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] Min 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,[2] Mariza de Andrade,[2] Yanhong Wu,[2] 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. Anderson[3]

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 6q 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

Translational Relevance

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.

segregation analyses (6–11). Despite increasing knowledge of genetic influences on lung cancer susceptibility, no specific causal genes have been identified. A recent genome-wide linkage study by the Genetic Epidemiology of Lung Cancer Consortium (GELCC) mapped a major susceptibility locus to 6q23-25 (12). An analysis of 52 extended pedigrees with ≥3 first-degree relatives with lung cancer produced a maximum heterogeneity logarithm of the odds (LOD) score of 2.79 at 155 cM (marker D6S2436) on chromosome 6q. Further analysis of a subset of 23 multigenerational pedigrees containing ≥5 family members affected with lung cancer yielded a multipoint heterogeneity LOD score of 4.26 at the same position.

To identify the causal gene in the *6q* susceptibility locus, the present study employed a combination of linkage fine mapping and region-wide single nucleotide polymorphism (SNP) association analysis. We have identified common variants in *RGS17* that associate with familial lung cancer and validated these in two independent samples of unrelated familial cases and controls. This established *RGS17* as a candidate familial lung cancer susceptibility gene for this major locus on chromosome 6q. *RGS17* encodes a recently identified member of the regulator of G-protein signaling (RGS) family. RGS proteins negatively regulate G-protein related signaling at least in part by accelerating the GTPase activity of Gα subunits. We showed that *RGS17* is highly expressed in tumor tissues and that loss of *RGS17* transcript inhibits the growth of xenografted tumors and the proliferation of tumor cells, whereas overexpression of *RGS17* increases the rate of proliferation of tumor cells.

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, C6S1848, D6S971, G15833, D6S960, D6S495, D6S2442, D6S2436, D6S442, D6S969, D6S1035, D6S955, 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 Fernald 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 Fernald 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 62.9 y (±9.5). In the controls, 65.2% and 19.7% are current/former smokers and non-smokers, 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
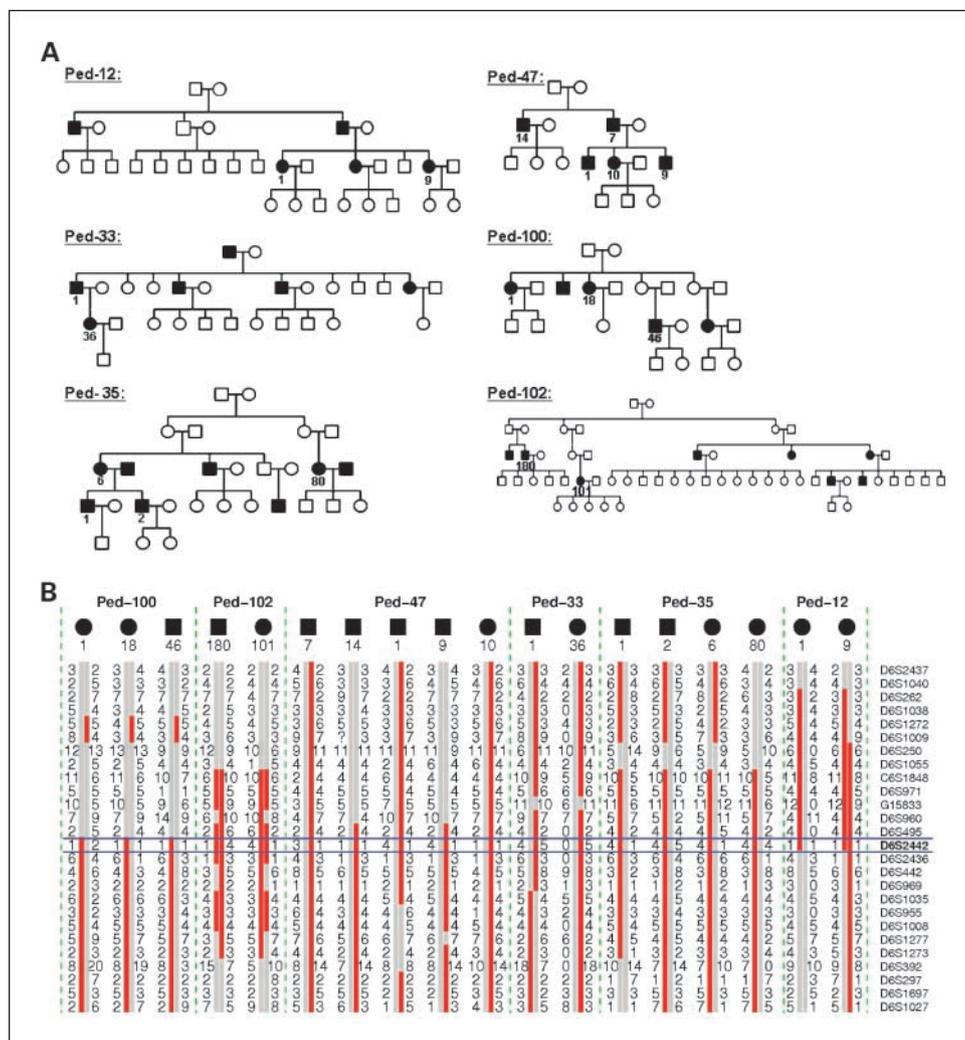
Fig. 1. Familial lung cancer pedigrees and linkage fine mapping. *A,* six familial lung cancer pedigrees used for fine mapping analysis. Filled circles (females) or squares (males) represent affected members within lung cancer families. The numbers below each individual correspond to the sample identifiers 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 ~3 cM centering on the marker D6S2442 154.10 cM), spanning 152.0 to 154.2 Mb on chromosome 6q.

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 × 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-Hazenszel 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 volume 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 *β-actin* was used as an internal control to compute the relative expression level (ΔC_T) for each sample. The fold change of gene expression in tumor tissues as compared with the paired normal
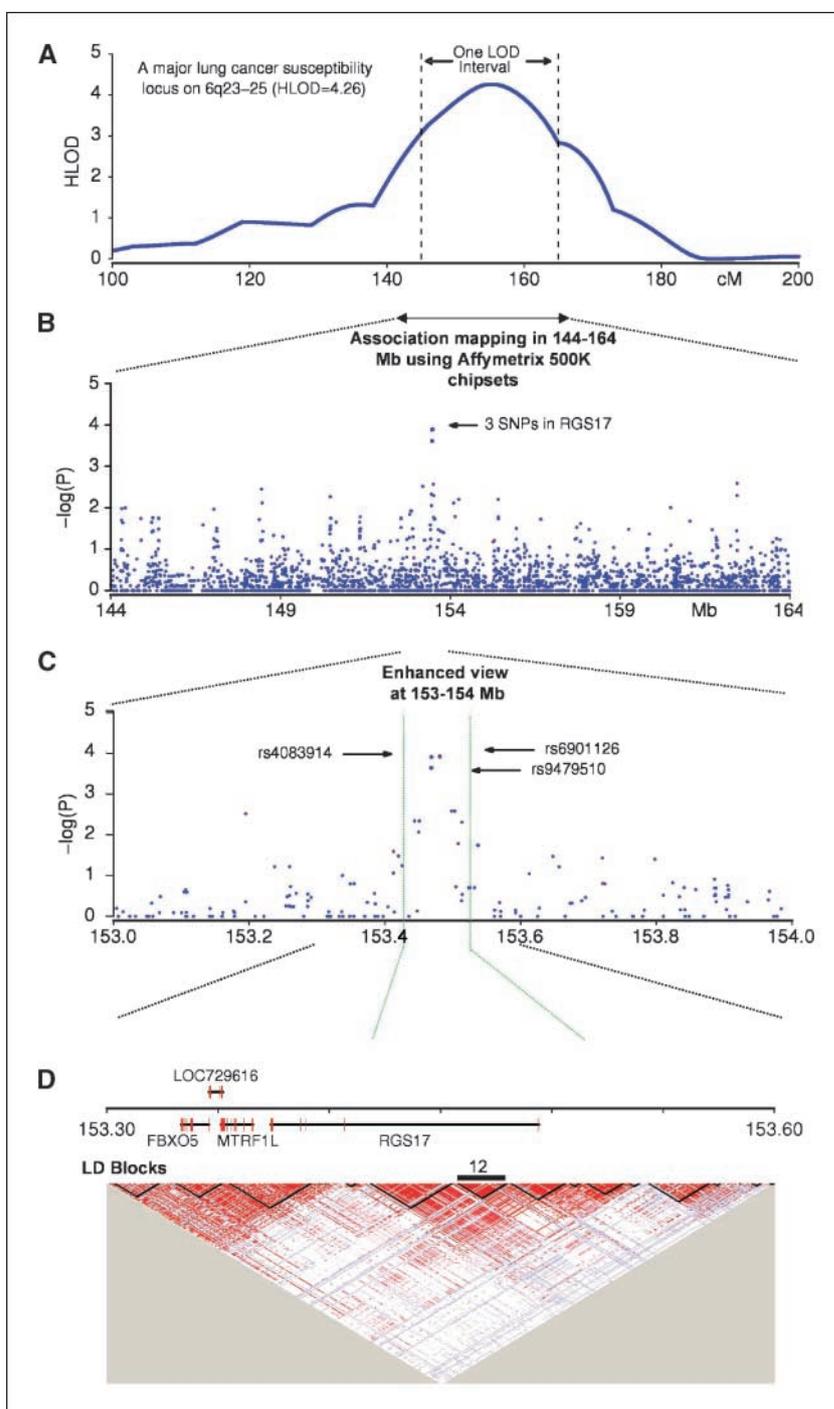
Fig. 2. Schematic view of linkage and association mapping of familial lung cancer on chromosome 6q23-25. *A,* linkage scan results for 6q23-25 from the GELCC lung cancer families (12). *B,* association mapping in 24 independent 6q-linked cases and 72 controls using 3,957 SNPs from Affymetrix 500K chipsets. *C,* enhanced view of association mapping in the 153-154 Mb region. All three significant SNPs were located in the first intron of *RGS17. D,* physical map and linkage disequilibrium blocks. Pairwise linkage disequilibrium, measured as D', was calculated from the HapMap Centre d'Etude du Polymorphisme Humain collection (CEU) using the methods of Gabriel (34) as implemented in Haploview (35). Shading represents the magnitude and significance of pairwise linkage disequilibrium, with a white-to-red gradient reflecting lower to higher linkage disequilibrium values. All three SNPs, rs4083914, rs9479510 and rs6901126, were located within block 12 with a size of 43 kb.

tissues was calculated as 2^d, where d = $\Delta C_{T\,normal}$ - $\Delta C_{T\,tumor}$. 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 $\log_2$

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 ug/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 for 2 to 4 d, until mock-infected cells were dead. Surviving cells were pooled and plated at the indicated densities.

For overexpression in H1299 cells, an NH_2-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 ug 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 dishes at a density of 500 cells per well. Cells were assayed for viable cell numbers using the CellTiter 96 Non-Radioactive Cell Proliferation Assay kit (Promega) periodically over 10 d in culture. *P* values were determined by one-tailed Student's *t*-test.

Colony formation assay. Cells stably expressing shRNA vectors as described above were seeded onto 100 mm/L tissue culture dishes 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.

Nude mouse tumorigenesis assay. Tumor cells stably expressing shRNA vectors as described above were cultured, counted, and resuspended in serum-free media at a concentration of 1.5×10^7 cells/mL. A volume of 200 μL (3×10^6 cells respectively) were injected s.c. into the right (vector) or left (shRNA) flank of athymic nude mice at an age of 4 to 6 wk. The health of these mice was monitored three times weekly and tumor sizes were measured periodically until sacrifice at 2 to 4 wk postinjection depending on the growth rate of the tumors. Tumor volume was determined by the formula (l × w × h). *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 at 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

| Samples/SNPs	Allele frequency			*P*	OR (95% CI)*
	Risk allele	Cases	Controls		
Initial screen (24 independent 6q-linked cases and 72 controls) [†]					
rs6901126	C	0.646	0.366	1.27×10^{-4}	
rs4083914	G	0.646	0.368	1.31×10^{-4}	
rs9479510	C	0.630	0.371	2.43×10^{-4}	
GELCC (226 independent familial cases and 313 controls)					
rs6901126	C	0.512	0.399	0.005	1.76 (1.17-2.68)
rs4083914	G	0.495	0.395	0.021	1.62 (1.07-2.41)
rs9479510	C	0.477	0.384	0.031	1.53 (1.06-2.26)
Mayo Clinic (154 independent familial cases and 325 controls)					
rs6901126	C	0.461	0.472	0.322	1.28 (0.81-2.05)
rs4083914	G	0.461	0.428	0.033	1.62 (1.03-2.58)
rs9479510	C	0.455	0.419	0.035	1.60 (1.02-2.55)
Combined (404 independent familial cases and 710 controls)					
rs6901126	C	0.500	0.430	1.56×10^{-4}	1.73 (1.30-2.30)
rs4083914	G	0.473	0.435	3.75×10^{-5}	1.80 (1.36-2.39)
rs9479510	C	0.477	0.399	8.56×10^{-5}	1.73 (1.31-2.28)

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 (*D+*) against the risk due to the wild-type genotype (*dd*). 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.

Fig. 3. *RGS17* mRNA expression is increased in lung tumor tissue, and RNAi knockdown of RGS17 transcript inhibits tumor cell proliferation and tumor growth. *A,* expression of *RGS17* in human lung adenocarcinomas and adenosquamous carcinomas relative to patient-matched normal lung tissue controls. Green, normal tissues; red, tumor tissues. *B, RGS17* knockdown inhibits proliferation. Cell growth is measured by MTT viable cell staining over 10 d. shRNA knockdown of RGS17 transcript in H1299 human lung tumor cells was measured by quantitative real-time PCR (*inset*). *C, RGS17* knockdown inhibits colony formation. Colonies were stained after 10 d in culture. *D,* nude mouse tumorigenesis assays. 3×10^6 cells were injected s.c. on the left (H1299-shRGS) and right (H1299-vector) flanks. *E,* tumor volume was monitored until the mice were sacrificed 28 d postinjection.
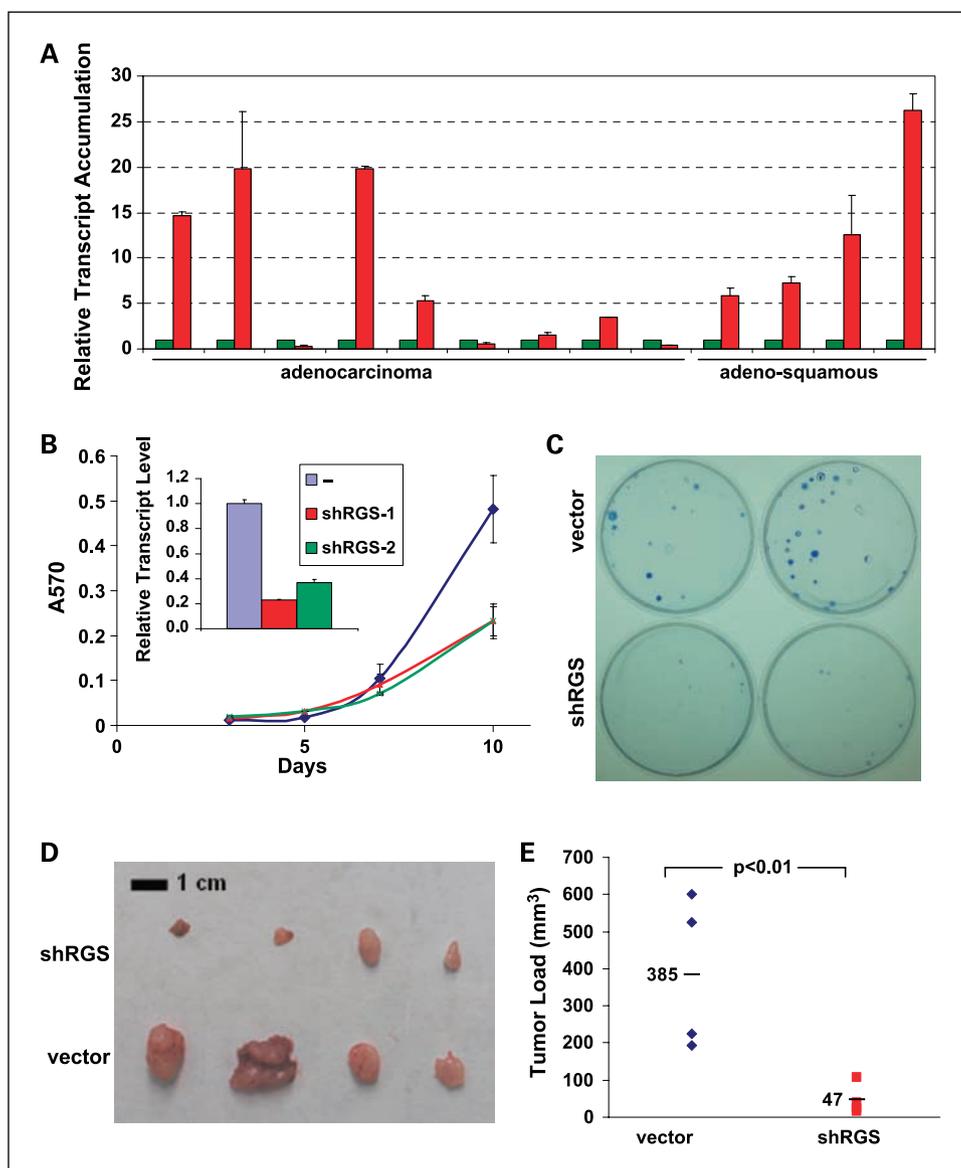
2.4 cM per marker. After genotyping, LOD scores for individual families were estimated with Simwalk2 under the autosomal dominant model as used previously (12). Haplotypes were inferred with Simwalk2 for all genotyped affected members from each of the six families, with the largest common haplotypes 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 ~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 24 6q-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 ($P \le 0.01$) 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 Gα subunits (23, 24).

Replication in GELCC 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 GELCC and Mayo Clinic. These lung cancer cases in the GELCC and Mayo Clinic collections were derived from families that have three or more first-degree relatives with lung cancer. In the GELCC 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 GELCC 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 underlie 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 ($P = 1.36 \times 10^{-9}$) 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

6q23-25. 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/Gα_i-mediated inhibition of cyclic adenosine monophosphate (cAMP) formation and abolishes thyrotropin-releasing hormone receptor/Gα_q-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 lung cancers of both non–small cell 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 providing a growth regulatory loop that is antagonized by nicotine (31, 32). Recently, positron emission tomography imaging studies have provided *in vivo* evidence for the presence of ä and μ opioid receptor types in small cell, squamous, and adenocarcinomas of the lung (33). Because of the negative action 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* overexpresion. 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

References

1. Friberg L, Kaij L, Dencker SJ, Jonsson E. Smoking habits of monozygotic and dizygotic twins. Br Med J 1959;1:1090–2.
2. Heath CW. Differences between smokers and non-smokers. AMA Arch Intern Med 1958;101:377–88.
3. Motulsky AG. Drug reactions enzymes, and biochemical genetics. J Am Med Assoc 1957;165:835–7.
4. Tokuhata GK, Lilienfeld AM. Familial aggregation of lung cancer in humans. J Natl Cancer Inst 1963;30:289–312.
5. Fraumeni JF, Wertelecki W, Blattner WA, Jensen RD, Leventhal BG. Varied manifestations of a familial lymphoproliferative disorder. Am J Med 1975;59:145–51.
6. Sellers TA, Bailey-Wilson JE, Elston RC, et al. Evidence for mendelian inheritance in the pathogenesis of lung cancer. J Natl Cancer Inst 1990;82:1272–9.
7. Bailey-Wilson JE, Elston RC, Sellers TA, Rothschild H.

Segregation analysis of lung cancer using class A regressive models. Am J Hum Genet 1992;51:A145.
8. Sellers TA, Bailey-Wilson JE, Elston RC, Rothschild H. Evidence for Mendelian factors in early onset lung cancer. In: Brugge J, Curran T, Harlow E, McCormick F, editors. Origins of human cancer: a comprehensive review. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1991. p. 775–80.
9. Sellers TA, Potter JD, Folsom AR. Association of incident lung cancer with family history of female reproductive cancers: the Iowa Women's Health Study. Genet Epidemiol 1991;8:199–208.
10. Sellers TA, Chen PL, Potter JD, Bailey-Wilson JE, Rothschild H, Elston RC. Segregation analysis of smoking-associated malignancies: evidence for Mendelian inheritance. Am J Med Genet 1994;52:308–14.
11. Yang P, Schwartz AG, McAllister AE, Swanson GM,

Aston CE. Lung cancer risk in families of nonsmoking probands: heterogeneity by age at diagnosis. Genet Epidemiol 1999;17:253–73.
12. Bailey-Wilson JE, Amos CI, Pinney SM, et al. A major lung cancer susceptibility locus maps to chromosome 6q23–25. Am J Hum Genet 2004;75:460–74.
13. Sobel E, Lange K. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 1996;58:1323–37.
14. Sobel E, Sengul H, Weeks DE. Multipoint estimation of identity-by-descent probabilities at arbitrary positions among marker loci on general pedigrees. Hum Hered 2001;52:121–31.
15. Liu P, Vikis HG, Wang D, et al. Familial aggregation of common sequence variants on 15q24–25.1 in lung

cancer. J Natl Cancer Inst 2008;100:1326 – 30. Epub 2008 Sep 9.

16. Yang P, Allen MS, Aubry MC, et al. Clinical features of 5,628 primary lung cancer patients: experience at Mayo Clinic from 1997 to 2003. Chest 2005;128: 452 – 62.

17. Taniguchi K, Yang P, Jett J, et al. Polymorphisms in the promoter region of the neutrophil elastase gene are associated with lung cancer development. Clin Cancer Res 2002;8:1115 – 20.

18. Wen W, Shu XO, Gao YT, et al. Environmental tobacco smoke and mortality in Chinese women who have never smoked: prospective cohort study. BMJ 2006; 333:376.

19. Rabbee N, Speed TP. A genotype calling algorithm for affymetrix SNP arrays. Bioinformatics 2006;22:7 – 12.

20. Schaid DJ, Batzler AJ, Jenkins GD, Hildebrandt MA. Exact tests of Hardy-Weinberg equilibrium and homogeneity of disequilibrium across strata. Am J Hum Genet 2006;79:1071 – 80.

21. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719 – 48.

22. Chaparro J, Reeds DN, Wen W, et al. Alterations in thigh subcutaneous adipose tissue gene expression in protease inhibitor-based highly active antiretroviral therapy. Metabolism 2005;54:561 – 7.

23. Nunn C, Mao H, Chidiac P, Albert PR. RGS17/ RGSZ2 and the RZ/A family of regulators of G-protein signaling. Semin Cell Dev Biol 2006;17:390 – 9.

24. Mao H, Zhao Q, Daigle M, Ghahremani MH, Chidiac P, Albert PR. RGS17/RGSZ2, a novel regulator of Gi/o, Gz, and Gq signaling. J Biol Chem 2004;279:26314 – 22.

25. Petitjean A, Achatz MI, Borresen-Dale AL, Hainaut P, Olivier M. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. Oncogene 2007;26:2157 – 65.

26. Gu J, Wu X, Dong Q, et al. A nonsynonymous single-nucleotide polymorphism in the PDZ-Rho guanine nucleotide exchange factor (Ser1416Gly) modulates the risk of lung cancer in Mexican Americans. Cancer 2006;106:2716 – 24.

27. Berman DM, Wang Y, Liu Z, et al. A functional polymorphism in RGS6 modulates the risk of bladder cancer. Cancer Res 2004;64:6820 – 6.

28. Senogles SE. D2 dopamine receptor-mediated antiproliferation in a small cell lung cancer cell line, NCI-H69. Anticancer Drugs 2007;18:801 – 7.

29. Ishibashi M, Fujisawa M, Furue H, Maeda Y, Fukayama M, Yamaji T. Inhibition of growth of human small cell lung cancer by bromocriptine. Cancer Res 1994;54:3442 – 6.

30. Ajit SK, Ramineni S, Edris W, et al. RGSZ1 interacts with protein kinase C interacting protein PKCI-1 and modulates ì opioid receptor signaling. Cell Signal 2007;19:723 – 30.

31. Maneckjee R, Minna JD. Opioid and nicotine receptors affect growth regulation of human lung cancer cell lines. Proc Natl Acad Sci U S A 1990;87:3294 – 8.

32. Maneckjee R, Minna JD. Opioids induce while nicotine suppresses apoptosis in human lung cancer cells. Cell Growth Differ 1994;5:1033 – 40.

33. Madar I, Bencherif B, Lever J, et al. Imaging ä- and ì-opioid receptors by PET in lung carcinoma patients. J Nucl Med 2007;48:207 – 13.

34. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225 – 9.

35. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263 – 5.

36. James MA, Lu Y, Liu Y, Vikis HG, You M. RGS17, an overexpressed gene in human lung and prostate cancer, induces tumor cell proliferation through the cyclic AMP-PKA-CREB pathway. Cancer Res 2009;69:2108 – 16.

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

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

Updated version	Access the most recent version of this article at: http://clincancerres.aacrjournals.org/content/15/8/2666
Supplementary Material	Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2009/04/21/1078-0432.CCR-08-2335.DC1

Cited articles	This article cites 35 articles, 10 of which you can access for free at: http://clincancerres.aacrjournals.org/content/15/8/2666.full#ref-list-1
Citing articles	This article has been cited by 4 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/15/8/2666.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://clincancerres.aacrjournals.org/content/15/8/2666. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.