Familial Aggregation of Common Sequence Variants on 15q24-25.1 in Lung Cancer


Three recent genome-wide association studies identified associations between sequence variants on 15q24-25.1 and the risk of lung cancer. We conducted genome-wide association analysis to investigate associations between single-nucleotide polymorphisms (SNPs) and the risk of lung cancer, in which we used blood DNA from 194 case patients with familial lung cancer and 219 cancer-free control subjects. We identified associations between common sequence variants at 15q24-25.1 (that spanned LOC123688 [a hypothetical gene], PSMA4, CHRNA3, CHRNA5, and CHRN4) and lung cancer. The risk of lung cancer was more than five-fold higher among subjects who had both a family history of lung cancer and two copies of high-risk alleles rs8034191 (odds ratio [OR] = 7.20, 95% confidence interval [CI] = 2.21 to 23.37) or rs1051730 (OR = 5.67, CI = 2.21 to 14.60, both of which were located in the 15q24-25.1 locus, than among control subjects. Thus, further research to elucidate causal variants in the 15q24-25.1 locus that are associated with lung cancer is warranted.


Lung cancer can occur sporadically in people with no known family history of lung cancer or it can occur in multiple members of the same family and be designated as familial lung cancer. Evidence of a genetic basis for susceptibility to lung cancer has been demonstrated through genome-wide association studies (1–3) and segregation analyses (4–9). We conducted a genome-wide association study among individuals with a familial history of lung cancer. These individuals are members of families with three or more first-degree relatives with lung cancer that were collected as part of the Genetic Epidemiology of Lung Cancer Consortium (GELCC).

For the genome-wide association study, we genotyped 194 case patients with familial lung cancer and 219 cancer-free control subjects by use of the Affymetrix 500K or Affymetrix Genome-Wide Human SNP Array 6.0 (Santa Clara, CA) (Supplementary Table 1, available online). To ensure genetic independence among subjects, one case patient with familial lung cancer was chosen from each high-risk lung cancer family. Noncancer control subjects were obtained from families of unaffected individuals from the GELCC families (n = 36) and the Fernald Medical Monitoring Program (Fernald, OH) (n = 117). These control subjects had no blood relationship with any selected case patients. To minimize possible effects of cigarette smoking and age, smokers with an older age were selected mainly as control subjects, except for spousal control subjects (Supplementary Table 1, available online). To maintain homogeneous population samples, only Caucasian populations from the GELCC, Fernald Institute for Medical Research, and the Fernald Medical Monitoring Program were used for the association analysis. Basic characteristics of the GELCC subjects are presented in Supplementary Table 1 (available online). The Texas and UK lung cancer cohorts have been described previously (1). Written informed consent and institutional review board approval were obtained for all the subjects involved in this study.

Single-nucleotide polymorphism (SNP) genotyping was done by the Vanderbilt University Microarray Shared Resource following the Affymetrix protocol (www.affymetrix.com). We used the Affymetrix 500K SNP 6.0 array to genotype the remaining 137 case patients and 164 control subjects. The samples used with the two platforms were randomly chosen from the GELCC collections. The Affymetrix 6.0 array includes more than 906,600 SNPs and covers 97.2% of SNPs presented in the Affymetrix 500K SNP 6.0 array to genotype the remaining 137 case patients and 164 control subjects. For the Affymetrix 500K SNP 6.0 array, a confidence score of 0.33 was used for genoty-
Prior knowledge
Genome-wide association studies have identified associations between markers in the chromosomal region 15q24-25.1 and the risk of lung cancer.

Study design
A genome-wide case–control association analysis was used to investigate relationships between single-nucleotide polymorphisms (SNPs) and the risk of familial lung cancer.

Contribution
Subjects with both a family history of lung cancer and two copies of either of two high-risk alleles in 15q24-25.1 had a higher risk of lung cancer than control subjects.

Implications
Additional research is required to identify which genetic variants in the 15q24-25.1 region are associated with a high risk of lung cancer.

Limitations
Associations of risk alleles with nicotine dependence were not directly tested because the data were not available. Smoking quantity was available; however, no association between smoking quantity and the high-risk alleles was found. The small sample size may have limited the ability to detect a smaller effect size for risk alleles among heterozygotes with familial lung cancer.

From the Editors

determination, by using the Birdseed algorithm (available at http://www.broad.mit.edu/mpg/birdsuite/birdseed.html). We excluded four samples with unexpected genetic relatedness that was detected by PLINK software (http://pngu.mgh.harvard.edu/~purcell/plink/) and 33 samples with a genotype call rate of less than 86%. Thus, a total of 413 samples were available for our analysis (Supplementary Table 1, available online)—57 case patients and 56 control subjects genotyped with the Affymetrix 500K chip-set, and 137 case patients and 163 control subjects genotyped with the Affymetrix SNP 6.0 array. When the frequency distribution of the overlapping SNPs in the two genotyping platforms was analyzed, no heterogeneity between the platforms was found (Supplementary Figure 1, available online). Among the 413 subjects, we detected no population stratification, as shown by linkage agglomerative clustering implemented in PLINK software. Hardy–Weinberg equilibrium for each SNP was examined with the software hweStrata, by use of an exact test as proposed previously (11) (stratum number K = 1). All statistical tests were two-sided. SNPs with an exact P value of .0001 or less from Hardy–Weinberg equilibrium tests and with a minor allele frequency of less than 1% among control subjects were excluded from association analyses. Thus, in the final association analysis of 413 samples, we retained 722 376 SNPs for the 300 samples that were genotyped by SNP Array 6.0 and 399 377 SNPs for the 113 samples that were genotyped by Affymetrix 500K. The statistical significance of the association between SNP allele and disease status was assessed primarily with Fisher exact tests (12) (Supplementary Figure 2, available online). The odds ratios (ORs) of lung cancer associated with each SNP and 95% confidence intervals (CIs) were estimated by allele and by genotype. We evaluated the cumulative distribution of P values from Fisher exact tests in our genome-wide association study. As shown in Supplementary Figure 3 (available online), the distribution of observed P values was similar to the expected uniform distribution [0, 1], indicating no inflation of test statistics from population structure or any other form of bias.

Our genome-wide association study identified several strong associations of SNPs on chromosomes 1, 3, 6, 9, 12, 15, and 20 with lung cancer. One of the most strongly associated clusters of SNPs was found in a 160-kb region at 15q24-25.1 (Figure 1, A and B, and Supplementary Figure 1, available online) that had strong linkage disequilibrium and contained multiple discrete haplotype blocks (ie, regions with no evidence of historical recombination), as defined by examining the HapMap data from Utah residents with ancestry from northern and western Europe (http://www.hapmap.org/) (Figure 1, C). These familial data confirm the recently described association between 15q25.1 and sporadic lung cancer (1–3). At least two common variants on 15q24-25.1, SNPs rs8034191 and rs1051730, were strongly associated with both familial (P = 3.74 × 10⁻⁴ for rs8034191 and P = 2.90 × 10⁻⁴ for rs1051730) and sporadic (from the Texas study, P = 9.71 × 10⁻⁹ for rs8034191 and P = 1.32 × 10⁻⁷ for rs1051730; from the UK study, P = 9.90 × 10⁻⁹ for rs8034191 and P = 2.84 × 10⁻⁸ for rs1051730) lung cancer (Table 1). These SNPs are in high linkage disequilibrium (r² = 0.87) and are located within intronic regions of hypothetical gene LOC123688 (rs8034191) and of the gene CHRN43 (rs1051730). Risk of lung cancer among homozygous carriers was statistically significantly different from risk among noncarriers for the two SNPs—rs8034191 and rs1051730—in large sporadic lung cancer samples previously reported from Texas and from the United Kingdom (1).

It is worth noting that the effect size of the risk allele at 15q24-q25.1 observed in familial lung cancer (for carrying two copies of risk alleles of rs8034191 and rs1051730, respectively, OR = 3.84, 95% CI = 1.75 to 8.84, and OR = 3.43, 95% CI = 1.66 to 7.37) was larger than that observed in sporadic lung cancer (for carrying two copies of the same risk alleles, respectively, OR = 1.76, 95% CI = 1.42 to 2.18, and OR = 1.75, 95% CI = 1.42 to 2.17, in the Texas study; and OR = 2.09, 95% CI = 1.60 to 2.73, and OR = 2.05, 95% CI = 1.57 to 2.70, in the UK study) (Table 1). We therefore tested heterogeneity between familial and sporadic lung cancers by Woolf’s test (13). Statistically significant heterogeneity was observed between familial and the Texas sporadic lung cancer samples (P = .04) but not between familial and UK sporadic lung cancer samples (P = .12).

To rule out confounding effects of smoking behavior on lung cancer risk, the association analysis was adjusted by sex, age, and pack-years of cigarette exposure, as continuous variables (Supplementary Table 2, available online). Association of the 15q24-25.1 locus with familial lung cancer remained statistically significant (P = 1.03 × 10⁻⁸ for rs8034191 and P = 3.10 × 10⁻⁴ for rs1051730) after this adjustment. The adjusted associations were even stronger among those in lung cancer families carrying two copies of risk alleles of rs8034191 (OR = 7.20, 95% CI = 2.21 to 23.37) or rs1051730 (OR = 5.67, 95% CI = 2.21 to 14.60). Thus, among lung cancer patients with a family history, alleles associated with a high risk of lung cancer appear to be located on the q arm of chromosome 15.
We next investigated whether the SNPs on 15q24-q25.1 that were strongly associated with lung cancer acted additively, recessively, or dominantly by use of logistic regression and Bayesian information criteria. Given any two estimated models, the model with the lower value of Bayesian information criteria is the one to be preferred. The strongest associations were found with the recessive model (Table 1 and Supplementary Table 3, available online). In the recessive model, an increased risk of lung cancer was associated with having two copies of the high-risk A allele of rs1051730, compared with having no copies (OR = 3.43, 95% CI = 1.66 to 7.37, \( P \) value = 2.90 \( \times \) 10\(^{-4} \)). However, an additive model gave similar levels of statistical significance for many of the most statistically significant SNPs in this region. The homozygous risk genotype appeared to occur more frequently and to have larger effect size among familial lung cancer samples (OR = 3.84 for rs1051730) than among sporadic lung cancer samples (OR = 1.75 for rs1051730 from the Texas study and OR = 2.09 for rs1051730 in the UK study).

The smaller sample size of the familial study was less powerful to detect the smaller risks among heterozygotes. This limitation and the higher frequency of homozygotes in the highly ascertained families with familial lung cancer may account for the difference in best fitting models at the 15q24 locus in familial vs sporadic data sets. A genome-wide association study on a larger sample of familial case patients would provide increased power. However, both familial and sporadic analyses support the presence of a risk allele at 15q24-25.1.

In addition to analyses of single SNPs, we performed a haplotype-based association analysis on the 15q24-q25.1 locus. Haplotypes at the 15q24-25.1 locus region were inferred by use of the fastPHASE computer program (14). To exhaustively exploit haplotype information, we then subject alleles (contiguous sets of markers) from sliding windows of all sizes to haplotype association tests (15). In the haplotype analysis, a common haplotype with a frequency larger than 5% was first identified and treated as one category and the other variants as another category. The Fisher exact test was then applied to determine statistical significance (12). Multiple haplotypes at the 15q24-25.1 locus were statistically significantly associated with familial lung cancer risk (Supplementary Table 4, available online). The most statistically significant association was with the haplotype A-T consisting of both rs7163730 and rs4461039 (\( P = 1.03 \times 10^{-4} \)). In the single-marker analyses, the two SNPs with the most statistically significant association with lung cancer were rs7163730 (\( P = 8.71 \times 10^{-5} \)) and rs4461039 (\( P = 8.67 \times 10^{-5} \)).

Using GELCC familial lung cancer samples and a genome-wide association study, we identified several common variants at 15q24-25.1 and confirmed the association between genetic variants on the q arm of chromosome 15 and sporadic lung cancer (1–3). When we adjusted for smoking and other covariates, a statistically significant association remained (Supplementary Table 2, available online). This observation is consistent with previous analyses (1–3), in which a direct role for these variants in lung cancer was postulated. In those analyses, the
Fisher exact tests were performed. All statistical tests were two-sided.

rs1051730 was genotyped by Human Genetics Division Genotyping Core.

The GELCC study had a total of 194 familial case patients and 219 control subjects. The Texas study had a total of 1865 sporadic case patients and 1769 control subjects. The UK study had a total of 2013 sporadic case patients and 3082 control subjects. The Texas and UK sporadic lung cancer samples are from the study by Amos et al. (1).

† Fisher exact tests were performed. All statistical tests were two-sided.

§ rs1051730 was genotyped by Human Genetics Division Genotyping Core.

number of cigarettes consumed per day, pack-years of exposure, and time to first cigarette were not associated with SNPs in the region (1,2). In contrast, Thorgerisson et al. (3) reported that the SNP rs1051730 was statistically significantly associated with lung cancer and that each copy of the T allele was associated with an increase in smoking of one cigarette per day. In addition, Saccone et al. (16) identified the SNPs rs16969968 (missense in CHRNA3) and rs578776 (in the 3′-untranslated region of CHRNA3) as functional variants that were statistically significantly associated with nicotine dependence. When we genotyped rs16969968 and rs578776 in our familial lung cancer population, we found that both were statistically significantly associated with lung cancer (for rs16969968, \( P = 2.29 \times 10^{-3} \); for rs578776, \( P = 4.47 \times 10^{-3} \)). Thus, polymorphisms in these genes may affect nicotine dependence and propensity to smoke and to develop lung cancer.

Our study had several limitations. First, the association of risk variants with nicotine dependence (as measured by the Fagerstrom test for nicotine dependence) was not directly tested because these data were unavailable. However, smoking quantity (a component of the Fagerstrom test) was available. We did not find an association between smoking quantity and the statistically significant variants associated with lung cancer. Second, the sample size in our familial genome-wide association analysis was small, and so we might not have been able to detect the smaller effect size of risk alleles located in 15q24-25.1 for heterozygotes in familial lung cancer.

The identity of variants in the 15q24-25.1 region that are most strongly associated with lung cancer remains unknown because many SNPs are in strong linkage disequilibrium with each other. In addition, whether the variants associate with a direct or indirect mode for lung cancer remains unresolved. From results of this and three other genome-wide association studies, the candidate genes include IREB2, LOC123688, PSMA4, CHRNA5, CHRNA3, and CHRNA4 (1–3). CHRNA5, CHRNA3, and CHRNA4 are strong candidates primarily because they encode subunits of the nicotinic acetylcholine receptors. These genes may participate in nicotinic addiction through reward pathways in the brain; however, evidence exists that nicotinic acetylcholine receptors are associated directly with lung carcinogenesis (17). Nicotinic acetylcholine receptors are expressed in lung cancers of both non–small cell and small cell subtypes (18), and treatment of lung cancer cell lines with nicotine can inhibit proapoptotic pathways initiated by opioids (19,20). Therefore, determination of a likely single candidate gene and further delineation of whether variants affect lung cancer directly or indirectly or both are warranted.

### Table 1. Associations of the single-nucleotide polymorphisms rs8034191 and rs1051730 genotypes with lung cancer*

<table>
<thead>
<tr>
<th>Samples†</th>
<th>Genotype</th>
<th>Frequency, %</th>
<th>Control subjects</th>
<th>Case patients</th>
<th>( P ) value‡</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GELCC</strong></td>
<td>rs8034191</td>
<td>AA</td>
<td>50.00</td>
<td>36.64</td>
<td>.362</td>
<td>1.30 (0.75 to 2.24)</td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>41.67</td>
<td>39.69</td>
<td>( 3.74 \times 10^{-4} )</td>
<td>3.84 (1.75 to 8.84)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>8.33</td>
<td>23.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs1051730</td>
<td>GG</td>
<td>48.11</td>
<td>37.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>44.81</td>
<td>43.35</td>
<td>.377</td>
<td>1.23 (0.78 to 1.96)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>7.08</td>
<td>19.08</td>
<td>2.90 \times 10^{-4}</td>
<td>3.43 (1.66 to 7.37)</td>
<td></td>
</tr>
<tr>
<td><strong>Texas</strong></td>
<td>rs8034191</td>
<td>AA</td>
<td>44.10</td>
<td>37.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>44.85</td>
<td>46.68</td>
<td>3.09 \times 10^{-3}</td>
<td>1.24 (1.07 to 1.43)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>11.05</td>
<td>16.32</td>
<td>9.71 \times 10^{-6}</td>
<td>1.76 (1.42 to 2.18)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs1051730</td>
<td>GG</td>
<td>44.31</td>
<td>36.82</td>
<td>( 1.04 \times 10^{-3} )</td>
<td>1.27 (1.10 to 1.47)</td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>44.54</td>
<td>46.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>11.15</td>
<td>16.23</td>
<td>3.12 \times 10^{-7}</td>
<td>1.75 (1.42 to 2.17)</td>
<td></td>
</tr>
<tr>
<td><strong>United Kingdom</strong></td>
<td>rs8034191</td>
<td>AA</td>
<td>46.67</td>
<td>36.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>43.23</td>
<td>46.86</td>
<td>1.72 \times 10^{-4}</td>
<td>1.38 (1.17 to 1.64)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>10.10</td>
<td>16.65</td>
<td>9.90 \times 10^{-6}</td>
<td>2.09 (1.60 to 2.73)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs1051730</td>
<td>GG</td>
<td>46.55</td>
<td>37.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>43.72</td>
<td>46.34</td>
<td>1.40 \times 10^{-3}</td>
<td>1.31 (1.11 to 1.56)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>9.73</td>
<td>16.12</td>
<td>2.84 \times 10^{-8}</td>
<td>2.05 (1.57 to 2.70)</td>
<td></td>
</tr>
</tbody>
</table>

* GELCC = Genetic Epidemiology of Lung Cancer Consortium; OR = odds ratio; CI = confidence interval.
† The GELCC study had a total of 194 familial case patients and 219 control subjects. The Texas study had a total of 1865 sporadic case patients and 1769 control subjects. The UK study had a total of 2013 sporadic case patients and 3082 control subjects. The Texas and UK sporadic lung cancer samples are from the study by Amos et al. (1).
‡ Fisher exact tests were performed. All statistical tests were two-sided.
§ rs1051730 was genotyped by Human Genetics Division Genotyping Core.

### References


### Funding

This work was supported in part by the following: National Institutes of Health (NIH) grants U01CA76293 (Genetic Epidemiology of Lung Cancer Consortium), R01CA058554, R01CA093643, R01CA099147, R01CA099187, R01ES012063, R01ES013340, R03CA77118, R01CA80127, P30ES06096, P50CA70907 (Specialized Program of Research Excellence), N01HG65404, N01-PC35145, P30CA22453, R01CA63700, DE-FGB-95ER62060, R01CA55769, R01CA121197, R01CA133996, Mayo Clinic intramural research funds, and Department of Defense VITTAL grant. This study was supported in part by NIH, the Intramural Research Programs of the National Cancer Institute, and the National Human Genome Research Institute. The authors had full responsibility of design of the study, the collection of the data, the analysis and interpretation of the data, the decision to submit the manuscript for publication, and the writing of the manuscript.

### Notes

Marshall W. Anderson and Ming You contributed equally to this work.

We thank the Fernald Medical Monitoring Program for sharing their biospecimens and data with us. We are grateful to the lung cancer families who participated in this research, and for the high caliber service of Vanderbilt University Microarray Shared Resource, Human Genetics Division Genotyping Core, and to Qiong Chen, Shaw Levy, Victor Guallar, and Jennifer Baker for their assistance in various aspects of this work.

Manuscript received April 2, 2008; revised June 30, 2008; accepted July 2, 2008.