Fine Mapping of Chromosome 6q23-25 Region in Familial Lung Cancer Families Reveals RGS17 as a Likely Candidate Gene

Ming You,1 Daolong Wang,1 Pengyuan Liu,1 Haris Vikis,1 Michael James,1 Yan Lu,1 Yian Wang,1 Qiong Chen,1 Dongmei Jia,1 Yan Liu,1 Weidong Wen,1 Ping Yang,2 Zhifu Sun,2 Susan M. Pinney,3 Wei Zheng,4 Xiao-Ou Shu,4 Jirong Long,4 Yu-Tang Gao,5 Yong-Bing Xiang,5 Wong-Ho Chow,6 Nat Rothman,6 Gloria M. Petersen,6 Mariza de Andrade,6 Yanhong Wu,7 Julie M. Cunningham,2 Jonathan S. Wiest,6 Pamela R. Fain,8 Ann G. Schwartz,9 Luc Girard,10 Adi Gazdar,10 Colette Gaba,11 Henry Rothschild,12 Diptasri Mandal,12 Teresa Coons,13 Juwon Lee,3 Elena Kupert,3 Daniela Seminara,6 John Minna,10 Joan E. Bailey-Wilson,7 Christopher I. Amos,14 and Marshall W. Anderson3

Abstract

Purpose: We have previously mapped a major susceptibility locus influencing familial lung cancer risk to chromosome 6q23-25. However, the causal gene at this locus remains undetermined. In this study, we further refined this locus to identify a single candidate gene, by fine mapping using microsatellite markers and association studies using high-density single nucleotide polymorphisms (SNP).

Experimental Design: Six multigenerational families with five or more affected members were chosen for fine-mapping the 6a linkage region using microsatellite markers. For association mapping, we genotyped 24 6q-linked cases and 72 unrelated noncancer controls from the Genetic Epidemiology of Lung Cancer Consortium resources using the Affymetrix 500K chipset. Significant associations were validated in two independent familial lung cancer populations: 226 familial lung cases and 313 controls from the Genetic Epidemiology of Lung Cancer Consortium, and 154 familial cases and 325 controls from Mayo Clinic. Each familial case was chosen from one high-risk lung cancer family that has three or more affected members.

Results: A region-wide scan across 6q23-25 found significant association between lung cancer susceptibility and three single nucleotide polymorphisms in the first intron of the RGS17 gene. This association was further confirmed in two independent familial lung cancer populations. By quantitative real-time PCR analysis of matched tumor and normal human tissues, we found that RGS17 transcript accumulation is highly and consistently increased in sporadic lung cancers. Human lung tumor cell proliferation and tumorigenesis in nude mice are inhibited upon knockdown of RGS17 levels.

Conclusion: RGS17 is a major candidate for the familial lung cancer susceptibility locus on chromosome 6q23-25.

Lung cancer can occur sporadically in people with no known family history of lung cancer or it can be familial, occurring in multiple members of the same family. Initial evidence of a genetic basis for susceptibility to lung cancer came from observations of individual differences in susceptibility to the same environmental risk factors (1–3), familial aggregation of lung cancer after accounting for personal smoking (4), increased risk of lung cancer mortality in siblings (5), and...
Lung cancer is the leading cause of cancer morbidity and mortality in developed nations. Although tobacco smoke is the main environmental influence, a genetic component to susceptibility also exists. A previous study has mapped a major susceptibility locus influencing familial lung cancer risk to chromosome 6q23–25. However, the susceptibility gene at this locus remains unresolved. Through a combination of genetic fine mapping and association studies we identified RGS17 as the major candidate susceptibility gene. Knowledge of the molecular and genetic mechanisms of lung cancer will improve diagnosis and treatment in the future.

**Translational Relevance**

RGS17 is a candidate lung cancer susceptibility gene.

The goal of this study, which was to identify candidate genes for lung cancer susceptibility on chromosome 6q, was realized with the identification of RGS17 as a familial lung cancer associated gene. Furthermore, this study shows that RGS17 is commonly overexpressed in lung tumors and that expression of RGS17 induces a proliferative phenotype in lung tumor cells.

**Materials and Methods**

Fine mapping for the 6q linkage. Six multigenerational families (with five or more affected members) were chosen for fine-mapping the 6q linkage. The LOD scores for these families in the initial linkage study at the D6S2436 position were 0.83 for family 12, 0.94 for family 33, 0.871 for family 35, 0.678 for family 47, 0.24 for family 100, and 0.6 for family 102 (12). The 26 microsatellite markers (including 7 from the original linkage study) used for mapping were D6S2437, D6S1040, D6S262, D6S1038, D6S1272, D6S1009, D6S250, D6S1055, D6S1848, D6S971, G15833, D6S960, D6S495, D6S2442, D6S2436, D6S442, D6S969, D6S1033, D6S953, D6S1008, D6S1277, D6S1273, D6S392, D6S297, D6S1697, and D6S1027. Genotyping was done as previously described, and LOD scores for individual families were estimated with Simwalk2 under the autosomal dominant model as used previously (12). Haplotypes were inferred with Simwalk2 (13, 14) for all genotyped affected members from each of the six families, with the largest common haplotypes indicated. The haplotype shared by affected members within families varied in length and position.

**Study samples.** For high-density SNP association mapping, we used 24 6q-linked cases (from pedigrees with a positive LOD at 155 cm) and 72 unrelated noncancer controls (both spouses and nonspouses) from the GELCC collection. To ensure genetic independence among subjects, one case was selected from each family, although multiple controls from the same families were allowed as long as they had no blood relationship with any selected cases from the relevant families. In this initial screen, 87.5% and 12.5% of the cases are current/former smokers and nonsmokers, respectively, with an average age of 61.2 y (±11.5). In controls, 67.5% and 29.7% of the subjects are current/former smokers and nonsmokers, respectively, with an average age of 70.9 y (±12.8).

For the GELCC familial validation study, we genotyped a separate sample of 226 familial lung cases and 313 controls from GELCC resources. Each familial case was chosen from one high-risk lung cancer family that has three or more affected members. Most of these families did not have linkage information ascertained due to the paucity of biospecimens. In these samples, only a blood sample of one affected member was collected. It is anticipated that some of these families are not 6q-linked, which may dilute the association. Noncancer controls were obtained from a combination of GELCC resources, the Coriell Institute for Medical Research (Camden, NJ), and the Familial Medical Monitoring Program. In order to minimize the possible effects of cigarette smoking and age, we selected mainly smokers of older age as controls, except spouse controls. In the cases, 83.6% and 15.5% of the subjects are current/former smokers and nonsmokers, respectively, with an average age of 61.6 y (±10.6). In the controls, 74.7% and 22.4% of the subjects are current/former smokers and nonsmokers, respectively, with an average age of 75.0 y (±9.4). To maintain the homogeneous population samples, only Caucasians from the GELCC, the Coriell Institute for Medical Research, and the Familial Medical Monitoring Program collections were used for the association analysis in the initial screen and the validation study. We detected no population stratification in these GELCC subjects, as shown by linkage agglomerative clustering implemented in PLINK software (15). For the Mayo Clinic validation study, we genotyped 154 familial lung cancer cases and 325 noncancer controls from Mayo Clinic (Rochester, MN). These familial lung cancer cases were chosen from families that have three or more first-degree relatives with lung cancer. These samples are part of the Mayo Clinic Lung Cancer Cohort (MCLCC; NIH CA77118, CA80127, and CA84354) collected from an ongoing case-control study (16). In the cases, 85.7% and 14.3% are current/former smokers and nonsmokers, respectively, with an average age of 72.9 y (±9.5). In the controls, 65.2% and 19.7% are current/former smokers and nonsmokers, respectively, with an average age of 75.5 y (±7.3). The basic characteristics of the study subjects with familial lung cancer and the corresponding controls are detailed in Supplemental Table S1.

For the sporadic validation studies, samples from the MCLCC study (Mayo) were used, including 553 sporadic cases and 627 controls (16, 17). Shanghai samples were collected from the Shanghai Women’s Health Study (Shanghai, China), an ongoing prospective cohort study of approximately 75,000 adult women, including 197 sporadic female cases and 410 female controls (18). All Chinese samples are female.
nonsmokers, which may be particularly informative in studying the main gene effects. The basic characteristics of the Mayo Clinic and the Shanghai study subjects with sporadic lung cancer and the corresponding controls are detailed in Supplemental Table S2.

**SNP genotyping.** The Affymetrix 500K SNP chipset, including two chips (Nsp and Sty), was used to genotype 24 cases and 72 controls. SNP genotyping was done by the Vanderbilt University Microarray Shared Resource at Nashville, TN, following the Affymetrix protocol (www.affymetrix.com). A confidence score of 0.33 was used for genotype calling, using the Bayesian Robust Linear Model with Mahalanobis algorithm (19), at which an average call rate of 96.9% was obtained across all case and control samples. The SNP-Genotyping Core at Washington University using the Sequenom platform carried out genotyping for the validation samples.

**Statistical analysis.** Hardy-Weinberg equilibrium for each SNP was examined with the software hweStrata implementing an exact test method proposed by Schaid and colleagues (20), under stratum number K = 1. SNPs with an exact \( P \leq 0.01 \) in controls were excluded in the association analysis. Due to the relatively small sample sizes, Fisher’s exact test was used to assess associations between SNPs and lung cancer. Conforming to the linkage study (12), the autosomal dominant model was used for coding marker genotypes, which involved two steps: (a) identifying, for each SNP, the putative “disease allele” (denoted as \( D \)) as the one that is more common in cases than in controls; and (b) forming a \( 2 \times 2 \) contingency table by combining genotypes \( DD \) and \( Dd \) into one group and taking genotype \( dd \) as the other group. For validation samples, the same Fisher’s exact test with the autosomal dominant model described above was applied to the association analysis. Results from multiple case-control groups were combined using a Mantel-Haenszel model in which the groups were allowed to have different population frequencies for alleles and genotypes but were assumed to have common relative risks (21).

**Gene expression study by quantitative real-time-PCR.** The GELCC has not collected paired tumor and normal tissues from GELCC cases and thus does not have RNA for quantitative real-time PCR; therefore, we employed available sporadic adenocarcinoma and adenosquamous carcinoma tumor samples for \( RGS17 \) expression analysis. RNA from 13 paired lung tumors and normal tissues was obtained from the Tissue Procurement Core at Washington University, St. Louis. cDNA from normal human tissues was obtained from BD Biosciences. Quantitative real-time PCR was conducted using the method as described previously (22). Briefly, two micrograms of total RNA per sample were converted to cDNA using the SuperScript First-Strand Synthesis system for real-time PCR (Invitrogen). Quantitative real-time PCR assay was done using the SYBR Green PCR Master Mix (Applied Biosystems). One microliter of cDNA was added to a 25-μL total reaction mixture containing water, SYBR Green PCR Master Mix, and primers. Each real-time assay was done in duplicate on a BioRad MyIQ machine. Data were collected and analyzed with Stratagene Mx3000 software. Gene \( \beta\)-actin was used as an internal control to compute the relative expression level (\( \Delta\Delta C_T \)) for each sample. The fold change of gene expression in tumor tissues as compared with the paired normal group.

---

**Fig. 1. Familial lung cancer pedigrees and linkage fine mapping.** A, six familial lung cancer pedigrees used for fine mapping analysis. Filled circles (females) or squares (males) represent affected members within lung cancer families. The numbers below each individual correspond to the sample identifiers in the pedigree used for fine mapping studies. B, fine linkage mapping with six multigenerational pedigrees at an increased average density of 2.4 cM per marker. LOD scores for individual families were estimated with Simwalk2 under the autosomal dominant model as used previously. The region of haplotype sharing by affected members within each of the families covers a region of \( \sim 3 \) cM centering on the marker D6S2442, 154.10 cM), spanning 152.0 to 154.2 Mb on chromosome 6q.
tissues was calculated as $2^d$, where $d = DCT_{normal} - DCT_{tumor}$. Pairwise Wilcoxon signed-rank test was carried out to assess the overall statistical significance of the difference in gene expression levels between the paired tumor and normal tissues.

Cell lines and tissues used and microarrays. Non–small cell lung cancer cell lines ($n = 56$) and normal controls ($n = 8$; NHBEC, SAEC, HBEC2-5) were from our Hamon Center collection at University of Texas Southwestern Medical Center. They were tested and found free of mycoplasma contamination and their identities were verified by DNA fingerprinting. Normal lung tissues ($n = 29$) were obtained from David Lam at Hong Kong University and from William Gerald at Memorial Sloan Kettering Cancer Center in New York. RNAs were labeled and hybridized to Affymetrix GeneChips according to the manufacturer’s protocol (http://www.affymetrix.com). Microarrays used were HG-U133-Plus2 (54,675 elements; 29,180 unique genes) and HG-U133A and HG-U133B (together 44,928 elements; 23,583 unique genes). When comparing different GeneChips, U133A and U133B were pooled together using their 100 common control genes and U133-Plus2 and U133A&B were analyzed using their 43,680 common genes. Microarray analysis was done using in-house Visual Basic software MATRIX 1.41, which incorporates a connection to the statistical programming language R. Array data were first processed with the mas method of the R package affy which yield PM-corrected signals similar to MAS5 (MicroArray Suite; Affymetrix) processing. The data were then log-transformed, thresholded to a log signal value of 5, and quantile-normalized. Classes of samples were compared by calculating log2 transformation.
ratios and t tests for each gene. All genes on the arrays were BLAST-verified and annotated using recent versions of public National Center for Biotechnology Information (NCBI) databases.

**Cell culture, overexpression, and shRNA knockdown.** The tumor cell line H1299 was cultured in RPMI-1640 plus 10% fetal bovine serum (Gibco). Cells were transduced with a lentiviral short-hairpin RNA (shRNA) construct based on the pLKO1 vector and designed to specifically target RGS17 transcripts (Open Biosystems). Vectors with no knockdown or vectors knocking down the RGS17 transcript were first made in Phoenix cells (Orbigen) and then transduced into H1299 cells with 8 μg/mL polybrene (Sigma). Media were replaced 24 h after transduction, and cells were split 1:4 48 h after transduction. At 72 h posttransduction, cells harboring lentiviral constructs were selected with 1 μg/mL puromycin. Cells were selected for 3 d until mock-infected cells were dead. Surviving cells were pooled and plated at the indicated densities.

For overexpression in H1299 cells, an NH2-terminal three hemagglutinin–tagged (HA3) full-length RGS17 cDNA was cloned into pCDNA3.1 for expression of HA-tagged RGS17 protein. Cells at ~60% confluence were transfected with 8 μg of vector with 3HA alone or that expressing 3HA-RGS17 using Geneporter II reagents (Gene Therapy Systems). Twenty-four hours posttransfection cells were split 1:4 and placed on 2 μg/mL puromycin. Cells were selected for 3 d until mock-transfected cells were dead. Cells were pooled and plated at 500 cells per well in a 12-well tissue culture dish. P values were determined by one-tailed Student’s t-test. Cells were quantitatively assayed for viable cell numbers in triplicate.

**MTT proliferation assay.** Cells stably expressing shRNA vectors or HA3-RGS17 as described above were seeded onto 6-well tissue culture plates at a density of 100 cells per dish. Colonies were stained with crystal violet after 10 d in culture. Visible colonies were counted. P values were determined by one-tailed Student’s t-test.

**Results**

**Fine mapping of the 6q23-25 linkage region.** To further narrow down the broad 6q23-25 linkage region, fine linkage mapping using additional microsatellite markers was employed. Mapping with increased marker density can capture more recombination events, and thus add resolution to the linkage region. We chose six multigenerational families with five or more affected family members for fine mapping studies (Fig. 1A and B). Of the 52 GELCC families used in the original linkage mapping, these 6 families were chosen for further microsatellite mapping because they had the largest number of affected per pedigree (i.e. ≥5) and they exhibited the strongest linkage marker D6S2436 (peak marker in 6q23-25 region). We estimate that these would be the most representative of informative families linked to the 6q susceptibility gene. The LOD scores for each individual high risk family ranged from 0.24 to 0.94 at marker D6S2436 (12). Twenty-six microsatellite markers (including seven markers used in the previous linkage study) selected from the Marshfield Map were genotyped in these six families with an increased marker density averaging

**Table 1. Association results for three RGS17 SNPs and familial lung cancer**

<table>
<thead>
<tr>
<th>Samples/SNPs</th>
<th>Allele frequency</th>
<th>P</th>
<th>OR (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial screen (24 independent 6q-linked cases and 72 controls)†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6901126C</td>
<td>0.646</td>
<td>0.366</td>
<td>1.27 × 10⁻⁴</td>
</tr>
<tr>
<td>rs4083914G</td>
<td>0.364</td>
<td>0.368</td>
<td>1.31 × 10⁻⁴</td>
</tr>
<tr>
<td>rs9479510C</td>
<td>0.630</td>
<td>0.371</td>
<td>2.43 × 10⁻⁴</td>
</tr>
<tr>
<td>GELCC (226 independent familial cases and 313 controls)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6901126C</td>
<td>0.512</td>
<td>0.399</td>
<td>0.005</td>
</tr>
<tr>
<td>rs4083914G</td>
<td>0.495</td>
<td>0.395</td>
<td>0.021</td>
</tr>
<tr>
<td>rs9479510C</td>
<td>0.477</td>
<td>0.384</td>
<td>0.031</td>
</tr>
<tr>
<td>Mayo Clinic (154 independent familial cases and 325 controls)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6901126C</td>
<td>0.461</td>
<td>0.472</td>
<td>0.322</td>
</tr>
<tr>
<td>rs4083914G</td>
<td>0.461</td>
<td>0.428</td>
<td>0.033</td>
</tr>
<tr>
<td>rs9479510C</td>
<td>0.455</td>
<td>0.419</td>
<td>0.035</td>
</tr>
<tr>
<td>Combined (404 independent familial cases and 710 controls)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6901126C</td>
<td>0.500</td>
<td>0.430</td>
<td>1.56 × 10⁻⁴</td>
</tr>
<tr>
<td>rs4083914G</td>
<td>0.473</td>
<td>0.435</td>
<td>3.75 × 10⁻⁵</td>
</tr>
<tr>
<td>rs9479510C</td>
<td>0.477</td>
<td>0.399</td>
<td>8.56 × 10⁻⁵</td>
</tr>
</tbody>
</table>

Abbreviations: OR, odds ratio; 95% CI, 95% confidence interval.
*Dominant genetic model was used for coding SNP genotypes, involving two steps: (a) identifying SNP allele associated with the putative disease allele that has higher frequency in cases than in controls, and (b) forming two genotype groups: DD/DD and dd only. A 2 × 2 contingency table is formed by considering cases and controls and the Fisher’s exact test was used to produce P values displayed in this table. OR (odds ratios) and their 95% confidence intervals (CI) were estimated comparing the risk due to genotype (0+) against the risk due to the wild-type genotype (00). Results from multiple case-control groups were combined using a Mantel-Haenszel model (21).
† OR estimates were greatly inflated due to small counts in one of the cells of the contingency table from the initial screen, and thus are not present in the table.

**Association results for three RGS17 SNPs and familial lung cancer**

**Fig. 1A.** Distribution of LOD scores for 6q23-25 region with 8 additional microsatellite markers. **Fig. 1B.** Pedigree of an affected five-generation family (A1) from a lung cancer casecontrol study (A2). A3, A4, and A5 are the nonafflicted father and two brothers of the proband. A6 is the father of the proband. A7 and A8 are the nonafflicted father and the mother of the proband. A9 and A10 are the nonafflicted father and the mother of the proband. B1, B2, B3, B4, B5, B6, B7, B8, B9, and B10 are the nonafflicted father and the mother of the proband. C1, C2, C3, C4, C5, C6, C7, C8, C9, and C10 are the nonafflicted father and the mother of the proband. D1, D2, D3, D4, D5, D6, D7, D8, D9, and D10 are the nonafflicted father and the mother of the proband. E1, E2, E3, E4, E5, E6, E7, E8, E9, and E10 are the nonafflicted father and the mother of the proband. F1, F2, F3, F4, F5, F6, F7, F8, F9, and F10 are the nonafflicted father and the mother of the proband. G1, G2, G3, G4, G5, G6, G7, G8, G9, and G10 are the nonafflicted father and the mother of the proband. H1, H2, H3, H4, H5, H6, H7, H8, H9, and H10 are the nonafflicted father and the mother of the proband.
2.4 cM per marker. After genotyping, LOD scores for individual families were estimated with Simwalk2 under the autosomal dominant model as used previously (12). Haplotypes were inferred with Simwalk2 for all genotyped affected members from each of the six families, with the largest common haplotypes indicated (13, 14). The haplotype shared by affected members within families varied in length and position. The common region of haplotype sharing by affected members across all the families covers a region of \( f_3 \) cM centering on the marker D6S2442, spanning 152.0 to 154.2 Mb on chromosome 6q (Fig. 1B). As a result, this region of haplotype sharing includes 12 annotated genes (NCBI Build 36.3).

Simultaneously, we used the Affymetrix 500K chipset to ascertain SNP information on 246q-linked unrelated cases and 72 unrelated noncancer controls. To minimize the chance of missing the causal gene(s), the association analysis was done on an expanded region of one-heterogeneity LOD support interval, spanning 144 Mb to 164 Mb on 6q23-25 (Fig. 2A). A total of 114 annotated genes reside in this region (NCBI Build 36.3). Each case is Caucasian and was chosen from one pedigree with a positive LOD at 155 cM. Although such a small sample size was insufficient for a genome-wide scan, prior linkage evidence to 6q permitted a region-wide threshold across 6q to be used (12). A total of 3,957 SNPs were extracted from the Affymetrix 500K chipset for the 6q region. After exclusion of SNPs significantly deviating from Hardy-Weinberg equilibrium \( (P < 0.01) \) in the Caucasian control sample, or with a minor allele frequency \( < 0.05 \), a total of 3,169 polymorphic SNPs were retained for association analysis. The average SNP coverage was 1 SNP per 6.3 kb. Under an autosomal dominant model as described and utilized in the previous linkage study (12), we identified three SNPs with the strongest association on 6q23-25: rs6901126 \( (P = 1.27 \times 10^{-4}) \), rs4083914 \( (P = 1.31 \times 10^{-4}) \), and rs9479510 \( (P = 2.43 \times 10^{-4}) \); Fig. 2B and C). These SNPs reside in a linkage disequilibrium block of 43 kb within the first intron of the RGS17 gene (Fig. 2D) and support our fine linkage mapping observations where specific haplotypes are shared by affected members within each of the families spanning the interval from 152.0 to 154.2 Mb on 6q25 (NCBI Build 36.3; Fig. 1B). RGS17 encodes a recently identified
member of the RGS family. RGS proteins negatively regulate G-protein related signaling at least in part by accelerating the GTPase activity of \( \alpha \) subunits (23, 24).

**Replication in GEIICC and Mayo familial lung cancer samples.** To validate the association signal from the 6q-linked case/control samples, we genotyped three SNPs with the lowest \( P \) values from the initial screen in two independent familial lung cancer samples from the GEIICC and Mayo Clinic. These lung cancer cases in the GEIICC and Mayo Clinic collections were derived from families that have three or more first-degree relatives with lung cancer. In the GEIICC samples, there were 226 familial lung cancer cases and 313 noncancer controls, all of Caucasian descent. All three SNPs were detected to be significantly associated with lung cancer in the GEIICC samples (Table 1), and the observed risk allele for the three SNPs was the same as in the initial screening sample. In the Mayo Clinic samples, there were 154 familial cases and 325 controls and we observed that two of the three SNPs (rs4083914, \( P = 0.033 \); and rs9479510, \( P = 0.035 \)) to be significantly associated with familial lung cancer. In the combined data sets (404 cases and 710 controls), all three SNPs showed significant association with lung cancer with odds ratios of around 1.8 (Table 1).

**Significant SNPs do not associate with sporadic lung cancer.** Less than 5% of lung cancer cases diagnosed were familial in origin, which prohibits extensive validation of putative genetic risk factors for familial lung cancer because of limited biospecimen availability. We instead sought to determine whether this gene is also associated with sporadic lung cancer cases, and genotyped 553 sporadic cases and 627 controls of Caucasian descent from Mayo Clinic and 197 Chinese sporadic female cases and 410 female controls from Shanghai, China. However, no RGS17 SNPs showed significant associations in these sporadic case populations (rs6901126, \( P = 0.403 \); rs4083914, \( P = 0.804 \); rs9479510, \( P = 0.951 \); Supplemental Table S3). This suggests a role for RGS17 in lung cancer that may be very different in sporadic cases versus familial cases, as in the case with the susceptibility gene \( p53 \) (25). This gene is mutated both somatically and in the germline, each of which plays roles in the development of various cancers. Hence, the specific mechanism of RGS17 dysfunction as it relates to cancer may be very different in sporadic cases as opposed to familial cases. This familial specificity of RGS17 was also supported by a comparative linkage analysis of different risk lung cancer families. A high LOD score was detected in 6q when analyzing families with five or more affected family members with lung cancer, whereas a strong positive 6q linkage signal was not detectable in families with three or fewer affected family members with lung cancer (Supplemental Fig. S1).

**RGS17 is overexpressed in lung cancer tumors and cell lines.** In order to investigate possible pathogenic changes in RGS17 in our familial group, we did direct sequencing of the RGS17 protein coding sequences in 10 familial lung cancer cases containing risk alleles implicated in the above association analysis. We did not uncover any coding mutations in RGS17, which might suggest that another mechanism such as changes in gene expression underneath disease susceptibility. RGS17 has been shown to be expressed in both central nervous and peripheral tissues, including the lung (23). We examined the gene expression levels of RGS17 in paired tumor and normal tissues from 13 sporadic lung cancer patients (9 adenocarcinoma and 4 adenosquamous carcinoma) using quantitative real-time PCR, and observed significant overexpression in the tumor tissue (Fig. 3A). Of these 13 paired tumors, 10 (77%) exhibited increased expression of RGS17 and the average difference in expression of RGS17 in these tumor tissues versus matched normal controls was 9.1-fold (pairwise Wilcoxon signed-rank test \( P = 0.009 \)). The smoking status of these samples was unavailable. In an expanded set of 61 sporadic lung tumors of various pathologies RGS17 transcript was increased in 80% of lung tumors over matched normal lung tissue tumors by an average of 8.3-fold (\( P = 1.36 \times 10^{-7} \)) confirming our observations in the original 13 samples (see Supporting Data - Section #1). Among these samples there is no statistical difference in RGS17 induction between adenocarcinomas (\( n = 18 \)) and adenosquamous carcinomas (\( n = 8 \)) as measured by a Student’s \( t \)-test (\( P > 0.1 \)).

We did Affymetrix gene chip expression analysis on 56 non–small cell lung cancer cell lines and 37 normal samples (29 normal lung tissue and 8 normal lung cell lines). These data revealed a significant increase in RGS17 expression as determined by Wilcoxon test, \( P = 8.1 \times 10^{-7} \) (Supplemental Fig. S2).

**RGS17 expression levels modulate cancer cell proliferation.** In order to evaluate the effect of RGS17 on the growth properties of tumor cells, lentiviral shRNA constructs were utilized to stably knock-down RGS17 transcript levels in a human lung tumor cell line. Human H1299 non–small cell lung cancer cells were chosen due to the high expression of RGS17 in this cell line as measured by expression microarray (Supplemental Fig. S2). RGS17 transcript accumulation was effectively reduced in H1299 human lung tumor cells and knock-down of RGS17 transcript resulted in a decrease in the proliferative rate of these cells in culture as measured by a 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) cell proliferation assay over 10 days (Fig. 3B). Showing similar proliferative effects using two distinct shRNA constructs in H1299 cells minimized the possibility of off-target effects with shRNA knockdown. Colony formation assays also clearly show decreased proliferative capacity of H1299 cells with RGS17 knockdown (Fig. 3C). Knockdown in cancer cell lines Hct116 (colon carcinoma) and DU145 (prostate carcinoma) has resulted in decreased proliferative rates in both cell lines (see Supporting data - Section #2). Knockdown was also attempted in the A549 lung cancer cell line, but sufficient knockdown of RGS17 transcript was not achieved in these cells. The in vivo significance of the proliferative effects of RGS17 knockdown was further established using an athymic nude mouse tumorigenesis assay. Mice were injected s.c. with H1299 human lung tumor cells stably expressing shRNA as described above and monitored for tumor growth over 4 weeks. The rate of growth and tumor load was decreased significantly by RGS17 knockdown (Fig. 3D and E). Average tumor weight was reduced from 148 mg to 23 mg (\( P = 0.03 \)), and average tumor volume was reduced from 385 to 47 mm\(^3\) (\( P < 0.01 \)) with RGS17 knockdown (Fig. 3E). Furthermore, exogenous overexpression of HA-tagged RGS17 in H1299 cells enhanced cell proliferation consistent with a role of RGS17 in tumor cell proliferation (Supplemental Fig. S3).

**Discussion**

Our statistical and biological analyses have strongly implicated RGS17 as a candidate for the lung cancer susceptibility locus at...
Although we were not able to analyze RGS17 expression levels in our familial lung cancer cases because of limited biospecimen availability, our data did indicate high RGS17 transcript up-regulation in sporadic lung tumors and cell lines, and strong effects on cell proliferation through knockdown and overexpression in a lung cancer cell line. We hypothesize that there exists a rare variant or variants, which lie on the same haplotype detected and defined by the significantly associated SNPs. This rare, highly penetrant genetic lesion is postulated to affect RGS17 expression and lung cancer susceptibility.

Our future efforts to identify causal variants are currently focused on a resequencing analysis of the linkage disequilibrium block 12 (Fig. 2D). This 43-kb region contains the core promoter, noncoding exon 1 and part of intron 1. Several CpG islands are clustered around exon 1. We intend to address the methylation status of four CpG islands located around exon 1, in sporadic lung tumor expression. Because expression data cannot be obtained from familial samples due to biospecimen availability, it will be necessary to determine meaningful changes in the RGS17 gene in familial germline DNA using these types of experiments until appropriate familial specimens become available for the analysis of expression. The elucidation of specific mechanisms by which RGS17 confers accelerated growth and other tumor phenotypes using cell and molecular biology studies must be pursued, and will be seminal in influencing the diagnostic, preventive, and therapeutic applications of this research. These studies constitute ongoing and long-term goals generated by the work presented here, and will elucidate the mechanism by which RGS17 affects familial lung cancer susceptibility.

Recent reports have linked RGS domain containing genes to cancer. One study describes SNPs in PDZ-RhoGEF, containing an RGS domain, which modulates the risk of lung cancer in Mexican Americans (26). Another such study describes the identification of a functional polymorphism in the 3’ untranslated region of RGS6 that is associated with bladder cancer risk, and was shown to affect protein translation (27). There is also evidence that RGS17 reduces dopamine-D2/Gαi-mediated inhibition of cyclic adenosine monophosphate (cAMP) formation and abolishes thyrotropin-releasing hormone receptor/Gαi-mediated calcium mobilization (24). D2 dopamine receptor agonists exhibit antiproliferative effects in lung tumors and lung cancer cell lines associated with decreased cAMP accumulation (28, 29). It is possible that RGS17 is involved in addiction/reward signaling pathways, as the D2 dopamine gene (DRD2) seems to be associated with smoking addiction. A recent paper also found that RGS17 may work on opioid receptor function (30). Previous work has shown that lungs of both non-smokers and small cell histologic types can express high affinity opioid receptors including μ receptors, opioid peptides, and also nAChR receptors (31). Lung cancer growth is inhibited and apoptosis induced by opioids including μ agonists whereas nicotine acting through nAChRs antagonizes this effect. Our recent work has shown that overexpression of RGS17 on opioid receptors it is possible that overexpression of RGS17 could act through inhibition of a growth regulatory pathway provided by opioid receptors. In a recent study, we show that RGS17 induces cAMP response element binding (CREB) phosphorylation and CREB responsive gene expression (36). RGS17 also enhances forskolin-mediated cAMP production, forskolin-induced gene expression, and forskolin-induced proliferation. Furthermore, protein kinase A (PKA) inhibition causes growth arrest of lung tumor cells, which is partially restored by RGS17 overexpression. Thus, RGS17 seems to be a potential oncoprotein promoting proliferation through cAMP-PKA-CREB signaling. The identification of RGS17 as the major candidate gene for familial lung cancer susceptibility on chromosome 6q has important implications for the diagnosis and treatment of lung cancer as well as delineation of the mechanisms underlying both familial and sporadic lung cancer.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

We thank the Familial 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, Washington University Genotyping Core, Mayo Clinic Genotyping Shared Resource (supported in part by P30CA1082), and to Julie Clark, Qiong Chen, Scott Hay, Mark Watson, and Jennifer Baker for their assistance in various aspects of this work. We also thank David Lam (University of Hong Kong) and William Gerald (Memorial Sloan Kettering Cancer Center) for microarray data on the normal lung tissue samples.

**References**

## Fine Mapping of Chromosome 6q23-25 Region in Familial Lung Cancer Families Reveals \textit{RGS17} as a Likely Candidate Gene

Ming You, Daolong Wang, Pengyuan Liu, et al.


<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: <a href="http://clincancerres.aacrjournals.org/content/15/8/2666">http://clincancerres.aacrjournals.org/content/15/8/2666</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at: <a href="http://clincancerres.aacrjournals.org/content/suppl/2009/04/21/1078-0432.CCR-08-2335.DC1">http://clincancerres.aacrjournals.org/content/suppl/2009/04/21/1078-0432.CCR-08-2335.DC1</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cited articles</th>
<th>This article cites 35 articles, 14 of which you can access for free at: <a href="http://clincancerres.aacrjournals.org/content/15/8/2666.full.html#ref-list-1">http://clincancerres.aacrjournals.org/content/15/8/2666.full.html#ref-list-1</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citing articles</td>
<td>This article has been cited by 11 HighWire-hosted articles. Access the articles at: <a href="http://clincancerres.aacrjournals.org/content/15/8/2666.full.html#related-urls">http://clincancerres.aacrjournals.org/content/15/8/2666.full.html#related-urls</a></td>
</tr>
</tbody>
</table>

**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, contact the AACR Publications Department at permissions@aacr.org.