Seizure 6-Like (SEZ6L) Gene and Risk for Lung Cancer

Ivan P. Gorlov,1 Peter Meyer,2,3,4 Triantafillos Liloglou,4 Jonathan Myles,7 Melanie Barbara Boettger,2,3 Adrian Cassidy,6 Luca Girard,5 John D. Minna,1 Reiner Fischer,7 Stephen Duffy,7 Margaret R. Spitz,1 Karl Haeussinger,2 Stefan Kammerer,5 Charles Cantor,5 Rainer Dierkesmann,1 John K. Field,6 and Christopher I. Amos1,4

1Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas; 2Department for Respiratory Medicine, Asklepios Specialist Hospitals, Munich-Gauting, Germany; 3Genefinder Technologies Ltd.; 4Institute of Molecular Medicine, Hospital, Stuttgart-Gerflingen, Germany; and 5Hamon Center for Therapeutic Oncology Research, Simmons Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas

Abstract

DNA pooling in combination with high-throughput sequencing was done as a part of the Sequenom-Geneﬁnder project. In the pilot study, we tested 83,715 single nucleotide polymorphisms (SNP), located primarily in gene-based regions, to identify polymorphic susceptibility variants for lung cancer. For this pilot study, 369 male cases and 287 controls of both sexes (white Europeans of Southern German origin) were analyzed. The study identiﬁed a candidate region in 22q12.2 that contained numerous SNPs showing signiﬁcant case-control differences and that coincides with a region that was shown previously to be frequently deleted in lung cancer cell lines. The candidate region overlies the seizure 6-like (SEZ6L) gene. The pilot study identiﬁed a polymorphic Met430lle substitution in the SEZ6L gene (SNP rs663048) as the top candidate for a variant modulating risk of lung cancer. Two replication studies were conducted to assess the association of SNP rs663048 with lung cancer risk. The M. D. Anderson Cancer Center study included 289 cases and 291 controls matched for gender, age, and smoking status. The Liverpool Lung Project (a United Kingdom study) included 248 cases and 233 controls. Both replication studies showed an association of the rs663048 with lung cancer risk. The homozygotes for the variant allele had more than a 3-fold risk compared with the wild-type homozygotes [combined odds ratio (OR), 3.32; 95% conﬁdence interval (CI), 1.81–7.21]. Heterozygotes also represented nearly a 2-fold elevated and statistically signiﬁcant (P = 0.004) level of SEZ6L expression in tumor samples compared with normal lung tissues. In conclusion, the results of these studies representing 906 cases compared with 811 controls indicate a role of the SEZ6L Met430lle polymorphic variant in increasing lung cancer risk. [Cancer Res 2007;67(17):8406–11]

Introduction

Although up to 90% of lung cancers are attributable to smoking, only a small fraction of smokers develop lung cancer over their lifetimes (1), suggesting that genetic variation may contribute to lung cancer susceptibility (2). Results of segregation analyses suggest that rare autosomal dominant polymorphisms may explain susceptibility to early-onset lung cancer; however, only a minority of lung cancer cases can be explained by the presence of such variants (3–6). There is also a rapidly expanding body of literature on the association of common, low-penetrance genes with lung cancer risk (5–8). According to the latest update of Cancer Genetics Web database, more than forty genes may be involved in susceptibility to, and progress of lung cancer.

Identification of genetic factors modulating lung cancer risk requires a combination of effective genotyping technologies with an appropriate and efﬁcient study design. Sequenom (San Diego, CA) has developed a DNA analysis platform, capable of high-throughput genotyping with pooled DNA allele frequency analysis. Using this approach, Sequenom implemented a Genetics Discovery platform with dense genome-wide single nucleotide polymorphism (SNP) markers (7, 8). A hypothesis-free approach using allele frequency estimates of many thousands (for lung cancer, 83,715 SNPs) of SNPs was used as a first step (pilot study) in identifying potentially relevant genetic variants. Significant SNPs identiﬁed in this first step were then individually genotyped and validated in replication studies using independent samples. The efﬁciency of this strategy has been shown by the rediscovery of genes shown
Materials and Methods

Sequenom-Genefinder pilot study (Southern Germany). Lung cancer cases for the pilot study were recruited from the Departments for Respiratory Medicine and Thoracic Surgery, Schillerhöhe Specialist Hospital, and the Department for Respiratory Medicine, Asklepios Specialist Hospitals. Controls were sampled from patients with nonmalignant disease at the same hospitals. The final sample consisted of 369 male cases and 287 controls of both genders, all with a positive history of tobacco smoke exposure. A total of 83,715 SNPs, mostly in gene-based regions, were used in the analysis. Epidemiologic data were collected by personal interview. Table 1 provides a description of cases and controls used in the pilot study.

The M. D. Anderson Cancer Center replication study (United States). Cases and controls for the M. D. Anderson Cancer Center (MDACC) study were recruited from an ongoing lung cancer case-control study enrolling patients with newly diagnosed, untreated lung cancer at The University of Texas MDACC (Houston, TX). The study has been described in detail previously (14). Control subjects were recruited from the largest private multispecialty physician group in the Houston metropolitan area. The controls did not have a previous diagnosis of any type of cancer and were frequency matched to the cases on age (±5 years), sex, ethnicity, and smoking status (current, former, and never). There were no never smokers in this subgroup (defined as a person who smoked fewer than 100 cigarettes in his/her lifetime). A former smoker had quit smoking for at least 1 year before the interview. Pack-years were defined as the number of cigarettes smoked per day divided by 20 and then multiplied by the number of years smoked. Epidemiologic data were collected by personal interview. In total, 289 lung cancer cases and 291 controls were included in the analysis. Table 1 provides a description of cases and controls used in the MDACC.

The Liverpool Lung Project replication study (United Kingdom). The lung cancer case-control data were derived from the Liverpool Lung Project (LLP) from an ongoing molecular epidemiologic study of lung cancer in Liverpool, United Kingdom (12). Histologically or cytologically confirmed lung cancer cases with primary tumors were recruited from participating chest clinics. Population controls were selected from registers of General Practitioners in Liverpool to ensure similar age-sex distributions to the cases. In all studies, a standardized questionnaire was used to determine basic demographic characteristics in addition to details on smoking history, lifetime residence and occupation, history of lung diseases, family history of cancer in first-degree relatives, and exposure to environmental tobacco smoke. Smoking status was defined as in the U.S. study. In total, 248 lung cancer cases and 233 controls were included in this analysis. Table 1 provides a description of cases and controls used in the LLP study.

SNP markers and genotyping. Genomic DNA was extracted from blood peripheral leukocytes by using the Qiagen DNA blood mini kit (Qiagen) according to the manufacturer's instruction. DNA pools were formed by combining equimolar amounts of individual samples as described elsewhere (13). For the pilot study, one pool of 369 cases and one pool of 287 controls, respectively, were constructed. For assays carried out on sample pools, 25 ng of a 5-ng/L pool were used for PCRs. All PCR and MassEXTEND reactions were conducted using standard conditions. Relative allele frequency estimates were derived from calculations based on the area under the peak of mass spectrometry measurements from four analyte aliquots (14). Tests of association between disease status and each SNP were carried out as previously discussed (15). When three or more replicate measurements of a SNP were available, the corresponding variance component was estimated from the data. Otherwise, the following historical laboratory averages were used to calculate sources of variability: pool formation, 5.0 × 10⁻²; PCR/mass extension, 1.7 × 10⁻²; and chip measurement, 1.0 × 10⁻⁴. The same procedure was used for individual genotyping except 2.5 ng DNA was used and only one mass spectrometry measurement was taken. The following gene-specific primers were used to genotype rs6636048: the forward PCR primer was 5'-TGGGCTATGAGCTC-CAGGG-3'; the reverse PCR primer was 5’-TGGCGCTTGGAGCATTGAT-3'; and extend primer was 5’-GGCTCTAGGAGGCTAAGAT-3’. The Sequenom-Genefinder pilot study included 83,715 SNPs selected based on their location within a gene region (including the coding region plus additional 10 kb at the both ends) and minor allele frequency (MAF) from a total of 125,799 experimentally validated polymorphic variations (7, 8, 10). In the first step, one PCR and primer extension reaction was carried out for 83,715 SNPs on each pool (case and control). In the second step, 4,293 SNPs (-5%) with the most statistically significant associations were reremasured in triplicate on each DNA pool. In the third step, the 301 most significant SNPs (-7%) from step two were individually genotyped in each sample. A total of 160 SNP markers were identified with statistically significant differences between cases and controls (P < 0.05) after individual genotyping in the German pilot study and were then genotyped in the MDACC and LLP replication samples.

Expression of the seizure 6-like gene. Besides analyzing the effect of the Met430ile variant on lung cancer risk, we also compared the seizure 6-like (SEZ6L) expression level in normal versus cancer cell lines (our data) and in primary tumors versus normal lung tissues [data from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database]. Affymetrix HG-U133A and HG-U133B high-density oligonucleotide microarrays were also used to evaluate expression of the SEZ6L gene in cell lines. Gene expression profiling was done on a panel of

| Table 1. Description of the cases and controls from the pilot and the two replication studies |
|----------------------------------|--------|--------|--------|--------|
|                                  | Sequenom-Genefinder pilot study (Southern Germany) | MDACC (United States) | LLP (United Kingdom) |
|                                  | Cases, n = 369 (%) | Controls, n = 287 (%) | Cases, n = 289 (%) | Controls, n = 291 (%) | Cases, n = 248 (%) | Controls, n = 233 (%) |
| Gender                          | Male 369 (100) | 196 (68.29) | 164 (56.75) | 158 (54.30) | 159 (64.11) | 158 (67.81) |
|                                | Female 0 | 91 (31.71) | 125 (43.25) | 133 (45.70) | 89 (35.89) | 75 (32.19) |
| Smoking status                 | Never 0 | 0 | 0 | 0 | 13 (5.24) | 66 (28.33) |
|                                | Former 0 | 0 | 168 (58.13) | 173 (59.45) | 105 (42.34) | 121 (51.93) |
|                                | Current 368 (100) | 287 (100) | 121 (41.87) | 118 (40.55) | 130 (52.42) | 46 (19.74) |
| Age (y)                        | Mean (SD) 65.4 (6.3) | 63.0 (8.3) | 62.7 (9.94) | 62.2 (10.09) | 65.9 (9.2) | 65.4 (8.4) |
We found that the position and size of the candidate region coincides well with the position and size of the SEZ6L gene. The candidate region occupied the distal part of the 428-kb region frequently deleted in lung cancer cell lines (18). The deletion contains two genes: SEZ6L and MYO18B (18). No significant SNPs were detected in the MYO18B region (Fig. 1).

The MDACC replication study. SNP markers with allelic frequencies found to be significantly different between cases and controls in the pilot study were validated in the MDACC replication study. The total number of SNP markers genotyped in the MDACC study was 147. Thirteen SNPs exhibited significant (P < 0.01) deviation from Hardy-Weinberg equilibrium in controls and were excluded from the analysis. The remaining 133 SNPs were analyzed to estimate their association with lung cancer. Analysis of all histologic types together enabled an identification of six cancer-associated SNP markers. In NSCLC, the same SNP markers were identified as showing significant associations with lung cancer as well as two additional SNPs (Table 2).

To further prioritize the significant SNPs listed in Table 2, we annotated the SNPs by using available data that potentially can help to identify causal SNPs (Table 3). The SNP rs663048 was the only nonsynonymous (Met430Ile) SNP in the list. Sorting Intolerant from Tolerant (SIFT; ref. 19) and Polymorphism Phenotyping (PolyPhen; ref. 20) software were used to predict the functional effect of the Met430Ile amino acid substitution. Both algorithms predicted that the Met430Ile is a functional (protein disturbing) variant; therefore, our further analysis was concentrated on rs663048 polymorphism.

To estimate the relative risk associated with the significant SNPs, we used a logistic regression model. Table 4 shows the predicted OR adjusted for age, intensity of smoking (pack-year), and gender. There was no significant deviation from Hardy-Weinberg equilibrium in both replication samples neither in controls nor in cases. Homozygotes for the more frequent allele were used as reference groups. OR values ranged from 1.29 to 1.72, with four of them being statistically significant. The OR for the SNP rs663048 was 1.48 (P = 0.03) with rare genotypes being associated with increased risk for NSCLC.

The LLP replication study (United Kingdom). The LLP replication sample was used to further validate the association

---

between rs663048 and lung cancer risk. In this study, we also found a significant association between Met430Ile polymorphism and lung cancer with cases having a higher frequency of the variant T/T genotype (8.9 ± 1.9%) than the controls (3.0 ± 1.1%; \( \chi^2 = 7.4; \text{df} = 1; P = 0.006 \)). We observed a 3.8-fold increased risk of lung cancer in individuals who carried the rs663048 null genotype T/T compared with those who carried the rs663048 more common (wild-type) genotype G/G (95% CI, 1.40–10.42) after adjusting for age, sex, and smoking (Table 4).

Expression of the SEZ6L in normal and lung tumor cell lines. Supplementary Fig. S1 shows the expression level of SEZ6L in 9 normal lung cell lines, 54 NSCLC, and 22 SCLC cell lines. The average expression signal in controls was 6.2 (\( n = 9 \)) compared with the average expression signal in the NSCLC cell lines of 7 (\( n = 54 \)) and in SCLC cell lines of 8.8 (\( n = 22 \)). A nonparametric Mann-Whitney U test was significant for all pairwise comparisons. The smallest (but still very significant) difference was found between the normal lung cell lines and NSCLC cell lines (\( Z = 3.3; P = 0.001 \)). We also found significant differences in expression level of GATA2, with the expression being higher in cancer lines compared with controls (Mann-Whitney \( U \) test, \( Z = 3.9; P = 0.0005 \)). Other candidate genes listed in Table 3 did not show differential expression between lung cancer and normal lung cell lines. SNP genotypes were not available for this analysis.

Expression of the SEZ6L in primary tumor and normal lung tissues. We used the NCBI GEO database containing microarray data on the genome wide assessment of gene expression. Using the key words “lung AND cancer OR tumor,” we have identified several entries with data on gene expression. The GDS619 data set (platform GPL962: CHUGAH1K) was most appropriate for our goal to compare SEZ6L expression in normal and tumor tissues. This data set contains data on gene expression in SCLC, adenocarcinomas, and normal tissues. We found that the average log 2-transformed SEZ6L expression value was significantly higher in adenocarcinoma compared with normal lung tissues (0.34 ± 0.13 versus –0.38 ± 0.15; two sided \( t \) test = 3.1; \( df = 29; P = 0.004 \)). The expression of SEZ6L was also insignificantly higher in SCLC compared with normal tissues (Student’s \( t \) test = 1.4; \( df = 43; P = 0.15 \)). The variance of expression values in NSCLC sample was significantly higher compared with the normal tissues (Var = 0.07 among controls versus Var = 0.42 among SCLCs; \( F = 6.2; P = 0.001 \)). This result together with our analysis of the SEZ6L expression in cell lines suggests an increased level of the SEZ6L expression in lung tumor compared with normal lung tissues.

Table 2. SNP markers found to be significant by analysis of all histologic types or NSCLCs (MDACC)

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Chromosome</th>
<th>Alleles</th>
<th>n</th>
<th>MA</th>
<th>MAF</th>
<th>SE</th>
<th>( \chi^2 )</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1037023</td>
<td>13</td>
<td>C/T</td>
<td>288</td>
<td>C</td>
<td>0.2</td>
<td>0.02</td>
<td>250</td>
<td>C</td>
</tr>
<tr>
<td>rs1906520</td>
<td>3</td>
<td>G/T</td>
<td>278</td>
<td>T</td>
<td>0.16</td>
<td>0.02</td>
<td>251</td>
<td>T</td>
</tr>
<tr>
<td>rs190697</td>
<td>14</td>
<td>A/C</td>
<td>286</td>
<td>A</td>
<td>0.08</td>
<td>0.01</td>
<td>241</td>
<td>A</td>
</tr>
<tr>
<td>rs3803</td>
<td>3</td>
<td>C/T</td>
<td>273</td>
<td>T</td>
<td>0.23</td>
<td>0.02</td>
<td>253</td>
<td>T</td>
</tr>
<tr>
<td>rs663048</td>
<td>22</td>
<td>G/T</td>
<td>287</td>
<td>T</td>
<td>0.21</td>
<td>0.02</td>
<td>245</td>
<td>T</td>
</tr>
<tr>
<td>rs422679</td>
<td>17</td>
<td>C/T</td>
<td>281</td>
<td>T</td>
<td>0.39</td>
<td>0.02</td>
<td>250</td>
<td>T</td>
</tr>
<tr>
<td>rs765707</td>
<td>12</td>
<td>C/T</td>
<td>277</td>
<td>T</td>
<td>0.06</td>
<td>0.01</td>
<td>247</td>
<td>T</td>
</tr>
<tr>
<td>rs1057050</td>
<td>10</td>
<td>G/A</td>
<td>288</td>
<td>A</td>
<td>0.05</td>
<td>0.01</td>
<td>254</td>
<td>A</td>
</tr>
</tbody>
</table>

Abbreviation: MA, minor allele.
*SNPs were significant only in NSCLCs.

Table 3. Annotations of the SNPs that have been shown to be significantly associated with lung cancer risk

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>SNP location</th>
<th>Chromosome</th>
<th>Chromosome position</th>
<th>Gene</th>
<th>Description</th>
<th>Cancer-related*</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1037023</td>
<td>Intron</td>
<td>13q14</td>
<td>24854588</td>
<td>FLJ10094</td>
<td>Hypothetical protein</td>
<td>No</td>
</tr>
<tr>
<td>rs1906520</td>
<td>7 kb Downstream CD200</td>
<td>3q13</td>
<td>113580534</td>
<td>CD200</td>
<td>Immunoglobulin</td>
<td>No</td>
</tr>
<tr>
<td>rs190697</td>
<td>2 kb Upstream SETP2</td>
<td>1q22</td>
<td>50961888</td>
<td>SETP2</td>
<td>Pseudogene</td>
<td>No</td>
</tr>
<tr>
<td>rs3803</td>
<td>5'-UTR</td>
<td>3q21</td>
<td>129682070</td>
<td>GATA2</td>
<td>Zinc finger, hematopoiesis</td>
<td>No</td>
</tr>
<tr>
<td>rs663048</td>
<td>Missense</td>
<td>22q12</td>
<td>25925077</td>
<td>SEZ6L</td>
<td>Transmembrane protein</td>
<td>Yes</td>
</tr>
<tr>
<td>rs422679</td>
<td>Intron</td>
<td>17p13</td>
<td>8226123</td>
<td>RPL26</td>
<td>Ribosomal protein L26</td>
<td>No</td>
</tr>
<tr>
<td>rs765707</td>
<td>2 kb Downstream OR7A19P</td>
<td>12q12</td>
<td>45288613</td>
<td>OR7A19P</td>
<td>Olfactory receptor</td>
<td>No</td>
</tr>
<tr>
<td>rs1057050</td>
<td>3'-UTR</td>
<td>10q24</td>
<td>104132284</td>
<td>GFB1</td>
<td>Guanine exchange factor</td>
<td>No</td>
</tr>
</tbody>
</table>

Abbreviation: UTR, untranslated region.
*An exhaustive search of literature was conducted to estimate potential relationship of the SNP to cancer.
Expression of \textit{SEZ6L} had different levels of expression of \textit{SEZ6L} with increased risk for lung cancer, suggesting that loss of normal \textit{SEZ6L} function may be a risk factor for lung cancer. This is consistent with the finding that \textit{SEZ6L} region is often deleted in lung cancer cell lines (18). On the other hand, we found that expression of \textit{SEZ6L} is elevated in lung cancer cell lines. One explanation might be that \textit{SEZ6L} is both a tumor marker and a variant affecting lung cancer susceptibility. Loss of normal \textit{SEZ6L} function is a risk factor for development of lung cancer; however, when lung cancer is caused by factors other than loss-of-\textit{SEZ6L} function, expression of the \textit{SEZ6L} is adaptively up-regulated to suppress tumorigenesis.

Several lines of evidence support the hypothesis that \textit{SEZ6L} might modulate lung cancer risk. First, frequent allelic losses on 22q in NSCLCs have been reported, indicating the presence of tumor suppressor gene(s) on that chromosome arm (18). Cloning of the breakpoints revealed a 400-kb deletion containing the \textit{SEZ6L} and \textit{MYO18B} genes (18, 27). A study conducted by Suzuki et al. (28) suggests that \textit{SEZ6L} gene may also influence development and progression of colorectal cancer. The authors found that \textit{SEZ6L} was one of the few genes highly hypermethylated in primary colorectal tumors.

In the pilot study, we populated the candidate region with 35 SNPs and found that markers located in the \textit{SEZ6L} gene region show a strong association with lung cancer; however, no significant associations were found in the neighboring \textit{MYO18B} gene. Applying a sliding window of five neighboring SNPs revealed a peak of $-\log_{10}(P)$ values that coincides with the position of the \textit{SEZ6L} gene. We found that the principal contributor to the peak was rs663048. The association of this SNP with lung cancer risk was verified in two independent replication studies. The rs663048 SNP is a Met430lle amino acid substitution that has been predicted to be functional by both SIFT and PolyPhen, suggesting that this amino acid substitution is protein disturbing.

Nishioke et al. (18) found that 95% (43 of 45) of primary tumor samples carry the Met430lle mutation. The authors did not estimate the frequency of the variant in controls. We found that 38% of controls and 48% of cases carry at least one variant allele. This suggests that $\sim$ 40% of tumors may carry a somatic Met430lle mutation. If we consider that according to the HapMap the frequency of the Met430lle polymorphism is lower in Japanese than in Caucasians, the percentage of accumulated somatic Met430lle may actually be higher.

We found that homozygotes for the variant allele had 3-fold higher lung cancer risk compared with the normal variant homozygotes. Lung cancer risk was also significantly elevated in

### Table 4. Risk estimates for lung cancer associated with rs663048 genotype in MDACC and LLP replication samples

<table>
<thead>
<tr>
<th>Sample</th>
<th>Genotype</th>
<th>Case, n (%)</th>
<th>Control, n (%)</th>
<th>OR* (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDACC</td>
<td>GG</td>
<td>137 (51.7)</td>
<td>176 (62.4)</td>
<td>1 (Reference)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TG</td>
<td>102 (38.5)</td>
<td>100 (34.8)</td>
<td>1.18 (0.87–2.04)</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>TT</td>
<td>26 (9.8)</td>
<td>11 (3.8)</td>
<td>2.84 (1.36–8.54)</td>
<td>0.008</td>
</tr>
<tr>
<td>LLP</td>
<td>GG</td>
<td>125 (50.4)</td>
<td>133 (57.1)</td>
<td>1 (Reference)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TG</td>
<td>91 (36.7)</td>
<td>85 (36.5)</td>
<td>1.23 (0.78–1.96)</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>TT</td>
<td>22 (8.9)</td>
<td>7 (3.0)</td>
<td>3.81 (1.40–10.42)</td>
<td>0.009</td>
</tr>
</tbody>
</table>

*ORs were adjusted for age, pack-year, and gender.

## Discussion

There are two major approaches used to identify cancer-related genes: (a) candidate gene approach and (b) genome-wide scan approach. The candidate gene approach is based on prior data indicating that a gene(s) is involved in cancer. For example, it has been shown that increased lung cancer risk correlates with decreased nucleotide excision repair (NER) capacity (21–23). Targeting genes involved in NER has identified polymorphisms in NER genes associated with elevated lung cancer risk (24–26). The advantage of the candidate gene approach is that the number of candidate genes is limited; therefore, the number of false positives among significant associations is also expected to be low. A disadvantage of the approach is that genes without prior data on associations will not be included in the analysis and, therefore, cannot be identified by the candidate gene approach. In the genome-wide scan approach, potentially all genes in the genome are targeted, allowing the identification of novel cancer-related genes. A disadvantage of the genome-wide scan is the large number of independent tests so that a large sample size is required to distinguish between true- and false-positive associations.

In our study, we have combined elements of both approaches. At the first step, more than 83,000 SNP markers covering the coding region of the whole genome were analyzed. This analysis yielded many candidate SNPs showing significant associations with lung cancer. Significant SNPs identified in the pilot study are a mix of false positives and true associations. For selected SNPs in \textit{SEZ6L} candidate region, two independent replication studies were then conducted to identify true associations. The MDACC replication study yielded eight SNPs showing associations with lung cancer risk. We then used additional SNP-related information to identify most promising genes. As a result of this analysis, the \textit{SEZ6L} gene emerged as a top candidate gene to be associated with lung cancer. It is interesting that based only on the significance of the $\chi^2$ test, this gene was number five in the list. The LLP replication study further validated the association of SNP rs663048 in \textit{SEZ6L} with lung cancer risk.

An analysis of the expression of \textit{SEZ6L} gene showed different expression of \textit{SEZ6L} in normal and NSCLC SCLC cell lines. Interestingly, we found that the two histologic types of lung cancer had different levels of expression of \textit{SEZ6L}. The average expression signal in NSCLC was $7.0 \pm 0.1$ and in SCLC $8.8 \pm 0.2$ (Mann-Whitney $U$ test, $Z = 6.0; P < 0.001$). It needs to be noted that there is an apparent inconsistency between the results of the analysis of the expression of \textit{SEZ6L} and the results of the association studies. In the MDACC and United Kingdom samples, the variant allele (that was predicted to be protein disturbing based on analysis of the protein structure and evolutionary conservation) was associated with increased risk for lung cancer, suggesting that loss of normal \textit{SEZ6L} function may be a risk factor for lung cancer. This is consistent with the finding that \textit{SEZ6L} region is often deleted in lung cancer cell lines (18). On the other hand, we found that expression of \textit{SEZ6L} is elevated in lung cancer cell lines. One explanation might be that \textit{SEZ6L} is both a tumor marker and a variant affecting lung cancer susceptibility. Loss of normal \textit{SEZ6L} function is a risk factor for development of lung cancer; however, when lung cancer is caused by factors other than loss-of-\textit{SEZ6L} function, expression of the \textit{SEZ6L} is adaptively up-regulated to suppress tumorigenesis.

Several lines of evidence support the hypothesis that \textit{SEZ6L} might modulate lung cancer risk. First, frequent allelic losses on 22q in NSCLCs have been reported, indicating the presence of tumor suppressor gene(s) on that chromosome arm (18). Cloning of the breakpoints revealed a 400-kb deletion containing the \textit{SEZ6L} and \textit{MYO18B} genes (18, 27). A study conducted by Suzuki et al. (28) suggests that \textit{SEZ6L} gene may also influence development and progression of colorectal cancer. The authors found that \textit{SEZ6L} was one of the few genes highly hypermethylated in primary colorectal tumors.

In the pilot study, we populated the candidate region with 35 SNPs and found that markers located in the \textit{SEZ6L} gene region show a strong association with lung cancer; however, no significant associations were found in the neighboring \textit{MYO18B} gene. Applying a sliding window of five neighboring SNPs revealed a peak of $-\log_{10}(P)$ values that coincides with the position of the \textit{SEZ6L} gene. We found that the principal contributor to the peak was rs663048. The association of this SNP with lung cancer risk was verified in two independent replication studies. The rs663048 SNP is a Met430lle amino acid substitution that has been predicted to be functional by both SIFT and PolyPhen, suggesting that this amino acid substitution is protein disturbing.

Nishioke et al. (18) found that 95% (43 of 45) of primary tumor samples carry the Met430lle mutation. The authors did not estimate the frequency of the variant in controls. We found that 38% of controls and 48% of cases carry at least one variant allele. This suggests that $\sim$ 40% of tumors may carry a somatic Met430lle mutation. If we consider that according to the HapMap the frequency of the Met430lle polymorphism is lower in Japanese than in Caucasians, the percentage of accumulated somatic Met430lle may actually be higher.

We found that homozygotes for the variant allele had 3-fold higher lung cancer risk compared with the normal variant homozygotes. Lung cancer risk was also significantly elevated in
References


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.
Seizure 6-Like (SEZ6L) Gene and Risk for Lung Cancer

Ivan P. Gorlov, Peter Meyer, Triantafillos Liloglou, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/67/17/8406

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2007/08/28/67.17.8406.DC1

Cited articles
This article cites 27 articles, 8 of which you can access for free at:
http://cancerres.aacrjournals.org/content/67/17/8406.full#ref-list-1

Citing articles
This article has been cited by 4 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/67/17/8406.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerres.aacrjournals.org/content/67/17/8406.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.