Promoter Hypermethylation of Multiple Genes in Sputum Precedes Lung Cancer Incidence in a High-Risk Cohort

Steven A. Belinsky,[1] Kieu C. Liechty,[1] Frederick D. Gentry,[1] Holly J. Wolf,[2] Justin Rogers,[2] Kieu Vu,[2] Jerry Haney,[2] Tim C. Kennedy,[2] Fred R. Hirsch,[2] York Miller,[2] Wilbur A. Franklin,[2] James G. Herman,[3] Stephen B. Baylin,[3] Paul A. Bunn,[2] and Tim Byers[2]

[1]Lovelace Respiratory Research Institute, Albuquerque, New Mexico; [2]University of Colorado Health Sciences Center, Denver, Colorado; and [3]Johns Hopkins University, Baltimore, Maryland

Abstract

A sensitive screening approach for lung cancer could markedly reduce the high mortality rate for this disease. Previous studies have shown that methylation of gene promoters is present in exfoliated cells within sputum prior to lung cancer diagnosis. The purpose of the current study is to conduct a nested case-control study of incident lung cancer cases from an extremely high-risk cohort for evaluating promoter methylation of 14 genes in sputum. Controls (n = 92) were cohort members matched to cases (n = 98) by gender, age, and month of enrollment. The comparison of proximal sputum collected within 18 months to >18 months prior to diagnosis showed that the prevalence for methylation of gene promoters increased as the time to lung cancer diagnosis decreased. Six of 14 genes were associated with a >50% increased lung cancer risk. The concomitant methylation of three or more of these six genes was associated with a 6.5-fold increased risk and a sensitivity and specificity of 64%. This is the first study to prospectively examine a large panel of genes for their ability to predict lung cancer and shows the promise of gene promoter hypermethylation in sputum as a molecular marker for identifying people at high risk for cancer incidence. (Cancer Res 2006; 66(6): 3338-44)

Introduction

Lung cancer is the leading cause of cancer mortality in the U.S. and 1.5 million deaths are projected worldwide from this disease by 2010 (1, 2). The overall 5-year survival rate for lung cancer is <15%, due largely to the late stage at which most patients are diagnosed and the lack of effective treatments for this systemic disease. A validated screening approach for lung cancer could substantially reduce the high mortality rate from this disease. The benefit of early detection is seen in patients with stage I tumors (<3 cm) in which surgical resection is the treatment option of choice, and the rate of recurrence within 5 years is <50% (3–5). With the adoption of adjuvant chemotherapy for these early stage lung cancer patients, the rate of recurrence should decline further (6).

Cytologic and/or genetic biomarkers for lung cancer risk detected in sputum could complement radiological imaging and bronchoscopy for detecting early lung tumors (7). Previous studies have shown that cytologic atypia present in epithelial cells exfoliated into sputum precedes lung cancer diagnosis (8). Transcriptional silencing of genes by CpG island methylation is now recognized as a crucial component in lung cancer initiation and progression (9, 10). The development of the methylation specific PCR (MSP) assay has allowed for the assay of methylation of specific genes in sputum where epithelial cells comprise only a fraction of the cellular content (11). In a small proof-of-concept study, methylation of the CDKN2A (*p16*) or O^6-methylguanine-DNA methyltransferase (*MGMT*) gene promoters was detected in sputum up to 3 years prior to the diagnosis of squamous cell carcinoma (12). The evaluation of cancer-free individuals who were at risk for lung cancer because of smoking and/or exposure to radon through uranium mining revealed methylation of the *p16* and *MGMT* genes in 15% and 25% of sputum samples, respectively. The most striking difference seen between lung cancer cases and controls was the detection of both methylated genes in 48% of sputum samples from cases but in only 3% of controls. Follow-up revealed three lung cancers diagnosed in controls 1 to 3 years after sputum collection with the *MGMT* gene being methylated in two of these individuals.

Several subsequent small studies support the promise of gene-specific methylation as a biomarker for predicting lung cancer (13–16). In those studies, methylation of *p16* and other genes was assessed in sputum from people at high risk for lung cancer. Follow-up conducted in two of those studies revealed that of the eight people positive for methylation of *p16* in their sputum, three developed lung cancer ~1 year after sputum collection (13). In addition, of the five people positive for methylation of the *ras* effector homologue 1 (*RASSF1A*) gene, three developed lung cancer 12 to 14 months after sputum collection (16). Methylation also persists after smoking cessation based on similar prevalence in sputum from current versus former smokers (17). Together, these studies indicate that the detection of gene promoter methylation in sputum could be a powerful approach to population-based screening for the early detection of lung cancer.

The well-documented field cancerization seen in lungs from smokers, stemming from the exposure of the entire respiratory tract to inhaled carcinogens within cigarette smoke, presents an obstacle to the early detection of lung cancer (18). The generation of multiple, independently initiated sites throughout the lungs of

Conflict of Interest Statement: S.A. Belinsky, S.B. Baylin, and J.G. Herman are consultants to OncoMethylome Sciences. Under a licensing agreement between Lovelace Respiratory Research Institute and OncoMethylome Sciences, nested methylation-specific PCR was licensed to OncoMethylome Sciences and S.A. Belinsky is entitled to a share of the royalties received by the Institute from sales of the licensed technology. The Institute, in accordance with its conflict-of-interest policies, is managing the terms of these arrangements. The commercial rights to the standard MSP technique was also licensed to OncoMethylome Sciences and S.B. Baylin and J.G. Herman are entitled to royalties from any commercial use of this procedure.

Requests for reprints: Steven A. Belinsky, Lung Cancer Program, Lovelace Respiratory Research Institute, 2425 Ridgecrest Drive Southeast, Albuquerque, NM 87108. Phone: 505-348-9465; Fax: 505-348-4990; E-mail: sbelinsk@LRRI.org.

doi:10.1158/0008-5472.CAN-05-3408

people with a long history of smoking likely accounts for detecting methylation of genes such as *p16* that is inactivated in the earliest stages of preinvasive disease (19). The use of promoter methylation as a biomarker for early detection of lung cancer may therefore require a panel of genes whose presence in sputum confer a high enough sensitivity and specificity for distinguishing very advanced dysplasia or early lung cancer from the large "at risk" population (20). A critical step in the development of a gene panel requires testing the biomarker in a prospective population-based study (21). The University of Colorado Specialized Program of Research Excellence in Lung Cancer initiated a cohort study in 1993 to assess whether sputum cytology or other molecular markers could be used to predict early lung cancer (17). The purpose of the current investigation was to conduct a nested, case-control study of incident lung cancer cases from this cohort to evaluate promoter methylation of 14 genes in sputum.

Materials and Methods

Study population. The University of Colorado Cancer Center Sputum Screening Cohort Study is an ongoing prospective study initiated in 1993 to determine whether biomarkers identified within sputum could predict future lung cancer development. The study methodology has been described previously (8). Briefly, subjects were recruited from community and academic pulmonary clinics primarily in the Denver, CO metropolitan area. At enrollment, subjects were 25 years or older with a cigarette smoking history of ≥30 pack years, and with pulmonary air flow obstruction documented by a spirometry finding of forced expiratory volume in 1 second (FEV1) of 75% or lower than predicted for age, and an FEV1/FVC ratio of ≤0.75. Subjects who had a diagnosis of cancer within 5 years prior to the time of recruitment (excluding non–melanoma skin cancer), a current acute respiratory infection, or who were judged by their physician to have a life expectancy of <5 years were excluded from the study. Participants were provided with two containers filled with a fixative solution of 2% carbowax and 50% alcohol (Saccomanno's fixative) and were instructed to collect an early morning, spontaneous cough sputum specimen for 6 consecutive days (3 days' collection into the first container and 3 days' collection into the second container). Material from the second 3-day pooled sputum specimen was sampled for this study (22).

Cohort members were contacted once a year to ask for their continued participation and to provide another sputum sample. Other cohort members were followed by active methods such as mail and by passive methods such as the matching to the Colorado Department of Public Health and Environment Vital Statistics records (through February 2005), the National Death Index (searched through December 2001), and the Colorado Central Cancer Registry (searched through December 2004). Among the 3,259 cohort members in this analysis, there were 1,353 documented deaths and 182 documented lung cancers. Cases in this analysis were those cohort members who contributed a sputum sample, who were not known or suspected to have a cancer at enrollment, and who were subsequently diagnosed with incident lung cancer after enrollment. Controls were cohort members matched to cases by gender, age, and month of enrollment. There were 121 cases and 120 matched controls selected for this study. Among those, DNA from 50 subjects was of insufficient quality to obtain results in the MSP assays. These subjects did not differ from the cases and controls that were evaluated for methylation with respect to demographics, clinical history, sputum cytology, or smoking history ($P > 0.05$). Two controls were found to be incident cases with ongoing follow-up during the course of the study. In addition, another case was determined to be a carcinoma *in situ* and was excluded. Therefore, the final analysis set included 98 cases and 92 controls. The Colorado Multi-Institutional and Lovelace Respiratory Research Institute (LRRI) Review Boards approved this study. Participants previously gave their consent for the use of sputum samples to investigate potential biomarkers in the University of Colorado Cancer Center Sputum Screening Cohort Study.

Sputum processing and MSP. Sputum samples were stored in Saccomanno's fixative. Four slides were prepared from sputum samples and stained using the Papanicolaou technique as described elsewhere (23). Slides were independently screened by cytotechnologists as: not adequate for diagnosis, normal, squamous metaplasia, mild atypia, moderate atypia, severe atypia, or carcinoma (23, 24). An aliquot from all sputum samples, irrespective of adequacy, for each study subject was taken, and DNA was isolated by protease digestion followed by phenol chloroform extraction and ethanol precipitation. DNA (6 μg) was sent to investigators at the LRRI for methylation assays. Samples were labeled only with study-specific coded identifiers to blind the LRRI investigators from case or control status. Assays were done with both cases and controls included in each batch.

Fourteen genes were selected for evaluation based on their prevalence in lung tumors (≥25%), diversity of function, and timing for inactivation during lung cancer development when known (19, 25–33). The *BETA3* and helix-loop-helix (*HLHP*) genes (34, 35) were identified through a genome-wide screening approach for methylation; however, the prevalence for methylation of these genes in lung cancer has not been reported yet.[4] Nested MSP was used to detect methylated alleles in DNA recovered from the sputum samples. We used our nested MSP assay described previously (12, 17) because of its increased sensitivity for the detection of promoter hypermethylation in biological fluids. In order to conserve DNA and effort, stage 1 multiplex PCR reactions amplifying four genes at once were done. DNA (120-150 ng) was used for stage 1 PCR following modification with bisulfite. PCR primers for stage 1 and 2 amplification of the *p16, MGMT*, death associated protein kinase (*DAPK*), *RASSF1A, PAX5* α, and *PAX5* β genes have been described elsewhere (12, 17, 25). PCR primers and conditions used for amplification of the *GATA4, GATA5*, secreted frizzle like protein 1 (*SFRP1*), laminin C2 (*LAMC2*), insulin-like growth factor receptor 3 (*IGFBP3*), H-cadherin, *BETA 3*, and *HLHP* genes are available from the corresponding author. Conditions for all stage 1 multiplexes were optimized through primer design and PCR conditions to achieve equal product intensity. These optimal conditions ensured a similar sensitivity for the detection of methylated alleles across genes in the stage 2 MSP assay. All stage 2 PCR reactions were conducted at annealing temperatures (68-70°C) that exceed the melting temperature of the primers to ensure the highest specificity for amplification of only methylated alleles present in the DNA sample. Cell lines positive and negative for methylation of these 14 genes and water blanks (bisulfite-modified and unmodified water) were used as controls for the MSP assays.

Because inflammatory, epithelial, and oral cells from the entire aerodigestive tract comprise the sputum sample, DNA from epithelial cells from the lower respiratory tract can be only a very small proportion of the total DNA in the sample. This scarcity of epithelial cell-based DNA can effect sampling to detect a rare methylation event because of the exfoliation of a minimal number of epithelial cells harboring a methylation change or the presence of a much higher number of inflammatory cells. To address the issue of stochastic sampling for this study and its effects on the sensitivity of the MSP assay, mixing experiments were conducted between cell lines positive and negative for methylation of the *p16* gene. DNA isolated from Calu6 cells (p16 methylated) was added to DNA isolated from H2009 (p16 unmethylated) at dilutions ranging from 1:10 to 1:50,000. Bisulfite modification (1 μg) was done on each sample and stage 1 MSP was conducted in triplicate, thereby sampling the modified DNA thrice. Stage 2 MSP for p16 was conducted on each sample. Methylation of the *p16* gene was always detected in samples containing 1:10 to 1:10,000 methylated alleles. In contrast, methylation was detected in two of three samples containing 1:20,000 and one of three samples containing 1:50,000 methylated alleles (data not shown). Based on these results, we evaluated each sputum sample in duplicate beginning with the bisulfite modification step.

Two separate aliquots (1 μg each) of DNA from each sputum sample were modified by bisulfite and subjected to MSP analysis for the 14 genes. A sample was called positive for methylation of a specific gene if either of the MSP assays were positive. The concordance between duplicate assays

[4] S.A. Belinsky, S.B. Baylin, and W.A. Palmisano, unpublished.

ranged between 65% and 95% for the 14 genes evaluated. As expected, the concordance between duplicate assays was inversely related to the prevalence for gene methylation in the sputum samples. A subset of samples (20%) that gave positive methylation products also was analyzed by methylation-sensitive restriction enzyme digestion of the resulting PCR product. The restriction digestion allows one to examine the methylation-state of CpGs within the amplified PCR product and serves as a control for false priming (17). Digestion within at least one of the restriction sites was seen for all samples positively confirming methylation.

Data analysis. The data were summarized using frequencies and percents for categorical variables. Differences in distribution between cases and controls were first examined with χ^2. Logistic regression models were then examined to assess the association between moderate or more severe atypia and other risk factors. First, univariate models were examined, then multivariate models were developed that adjusted for the most important covariates. Because the matching broke down with the exclusion of some of the cases and controls, unconditional methods for multivariate analysis were used whereas retaining all matching variables in the models as covariates. Age was considered as a continuous variable in the analyses. Former smokers were defined as those individuals who had quit smoking 1 year or more at the time of questionnaire completion. Pack years of cigarette smoking at enrollment was defined as the average number of packs smoked per day multiplied by the number of years of smoking. FEV-1 was the value assessed at the time of cohort enrollment.

Associations were expressed as odds ratios and their corresponding 95% confidence intervals (CI). Analyses were conducted to assess the association with lung cancer for each gene separately, then for different combinations of genes together. Multiplicity of methylated genes was determined as the number of genes methylated in the sample collected closest to lung cancer diagnosis among a panel of six genes with the highest individual odds ratio (*p16, PAX5-β, MGMT, DAPK, GATA5*, and *RASSF1A*). Those analyses were also stratified by the length of time between the sputum collection and lung cancer diagnosis (>18 months versus within 18 months of diagnosis). The relationship between cytologic atypia and methylation was assessed, as was the relationship between methylation and the histologic type of lung cancer. Statistical significance was expressed by *P* values and 95% CI. However, because these were exploratory analyses, the type 1 error rate for some of the associations may be underestimated by the *P* values and 95% CI expressed in the tables, especially in the analysis of multiplicity among the genes in panels constructed based on the strength of association with lung cancer. All analyses were carried out using Statistical Analysis Software (version 8.1, SAS Institute, Inc., Cary, NC).

Results

Exposure history, sputum cytology, and pathology. Key demographic variables for cases and controls are summarized in Table 1. The only significant difference seen was a greater proportion of former smokers among controls than cases. With respect to longitudinal follow-up, less than two sputum samples were available for evaluation from each case and control (Table 1). Therefore, sputum samples collected proximal to lung cancer diagnosis for the incident cases and controls were examined for gene promoter methylation. Sputum adequacy defined as the presence of deep lung macrophages or Curschmann's spiral (23) was observed for 93% of the proximal sputum collected from cases and controls. Moderate atypia was present in sputum from 35 subjects and severe atypia from 10 subjects. The distribution of tumor histology among the 98 cases was 20% squamous cell carcinoma, 28% adenocarcinoma, 7% large cell, 8% small cell, and 37% carcinoma (subtype not specified).

Gene promoter methylation in sputum. Examples of the first and second stages of the nested PCR assay used to evaluate promoter methylation status for the *p16, MGMT, DAPK,* and *RASSF1A* genes and results for a stage 2 PCR for detection of a methylated *p16* gene are shown in Fig. 1. The individual odds ratio in the proximal sputum sample ranged from 0.7 to 1.9 for detecting methylation of a specific gene in cases versus controls after adjustment for age, sex, FEV1, pack years, and current smoking status (Table 2). The strongest association was the ~2-fold increased risk seen for the *p16* gene. In addition, the detection of cytologic atypia (moderate or severe) was associated with a 1.7-fold increased risk for lung cancer. Because the proximal sputum was collected from 15% and 25% of cases >2 and 4 years (up to 72 months), respectively, prior to cancer diagnosis, we evaluated whether the prevalence for methylation of these biomarkers increased in the months leading up to lung cancer diagnosis. An increase in odds for methylation ranging from 1.5- to 2.2-fold was seen for each of the *p16, PAX5 β, MGMT, DAPK, RASSF1A, GATA4*, and *GATA5* genes in cases compared with controls when comparing methylation in sputum samples collected within 18 months of diagnosis (Table 3). The odds ratio for cytologic atypia also increased to 2.0. In contrast, only the *p16* gene was associated with an adjusted odds ratio that exceeded 1.5 in cases compared with controls when comparing methylation in sputum samples collected >18 months from cancer diagnosis (Table 3).

The relationships between gene-specific promoter methylation in sputum and tumor histology and cytology were examined for the *p16, MGMT, PAX5 α, PAX5 β, DAPK, GATA4, GATA5*, and *RASSF1A* genes. These genes were selected because they exhibit the largest increase in odds ratio when examining all proximal sputum or specimens collected within 18 months (Tables 2 and 3). Patterns of gene methylation were not different between specific histologic types of lung cancer (data not shown). Only the prevalence for methylation of the *MGMT* gene was positively associated with increasing cytologic changes (moderate atypia, severe dysplasia; *P* value for trend <0.04). The presence of cytologic atypia in sputum samples did not predict for specific tumor histology (data not shown).

Table 1. Summary of selected variable by case-control status

Variable	Cases (*n* = 98)	Controls (*n* = 92)	*P*
Number (%)			
Age (y)			0.08
30-59	22 (22)	15 (16)	
60-69	37 (38)	41 (45)	
70+	39 (40)	36 (39)	
Sex			0.81
Male	76 (78)	70 (76)	
Female	22 (22)	22 (24)	
Smoking status			0.03
Current	42 (43)	26 (28)	
Former	56 (57)	66 (72)	
Pack years			0.24
≤50	32 (33)	35 (38)	
51-74	25 (26)	30 (33)	
>75	41 (42)	27 (29)	
Average no. of sputum samples	1.6	1.7	0.38

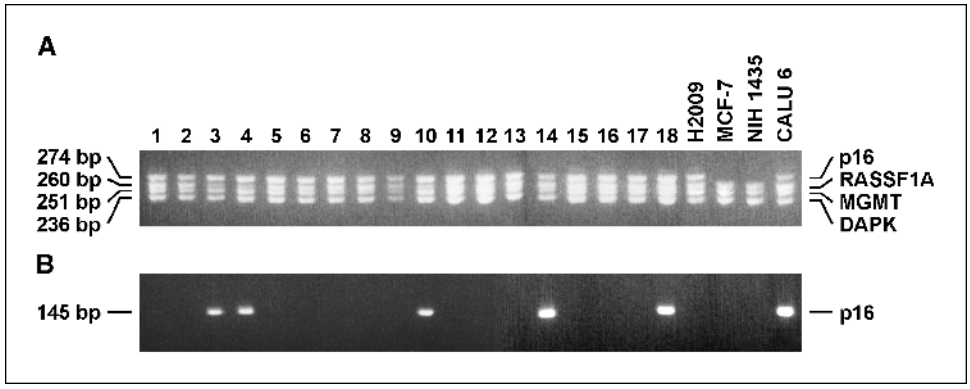

Figure 1. Nested MSP for amplification of the *p16, MGMT, DAPK*, and *RASSF1A* genes. DNA from 18 sputum samples and four cell lines were amplified. Stage I products of equivalent intensity are seen for all sputum samples, Calu6, and H2009. The *p16* gene is deleted in the MCF-7 and NIH 1435 cell lines, thus, only three PCR products are seen. The stage I product is diluted 1:50 and 5 μL used in a stage II amplification with primers specific to methylated *p16*. Methylation of *p16* is detected in five sputum samples (*lanes 3, 4, 10, 14,* and *18*) and in the Calu6 cell line.

Field cancerization effects in high-risk smokers will likely dictate the development of a gene panel for early detection of lung cancer that focuses on assessing the prevalence for promoter hypermethylation of multiple genes in individual sputum specimens. This was evaluated by using a panel of genes that included *p16, MGMT, DAPK, RASSF1A, PAX5 β*, and *GATA5*. These genes were selected because their individual odds ratios were >1.5 in proximal sputum samples collected within 18 months of lung cancer diagnosis (Table 3). This approach yielded the most important finding of our study: cases with methylation of three or more genes in their sputum that was collected within 18 months of diagnosis had a 6.5-fold increased risk for lung cancer (95% CI, 1.2-35.5; Table 4). This represents a sensitivity and specificity for predicting lung cancer of 64%. Additional analyses (data not shown) of four or five positive markers in this panel did not show any greater discrimination for case status than the three or more markers described above. The addition of cytology as a biomarker to the multivariate models containing the methylation markers did not substantially alter either the methylation odds ratios or the cytologic atypia odds ratio, as methylation and cytologic atypia were not associated in these samples. For sputum samples collected >18 months before cancer diagnosis, a 1.5-fold greater odds for methylation of three or more genes was seen in cases compared with controls (Table 4).

Discussion

This is the first study to prospectively examine a large panel of genes for their sensitivity and specificity for predicting lung cancer development in a cohort of very high-risk subjects. A key finding was the increase in detection of some genes in sputum samples collected within 18 months versus >18 months prior to cancer diagnosis. Furthermore, methylation of three or more genes in sputum collected within 18 months of diagnosis was associated with a 6.5-fold increase in the risk for lung cancer. This significant finding supports our hypothesis of an expanding field of precancerous changes throughout the aerodigestive tract that exfoliates cells with increasing frequencies of genetic and epigenetic changes as cancer develops. The fact that gene promoter methylation in sputum was not associated with a particular histologic diagnosis of lung cancer also substantiates the utility of gene promoter methylation as a biomarker for predicting lung

Table 2. Prevalence and odds for gene promoter methylation and cytologic atypia in proximal sputum samples

Biomarker	Cases, % (*n* = 98)	Controls, % (*n* = 92)	Odds ratio (CI)	Adjusted odds ratio* (CI)
P16	39 (40)	25 (27)	1.8 (1.0-3.3)	1.9 (1.0-3.7)
PAX5 β	41 (42)	32 (35)	1.3 (0.7-2.4)	1.4 (0.7-2.5)
PAX5 α	28 (29)	24 (26)	1.1 (0.6-2.1)	1.2 (0.6-2.3)
MGMT	23 (24)	22 (24)	1.0 (0.5-1.9)	0.9 (0.5-1.8)
DAPK	42 (43)	30 (33)	1.6 (0.9-2.8)	1.5 (0.8-2.7)
GATA5	34 (35)	26 (28)	1.3 (0.7-2.5)	1.4 (0.7-2.7)
GATA4	48 (49)	42 (46)	1.1 (0.6-2.0)	1.2 (0.6-2.1)
RASSF1A	12 (12)	6 (7)	2.0 (0.7-5.6)	1.6 (0.6-4.7)
SFRP1	68 (69)	71 (77)	0.7 (0.4-1.3)	0.7 (0.4-1.4)
HLHP	42 (43)	36 (39)	1.2 (0.7-2.1)	1.2 (0.6-2.1)
BETA3	12 (12)	11 (12)	1.1 (0.5-2.8)	1.5 (0.6-2.1)
IGFBP3	25 (26)	30 (33)	0.7 (0.4-1.3)	0.7 (0.3-1.3)
HCAD	27 (28)	23 (25)	1.1 (0.6-2.2)	1.2 (0.6-2.4)
LAMC2	72 (74)	70 (76)	0.9 (0.5-1.7)	1.0 (0.5-2.0)
Atypia †	27 (28)	12 (20)	1.6 (0.8-31)	1.7 (0.9-3.1)

*Adjusted for age, sex, FEV1, pack years, and current smoking status.
†Moderate atypia or worse.

Table 3. Prevalence and odds for gene promoter methylation and cytologic atypia in proximal sputum samples obtained 3 to 18 and 19 to 72 months prior to cancer diagnosis

Biomarker	Cases (%)*	Controls (%)*	Odds ratio (CI)	Adjusted odds ratio † (CI)
3-18 Months prior to cancer diagnosis				
P16	22 (42)	13 (29)	1.8 (0.9-4.5)	2.2 (0.9-5.2)
PAX5 β	24 (46)	16 (34)	1.7 (0.7-3.7)	1.9 (0.8-4.3)
MGMT	17 (33)	10 (21)	1.8 (0.7-1.9)	1.7 (0.7-4.5)
DAPK	24 (46)	16 (34)	1.7 (0.7-3.7)	1.6 (0.7-3.7)
GATA5	18 (35)	12 (26)	1.5 (0.6-3.7)	1.9 (0.7-5.1)
GATA4	26 (50)	20 (43)	1.4 (0.6-3.0)	1.5 (0.6-3.6)
RASSF1A	7 (14)	3 (6)	2.3 (0.6-9.4)	1.7 (0.4-7.6)
Atypia ‡	19 (37)	10 (21)	2.1 (0.9-5.2)	2.0 (0.8-5.2)
19-72 Months prior to cancer diagnosis				
P16	17 (37)	12 (27)	1.6 (0.7-3.9)	1.8 (0.7-5.0)
PAX5 β	17 (37)	16 (36)	1.1 (0.5-2.5)	1.0 (0.4-2.6)
MGMT	6 (13)	12 (27)	0.4 (0.1-1.2)	0.4 (0.1-1.3)
DAPK	18 (39)	14 (31)	1.4 (0.6-3.4)	1.3 (0.5-3.1)
GATA5	16 (35)	14 (31)	1.2 (0.5-2.8)	1.3 (0.5-3.1)
GATA4	22 (48)	22 (49)	1.0 (0.4-2.2)	1.0 (0.4-2.5)
RASSF1A	5 (11)	3 (7)	1.7 (0.4-7.6)	1.2 (0.3-6.0)
Atypia ‡	8 (17)	8 (18)	1.0 (0.3-2.9)	0.9 (0.3-2.9)

*Cases (n = 52) and controls (n = 47) comprised the study group for sputum samples collected within 18 months of cancer diagnosis. Cases (n = 46) and controls (n = 45) comprised the study group for sputum samples collected >18 months prior to cancer diagnosis.
†Adjusted for age, sex, FEV1, pack years, and current smoking status.
‡Moderate atypia or worse.

cancer. It is highly unlikely that asymptomatic subjects with early stage lung cancer will exfoliate only cells specific to a small (<1 cm) lesion. Rather, as the field cancerization process grows in the lungs, more cells are exfoliated containing methylated genes, and the probability for an early lung tumor is increased (20). As expected, the proportion of controls with three or more methylated genes was similar in sputum collected within or >18 months before diagnosis of cancer in the matched case. Furthermore, multiplicity for methylation in sputum collected from cases >18 months from cancer diagnosis was also quite similar to both groups of controls. Additional longitudinal studies of this type would be required to continue evaluating how the prevalence for gene methylation in sputum changes closer to cancer diagnosis. Detecting lung cancer 12 to 18 months prior to clinical symptoms will likely affect the

Table 4. Prevalence and odds for multiple gene promoter methylation events in proximal sputum samples obtained 3 to 18 and 19 to 72 months prior to cancer diagnosis

No. of genes methylated*	Cases (%) †	Controls (%) †	Odds ratio (CI)	Adjusted odds ratio ‡ (CI)
3-18 Months prior to cancer diagnosis				
0	3 (6)	7 (15)	Reference	Reference
1	7 (14)	13 (47)	1.3 (0.2-6.4)	3.5 (0.3-40.8)
2	9 (17)	10 (21)	2.1 (0.4-10.7)	4.3 (0.5-36.7)
≥3	33 (64)	17 (36)	4.5 (1.0-19.8)	6.5 (1.2-35.5)
P for trend			0.004	0.02
19-72 Months prior to cancer diagnosis				
0	7 (15)	8 (18)	Reference	Reference
1	7 (15)	9 (20)	0.9 (0.2-3.7)	0.5 (0.1-2.8)
2	13 (28)	11 (24)	1.4 (0.4-4.9)	1.5 (0.3-6.4)
≥3	19 (41)	17 (38)	1.3 (0.4-4.3)	1.5 (0.4-6.3)
P for trend			0.56	0.42

*The genes examined included *p16, MGMT, PAX5 β, DAPK, GATA5*, and *RASSF1A*.
†Cases (n = 52) and controls (n = 47) comprised the study group for sputum samples collected within 18 months of cancer diagnosis. Cases (n = 46) and controls (n = 45) comprised the study group for sputum samples collected >18 months prior to cancer diagnosis.
‡Adjusted for age, sex, FEV1, pack years, and current smoking status.

management of this disease through "stage reduction" that will allow for an increase in the number of cases for which surgical resection can be done.

The six genes that showed the highest individual odds ratios for distinguishing incident cases from controls are inactivated at different stages of lung cancer development. Methylation of *p16* is one of the earliest epigenetic changes seen in the lungs of smokers. *P16* is detected in histologically normal bronchial epithelium from some smokers, in alveolar hyperplasia, and in lesions that are precursors to squamous cell carcinoma (17, 20, 36). In contrast, inactivation of *MGMT* is rare (8%) in hyperplasia, metaplasia, and dysplasia within the central airways (26). Methylation of *DAPK* is seen in alveolar hyperplasia in a murine model of lung cancer (37). The lack of methylation of *RASSF1A* in bronchial epithelium from smokers and the low prevalence of methylation seen in sputum from this and other studies support the inactivation of this gene as a later event in the development of preinvasive cancer (17, 38). The timing for methylation of the *PAX5 β* and *GATA5* genes in lung cancer development has not been characterized. The genes examined also have diverse cellular functions that include cell cycle regulation (*p16* and *PAX5 β*), apoptosis (*DAPK* and *RASSF1A*), signal transduction (*GATA5*), and DNA repair (*MGMT*; refs. 39–45). These functions may provide a biological explanation for their combined ability in predicting lung cancer.

The panel of six genes most associated with lung cancer risk has both sensitivity and specificity of 65%. This level of specificity is not high enough for prospective screening studies of the general population at risk for lung cancer development. However, considering the very high-risk status of the cohort studied, our results are encouraging and may be somewhat understated. In the present nested case-control study, all patients have chronic obstructive pulmonary disease (COPD) that increases risk for lung cancer by ~2.5-fold, and 65% of subjects have a smoking history of ≥50 pack years. This has resulted in a 1% to 2% incidence of new lung cancer cases each year in the cohort, and in fact, two of the controls selected became cases during the course of this study. This is in contrast to a yearly overall risk of 0.3% for all current and former smokers. The high rate of comorbidity (60% of selected controls are now deceased) makes it impossible to determine through follow-up how many of the selected controls may become a case. The controls from this study had exposures (smoking history, COPD) that were greater than the cancer-free smokers examined for methylation of the *p16* and *MGMT* genes in our original study of squamous cell carcinoma (12). This likely accounts for the higher prevalence of gene methylation in sputum from controls in the current study. We also recently conducted a study examining methylation of the *p16, MGMT, DAPK, RASSF1A, H-cadherin, PAX5 α*, and *PAX5 β* genes in sputum from 121 cancer-free women with an average smoking history of 40 pack years and a 25% incidence of COPD. Compared with cancer-free controls in the current study, a similar prevalence for methylation was seen for methylation of the *p16, RASSF1A*, and *H-cadherin* genes in sputum collected from these women, whereas the prevalence for methylation of the other four genes was 2- to 5-fold lower (38). Thus, the controls from the Colorado cohort may not reflect the average smoker's risk for lung cancer. In this regard, the present study sets the stage for large studies of smokers with more typical risk for lung cancer.

Our work also emphasizes a clear need to evaluate other promising genes in sputum from this ongoing prospective study for their potential to improve the sensitivity and specificity of this current gene panel (45–47). The optimized gene panel will ultimately be assessed for performance in prospective trials to characterize it's clinical utility. Validation will likely entail screening a high-risk population followed by clinical evaluation using spiral computed tomography and/or bronchoscopy for individuals whose methylation profile conveys the highest risk for early lung cancer. Future assessments will also include the application of machine learning approaches to the identification of combinations of biomarkers that optimize the screening variables of sensitivity and specificity. Our current findings highlight the promise that developing a panel of genes whose inactivation by promoter hypermethylation in sputum can be used to sensitively and specifically identify people at high risk for lung cancer incidence, allowing careful monitoring to achieve the earliest possible diagnosis and intervention for this deadly malignancy.

Acknowledgments

Received 9/22/2005; revised 11/28/2005; accepted 1/11/2006.

Grant support: NIH grants from the National Cancer Institute Specialized Program of Research Excellence P50 CA58187 and CA58184 and a research grant from OncoMethylome Sciences, Inc.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Randy Willink and Linda-Ann Crosby for their technical assistance, Sarah Braudrick and Jamie Kean, and the Health Statistics Unit and Colorado Central Cancer Registry of the Colorado Department of Public Health and Environment for epidemiologic follow-up of the cohort.

References

1. Jemal A, Thomas A, Murray T, Thun M. Cancer statistics. CA Cancer J Clin 2002;52:23–47.
2. Parkin DM, Bray FI, Devesa SS. Cancer burden in the year 2000. The global picture. Eur J Cancer 2001;37: S4–66.
3. Harpole DH, Herndon JE, Wolfe WG, Iglehart JD, Marks JR. A prognostic model of recurrence and death in stage I non-small cell lung cancer using presentation, histopathology and oncogene expression. Cancer Res 1995;55:51–6.
4. Read RC, Schaefer R, North N, Walls R. Diameter, cell type, and survival in stage I primary non-small cell lung cancer. Arch Surg 1988;123:446–9.
5. Rice D, Kim HW, Sabichi A, et al. The risk of second primary tumors after resection of stage I nonsmall cell lung cancer. Ann Thorac Surg 2003;76:1001–7.
6. Socinski MA. Adjuvant therapy of resected non-small cell lung cancer. Clin Lung Cancer 2004;63:162–9.
7. Henschke CI, Yankelevitz DF, Libby DM, et al. Early lung cancer action project: annual screening using single-slice helical CT. Ann N Y Acad Sci 2001;952: 124–34.
8. Prindiville SA, Byers T, Hirsch FR, et al. Sputum cytological atypia as a predictor of incident lung cancer in a cohort of heavy smokers with airflow obstruction. Cancer Epidemiol Biomarkers Prev 2003;1210:987–93.
9. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nat Rev Genet 2002;36: 415–28.
10. Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. N Engl J Med 2003;34921:2042–54.
11. Herman JG, Graff J, Myohannen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci U S A 1996;9318:9821–6.
12. Palmisano WA, Divine KK, Saccomanno G, et al. Predicting lung cancer by detecting aberrant promoter methylation in sputum. Cancer Res 2000;6021: 5954–8.
13. Kersting M, Friedl C, Kraus A, Behn M, Pankow W, Schuermann M. Differential frequencies of $p16^{INK4a}$ promoter hypermethylation, p53 mutation, and K-ras mutation in exfoliative material mark the development of lung cancer in symptomatic chronic smokers. J Clin Oncol 2000;1818:3221–9.
14. Destro A, Bianchi P, Alloisio M, et al. K-ras and $p16^{INK4a}$ alterations in sputum of NSCLC patients and in heavy asymptomatic chronic smokers. Lung Cancer 2004;441:23–32.
15. Zochbauer-Muller S, Lam S, Toyooka S, et al. Aberrant methylation of multiple genes in the upper

aerodigestive tract epithelium of heavy smokers. Int J Cancer 2003;1074:612–6.
16. Honorio S, Agathanggelou A, Schuermann M, et al. Detection of RASSF1A aberrant promoter hypermethylation in sputum from chronic smokers and ductal carcinoma in situ from breast cancer patients. Oncogene 2003;2219:147–50.
17. Belinsky SA, Palmisano WA, Gilliland FD, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer Res 2002;628:2370–7.
18. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer 1953; 5:963–8.
19. Belinsky SA, Nikula KJ, Palmisano WA, et al. Aberrant methylation of $p16^{INK4a}$ is an early event in lung cancer and a potential biomarker for early diagnosis. Proc Natl Acad Sci U S A 1998;9520:11891–6.
20. Belinsky SA. Gene promoter hypermethylation as a biomarker in lung cancer. Nat Rev Cancer 2004;49:707–17.
21. Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001;9314:1054–61.
22. Kennedy TC, Proudfoot SP, Piantadosi S, et al. Efficacy of two sputum collection techniques in patients with air flow obstruction. Acta Cytol 1999;434:630–6.
23. Saccomanno G. Diagnostic pulmonary cytology. American Society of Clinical Pathologists. Chicago (IL); 1978.
24. Petty TL, Tockman MS, Palcic B. Diagnosis of roentgenographically occult lung cancer by sputum cytology. Clin Chest Med 2002;231:59–64.
25. Palmisano WA, Crume KP, Grimes MJ, et al. Aberrant promoter methylation of the transcription factor genes PAX5 α and β in human cancers. Cancer Res 2003;6315: 4620–5.
26. Pulling LC, Divine KK, Klinge DM, et al. Promoter hypermethylation of the O^6-methylguanine-DNA methyltransferase gene: more common in lung adenocarcinomas from never-smokers than smokers and associated with tumor progression. Cancer Res 2003;6316:4842–8.
27. Dammann R, Li C, Yoon JH, Chin PL, Bates S, Pfeifer GP. Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. Nat Genet 2000;253:315–9.
28. Kim DH, Nelson HH, Wiencke JK, et al. Promoter methylation of DAP-kinase: association with advanced stage in non-small cell lung cancer. Oncogene 2001;2014: 1765–70.
29. Toyooka KO, Toyooka S, Virmani AK, et al. Loss of expression and aberrant methylation of the CDH13 (H-C cadherin) gene in breast and lung carcinomas. Cancer Res 2001;6111:4556–60.
30. Guo M, Akiyama Y, House MG, et al. Hypermethylation of the GATA genes in lung cancer. Clin Cancer Res 2004;1023:7917–24.
31. Sathyanarayana UG, Toyooka S, Padar A, et al. Epigenetic inactivation of laminin-5-encoding genes in lung cancers. Clin Cancer Res 2003;97:2665–72.
32. Chang YS, Wang L, Liu D, Mao L, Hong WK, Khuri FR, et al. Correlation between insulin-like growth factor-binding protein-3 promoter methylation and prognosis of patients with stage I non-small cell lung cancer. Clin Cancer Res 2002;812:3669–75.
33. Divine KK, Pulling LC, Marron-Terada PG, et al. Multiplicity of abnormal promoter methylation in lung adenocarcinomas from smokers and never smokers. Int J Cancer 2005;1143:400–5.
34. Xu ZP, Dutra A, Stellrecht CM, Wu C, Piatigorsky J, Saunders, GF. Functional and structural characterization of the human gene BHLHB5, encoding a basic helix-loop-helix transcription factor. Genomics 2002;803:311–8.
35. Peyton M, Stellrecht CM, Naya FJ, Huang HP, Samora PJ, Tsai MJ. BETA3, a novel helix-loop-helix protein, can act as a negative regulator of BETA2 and MyoD-responsive genes. Mol Cell Biol 1996;162:626–33.
36. Swafford DS, Middleton SK, Palmisano WA, et al. Frequent aberrant methylation of $p16^{INK4a}$ in primary rat lung tumors. Mol Cell Biol 1997;173:1366–74.
37. Pulling LC, Vuillemenot BR, Hutt JA, Devereux TR, Belinsky SA. Aberrant promoter hypermethylation of the death-associated protein kinase gene is early and frequent in murine lung tumors induced by cigarette smoke and tobacco carcinogens. Cancer Res 2004;6411: 3844–8.
38. Belinsky SA, Klinge DM, Dekker JD, et al. Gene promoter methylation in plasma and sputum increases with lung cancer risk. Clin Cancer Res 2005;1118:6505–11.
39. Lukas J, Parry D, Aagaard L, et al. Retinoblastoma-protein-dependent cell-cycle inhibition by the tumor suppressor p16. Nature 1995;375:503–6.
40. Kozmik Z, Wang S, Dorfler P, Adams B, Busslinger M. The promoter of the CD19 gene is a target for the B-cell-specific transcription factor BSAP. Mol Cell Biol 1992; 126:2662–72.
41. Cohen O, Feinstein E, Kimchi A. DAP-kinase is a Ca^{2+}/calmodulin-dependent, cytoskeletal-associated protein kinase, with cell death-inducing functions that depend on its catalytic activity. EMBO J 1997;165: 998–1008.
42. Cohen O, Inbal B, Kissil JL, et al. DAP-kinase participates in TNF-α and Fas-induced apoptosis and its function requires the death domain. J Cell Biol 1999; 1461:141–8.
43. Khokhlatchev A, Rabizadeh S, Xavier R, et al. Identification of a novel ras-regulated proapoptotic pathway. Curr Biol 2002;124:253–65.
44. Esteller M, Toyota M, Sanchez-Cespedes M, et al. Inactivation of the DNA repair gene O^6-methylguanine-DNA methyltransferase by promoter hypermethylation is associated with G to A mutations in K-ras in colorectal tumorigenesis. Cancer Res 2000;609:2368–71.
45. Eckfeld K, Hesson L, Vos MD, Blieche I, Latif F, Clark GJ. RASSF4/AD037 is a potential ras effector/tumor suppressor of the RASSF family. Cancer Res 2004;6423: 8688–93.
46. Hesson LB, Wilson R, Morton D, et al. CpG island promoter hypermethylation of a novel Ras-effector gene RASSF2A is an early event in colon carcinogenesis and correlates inversely with K-ras mutations. Oncogene 2005;2424:3987–94.
47. Yano M, Toyooka S, Tsukuda K, et al. Aberrant promoter methylation of human DAB2 interactive protein (hDAB2IP) gene in lung cancers. Int J Cancer 2005;1131:59–66.