Chr5:1286516 Intronic 1 of TERT</td>
<td>T/G</td>
<td>1.31</td>
<td>2.70</td>
<td>1.16</td>
<td>0.37</td>
</tr>
<tr>
<td>Squamous cell (n = 340)</td>
<td>rs401681 Chr5:1322087 Intronic 13 of CLPTM1L</td>
<td>T/C</td>
<td>1.12</td>
<td>0.12</td>
<td>1.10</td>
<td>0.99</td>
</tr>
<tr>
<td>Small cell (n = 97)</td>
<td>rs2844463 Chr6:31615167 Intronic 7 of BAT3</td>
<td>C/T</td>
<td>0.92</td>
<td>0.26</td>
<td>0.74</td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td>rs2736158 Chr6:31600304 Exon 16 of BAT2 (missense Ala!Gly)</td>
<td>C/G</td>
<td>1.06</td>
<td>0.61</td>
<td>0.64</td>
<td>1.80</td>
</tr>
</tbody>
</table>

aTop 3 associations for small-cell carcinoma are not included due to limited sample size. All 3 SNPs were located on 15q25.1.
bMinor allele listed second.
cOR for each additional copy of the minor allele, estimated in a logistic regression model adjusted for: age, sex, study site, % sub-Saharan African ancestry, and number of pack-years smoked.
dP-value in an allelic additive logistic regression model, adjusted for: age, sex, study site, % sub-Saharan African ancestry, and number of pack-years smoked.

Discussion

Our findings provide evidence that inherited variation in the 3 regions most frequently associated with lung cancer in European and East Asian populations are also associated with lung cancer in African-Americans. Despite this consistency, there remains heterogeneity across ethnicities in which particular SNPs have the strongest associations, particularly on 15q25.1 where the association peak in African-Americans is more strongly localized to CHRNA5 than it is among other ethnicities. We further show that SNPs which confer risk for specific...
lung cancer histologies are not necessarily risk factors for all types of lung cancer, as 5p15.33 associations were driven by risk for adenocarcinoma and 6p21.33 associations appear specific to risk for squamous cell carcinoma. Imputation of an additional 305 SNPs on 5p15.33 and 3,447 SNPs on 6p21.33 further support the histology-specific nature of these associations. Finally, a list of the most strongly associated SNPs in the analysis includes functional variants on 15q25.1 and 6p21.33, which may directly influence lung cancer risk in African-Americans.

We observed a moderate association between lung cancer and variants at 5p15.33 in \textit{TERT} and \textit{CLPTM1L}, expanding on previous reports of associations in this region in African-Americans (29). These associations appear largely to be confined to risk for adenocarcinoma, as several SNP associations in and near \textit{TERT} were greatly strengthened when analysis was conducted only on cases with adenocarcinoma, despite the reduction in sample size. The association in adenocarcinoma cases was well localized to the region immediately upstream and in the first 2 introns of \textit{TERT}, suggesting that \textit{TERT} is likely the gene in this region functionally involved in lung cancer pathogenesis.

In analysis of the full case-control dataset, variants in the HLA region on 6p22.1-p21.31 showed little association with lung cancer. However, when stratified by histology, a previously associated region in \textit{BAT3} was replicated in the squamous cell carcinoma subgroup and an associated missense variant in \textit{BAT2} was identified. Associations in the 6p21.33 region have not been consistently replicated, in part due to the European risk SNPs having low allele frequencies in East Asian populations. We offer another possible reason for the inconsistent associations of SNPs in this region – namely, that the associations appear to be histology specific and may be difficult to detect when squamous cell carcinomas are analyzed with other tumor histologies.
histologies. Just as the TERT associations are strongest in cases of adenocarcinoma, in African-Americans the 6p21.33 associations appear strongest in cases of squamous cell carcinoma. Indeed, of the 3 GWAS identifying cases of adenocarcinoma, in African-Americans the variants in African-Americans has allowed functional and tagging because of large LD blocks in the region, the reduced LD in difficulty mapping the 15q25.1 signal to a particular gene project. While studies in European populations have had additional variants using data from the 1000 Genomes genotyped many more SNPs in the 15q25.1 region than associated variants in the 5
may be functionally relevant, including strongly associated variants in the 5' UTR and in exon 5. We genotyped many more SNPs in the 15q25.1 region than previous studies of African-Americans (41), and imputed additional variants using data from the 1000 Genomes project. While studies in European populations have had difficulty mapping the 15q25.1 signal to a particular gene because of large LD blocks in the region, the reduced LD in African-Americans has allowed functional and tagging variants in CHRNA5 to emerge as the likeliest candidates in our analysis.

Controlling for the number of pack-years smoked modestly attenuated CHRNA5 associations, but they remained the strongest association signals in the analyses. Because SNPs in this region are associated with smoking behaviors, and pack-years is an imperfect measurement of smoking behavior, the effects observed here may be mediated by smoking. Conditioning association tests on missense SNP rs16969968 revealed additional association signals in CHRN4 and AGPHD. Whether these associations are due to the effects of additional genes on 15q25.1, or simply result from linkage between SNPs in CHRN4 and AGPHD with other CHRNA5 risk variants remains to be determined. Whether differences in the association patterns on 15q25.1 seen in African-Americans, compared with other ethnicities, are because of differences in LD patterns or differences in smoking behaviors remains to be determined.

The current analysis represents the largest genetic study of lung cancer conducted in African-Americans to date and has identified functional variants which may contribute to the pathogenesis of this devastating disease. While multicenter studies pose certain challenges, we believe that the close similarity of the percentage of sub-Saharan African ancestry estimates and the similar sample size across study sites is a particular strength of the study. Our study provides evidence that regions which are associated with lung cancer in Europeans and East Asians are also associated in African-Americans, but that there is incomplete overlap of the associated SNPs. We further conclude that tumor histology is an important consideration in genetic association analyses of lung cancer, especially in African-Americans where there is both a different distribution of tumor histology (42) and of SNP allele frequencies compared with European populations.

 Disclosure of Potential Conflicts of Interest

Laura J. Bierut served as a consultant for Pfizer Inc. in 2008 and is an inventor on the patent ‘‘Markers for Addiction’’ (US 20070258898) covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. No potential conflicts of interest were disclosed by the other authors.

 Authors’ Contributions

Conception and design: K.M. Walsh, C. Wei, M.F. Seldin, M.R. Wrensch, A.G. Schwartz, J.K. Wiencke, C.I. Amos
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): H.M. Hansen, M.R. Spitz, J.D. Sison, M.L. Frazier, M.R. Wrensch, A.G. Schwartz, J.K. Wiencke, C.I. Amos
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): H.M. Hansen, X. Wu, H. Zhang, A.S. Wenzlaff, J.D. Sison, M.R. Wrensch, J.K. Wiencke, C.I. Amos
Other (cross-population work on cigarette smoking): L.J. Bierut

Grant Support

This work was financially supported by NIH grants R01CA52689, R01ES06717, R01CA121197, R01CA121197S2, R01CA14176, R01CA066091, N01PC35145, HHSN261200100028C, and R25CA112355 to K.M. Walsh and R01DA021257 to L.J. Bierut.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 3, 2012; revised October 25, 2012; accepted November 14, 2012; published OnlineFirst December 5, 2012.

References

Walsh 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
