Inherited Variation at Chromosome 12p13.33, Including RAD52, Influences the Risk of Squamous Cell Lung Carcinoma

Jianxin Shi, Nilanjan Chatterjee, Melissa Rotunno, Yufei Wang, Angela C. Pesatori, Daria Consonni, Peng Li, William Wheeler, Peter Broderick, Marc Henrion, Timothy Eisen, Zhaoming Wang, Wei Chen, Qiong Dong, Demetrius Albanes, Michael Thun, Margaret R. Spitz, Pier Alberto Bertazzi, Neil E. Caporaso, Stephen J. Chanock, Christopher I. Amos, Richard S. Houlston, and Maria Teresa Landi

ABSTRACT

Although lung cancer is largely caused by tobacco smoking, inherited genetic factors play a role in its etiology. Genome-wide association studies in Europeans have only robustly demonstrated 3 polymorphic variations that influence the risk of lung cancer. Tumor heterogeneity may have hampered the detection of association signal when all lung cancer subtypes were analyzed together. In a genome-wide association study of 5,355 European ever-smoker lung cancer patients and 4,344 smoking control subjects, we conducted a pathway-based analysis in lung cancer histologic subtypes with 19,082 single-nucleotide polymorphisms mapping to 917 genes in the HuGE-defined “inflammation” pathway. We identified a susceptibility locus for squamous cell lung carcinoma at 12p13.33 (RAD52, rs6489769) and replicated the association in 3 independent studies totaling 3,359 squamous cell lung carcinoma cases and 9,100 controls (OR = 1.20, P_{combined} = 2.3 \times 10^{-8}).

INTRODUCTION

Lung cancer is a major cause of cancer death worldwide, causing more than 1 million deaths each year (1). The histologic classification separates small cell lung cancers from non–small cell lung cancers, the latter mostly comprising adenocarcinoma and squamous cell tumors. These lung cancer histologies have diverse molecular characteristics that reflect differences in carcinogenesis, etiology, and treatment (2, 3).

Although lung cancer is largely caused by tobacco smoking, studies have also implicated inherited genetic factors in disease etiology. Genome-wide association studies in European subjects have consistently identified 3 polymorphic variations at 15q25.1 (CHRNA5-CHRNA3-CHRNB4), 5p15.33 (TERT-CLPTM1), and 6p21.33 (BAT3-MSH5) that influence the risk of lung cancer (4–8). Interestingly, one single-nucleotide polymorphism (SNP), rs2736100, which localizes to the TERT gene, was distinctly associated with the risk of adenocarcinoma (9), suggesting that additional searches for histology-specific associations are likely to prove informative.

It is estimated that perhaps as much as 25% of all cancers are associated with chronic infection and inflammation (10). Because tobacco smoking can initiate or sustain chronic inflammation (11), often in concert with altered DNA repair and inflammatory response (12), we explored the impact of common genetic variation in inflammatory genes on the risk of lung cancer in smokers by using a genome-wide association study from the National Cancer Institute (NCI; ref. 9). The study included 5,355 smokers (both current and former) and 4,344 smoking control subjects of European ancestry. We performed a pathway analysis of the genes listed under the category of “inflammation” in the HuGE Navigator (version 1.4; ref. 13) to measure the collective effect of these genes and identify the SNPs with the strongest association for replication in independent samples. We analyzed the association with lung cancer risk overall and within the major lung cancer histology groups to explore the genetic basis of disease subtypes.

RESULTS

We first conducted quality control analysis for all genome-wide association study SNPs included in this study. Quantile-quantile plots of the negative logarithm of the genome-wide \( P \)-values and genomic control \( \lambda \) values computed on the basis of all genome-wide association study SNPs indicated no global variance inflation (genomic controls \( \lambda \) between 1 and 1.02 for all analyses, Supplementary Fig. S1), excluding the possibility of inflated type I error rates in the pathway analyses.

Among the 591,928 SNPs from the original genome-wide association studies, we analyzed 19,082 SNPs mapping to 917 genes from the HuGe navigator list (13). Given the large number of genes and the variable degree of involvement of these genes in the inflammation-related function, we used HuGE Literature Finder to assign each gene a score, reflecting the strength of evidence for association with inflammation (Supplementary Table S1 and Supplementary Methods), and tested the pathway association integrating this weighting for the genes. The list of the 917 genes and the corresponding SNP \( P \)-values are reported in Supplementary Table S1. After applying a Bonferroni correction, we found a strong association of the pathway for cases of squamous cell cancer (Fig. 1 and Supplementary Table S2; \( P = 0.0004 \), ever smokers), suggesting the existence of SNPs truly associated with risk in this histologic subtype. An analysis without the HuGE-based weighting provided similar results, although with weaker associations (squamous cell, \( P = 0.001 \); Supplementary Table S2). No statistically significant association was found for the risk of adenocarcinoma or small cell lung cancer when we used either the weighted or unweighted approach (Fig. 1 and Supplementary Table S2).

Therefore, we restricted further analyses to the squamous cell subtype. We chose the 55 SNPs in the 55 genes with the strongest evidence for association in squamous cell (gene-wise \( P \)-value < 0.05) and performed replication in 2 independent studies of European ancestry, including UK1 (8), with 592 squamous cell cases and 2,699 control subjects from the 1958 Birth Cohort (Wellcome Trust Case-Control Consortium; ref. 14), genotyped with the use of Illumina HumanHap550 arrays; and Texas (5), with 306 squamous cell cases and 1,137 control subjects, genotyped with the use of Illumina HumanHap300 arrays.

Among the 55 SNPs, only rs6489769 was consistently replicated in both studies, with a combined \( P = 6.48 \times 10^{-7} \) for the association with squamous cell risk (Table 1). The SNP marker rs6489769 maps to chromosome 12p13.33 (943,226 bps). The pathway analysis in NCI squamous cell data after the exclusion of the SNPs at 12p13.33 showed a pathway-level \( P = 0.0008 \). We then replicated the rs6489769 SNP in a third independent study, UK2 (15), with 1,038
squamous cell cases and 933 control cases genotyped with the use of the illumina infinium custom arrays. the association was confirmed also in the third sample set (table 2). although these case-control series were smaller than the discovery dataset, each had the statistical power to replicate the signal of the nci discovery set (or, 1.23) at one-sided \( P < 0.05 \) (statistical power for uk1 = 0.92, uk2 = 0.93, and texas = 0.70).

When we combined data from all 4 studies, we found the association was statistically significant on a genome-wide basis with \( P = 2.30 \times 10^{-8} \), 2 orders of magnitude below the bonferroni corrected \( P \)-value threshold for 19,082 snps (0.05/19,082 snps, \( P = 2.62 \times 10^{-8} \)) and or, 1.20 (95% confidence interval, 1.12–1.28; \( P_{	ext{het}} = 0.89, i^2 = 0\%; \) table 2).

We tested whether the association with squamous cell risk for this SNP was modified by pack-years of tobacco smoking in the nci genome-wide association studies but found very similar results across smoking strata (Supplementary Table S3). We also investigated in the environment and genetics in lung cancer etiology (EAGLE) study (441 squamous cell cases and 1,319 control subjects; ref. 16) whether the association between rs6489769 and squamous cell was confirmed by chronic obstructive pulmonary disease (COPD) status but found no major changes in the adjusted data (data not shown).

rs6489769 was not significantly associated with chronic obstructive pulmonary disease in lung cancer cases (\( P = 0.67, \text{OR} = 0.97 \)). Because only 131 of the control subjects had documented COPD, a larger study of lung-free patients with COPD is required to robustly examine the impact of this SNP on the risk of COPD.

To explore the 12p13.33 region further, we imputed unobserved genotypes in squamous cell cases and control subjects in nci squamous cell data by using HapMap Phase III and 1000 Genomes Project data but did not identify any stronger association at 12p13.33 than that provided by rs6489769. This locus harbors the RAD52 gene, which is involved in homologous recombination. We examined whether other genes in the homologous recombination pathway or the overall DNA repair pathway influence squamous cell risk. None of the other 17 homologous recombination genes (involving 142 SNPs) showed an association with gene-wise \( P < 0.05 \) (Supplementary Table S4). In the analysis of the overall DNA repair pathway, including 1,410 SNPs mapping to 136 genes (Supplementary Table S5), we observed a modest pathway-level association in the nci squamous cell data including (\( P = 0.006 \)) or excluding (\( P = 0.04 \)) the RAD52 SNPs, and the only SNP with a \( P \)-value < 0.001 was rs6489769.

**Discussion**

By using a pathway analysis in a genome-wide association study of lung cancer from European ancestry, we identified a susceptibility locus for squamous cell carcinoma risk on chromosome 12p13.33. The finding was replicated in 3 independent samples, did not appear to be modified by smoking quantity or personal history of COPD, and exceeded a genome-wide threshold for association.

The 12p13.33 locus has at least 31 alternatively spliced variants (AceView; ref. 17). Depending on the transcripts, rs6489769, the SNP most strongly associated with squamous cell risk, appears located approximately 13 kb centromeric to or within a plausible candidate gene RAD52 (yeast, homolog of RAD52; MIM 600392; fig. 2). At a 53-kb linkage disequilibrium interval from rs6489769 is also the gene encoding WNK1 (protein kinase, lysine deficient 1; MIM 605232), which plays a role in pseudohypaldosteronism and hereditary sensory neuropathy. At 27 kb from the same SNP also lies ERC1 (ELKS/RAB6-interacting/CATF family member 1, MIM 607127), a member of a family of RIM-binding proteins, but rs6489769 does not appear to correlate with SNPs in this gene (fig. 2).

High-fidelity replication of DNA, and its accurate segregation to daughter cells, is critical for maintaining genome stability and suppressing cancer. DNA-replication forks are stalled by many DNA lesions, and stalled forks may eventually collapse, producing a broken DNA end (18). In concert with BRCA2, RAD52 plays a pivotal role in repairing these DNA double-strand breaks through homologous recombination (19). RAD52 and BRCA2 seem to act in parallel pathways, and RAD52 provides an important alternate mechanism for repairing replication-associated damage by homologous recombination in the absence of BRCA2 (20). RAD52 interacts with DNA recombination protein RAD51 and participates in the regulation of its polymerization (21). Thus, variation in RAD52 may disrupt the double-strand breaks repair function of RAD51. RAD52 also cooperates with OGG1 to repair oxidative DNA damage, thereby enhancing cellular resistance to inflammatory-related oxidative stress (22). Because most therapeutic strategies for lung cancer create DNA-replication stress, inherited variation in replication stress response may affect treatment efficacy (18, 23).

Although double-strand breaks can arise during DNA replication, they can also be induced by smoking tobacco (24), chronic inflammation (25), and other agents (20). Thus, genetic variation in RAD52 could contribute to altered repair of tobacco-induced and microinflammatorysustained DNA damage a priori, providing additional support for the role of homologous recombination dysfunction in the development of cancer. In our data, the rs6489769-squamous cell association was not modified by levels of tobacco smoking. Although the locus may have an effect on squamous cell risk independent of smoking, our sample size had limited statistical power to detect small changes across smoking strata.

Our analysis also provides evidence for the collective role of inflammatory genes in squamous cell cancer, which may contribute to the development or maintenance of a carcinogenic inflammatory microenvironment. Because RAD52 is the plausible candidate gene, we verified whether other genes in the homologous recombination pathway or the overall DNA-repair pathway were associated with the risk of squamous cell cancer. Although there may be other homologous recombination pathways genes influencing the risk of squamous cell cancer, our analysis suggests that by far the primary common determinant is related to the
<table>
<thead>
<tr>
<th>Gene</th>
<th>Minor frequency</th>
<th>Minor allele</th>
<th>Risk allele</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-B</td>
<td>rs17258574</td>
<td>G</td>
<td>G</td>
<td>2.12 × 10^{-7}</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>rs13602249</td>
<td>A</td>
<td>A</td>
<td>1.41 × 10^{-9}</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>rs73721577</td>
<td>0.23</td>
<td>0.23</td>
<td>0.71 × 10^{-5}</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>rs308936</td>
<td>G</td>
<td>G</td>
<td>1.39 × 10^{-5}</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>rs4897799</td>
<td>A</td>
<td>A</td>
<td>1.15 × 10^{-5}</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>rs1382249</td>
<td>A</td>
<td>A</td>
<td>3.60 × 10^{-5}</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>rs7721577</td>
<td>0.23</td>
<td>0.23</td>
<td>0.71 × 10^{-5}</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>rs1382249</td>
<td>A</td>
<td>A</td>
<td>1.39 × 10^{-5}</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>rs73721577</td>
<td>0.23</td>
<td>0.23</td>
<td>0.71 × 10^{-5}</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>rs308936</td>
<td>G</td>
<td>G</td>
<td>1.39 × 10^{-5}</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>rs4897799</td>
<td>A</td>
<td>A</td>
<td>1.15 × 10^{-5}</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>rs1382249</td>
<td>A</td>
<td>A</td>
<td>3.60 × 10^{-5}</td>
<td>0.07</td>
</tr>
</tbody>
</table>

aThe SNP with the smallest P-value of association with cancer risk mapping to the gene.

bMinor allele frequency in the NCI controls.

cTwo-sided P-value.
<table>
<thead>
<tr>
<th>Gene</th>
<th>rsID</th>
<th>Allele</th>
<th>OR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD86</td>
<td>rs435</td>
<td>G</td>
<td>1.10 (1.09–1.12)</td>
<td>0.11</td>
</tr>
<tr>
<td>TFAP2B</td>
<td>rs20374029</td>
<td>A</td>
<td>1.10 (1.09–1.12)</td>
<td>0.11</td>
</tr>
<tr>
<td>TERT</td>
<td>rs406936</td>
<td>G</td>
<td>1.10 (1.09–1.12)</td>
<td>0.11</td>
</tr>
<tr>
<td>IL12A</td>
<td>rs12112229</td>
<td>A</td>
<td>1.10 (1.09–1.12)</td>
<td>0.11</td>
</tr>
<tr>
<td>IL20</td>
<td>rs18036</td>
<td>G</td>
<td>1.10 (1.09–1.12)</td>
<td>0.11</td>
</tr>
<tr>
<td>TNMD</td>
<td>rs3729535</td>
<td>A</td>
<td>1.10 (1.09–1.12)</td>
<td>0.11</td>
</tr>
</tbody>
</table>

**Minor allele frequency in the NCI controls:**

- rs435: 0.10
- rs20374029: 0.17
- rs406936: 0.11
- rs12112229: 0.14
- rs18036: 0.24
- rs3729535: 0.09

**Meta-analysis 2-sided P-value:**

- rs435: 0.01
- rs20374029: 0.02
- rs406936: 0.001
- rs12112229: 0.001
- rs18036: 0.001
- rs3729535: 0.001

**The SP with the smallest P-value of association with cancer risk mapping to the gene:**

- rs435: CD86
- rs20374029: TFAP2B
- rs406936: TERT
- rs12112229: IL12A
- rs18036: IL20
- rs3729535: TNMD
Rad52 association. Fine-mapping studies and functional analyses are required to determine the biological basis of the association.

The 12p13.33 (Rad52) locus was distinctly associated with the risk of squamous cell carcinoma. Together with the previous identification of variants at 5p15.33 (TERT) associated with the risk of adenocarcinoma (9), our findings underscore the importance of searching for histology-specific lung cancer risk variants. Studies in other tumors have also identified different genetic variation by tumor subtypes or phenotypic diversity (26, 27), confirming that studying specific disease subtypes can enhance the power for detecting susceptibility loci in genome-wide association studies (28). Moreover, evaluating the associations between susceptibility loci and tumor subtypes may improve risk assessment, and predicting the risk for specific tumor subtypes may lead to targeted early detection or prevention strategies.

Moreover, identifying histology-specific SNPs may refine mechanistic understanding of currently unknown origins of morphologic differences and may contribute to the ongoing search for personalized treatment for subtype specific lung cancer cases (29, 30). However, a challenge of this approach is the difficulty of accruing necessary sample sizes given the relative rarity of many such disease subtypes. Our pathway-based approach took advantage of previous knowledge of the disease etiology and substantially helped prioritizing the most relevant SNPs for replication even in a relatively small sample size. This suggests that the combination of a pathway-based approach and information on disease specific subtypes can greatly improve the identification of cancer susceptibility loci.

**METHODS**

**Discovery Samples**

**NCI genome-wide association studies** The 9,699 smokers used as subjects were drawn from one population-based case-control study and 3 cohort studies: EAGLE (case-control, Italy); Alpha-Tocopherol Beta-Carotene Cancer Prevention study (ATBC, cohort, Finland); Prostate, Lung, Colon, Ovary screening trial (PLCO, cohort, United States); and Cancer Prevention Study II (CPS-II, cohort, United States). All subjects were of European ancestry. The study included 3 main histologic subtypes, that is, adenocarcinoma, squamous cell, small cell lung cancer, and a small number of other lung cancer subtypes. Subjects were genotyped using the use of Illumina 1M, 610QUAD, 550K, and 317K+240S HumanHap arrays.

The distribution of subjects by histology, smoking status, and genotyping platforms is listed in Supplementary Table S6. The details of quality control were reported previously (9). To summarize in brief, SNPs with missing rate >5%, Hardy-Weinberg equilibrium $P < 10^{-7}$ in controls, and minor allele frequency less than 5% were excluded. Subjects were removed if they were the outliers in the ancestry plots.
had high missing genotype rates, or were duplicates or relatives of other subjects. The genomic control \( \lambda \) values were 1.03, 1.03, 1.01, and 1.01 in EAGLE, PLCO, CPS-II, and ATBC, respectively, suggesting no significant hidden population substructure or unadjusted confounding factors.

**Replication Samples**

**UK1** Cases with pathologically confirmed squamous cell were ascertained through the Genetic Lung Cancer Predisposition Study (GELCAPS), genotyped using HumanHap550. All subjects were British residents and self-reported to be of European Ancestry. Control subjects were from the 1958 Birth cohort. The genomic control \( \lambda \) value was 1.03. Details on the quality control procedures have been previously reported (8).

**UK2** In UK2 (8, 15) cases of lung cancer were ascertained through GELCAPS. The 933 healthy smoking individuals included in the analysis are part of the National Cancer Research Network genetic epidemiological studies (1,497 men, 1,539 women; mean age 61 years, SD 11), the National Study of Colorectal Cancer (NSCCG; 1999–2006; \( n = 541 \)), GELCAPS (1999–2004; \( n = 1,520 \)), and the Royal Marsden Hospital Trust/Institute of Cancer Research Family History and DNA Registry (1999–2004; \( n = 975 \)). These control subjects were the spouses or unrelated friends of patients with malignancies. None had a personal history of malignancy at time of ascertainment. All were British residents and self-reported to be of European Ancestry. Subjects were genotyped using Illumina Infinium custom array.

**Texas** Case-control study of subjects of European ancestry, including lung cancer cases newly diagnosed at the University of Texas MD Anderson Cancer Center since 1991 and controls from the Kelsey-Seybold clinics (the genome-wide association studies included only smokers and cases with non-small cell lung cancer). Controls were frequency matched to cases according to their smoking behavior, age, ethnicity, and sex. Former smoking controls were further frequency matched to former smoking cases according to the number of years since smoking cessation. Subjects were genotyped with the Illumina HumanHap300 chip. The genomic control \( \lambda \) value was 1.03. Details on the quality control procedures have been previously reported (5).

All subjects included in the discovery sample and replication samples signed an informed consent form. The studies were performed after approval by each local institutional review board.

**Selecting Genes and SNPs in the Inflammation Pathway**

We used the SNP list compiled by the HuGE Literature Finder (HuGE Navigator, version 1.4; ref. 13), corresponding to 970 genes potentially involved in inflammation and assigned each gene the HuGE score. The HuGE score was calculated on the basis of the frequency of reported associations with inflammation in HuGE literature (the larger the score, the stronger the association), including studies of genetic association in humans and animal studies. In particular, the number of all publications reporting on the association of a given gene with inflammation and whether the identified associations were determined by genome-wide analyses, meta-analyses, or genetic testing were taking into account. Publications in which the authors used animal models were also added to the final weighting system (Supplementary Methods). Among the 970 genes, 917 (Supplementary Table S1) were covered by the NCI genome-wide association studies, corresponding to 19,082 SNPs (within ~20 kb upstream of the start of transcription and approximately ~10 kb downstream of stop of transcription, NCBI build 36).

**Statistical Analysis**

**Testing association of single SNPs** For the NCI data, we performed an unconditional logistic regression analysis to test the additive effect of each SNP genotype on lung cancer risk by using PLINK software (31), adjusting for age (<= 50, 51-55, 56-60, 61-65, 66-70, 71-75, 76+), sex, study (ATBC, EAGLE, ATBC, CPS-II), 4 principal components derived based on EIGENSTRAT (32) to control population stratification and cigarettes smoked per day (<=10, 11-20, 21-30, 31-40, 41+), smoking duration in 10-year intervals, smoking status (former vs. current smoker) and, for former smokers, the number of years since quitting (1-5, 6-10, 11-20, 21-30, 30+). To assess the possibility of systematic bias caused by unadjusted confounding factors or cryptic relatedness, we computed the genomic control \( \lambda \) values on the basis of all SNPs passing the quality control filters in the NCI genome-wide association studies by histology and smoking status.

### Table 2. Summary data for the 12p13.33 SNP rs6489769 associated with the risk of squamous cell lung carcinoma

<table>
<thead>
<tr>
<th>Study</th>
<th>Case Minor allele frequency</th>
<th>Control Minor allele frequency</th>
<th>ORb</th>
<th>95% CI</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA*</td>
<td>AG*</td>
<td>GG*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCI</td>
<td>0.35 596 663 165</td>
<td>0.38 1,654 2,025 653</td>
<td>1.23</td>
<td>1.11–1.35</td>
<td>3.18 × 10⁻⁵</td>
</tr>
<tr>
<td>UK1</td>
<td>0.34 261 263 68</td>
<td>0.37 1,053 1,293 352</td>
<td>1.16</td>
<td>1.01–1.32</td>
<td>0.043</td>
</tr>
<tr>
<td>UK2</td>
<td>0.33 467 456 115</td>
<td>0.37 381 420 132</td>
<td>1.17</td>
<td>1.03–1.34</td>
<td>0.017</td>
</tr>
<tr>
<td>Texas</td>
<td>0.33 131 144 30</td>
<td>0.38 453 500 184</td>
<td>1.22</td>
<td>1.01–1.47</td>
<td>0.036</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
<td>1.20</td>
<td>1.12–1.28</td>
<td>2.30 × 10⁻⁸</td>
</tr>
</tbody>
</table>

Results from the NCI squamous cell carcinoma discovery phase, replication series (UK1, UK2, and Texas) and meta-analysis data are reported.

**a**Common homozygous, heterozygous, and rare homozygous genotypes.

**b**OR for the A allele.

**P**two-sided P value.

OR and P values were based on a fixed-effect model. Cochran’s Q test \( P_{\text{het}} = 0.89 \) and \( I^2 = 0.0 \).

---

**Statistical Analysis**

**Testing association of single SNPs** For the NCI data, we performed an unconditional logistic regression analysis to test the additive effect of each SNP genotype on lung cancer risk by using PLINK software (31), adjusting for age (<= 50, 51-55, 56-60, 61-65, 66-70, 71-75, 76+), sex, study (ATBC, EAGLE, ATBC, CPS-II), 4 principal components derived based on EIGENSTRAT (32) to control population stratification and cigarettes smoked per day (<=10, 11-20, 21-30, 31-40, 41+), smoking duration in 10-year intervals, smoking status (former vs. current smoker) and, for former smokers, the number of years since quitting (1-5, 6-10, 11-20, 21-30, 30+). To assess the possibility of systematic bias caused by unadjusted confounding factors or cryptic relatedness, we computed the genomic control \( \lambda \) values on the basis of all SNPs passing the quality control filters in the NCI genome-wide association studies by histology and smoking status.
Testing association between inflammation pathway and risk of lung cancer

We tested the pathway-level association by integrating the HuGE scores that reflected different strength of association with lung cancer risks for each gene. To summarize in brief, we first derived the gene-wise P-values \( P_1, \ldots, P_k \) for \( k = 917 \) genes, adjusting for the number of SNPs mapping to each gene region and the linkage disequilibrium. We converted the gene-wise P-values into quantities \( Q_i \) of the \( X_i^2 \) distribution, that is, \( X_i^2(Q) = 1 - P_i \). We then tested if \( P_1, \ldots, P_k \) as a set deviated from the uniform distribution (no pathway-level association) by using statistic 
\[
T = \sum_{i=1}^{k} \log\left(1 - w_i + w_i e^{Q_i}\right),
\]
with \( w_i \) being the weights converted from the HuGE scores. A large value of \( T \) indicated a pathway-level association. The statistical significance was evaluated on the basis of 5,000 random permutations. The details of the pathway testing procedure are in the Supplementary Methods. We replicated the same analysis without applying the HuGE score by setting identical \( w_i \) across the genes.

Selection of SNPs for replication

We selected the 66 genes with gene-wise \( P \)-value <5% in the NCI squamous cell data, and chose the min-P SNP for each gene. Because physically close genes may be in linkage disequilibrium or have identical min-P, for each pair of min-P SNPs with high linkage disequilibrium (\( r^2 > 0.5 \)), we removed the SNP with a weaker association signal in the NCI squamous cell data. We compiled a list of 59 SNPs for replication, among which 4 SNPs were not genotyped in any replication samples. In summary, we chose 55 SNPs for replication. Details of SNP selection within the major histocompatibility complex region with long range linkage disequilibrium are in the Supplementary Methods.

Meta-analysis

Standard fixed effect meta-analysis was performed for the 55 SNPs in NCI, UK1, and Texas to derive the odds ratios, 95% CIs, and \( P \)-values. For the most significant SNP rs6489769, we performed meta-analysis of the results from NCI, UK1, UK2, and Texas datasets. Cochran’s Q statistic was calculated to test for heterogeneity with \( P \)-value \( P_{\text{het}} \).

SNP imputation in 12p13.33 locus

Prediction of the untyped SNPs in the 12p13.33 locus was performed with the use of IMpute2 (33), based on CEU HapMap Phase III haplotypes release 2 and 1000 Genomes Project. Unconditional logistic regression was performed to test the association between the imputed genotypic dosage and the trait using \( R \) adjusting for age, sex, study, principal components, cigarettes smoked per day, smoking duration, smoking status, and, for former smokers, the number of years since quitting as in the single SNP analysis. Linkage disequilibrium metrics between HapMap SNPs and association \( P \)-values were plotted using SNAP (34).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank the investigators involved in the EAGLE study, listed in http://eagle.cancer.gov. We thank the clinicians who took part in the GELCAPS consortium. The U.K. studies made use of genotyping data on the 1958 Birth Cohort, and these data were generated and generously supplied to us by Panagiotis Deloukas of the Wellcome Trust Sanger Institute. A full list of the investigators who contributed to the generation of the data is available from www .wccrc.org.uk.

This study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov).

Grant Support

This study was supported by the Intramural Research Program of NIH, National Cancer Institute, Division of Cancer Epidemiology and Genetics.

The studies contributing to the NCI genome-wide association studies, including The Environment and Genetics in Lung cancer Etiology (EAGLE), Prostate, Lung, Colon, Ovary Screening Trial (PLCO), and Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) studies were supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute (NCI), Division of Cancer Epidemiology and Genetics.

ATBC was also supported by U.S. Public Health Service contracts (N01-CN-45165, N01-RC-45035, and N01-RC-37004) from the NCI.

PLCO was also supported by individual contracts from the NCI to the University of Colorado Denver (N01-CN-25514), Georgetown University (N01-CN-25522), the Pacific Health Research Institute (N01-CN-25515), the Henry Ford Health System (N01-CN-25512), the University of Minnesota, (N01-CN-25513), Washington University (N01-CN-25516), the University of Pittsburgh (N01-CN-25511), the University of Utah (N01-CN-25524), the Marshfield Clinic Research Foundation (N01-CN-25518), the University of Alabama at Birmingham (N01-CN-75022), Westar, Inc. (N01-CN-25476), and the University of California, Los Angeles (N01-CN-25404).

The Cancer Prevention Study-II (CPS-II) Nutrition Cohort was supported by the American Cancer Society.

The NIH Genes, Environment and Health Initiative (GEI) partly funded DNA extraction and statistical analyses (HG-06-033-NCT-01 and RO1HL091172-01), genotyping at the Johns Hopkins University Center for Inherited Disease Research (U01HG004438 and NIH HHSN268200782096C), and study coordination at the GENEVA Coordination Center (U01HG004446) for the EAGLE study and part of the PLCO. Genotyping for the remaining part of PLCO and all ATBC and CPS-II samples were supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics.

The “Texas” study was supported by NIH grants R01CA55769, R01 CA127219, R01CA133996, and U19CA121197.

The UK1 and UK2 work was supported by Cancer Research UK (C1298/A8780 and C1298/A8362- Bobby Moore Fund for Cancer Research UK). Y. Wang was supported by an extramural NIH grant U19 CA14812701. We are also grateful to National Cancer Research Network, Helen Rollason Heal Cancer Charity, and Sanofi-Aventis and the NHS funding for the Royal Marsden Biomedical Research Centre.

Received September 26, 2011; revised November 21, 2011; accepted December 5, 2011; published OnlineFirst December 7, 2011.

REFERENCES

14. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.
Inherited Variation at Chromosome 12p13.33, Including *RAD52*, Influences the Risk of Squamous Cell Lung Carcinoma

Jianxin Shi, Nilanjan Chatterjee, Melissa Rotunno, et al.

*Cancer Discovery* 2012;2:131-139. Published OnlineFirst December 7, 2011.

<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: doi:10.1158/2159-8290.CD-11-0246</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at: <a href="http://cancerdiscovery.aacrjournals.org/content/suppl/2011/12/07/2159-8290.CD-11-0246.DC1">http://cancerdiscovery.aacrjournals.org/content/suppl/2011/12/07/2159-8290.CD-11-0246.DC1</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cited articles</th>
<th>This article cites 32 articles, 7 of which you can access for free at: <a href="http://cancerdiscovery.aacrjournals.org/content/2/2/131.full#ref-list-1">http://cancerdiscovery.aacrjournals.org/content/2/2/131.full#ref-list-1</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citing articles</td>
<td>This article has been cited by 9 HighWire-hosted articles. Access the articles at: <a href="http://cancerdiscovery.aacrjournals.org/content/2/2/131.full#related-urls">http://cancerdiscovery.aacrjournals.org/content/2/2/131.full#related-urls</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E-mail alerts</th>
<th>Sign up to receive free email-alerts related to this article or journal.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reprints and Subscriptions</td>
<td>To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a>.</td>
</tr>
<tr>
<td>Permissions</td>
<td>To request permission to re-use all or part of this article, contact the AACR Publications Department at <a href="mailto:permissions@aacr.org">permissions@aacr.org</a>.</td>
</tr>
</tbody>
</table>