A SNP in a let-7 microRNA Complementary Site in the KRAS 3’ Untranslated Region Increases Non–Small Cell Lung Cancer Risk

Lena J. Chin,1 Elena Ratner,2 Shuguang Leng,3 Rihong Zhai,6 Sunitha Nallur,2 Imran Babar,1 Roman-Ulrich Muller,7 Eva Straka,7 Li Su,8 Elizabeth A. Burki,3 Richard E. Crowell,9 Rajeshvari Patel,1 Trupti Kulkarni,2 Robert Homer,4 Daniel Zelterman,3 Kenneth K. Kidd,2 Yong Zhu,5 David C. Christiani,67 Steven A. Belinsky,9 Frank J. Slack,1 and Joanne B. Weidhaas2

Departments of 1Molecular, Cellular and Developmental Biology, 2Therapeutic Radiology, 3Genetics, 4Pathology, and 5Epidemiology, Yale University, New Haven, Connecticut; and 6Department of Environmental Health, Harvard School of Public Health and 7Pulmonary and Critical Care Unit, Department of Medicine, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts; and
8Department of Internal Medicine, University of New Mexico and 9Lung Cancer Program, Lovelace Respiratory Research Institute, Albuquerque, New Mexico

Abstract

Lung cancer is the leading cause of cancer deaths worldwide, yet few genetic markers of lung cancer risk useful for screening exist. The let-7 family of-microRNAs (miRNA) are global genetic regulators important in controlling lung cancer oncogene expression by binding to the 3’ untranslated regions of their target mRNAs. The purpose of this study was to identify single nucleotide polymorphisms (SNP) that could modify let-7 binding and to assess the effect of such SNPs on target gene regulation and risk for non–small cell lung cancer (NSCLC). let-7 complementary sites (LCS) were sequenced in the KRAS 3’ untranslated region from 74 NSCLC cases to identify mutations and SNPs that correlated with NSCLC. The allele frequency of a previously unidentified SNP at LCS6 was characterized in 2,433 people (representing 46 human populations). The frequency of the variant allele is 18.1% to 20.3% in NSCLC patients and 5.8% in world populations. The association between the SNP and the risk for NSCLC was defined in two independent case-control studies. A case-control study of lung cancer from New Mexico showed a 2.3-fold increased risk (confidence interval, 1.1–4.6; \( P = 0.02 \)) for NSCLC cancer in patients who smoked \(<40\) pack-years. This association was validated in a second independent case-control study. Functionally, the variant allele results in KRAS overexpression in vitro. The LCS6 variant allele in a KRAS miRNA complementary site is significantly associated with increased risk for NSCLC among moderate smokers and represents a new paradigm for let-7 miRNAs in lung cancer susceptibility. [Cancer Res 2008;68(20):8535–40]

Introduction

MicroRNAs (miRNA) are recently identified gene regulators that are at abnormal levels and implicated in virtually all cancer subtypes studied (1). MiRNAs bind to the 3’ untranslated regions (UTR) of their target genes, regions which are evolutionarily highly conserved (2, 3), suggesting an important role for these regions in natural selection. Because miRNAs each regulate hundreds of mRNAs simultaneously (4), the potential of cellular transformation resulting from dysfunction of a single miRNA is high. The role of miRNA single nucleotide polymorphisms (SNP) in disease is just being defined. For example, SNPs in miRNAs important in cancer have been identified; mir-125a, shown to be altered in breast cancer (5, 6), has a variant allele at a SNP in the mature miRNA sequence that decreases expression (7). There is also evidence that SNPs in miRNA binding sites could be associated with disease, for example, a point mutation identified in several Tourette’s syndrome patients in the 3’UTR of SLITRK1 disrupts the binding of miR-189 (8). Furthermore, two recent papers report SNPs in miRNA target sites in human cancer genes (9) and show that allele frequencies vary between cancerous and normal tissues (10). Finally, a SNP identified in a miRNA binding site in the kit oncogene was associated with increased gene expression in papillary thyroid cancer (11).

The let-7 family of miRNAs seems to play a key role in lung cancer: they are at low levels in non–small cell lung cancer (NSCLC; refs. 12, 13); their lower levels are biomarkers of a poor outcome (14, 15); they regulate multiple important lung cancer oncogenes, including RAS (12, 16); and they inhibit growth of lung cancer cells in vitro (12, 14) and in vivo (17, 18). The purpose of this study was to identify SNPs that could potentially modify let-7 binding and to assess the effect of these SNPs on target gene regulation and the risk for NSCLC. Here, we show that a variant allele at a SNP in a LCS in the KRAS 3’UTR is associated with increased risk for NSCLC in moderate smokers. Furthermore, this variant allele leads to altered KRAS regulation in vitro, with higher KRAS expression in the presence of the variant allele. In addition, in tumors from patients harboring the variant allele, let-7s were lower than in nonvariant allele tumors, suggesting the SNP might be associated with NSCLC with a poor prognosis. The finding that a SNP disrupting let-7 miRNA regulation of a known oncogene can affect cancer predisposition to NSCLC is a new paradigm and supports further studies to identify similar SNPs in all cancer types.

Materials and Methods

Study populations. Lung tissue samples from patients with a diagnosis of NSCLC were collected following Yale University Human Investigation Committee approval. Cases were chosen based on the availability of frozen stored tissue from lung tumor resections from 1994 through 2003 and from recent cases with extra tissue available. Tissue was collected from 87...
patients. Seven patients were excluded due to other risk factors for lung cancer (e.g., immunosuppression and tuberculosis) and six were excluded due to their tumors being nonlung primary metastatic disease. Seventy-four patients were included in the analysis (see Supplementary Table S1).

To determine the frequency of the SNP alleles, 2,433 individuals were genotyped from a global sample of 46 populations. According to population ancestry and geographic locations, these 46 populations are categorized into 4 groups: European (including West Asia), African, Asian (including the Pacific), and Native American. Sample descriptions and samples sizes can be found in the Allele Frequency Database (ALFRED; 19) by searching for the population names. DNA samples were extracted from lymphoblastoid cell lines established and/or grown in the Yale University laboratory of K.K.K. The methods of transformation, cell culture, and DNA purification have been described. All volunteers were apparently normal and otherwise healthy adult males or females and samples were collected after receipt of appropriate informed consent.

Lung cancer cases (n = 325) for the New Mexico case-control study were recruited beginning in 2004 from Albuquerque through two local hospitals, the Veterans hospital and the University of New Mexico hospital. All stages and histologic types of lung cancer were included. Controls (n = 325) with no history of any prior cancer were recruited from two ongoing local smoker cohorts, the Veterans Smokers Cohort (mainly veterans from Albuquerque), and the Lovelace Smokers Cohort (general residents in Albuquerque). Those two cohorts started to recruit participants in 2001 to conduct longitudinal studies on molecular markers of respiratory carcinogenesis in biological fluids such as sputum from people at risk for lung cancer. A standardized questionnaire was used to collect information on medical, family, and smoking exposure history, and quality of life for both lung cancer cases and control cohort members. Controls were randomly matched to lung cancer cases after categorization into different age groups (5-y differences) by sex and cohort (see Supplementary Table S1). Controls were matched to lung cancer cases and control cohort members. Controls were recruited beginning in 2004 from Albuquerque through two local hospitals, the Veterans hospital and the University of New Mexico hospital. All stages and histologic types of lung cancer were included. Controls (n = 325) with no history of any prior cancer were recruited from two ongoing local smoker cohorts, the Veterans Smokers Cohort (mainly veterans from Albuquerque), and the Lovelace Smokers Cohort (general residents in Albuquerque). Those two cohorts started to recruit participants in 2001 to conduct longitudinal studies on molecular markers of respiratory carcinogenesis in biological fluids such as sputum from people at risk for lung cancer. A standardized questionnaire was used to collect information on medical, family, and smoking exposure history, and quality of life for both lung cancer cases and control cohort members. Controls were randomly matched to lung cancer cases after categorization into different age groups (5-y differences) by sex and cohort (see Supplementary Table S1). Cases with small cell lung cancer were excluded to more precisely assess the effect of the let-7 complementary site (LCS6) SNP on risk for NSCLC. Cases over ages 82 y (the maximum age in the control group), cases with any prior cancer history, never smokers, or cases with missing data on smoking-related covariates were also excluded in the data analysis, resulting in 218 cases in the analysis.

A second lung cancer case-control study was conducted to validate findings from the New Mexico study. The study population was derived from a large ongoing molecular epidemiologic study in Boston, MA, that began in 1992 and now has >2,205 NSCLC patients. Details of this case-control population have been described previously (21–23). For this study >3,700 samples were analyzed, which included smokers and nonsmokers. This study was approved by the Human Subjects Committees of Massachusetts General Hospital and Harvard School of Public Health, Boston, MA. Briefly, all histologically confirmed, newly diagnosed patients with NSCLC at Massachusetts General Hospital were recruited between December 1992 and February 2006. Before 1997, only early stage (stage I and II) patients were recruited. After 1997, all stages of NSCLC cases were recruited in this study. Controls were recruited at Massachusetts General Hospital from healthy friends and nonblood-related family members (usually spouses) of several groups of hospital patients: (a) patients with cancer, whether related or not related to a case; or (b) patients with a cardiothoracic condition undergoing surgery. No matching was performed. Importantly, none of the controls were patients. Potential controls with a previous diagnosis of any cancer (other than nonmelanoma skin cancer) were excluded from participation. Over 85% of eligible cases and over 90% of controls participated in this study and provided blood samples. A research nurse administered questionnaires on demographic information and detailed smoking history (see Supplementary Table S2B). To reduce potential variation in allele frequency by ethnicity, only Caucasians were considered in the analysis.

*Evaluation of 3′UTR sequences and the LCS6 SNP.* DNA was isolated from frozen and FFPE lung tissue using the DNeasy Blood and Tissue kit (Qiagen). Segments of the KRAS 3′UTR were amplified using PfuTurbo DNA polymerase (Stratagene) and DNA primers specific to this sequence (see Supplementary Table S3). PCR products were purified using the QiAquick PCR Purification kit or 96 PCR Purification kit (Qiagen) and sequenced using the same primers. The KRAS 3′UTR was sequenced in the same manner.

For high-throughput genotyping, the DNA isolated from lymphocytes, blood, or tumor samples was amplified using TaqMan PCR assays designed specifically to identify the T or G allele of the LCS6 SNP (Applied Biosciences).

**Determining the effect of the LCS6 variant allele on KRAS expression.** We generated a pGL3 derivative containing almost the entire KRAS 3′UTR (KRAS wild-type) as follows. KRAS wild-type includes 3,910 bp of the KRAS 3′UTR, which was amplified from human genomic DNA using the forward primer SMJ104 and reverse primer LCJ5 (see Supplementary Table S3). XhoI restriction sites were included on the 5′ ends of the primers for convenient cloning. The product was first cloned into the TOPO cloning vector (Invitrogen) and then subcloned into pGL3 (Ambion) for use in subsequent luciferase assays. The luciferase reporter with the variant LCS6 KRAS 3′UTR (KRAS mLCS6) was constructed through site-directed mutagenesis of KRAS wild-type using GeneTailor (Invitrogen). A549 cells were cultured in DMEM with 10% fetal bovine serum and penicillin/

**Figure 1.** Identification of a SNP in the KRAS 3′UTR. A, location of the putative LCSs in the KRAS 3′UTR. LCS1-LCS8 had been previously identified (12). LCSs where mutations were found are shown in red. The KRAS 3′UTR is 5,016 bp, and the markers are positioned every 1,000 bp. B, the sequence of LCS6 with either the reference allele or the variant allele (red). C, the variant allele in LCS6 was seen in 20.3% of the primary lung tumors and was present in the adjacent tissue when available (*)

\[\text{http://alfred.med.yale.edu/}\]
streptomycin (Invitrogen). A549 cells were transfected with 500 ng KRAS wild-type or KRAS mLCS6 and 50 ng pRL-TK (Promega) using Lipofectamine 2000 (Invitrogen) for 24 h. Reporter expression was analyzed using the Dual-Luciferase Reporter Assay (Promega) and Wallac Victor2 1420 (PerkinElmer). Two-tailed t tests were done to verify statistical significance of differences in luciferase expression using GraphPad Prism.

Measuring let-7 in vivo. RNA was isolated from normal or tumor tissue specimens using Ambion isolation kits. RTPCR was done for let-7a, b, d, and g using the Ambion RTPCR specific primers for eight tumor samples with the variant allele and eight samples without the variant allele, using the ABI 7900.

Statistical analysis. To calculate significance, a χ² test was used for categorical variables, a t test was used for continuous variables, and in some cases, a two-sided Fisher's exact test was used. Two-sided two-sample t tests, χ² analyses, and two-sided Wilcoxon rank-sum tests were performed, as appropriate to compare the demographic variables between cases and controls. An unconditional logistic regression model was used to calculate odds ratio (OR) and confidence intervals (CI) for the KRAS LCS6 SNP in the general population with a unique human genetic resource found at Yale University, a collection of genomic DNA from 2,433 healthy individuals from the HapMap (24) and dbSNP (25) databases, only one SNP, rs712 (−), is reported in an LCS. Tumors (and adjacent normal tissue when available) from 74 patients with NSCLC were evaluated for sequence variations in these 10 LCSs. SNPs in LCS1, LCS9, and LCS4 (see Supplementary Table S4) were randomly observed at low frequency in tumor or adjacent tissue, suggesting no specific involvement in NSCLC. However, a SNP (T to G, with G the less frequent variant) identified at the fourth nucleotide in LCS6 was found in 20.3% of tumor and corresponding adjacent normal tissues from NSCLC patients (Fig. 1B and C; Supplementary Fig. S1). Because this was a previously unreported SNP that was found frequently in our NSCLC patients, we hypothesized that this SNP might be a marker of an increased risk for NSCLC.

As a control, we sequenced (in the same NSCLC patients) the 3′UTR of NRAS, which is not associated with lung cancer but contains 9 putative LCSs (12), to look for similar SNPs. No SNPs were identified within the LCSs of the NRAS 3′UTR (data not shown), further supporting the idea that the identified SNP in the KRAS 3′UTR is likely an important change with respect to NSCLC.

Frequency of the variant allele across world populations. To determine if the prevalence of the variant allele in our NSCLC population was higher than expected, we determined the allele frequencies of the LCS6 SNP in the general population with a unique human genetic resource found at Yale University, a collection of genomic DNA from 2,433 healthy individuals from around the world. The frequencies of the reference allele (T) at the SNP locus were examined in 2,433 people. They represent 46 populations from around the world, which were categorized based on geography: Africa (blue); Europe, Southwest Asia, and Western Siberia (pink); South central Asia, East Asia, and the Pacific (green); and the Americas (orange). The frequency of the reference allele across all chromosomes tested is 97.1% and the frequency of the alternative allele (G) is 2.9%. The frequency of the G allele in African populations is 1.3%, in European populations is 7.6%, in Asian populations is 0.3%, and in Native American populations is 0.3%. The allele frequencies in the individual populations are in ALFRED.

Results

A SNP in a LCS in the KRAS 3′UTR. Human KRAS expression is regulated in a 3′UTR and let-7–dependent manner through 10 putative LCSs in its 3′UTR (ref. 12; Fig. 1A). Based on data from the HapMap (24) and dbSNP (25) databases, only one SNP, rs712 (−), is reported in an LCS. Tumors (and adjacent normal tissue when available) from 74 patients with NSCLC were evaluated for sequence variations in these 10 LCSs. SNPs in LCS1, LCS9, and LCS4 (see Supplementary Table S4) were randomly observed at low frequency in tumor or adjacent tissue, suggesting no specific involvement in NSCLC. However, a SNP (T to G, with G the less frequent variant) identified at the fourth nucleotide in LCS6 was found in 20.3% of tumor and corresponding adjacent normal tissues from NSCLC patients (Fig. 1B and C; Supplementary Fig. S1). Because this was a previously unreported SNP that was found frequently in our NSCLC patients, we hypothesized that this SNP might be a marker of an increased risk for NSCLC.

As a control, we sequenced (in the same NSCLC patients) the 3′UTR of NRAS, which is not associated with lung cancer but contains 9 putative LCSs (12), to look for similar SNPs. No SNPs were identified within the LCSs of the NRAS 3′UTR (data not shown), further supporting the idea that the identified SNP in the KRAS 3′UTR is likely an important change with respect to NSCLC.

Frequency of the variant allele across world populations. To determine if the prevalence of the variant allele in our NSCLC population was higher than expected, we determined the allele frequencies of the LCS6 SNP in the general population with a unique human genetic resource found at Yale University, a collection of genomic DNA from 2,433 healthy individuals from around the world. The frequencies of the reference allele (T) at the SNP locus were examined in 2,433 people. They represent 46 populations from around the world, which were categorized based on geography: Africa (blue); Europe, Southwest Asia, and Western Siberia (pink); South central Asia, East Asia, and the Pacific (green); and the Americas (orange). The frequency of the reference allele across all chromosomes tested is 97.1% and the frequency of the alternative allele (G) is 2.9%. The frequency of the G allele in African populations is 1.3%, in European populations is 7.6%, in Asian populations is 0.3%, and in Native American populations is 0.3%. The allele frequencies in the individual populations are in ALFRED.

![Figure 2. Frequency of the reference allele across the world.](image-url)
A global set of 46 populations (26). An extensive data base of genetic variations in these samples exists and can be found, along with the population descriptions in ALFRED (19). Using a TaqMan assay, we found that <3% of the 4,866 chromosomes, or 5.8% of the people tested, had the G allele (variant) at the LCS6 SNP site (Fig. 2). The frequency of this allele varied across geographic populations, with “European” populations exhibiting the variant allele most frequently (7.6% of the chromosomes tested); African populations less frequently (<2.0% of chromosomes tested); and “Asian” and Native American populations infrequently (<0.4% of chromosomes tested). Of note, over 85% of the patients in our retrospective patient cohort were of European descent. Importantly, these findings indicate that the prevalence of the variant allele (20.3%) in our NSCLC patient cohort is significantly higher than expected in any existing geographic population, further supporting the hypothesis that this variant allele is a marker of an increased risk to NSCLC.

The variant allele is associated with increased risk for NSCLC in moderate smokers. Lung cancer cases and controls from smokers enrolled in a lung cancer case-control study in New Mexico were genotyped for the LCS6 variant allele. The frequency of this allele varied across geographic populations, with “European” populations exhibiting the variant allele most frequently (7.6% of the chromosomes tested); African populations less frequently (<2.0% of chromosomes tested); and “Asian” and Native American populations infrequently (<0.4% of chromosomes tested). Of note, over 85% of the patients in our retrospective patient cohort were of European descent. Importantly, these findings indicate that the prevalence of the variant allele (20.3%) in our NSCLC patient cohort is significantly higher than expected in any existing geographic population, further supporting the hypothesis that this variant allele is a marker of an increased risk to NSCLC.

**Table 1. The KRAS allele in NSCLC in case-control studies**

<table>
<thead>
<tr>
<th>Group</th>
<th>Genotype</th>
<th>Crude OR</th>
<th>P</th>
<th>Adjusted OR*</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall (218 cases vs 325 controls)</td>
<td>TT</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TG+GG</td>
<td>1.4 (0.9–2.3)</td>
<td>0.13</td>
<td>1.5 (0.9–2.4)</td>
<td>0.15</td>
</tr>
<tr>
<td>Pack-years &lt;41 (75 cases vs 195 controls)</td>
<td>TT</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TG+GG</td>
<td>2.3 (1.1–4.4)</td>
<td>0.02</td>
<td>2.3 (1.1–4.6)</td>
<td>0.02</td>
</tr>
<tr>
<td>Pack-years ≥41 (143 cases vs 130 controls)</td>
<td>TT</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TG+GG</td>
<td>1.0 (0.5–1.9)</td>
<td>0.99</td>
<td>0.9 (0.5–1.8)</td>
<td>0.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Genotype</th>
<th>Crude OR</th>
<th>P</th>
<th>Adjusted OR†</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall (2,205 cases vs 1,497 controls)</td>
<td>TT</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TG+GG</td>
<td>1.11 (0.93–1.32)</td>
<td>0.23</td>
<td>1.17 (0.97–1.44)</td>
<td>0.15</td>
</tr>
<tr>
<td>Pack-years &lt;40 (956 cases vs 1,214 controls)</td>
<td>TT</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TG+GG</td>
<td>1.28 (1.03–1.60)</td>
<td>0.03</td>
<td>1.36 (1.07–1.73)</td>
<td>0.01</td>
</tr>
<tr>
<td>Pack-years ≥40 (1,249 cases vs 283 controls)</td>
<td>TT</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TG+GG</td>
<td>0.85 (0.61–1.29)</td>
<td>0.34</td>
<td>0.89 (0.63–1.25)</td>
<td>0.49</td>
</tr>
</tbody>
</table>

*Age, race, gender, and smoking status were adjusted in an unconditional logistic regression model.
†41 pack-years is the median of 543 smokers.
‡Age, gender, smoking status, pack-years of smoking and years since smoking cessation (if exsmoker) were adjusted in an unconditional logistic regression model.
§40 pack-years is the median of 3702 smokers.

a global set of 46 populations (26). An extensive data base of genetic variations in these samples exists and can be found, along with the population descriptions in ALFRED (19). Using a TaqMan assay, we found that <3% of the 4,866 chromosomes, or 5.8% of the people tested, had the G allele (variant) at the LCS6 SNP site (Fig. 2). The frequency of this allele varied across geographic populations, with “European” populations exhibiting the variant allele most frequently (7.6% of the chromosomes tested); African populations less frequently (<2.0% of chromosomes tested); and “Asian” and Native American populations infrequently (<0.4% of chromosomes tested). Of note, over 85% of the patients in our retrospective patient cohort were of European descent. Importantly, these findings indicate that the prevalence of the variant allele (20.3%) in our NSCLC patient cohort is significantly higher than expected in any existing geographic population, further supporting the hypothesis that this variant allele is a marker of an increased risk to NSCLC.

The variant allele is associated with increased risk for NSCLC in moderate smokers. Lung cancer cases and controls from smokers enrolled in a lung cancer case-control study in New Mexico were genotyped for the LCS6 variant allele. The frequency of the variant allele in the NSCLC cases was 18.8%, which was not significantly different from the frequency in the lung cancer patients studied at Yale (P = 0.20). Although the presence of the LCS6 variant allele did not predict NSCLC risk for the entire patient cohort, the variant allele was associated with increased NSCLC risk in smokers with a <41 pack-year smoking history (OR, 2.3; 95% CI, 1.1–4.6; P = 0.02; Table 1A), meaning a person has smoked the equivalent of 1 pack of cigarettes per day for 41 years, the median level of smoking in this population. The OR was adjusted for age, gender, smoking status, pack-years of smoking, and years since smoking cessation. This finding was then validated in a larger independent study of lung cancer cases and controls enrolled in Boston, MA. In that study, the frequency of the variant allele was 18.1%, and again, no association was seen between the allele and lung cancer risk in the entire population. However, after stratifying by the median smoking level (40 pack-years), a 1.4-fold increase in risk for lung cancer was seen in persons who smoked <40 pack-years (CI, 1.1–1.7; P = 0.01; Table 1B). These studies did not find an association between the LCS6 variant allele and tumor histology.

These findings confirm that the variant allele is a marker for an increased risk of NSCLC in patients with less cigarette exposure, which in these studies was less than the median smoking exposure of ~40 pack-years. Our finding that the LCS6 SNP only affects cancer risk for those with less cigarette exposure agrees with other studies showing a dose-dependent gene-environment interaction for smoking-induced lung cancer risk (23, 27, 28); with higher smoking exposure, any genetic predisposition is hypothesized to be overwhelmed by the extent of smoking-related damage.
The LCS6 variant allele affects \textit{KRAS} regulation and is associated with lower \textit{let-7}. To determine if LCS6 affects \textit{KRAS} regulation, we used a previously described luciferase reporter construct for \textit{KRAS} expression to determine the effect of the LCS6 variant on expression (12). We transfected a luciferase reporter with a full-length \textit{KRAS} 3′UTR containing the variant allele at LCS6 (\textit{KRAS} mLCS6) into A549 cells, a lung cancer cell line, with known low \textit{let-7} levels (14), and compared luciferase expression to A549 cells transfected with the luciferase reporter with a wild-type \textit{KRAS} 3′UTR (\textit{KRAS} wild-type). We found that there was a significant increase in luciferase activity in cells transfected with \textit{KRAS} mLCS6 versus the \textit{KRAS} wild-type (Fig. 3A and B). These findings support the hypothesis that the variant allele has altered the dependence of the \textit{KRAS} 3′UTR on LCS6, and allows increased \textit{KRAS} expression in the presence of this variant allele.

We were unable to measure \textit{KRAS} levels in patient samples due to limited tissue and only DNA access for case-control samples. However, we did evaluate the \textit{KRAS} gene in patient samples harboring the SNP for common activating mutations (in codons 12, 13, and 61) and did not find any activating mutations (data not shown). As \textit{KRAS} is rarely activated in squamous tumors and in only 30% of lung adenocarcinomas, these findings are not surprising, and additional studies will be required to fully understand \textit{KRAS} expression and/or mutations in tumors harboring the variant allele.

We next measured the levels of \textit{let-7} in available patient tumors with or without the variant allele, using reverse transcription-PCR. On average, we found that the levels of \textit{let-7}a, b, d, and g were lower in patients with the variant allele compared with patients without the variant allele (Fig. 3C). These findings suggest that the variant allele is associated with lower \textit{let-7}, at least in the NSCLC tumors in this study.

**Discussion**

We have identified a variant allele in a LCS in the \textit{KRAS} 3′UTR that alters \textit{let-7}–mediated regulation of \textit{KRAS} expression and is associated with a 1.4- to 2.3-fold increased risk for NSCLC among moderate smokers. This variant allele is found in 18.1% to 20.3% of lung cancer cases versus in only 5.8% of the world populations. The association of this allele to lung cancer was identified in the NM case-control study and validated in the MA case-control study, making this finding highly significant.

We find that this variant allele at the LCS6 site indeed affects regulation of \textit{KRAS} expression \textit{in vitro}, allowing increased levels of \textit{KRAS}. Interestingly, tumors containing the variant allele had lower \textit{let-7} levels than tumors without the variant allele, and as low \textit{let-7} has been associated with a poor prognosis in NSCLC (14, 15, 29), these findings raise the possibility that the subset of NSCLC patients harboring the variant allele may be those with an especially poor prognosis, a hypothesis requiring future studies to validate.

Although it is not possible at this time to further define the mechanism of cancer predisposition for patients who harbor this variant allele, the finding that the variant allele seems to allow increased \textit{KRAS} \textit{in vitro} suggests that \textit{KRAS} overexpression might be one plausible mechanism. An additional hypothesis could be that altered \textit{let-7} binding in the \textit{KRAS} 3′UTR could somehow lower cellular levels of \textit{let-7}, perhaps through a feedback loop, and lower \textit{let-7} levels could further result in increased cell growth, as \textit{let-7} is known to repress cell growth pathways (16). For example, recent evidence indicates that a negative feedback loop involving \textit{lin-28} specifically regulates cellular \textit{let-7} levels (30). In the background of the variant allele, a cycle creating lowered \textit{let-7} levels and increased \textit{KRAS} expression could in concert act as the first steps in oncogenesis. These hypotheses need testing in future studies.

Because lung cancer is so deadly when caught at later stages, screening programs have been initiated in current and ex-smokers:
The Early Lung Cancer Action Project found that a chest computed tomography (CT) scan is thirteens more sensitive in detecting early-stage lung cancer than a chest X-ray in “high-risk” populations (2.4% versus 0.7%; refs. 31, 32). Yet there remains considerable controversy over the use of lung CT scans because of the expense (estimated cost 2 billion dollars yearly in the United States alone) and the very low yield of cancers detected yearly (1.2%; ref. 33). One of the primary problems is that although smoking is the number one risk factor for developing lung cancer, only 10% of smokers ever develop lung cancer. With 44.5 million current smokers in the United States (20.9% of the population) and over 1.3 billion smokers worldwide, there is a clear need to identify markers whose genetic variation is associated with lung cancer risk that would help better prioritize who should be screened clinically and be offered chemopreventive agents.

Although numerous studies of carcinogen metabolism and detoxifying enzymes and DNA repair genes have identified sequence variations associated with risk for lung cancer, a meta-analysis of such polymorphisms in DNA repair pathways concluded that for any increased risk association between a single SNP and lung cancer, the risk fluctuation would likely be minimal, and only panels evaluating a collection of SNPs would ever successfully predict lung cancer risk (34). Supporting this hypothesis, 2 SNPs associated with lung cancer risk (OR, 1.19–1.8) were recently reported (34) and the very low yield of cancers detected yearly (1.2%; ref. 33). The Early Lung Cancer Action Project found that a chest computed tomography (CT) scan is thirteens more sensitive in detecting early-stage lung cancer than a chest X-ray in “high-risk” populations (2.4% versus 0.7%; refs. 31, 32). Yet there remains considerable controversy over the use of lung CT scans because of the expense (estimated cost 2 billion dollars yearly in the United States alone) and the very low yield of cancers detected yearly (1.2%; ref. 33). One of the primary problems is that although smoking is the number one risk factor for developing lung cancer, only 10% of smokers ever develop lung cancer. With 44.5 million current smokers in the United States (20.9% of the population) and over 1.3 billion smokers worldwide, there is a clear need to identify markers whose genetic variation is associated with lung cancer risk that would help better prioritize who should be screened clinically and be offered chemopreventive agents.

Although numerous studies of carcinogen metabolism and detoxifying enzymes and DNA repair genes have identified sequence variations associated with risk for lung cancer, a meta-analysis of such polymorphisms in DNA repair pathways concluded that for any increased risk association between a single SNP and lung cancer, the risk fluctuation would likely be minimal, and only panels evaluating a collection of SNPs would ever successfully predict lung cancer risk (34). Supporting this hypothesis, 2 SNPs associated with lung cancer risk (OR, 1.19–1.8) were recently reported (34) and the very low yield of cancers detected yearly (1.2%; ref. 33). The Early Lung Cancer Action Project found that a chest computed tomography (CT) scan is thirteens more sensitive in detecting early-stage lung cancer than a chest X-ray in “high-risk” populations (2.4% versus 0.7%; refs. 31, 32). Yet there remains considerable controversy over the use of lung CT scans because of the expense (estimated cost 2 billion dollars yearly in the United States alone) and the very low yield of cancers detected yearly (1.2%; ref. 33). One of the primary problems is that although smoking is the number one risk factor for developing lung cancer, only 10% of smokers ever develop lung cancer. With 44.5 million current smokers in the United States (20.9% of the population) and over 1.3 billion smokers worldwide, there is a clear need to identify markers whose genetic variation is associated with lung cancer risk that would help better prioritize who should be screened clinically and be offered chemopreventive agents.

Although numerous studies of carcinogen metabolism and detoxifying enzymes and DNA repair genes have identified sequence variations associated with risk for lung cancer, a meta-analysis of such polymorphisms in DNA repair pathways concluded that for any increased risk association between a single SNP and lung cancer, the risk fluctuation would likely be minimal, and only panels evaluating a collection of SNPs would ever successfully predict lung cancer risk (34). Supporting this hypothesis, 2 SNPs associated with lung cancer risk (OR, 1.19–1.8) were recently reported (34) and the very low yield of cancers detected yearly (1.2%; ref. 33). The Early Lung Cancer Action Project found that a chest computed tomography (CT) scan is thirteens more sensitive in detecting early-stage lung cancer than a chest X-ray in “high-risk” populations (2.4% versus 0.7%; refs. 31, 32). Yet there remains considerable controversy over the use of lung CT scans because of the expense (estimated cost 2 billion dollars yearly in the United States alone) and the very low yield of cancers detected yearly (1.2%; ref. 33). One of the primary problems is that although smoking is the number one risk factor for developing lung cancer, only 10% of smokers ever develop lung cancer. With 44.5 million current smokers in the United States (20.9% of the population) and over 1.3 billion smokers worldwide, there is a clear need to identify markers whose genetic variation is associated with lung cancer risk that would help better prioritize who should be screened clinically and be offered chemopreventive agents.
A SNP in a let-7 microRNA Complementary Site in the KRAS 3′ Untranslated Region Increases Non–Small Cell Lung Cancer Risk

Lena J. Chin, Elena Ratner, Shuguang Leng, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/68/20/8535

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2008/10/13/68.20.8535.DC1

Cited articles
This article cites 35 articles, 14 of which you can access for free at:
http://cancerres.aacrjournals.org/content/68/20/8535.full#ref-list-1

Citing articles
This article has been cited by 40 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/68/20/8535.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, contact the AACR Publications Department at permissions@aacr.org.