Defining a Gene Promoter Methylation Signature in Sputum for Lung Cancer Risk Assessment

Shuguang Leng¹, Kieu Do¹, Christin M. Yingling¹, Maria A. Picchi¹, Holly J. Wolf³, Timothy C. Kennedy³, William J. Feser³, Anna E. Baron³, Wilbur A. Franklin³, Malcolm V. Brock⁴, James G. Herman⁴, Stephen B. Baylin⁴, Tim Byers³, Christine A. Stidley², and Steven A. Belinsky¹

Abstract

**Purpose:** To evaluate the methylation state of 31 genes in sputum as biomarkers in an expanded nested, case–control study from the Colorado cohort, and to assess the replication of results from the most promising genes in an independent case–control study of asymptomatic patients with stage I lung cancer from New Mexico.

**Experimental Design:** Cases and controls from Colorado and New Mexico were interrogated for methylation of up to 31 genes using nested, methylation-specific PCR. Individual genes and methylation indices were used to assess the association between methylation and lung cancer with logistic regression modeling.

**Results:** Seventeen genes with ORs of 1.4 to 3.6 were identified and selected for replication in the New Mexico study. Overall, the direction of effects seen in New Mexico was similar to Colorado with the largest increase in case discrimination (ORs, 3.2–4.2) seen for the \( \text{PAX5} \), \( \text{GATA5} \), and \( \text{SULF2} \) genes. Receiver operating characteristic (ROC) curves generated from seven-gene panels from Colorado and New Mexico studies showed prediction accuracy of 71% and 77%, respectively. A 22-fold increase in lung cancer risk was seen for a subset of New Mexico cases with five or more genes methylated. Sequence variants associated with lung cancer did not improve the accuracy of this gene methylation panel.

**Conclusions:** These studies have identified and replicated a panel of methylated genes whose integration with other promising biomarkers could initially identify the highest risk smokers for computed tomographic screening for early detection of lung cancer. *Clin Cancer Res; 18(12); 3387–95.*

Introduction

Lung cancer remains the leading cause of cancer-related death in the United States. The development and implementation of low cost, noninvasive screening approaches to identify smokers at the highest risk for lung cancer is one approach that could reduce the mortality associated with this disease. Considerable excitement was generated over findings from the National Lung Screening Trial (NLST) that reported a 20% reduction in mortality from lung cancer compared with standard chest radiography (1). However during CT screening 39% of participants had at least one positive screening result with >96% of those findings ultimately being classified as false positive. This represents a sensitivity of 94%, specificity of 61%, and positive predictive value of 6% over the 3-year screening period. Moreover, the eligibility requirements for screening of 55 to 74 years of age with a minimum smoking history of 30 pack-years captures approximately 30% of the lung cancers currently being diagnosed (1). The addition of molecular biomarkers interrogated in accessible biologic fluids such as sputum and blood could identify smokers at highest risk for lung cancer and augment CT screening by reducing false-positive tests.

Our group has focused on developing gene promoter hypermethylation detection in sputum as a molecular marker for identifying people at high risk for cancer incidence (2). Gene silencing through methylation of cytosine adjacent to guanosine in CpG islands in conjunction with chromatin remodeling leads to the development of heterochromatin of the gene promoter region, which denies access to regulatory proteins needed for transcription (3). This epigenetically driven process is a major and causal event silencing hundreds of genes involved in all aspects of normal cellular function during lung cancer initiation and progression. Importantly, silencing of genes such as \( \text{CDKN2A} \) (\( \text{p16} \)), \( \text{O}^6\)-methylguanine-DNA methyltransferase...
Translational Relevance

The addition of molecular biomarkers detected in biologic fluids could identify smokers at high risk for lung cancer and augment computed tomographic (CT) screening by reducing false-positive tests. Our group has previously identified gene methylation in sputum as a promising biomarker for early lung cancer detection. We extended this work by evaluating 31 genes in a nested case–control study of incident lung cancer in the Colorado cohort, with replication of promising genes in a geographically independent case–control study of patients with stage I lung cancer from New Mexico. Receiver operating characteristic curves showed prediction accuracy up to 77% with a seven-gene panel. Furthermore, having at least five of seven genes methylated was associated with a 22-fold increase in lung cancer risk in the New Mexico cohort. Integrating our gene methylation panel with other classes of biomarkers could ultimately yield a molecular marker platform with specificity and sensitivity high enough for prospective screening of heavy smokers.

(MGMT), and adenomatous polyposis coli (APC) is detected in alveolar and bronchial epithelium of smokers, in precursor lesions to adenocarcinoma and squamous cell carcinoma, and the prevalence of gene methylation increases during disease progression (4, 5). These findings of epigenetic changes in epithelial cells and premalignant lesions in the lungs of smokers reflect the well-documented field cancerization present throughout the aerodigestive tract that also presents an obstacle for distinguishing early lung cancer from the large ‘at risk’ population (6).

On the basis of the silencing of key tumor suppressor genes in the lungs of smokers, we hypothesized that the detection of gene promoter hypermethylation in exfoliated epithelial cells in sputum would provide an assessment of the extent of field cancerization that in turn may predict early lung cancer. The development of methylation-specific PCR (MSP) and subsequent nested MSP enabled the sensitivity and specificity required to detect methylation of gene promoters in exfoliated epithelial cells that comprise only a fraction (<3%) of the cellular content present in sputum from patients with lung cancer and cancer-free smokers (7, 8). In a small proof-of-concept study, methylation of p16 or MGMT gene promoters was detected up to 3 years before the diagnosis of squamous cell carcinoma (8). This finding led us to conduct a nested case–control study of incident lung cancer cases from an extremely high-risk cohort (Colorado cohort) to evaluate whether a panel of genes could be identified whose methylation in sputum would predict lung cancer. Key findings from that study included the increased prevalence of gene promoter methylation detected in sputum as the time to lung cancer diagnosis decreased and that 6 of 14 genes were associated with a >50% increased lung cancer risk (9). Importantly having 3 or more of these 6 genes methylated was associated with a sensitivity and specificity of 64%.

These findings support the promise of gene promoter hypermethylation as one type of molecular marker for stratifying lung cancer risk, while also emphasizing the need to evaluate other genes commonly silenced through promoter hypermethylation in lung tumors to improve the sensitivity and specificity of this gene panel. The purpose of the current study was to evaluate 23 candidate genes along with the most promising 8 genes from our previous study (9) in an expanded nested, case–control study from the Colorado (CO) cohort and to assess the replication of results from the most promising genes in an independent cohort study of asymptomatic patients with stage I lung cancer from New Mexico (NM). In addition, we determined whether integrating results from genotyping sequence variants identified through genome-wide association studies (GWAS), as being associated with risk of lung cancer would improve the sensitivity and specificity of a refined gene methylation panel.

Materials and Methods

Study populations

Study populations for testing and replicating methylation biomarkers were from CO and NM, respectively. All participants signed a consent form and studies were Institutional Review Board–approved. CO participants were selected from the University of Colorado Cancer Center Sputum Screening Cohort Study, a prospective study initiated in 1993 to determine whether biomarkers identified within sputum can predict future lung cancer development. The study methodology has been described previously (9, 10). 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, gender, and height; and an FEV1/FVC (FVC, forced vital capacity) ratio of ≤0.75. Participants were provided with 2 containers filled with a fixative solution of 2% carboxax 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 (11). As most participants contributed less than 2 sputum samples in this prospective cohort, the cases selected for study were those who provided a sputum sample within 14 months before cancer diagnosis. This time frame was selected on the basis of our previous study that showed that the prevalence of methylation of gene promoters increased with samples collected within that time period compared with >18 months. There were 64 cases that met these criteria and they were matched to controls (n = 64) by gender, age, and month of enrollment.
NM cases and controls were selected from the Lung Cancer Cohort and Lovelace Smokers Cohort (LSC), respectively. The New Mexico Lung Cancer Cohort was established in 2005 to enroll and follow newly diagnosed patients with lung cancer irrespective of tumor stage and histology. Patients with Lung cancer were recruited through the Multidisciplinary Chest Clinic at the University of New Mexico (Albuquerque, NM). Tumor stage and histology are defined through clinical presentation and pathology. This study was restricted to stage I lung cancer cases that are generally asymptomatic for disease and undergo “curative” tumor resection. Cases were current or former smokers 45- to 75-year old and able to provide a sputum sample. Spontaneous sputum was collected at the time of enrollment (approximately 1–2 weeks before surgery) at home and in the morning as described for the CO cohort. At the time of this study, 328 lung cancer cases had enrolled into the cohort and 90 presented with stage I disease. Forty-five of these subjects produced sputum and 5 were excluded because of age resulting in a sample size of 40. Overall approximately 65% and 70% of lung cancer cases and smokers from LSC produced cytologic adequate sputum, respectively, as defined by Kennedy and colleagues (11). Controls were cancer-free smokers enrolled into the LSC to identify risk factors for gene methylation (12, 13). Enrollment, which is still ongoing, is restricted to current and former smokers ages 40 to 75 years with a minimum of 15 pack-years of smoking. Participants primarily are residents of the Albuquerque, NM metropolitan area and provide sputum and blood and undergo standard pulmonary function testing. Controls (n = 90) were frequency matched to cases (~2:1) by age (5-year interval) and gender.

Sputum processing and MSP

We observed in our original studies that long-term storage (>3 years) of CO sputum samples in Saccomanno’s fixative can lead to DNA degradation and implemented a protocol for NM that involved washing the sputum sample to reduce the mucous, followed by freezing the specimen as a pellet at –80°C within 6 months of collection. The sputum sample from CO was collected within 18 months of diagnosis, although confirmation of case status often takes several years postcollection. DNA was isolated from sputum by protease digestion followed by phenol chloroform extraction and ethanol precipitation. Samples were labeled only with study-specific coded identifiers to blind investigators from case or control status. Assays were done with both cases and controls included in each batch.

Twenty-three new genes (see Results), selected for evaluation based on their prevalence in lung tumors (≥25%), diversity of function, and timing for inactivation during lung cancer development when known, were studied in the CO cases and controls (4, 5 14–17). In addition, the top 8 genes associated with increased lung cancer risk (p16, MGMT, DAPK, RASSF1A, GATA4, GATA5, PAX5a, and PAX5b) from our previous study were also assessed (9). Our nested, MSP assay was used because of its increased sensitivity for detection of promoter methylation in sputum. Briefly, the method involves bisulfite modification of the DNA (9) followed by stage I PCR to simultaneously amplify 4 gene promoter regions. Stage I primers recognize bisulfite-modified template, but do not discriminate between methylated and unmethylated alleles. Stage I PCR product (5 μL of a 1:50 dilution of the stage I) is then subjected to stage II PCR in which primers specific for methylated template are used. All stage II PCR reactions are conducted at annealing temperatures (68°C–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. To accurately compare the prevalence of methylation in specimens across all persons with a sensitivity of 1 in 10,000 to 20,000, 100 to 150 ng of DNA was used for stage I PCR following modification with bisulfite. The methylation-specific primers for stage II for each gene promoter are located around the transcription start site where methylation is strongly correlated with gene silencing and are depicted (Supplementary Table S1) for the 12 genes moving forward for prospective studies. Sputum is a very heterogeneous specimen containing inflammatory, epithelial, and oral cells. The epithelial fraction generally comprises <3% of the sputum sample. This considerable cellular heterogeneity, which varies across subjects, limits the ability to quantitate methylation; thus, methylation was scored as positive or negative.

Because of the issue of stochastic sampling (discussed in detail in ref. 9), where the epithelial fraction containing the methylated genes is usually <3% of the sputum sample, assays are conducted in duplicate starting with the bisulfite modification of the DNA. Detection of gene methylation in either assay is scored as a positive for methylation of that specific gene (9). Quantitative MSP was used to assess methylation levels for a subset of genes in NM cases and controls to confirm or refute our strategy for nonquantitative assessment of methylation in sputum. Gene methylation levels were quantified by standard quantitative MSP (QMSP), with and without a nested amplification by our commercial collaborator MDxHealth, formally OncoMethylome Sciences. Methylation cutoff values were set by receiver operating characteristic (ROC) curves (18).

Single-nucleotide polymorphism genotyping

Five single-nucleotide polymorphisms (SNP) from chromosomes 15q25, 5p15, and 6p21 identified in GWAS to be associated with lung cancer were genotyped in cases and controls from both cohorts using the TaqMan allelic discrimination assay (19–23). Genomic DNA recovered from sputum samples and peripheral lymphocytes was used for genotyping in CO and NM cohorts, respectively. Genotypes measured in sputum samples and peripheral lymphocytes from a subset of NM patients showed 100% agreement. Ten percent of samples were repeated and 100% concordance was seen with regard to results from the initial genotyping. In addition, samples comprising known genotypes for each SNP were included in each batch of 96 samples to guide correct clustering.
Data analysis

Demographic, methylation, and genotype variables were summarized by case–control status with percents for categorical variables and means and SDs for continuous variables. Differences in demographic variables between cases and controls were assessed with the Fisher exact test for categorical variables and the Wilcoxon rank-sum test for continuous variables. The association between the methylation and genotype variables and case–control status was expressed as ORs and their corresponding 95% confidence intervals (CI) obtained from logistic regression models with adjustment for the design variables (age and gender) and other important covariates including smoking status and pack-years of smoking. Age and pack-years were entered as continuous variables. Pack-years of cigarette smoking were defined as the average number of packs smoked per day multiplied by the number of years of smoking. Former smokers were defined as those individuals who had quit smoking 1 year or more at the time of questionnaire completion.

Both individual genes and methylation indices were used to examine the association between methylation and lung cancer. All genes were individually assessed, with adjustment for the covariates listed earlier. Chronic obstructive pulmonary disease (COPD) was not included in the logistic regression models for CO because all subjects had COPD. With regard to NM, there was missing pulmonary function results for 30% of the cases, thus COPD was not included in the models as the sample size would have been reduced significantly. For subsets of the genes, multiplicity of methylated genes was defined as the number of genes methylated in the panel of genes. The resulting methylation index also was used as an independent variable in the logistic regression modeling. ROC curves were used to assess these models and to determine an optimal gene methylation panel. The relationship between the methylation indices and the histologic type of lung cancer was evaluated among cases within each cohort. In addition, the association between methylation index and sub-stage (IA vs. IB) of lung cancer also was assessed among the NM cases. Statistical significance was expressed by $P$ values. All analyses were carried out with Statistical Analysis Software (SAS, version 9.2; SAS Institute, Inc.).

Results

Exposure history and pathology

Key demographic variables for cases and controls from CO and NM are summarized in Table 1. Smoking history with regard to pack-years was higher in members of the CO cohort reflecting the enrollment criterion of a minimum of 30 versus 15 pack-years for CO and NM participants, respectively. The incidence of squamous cell and nonsquamous cell lung cancer (inclusive of adenocarcinoma, large cell, and carcinoma unspecified) was similar among CO cases. In contrast, 68% of cancer diagnosed in NM cases was nonsquamous lung cancer. All cases from NM were stage I with an equal distribution of IA and IB, whereas stage of disease for CO cases was not available.

Identification of methylated genes in sputum associated with risk for lung cancer

The gene methylation state of 23 new genes and 8 genes ($p16$, $MGMT$, $DAPK$, $RASSF1A$, $GATA4$, $GATA5$, $PAX5a$, $PAX5b$) originally showing increased odds for methylation in cases compared with controls were evaluated in sputum samples from CO. Investigators were blinded to case status and assays were conducted in batches of 30 to 35 samples starting with samples for which the largest amount of DNA was available. Because the amount of DNA was limiting for some samples, an interim analysis was conducted after approximately two thirds of cases and controls had been tested. Fourteen genes with prevalence of methylation of

### Table 1. Summary of selected variable by case–control status

<table>
<thead>
<tr>
<th>Variable</th>
<th>CO</th>
<th>NM</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y, mean(SD))</td>
<td>68.3(8.1)</td>
<td>67.6(7.2)</td>
<td>0.64</td>
</tr>
<tr>
<td>Gender (% female)</td>
<td>23.4</td>
<td>21.9</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>Smoking status (% current)</td>
<td>39.1</td>
<td>29.7</td>
<td>0.35</td>
</tr>
<tr>
<td>Pack-years (Mean (SD))</td>
<td>71.5(30.6)</td>
<td>68.5(45.0)</td>
<td>0.20</td>
</tr>
<tr>
<td>Ethnicity (%)</td>
<td>NHW 90.6</td>
<td>95.3</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Hispanic 0.0</td>
<td>0.0</td>
<td>27.5</td>
</tr>
<tr>
<td></td>
<td>Missing 9.4</td>
<td>4.7</td>
<td>12.5</td>
</tr>
</tbody>
</table>

Abbreviation: NHW, non-Hispanic white. $^aP$ values are from the Fisher exact test for categorical variables and the Wilcoxon rank-sum test for continuous variables. $^b$Test excludes missing ethnicity.
Differences were seen for in CO (Table 2). Overall, for NM subjects, significant fold increase in OR in NM compared with 1.4- to 1.8-fold were seen for 5 genes that included PAX5. The largest increases in case discrimination were for the controls. The largest increases in case discrimination genotypes showed little to no difference between cases and controls (DNA was exhausted from many of the CO samples). Methylation of CXCL14, but not CXCL12 was significantly associated with lung cancer (OR = 6.3, P < 0.004; Table 2), corroborating original studies suggesting this gene could be a potential biomarker for lung cancer detection (24).

The association of overall gene methylation seen in either cohort was not different between specific histologic types of lung cancer or between stage Ia and Ib patients in NM (not shown). QMSP was conducted on samples from NM to determine whether quantifying methylation in sputum would better distinguish cases from controls. Methylation of P16, DAPK, GATA4, and GATA5 was evaluated with nested QMSP and standard QMSP. Neither approach using methylation cutoff values improved the ability to classify cases and controls (not shown), a finding consistent with the fact that the sputum sample is highly heterogeneous with a variable epithelial component. Moreover, there is no reason to suspect that a small lung tumor (<1 cm) in the peripheral lung would directly contribute enough tumor cells into the sputum specimen to generate significantly

![Table 2. Prevalence and ORs for gene promoter methylation in sputum samples from CO and NM case-control studies](image-url)

Abbreviation: ND, not determined.

*Adjusted for age, gender, smoking status, and pack-years.
Clinical Cancer Research

The OR and 95% CI were calculated to further characterize the association between the methylation panels and risk for lung cancer. The greatest increase in risk was seen with the 7-gene panels. When the number of methylated genes was used as a continuous variable (0–7), a 1.5- and 2.0-fold increased risk for lung cancer was seen for each one-gene difference in methylation in the CO and NM groups, respectively (Table 3). The OR associated with lung cancer risk for having 3 or more methylated genes compared with <3 methylated genes from the 7-gene panels increased to 2.3 and 5.0 for CO and NM, respectively (Table 3). Moreover, 38% of NM cases compared with only 3% of controls showed methylation of 5 or more genes compared with those with fewer than 5 genes. This equates to a 22.3-fold increased risk for lung cancer.

Assessment of SNPs identified through GWAS for lung cancer risk

GWAS have identified 3 chromosomal regions at 15q25, 5p15, and 6p21 as being associated with risk for lung cancer (19–23). We genotyped 5 SNPs within these chromosomal regions in cases and controls from CO and NM to determine whether integrating a genetic index of sequence variants with our gene methylation panels would improve risk assessment. One SNP on chromosome 15 (nAChR gene cluster), one within 6p21.31 (HLA-DQA1), one within 6p21.33 (BAT3-MSH5), and 2 within 5p15.33 (CLPTM1L-TER1 locus) were interrogated. All SNPs were in Hardy–Weinberg equilibrium ($P > 0.05$) for the control group from each case–control study. No SNPs were associated with a significant increased risk, but 2 of the SNPs (rs1051730 Chr15q25 and rs3117582 Chr6p21) showed protection within the CO cohort, a finding dissimilar to the other cohort and likely to be spurious due to small sample size (Table 4). The addition of a genetic index that summed the risk alleles from the 5 loci in each individual to the gene methylation panels did not improve estimates of lung cancer risk in either cohort (not shown).

Discussion

This study has identified and replicated in 2 geographically independent cohorts, a gene methylation panel that shows improved sensitivity and specificity for detecting lung cancer over our original study (9). Moreover, among cases from NM, methylation of 5 or more genes in a 7-gene panel conferred a 22-fold increase in risk for lung cancer. The greater sensitivity and specificity for lung cancer detection seen in the NM cohort than in CO is consistent with our ongoing hypothesis that as the premalignant burden reflective of field cancerization increases and therefore the risk for lung cancer, so does the ability for detection of methylated genes in sputum (2). Lung cancer was detectable in the NM patients as stage I disease, whereas diagnosis occurred between 3 and 18 months following collection of the sputum sample in the CO patients. The inability of
sequence variants associated with lung cancer to improve the prediction accuracy for our gene methylation panel is likely the result of the low penetrance effect associated with these risk alleles that can only be observed in large population-based studies.

The moderate difference between the 7-gene panels for CO and NM make it difficult to reduce the 12-gene panel for subsequent validation studies. Six genes, p16, MGMT, DAPK, PAX5a, PAX5b, and GATA5 were identified in our initial study and have shown use as biomarkers for uncovering genetic determinants for methylation in the aero-digestive tract and identifying dietary factors associated with reduced gene promoter methylation (12, 13, 25). Dal1, an actin-binding protein that suppresses the growth of lung cancer cells in vitro, showed similar significant differences in distinguishing cases from controls in both studies (OR, 3.1–3.6). Methylation of this gene is detected in 57% of non–small cell lung carcinoma and increases in prevalence from early- to late-stage tumors (26). CXCL14, while only assessed in the NM cohort was strongly associated with lung cancer and its loss of function impacts the expression of cell cycle and proapoptosis genes (24). Overall, this gene panel is largely composed of genes that code for transcription factors (GATA5, PAX5a, PAX5b) and genes involved in regulating cell proliferation, adhesion, and apoptosis (p16, DAPK, DAL1, CXCL14, PCDH20, SULF2; ref. 14–17, 24, 26).

The challenge is considerable for identifying noninvasive biomarkers for a disease developing over 30 years, and diagnosed at an annual incidence of 1% to 2% in current and former smokers’ ages 45 to more than 80 with a smoking history of more than 30 pack-years. It is unlikely that any single biomarker platform will achieve the sensitivity and specificity required to move forward for

### Table 3. Association of gene methylation panels and risk for lung cancer

<table>
<thead>
<tr>
<th>Methylation panel</th>
<th>CO (N = 128)</th>
<th>NM (N = 130)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR² (95% CI)</td>
<td>P</td>
</tr>
<tr>
<td>Count²</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11- or 12-gene panel</td>
<td>1.3 (1.1–1.5)</td>
<td>0.006</td>
</tr>
<tr>
<td>7-gene panel</td>
<td>1.5 (1.1–1.9)</td>
<td>0.003</td>
</tr>
<tr>
<td>Categorical³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥3 genes—11- or 12-gene panel</td>
<td>2.3 (1.1–4.9)</td>
<td>0.03</td>
</tr>
<tr>
<td>≥3 genes—7-gene panel</td>
<td>2.3 (1.0–5.2)</td>
<td>0.048</td>
</tr>
</tbody>
</table>

²Adjusted for age, gender, smoking status, and pack-years.
³OR indicates the effect associated with a one-gene difference.
⁴OR compares high to low methylation groups.

### Table 4. Association of sequence variants with risk for lung cancer in CO and NM case–control studies

<table>
<thead>
<tr>
<th>SNP</th>
<th>Genotype</th>
<th>CO (OR² (95% CI))</th>
<th>P</th>
<th>NM (OR² (95% CI))</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1051730</td>
<td>G—A</td>
<td>0.45 (0.2–0.8)</td>
<td>0.01</td>
<td>1.2 (0.6–2.2)</td>
<td>0.48</td>
</tr>
<tr>
<td>Chr15q25</td>
<td>rs1742693</td>
<td>1.2 (0.6–2.6)</td>
<td>0.57</td>
<td>1.1 (0.6–2.3)</td>
<td>0.74</td>
</tr>
<tr>
<td>Chr6p21</td>
<td>T—G</td>
<td>0.3 (0.1–0.8)</td>
<td>0.02</td>
<td>0.6 (0.2–1.9)</td>
<td>0.41</td>
</tr>
<tr>
<td>rs3117582</td>
<td>A—C</td>
<td>1.5 (0.8–2.6)</td>
<td>0.57</td>
<td>0.9 (0.5–1.6)</td>
<td>0.68</td>
</tr>
<tr>
<td>Chr5p15</td>
<td>rs401681</td>
<td>1.5 (0.8–2.6)</td>
<td>0.20</td>
<td>0.8 (0.4–1.3)</td>
<td>0.60</td>
</tr>
</tbody>
</table>

²Letters to the right of the arrow represent the minor allele, with exception for rs401681 where the C allele is the major allele associated with increased risk for lung cancer (19–23).
³With the exception of rs3117582, an additive model was used so the OR is the effect per allele. Because there was only one homozygote rare for rs3117582, a dominant model was used. All models include adjustment for age, gender, smoking status, and pack-years.
prospective screening as an adjunct to CT. The fact that gene promoter methylation in sputum was not associated with a particular histologic diagnosis of lung cancer substantiates its use as one type of biomarker for predicting these tumor types. Integrating our gene methylation panel with other classes of biomarkers such as chromosomal anomaly detected by fluorescence in situ hybridization in sputum is one avenue currently being pursued (27). Serum-based assays to detect altered expression of miRNAs, that inherently are more stable in the circulation than large RNAs, are showing promise as biomarkers for lung cancer detection, as are a panel of autoantibodies recently developed by Qui and colleagues (28–30). A key for serum-based markers is the need to be disease specific and to show a large enough change in level compared with disease-free controls for establishing “classification cutoffs” that will survive the influence of variation in blood processing and sample preparation (e.g., RNA isolation and generation of cDNA). Nasal epithelium may represent another source of tissue for lung cancer screening with a few studies suggesting that the airway field of injury extends to the nose (reviewed in ref. 31). Implementing a multifaceted approach for validating and integrating the best DNA, RNA, and protein-based biomarkers interrogated across common specimens (blood, sputum, and nasal epithelium) in a large case-control study could ultimately yield a molecular marker platform with specificity and sensitivity high enough for prospective screening of heavy smokers.

Disclosure of Potential Conflicts of Interest
M.V. Brock and J.G. Herman have commercial research support for MDxHealth. J.G. Herman is the consultant/advisory board member for MDxHealth. No potential conflicts of interest were disclosed by other authors.

Grant Support
This work was supported by National Cancer Institute Specialized Program of Research Excellence P50 CA51817 and CA58184, R01 CA97156, a research grant from MDxHealth, formerly Oncometyme Sciences, Inc. and the State of New Mexico.

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

Received November 29, 2011; revised March 1, 2012; accepted April 8, 2012; published OnlineFirst April 17, 2012.

References
Defining a Gene Promoter Methylation Signature in Sputum for Lung Cancer Risk Assessment

Shuguang Leng, Kieu Do, Christin M. Yingling, et al.