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. Wolf2, Timothy C. Kennedy3, William J. Feser2, Anna E. Baron2, Wilbur A. Franklin2, Malcolm V. Brock3, James G. Herman3, Stephen B. Baylin3, Tim Byers2, Christine A. Stidley4, and Steven A. Belinsky1

1Lung Cancer Program
Lovelace Respiratory Research Institute
Albuquerque, NM

2University of Colorado Denver
Aurora, CO

3Johns Hopkins University
Baltimore, MD

4Department of Internal Medicine
University of New Mexico
Albuquerque, NM

Running title: Gene promoter methylation signature in sputum

Key words: gene methylation, sputum, lung cancer, biomarker

Correspondence should be addressed to S. A. Belinsky, Lovelace Respiratory Research Institute, 2425 Ridgecrest Dr. SE, Albuquerque, NM 87108, Phone: 505-348-9465, Fax: 505-348-4990, sbelinsk@LRRI.org.

Conflicts of Interest Statement: The authors declare no conflicts of interest.

Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).
**Translational relevance**

The addition of molecular biomarkers detected in biological fluids could identify smokers at high risk for lung cancer and augment 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 Stage I lung cancer patients from New Mexico. ROC 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.
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 Stage I lung cancer patients 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 odds ratios of 1.4 – 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 (odds ratios, 3.2 – 4.2) seen for the PAX5α, GATA5, and SULF2 genes. 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 tomography screening for early detection of lung cancer.
Introduction

Lung cancer remains the leading cause of cancer-related death in the U.S. The development and implementation of low cost, non-invasive 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 with low-dose computed tomography (CT) screening compared to 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 three-year screening period. Moreover, the eligibility requirements for screening of 55 – 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 biological 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 $CDKN2A$ ($p16$), $O^6$-methylguanine-DNA
methyltransferase (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).

Based on 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 lung cancer patients 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 prior to 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 six of 14 genes were associated with a >50% increased lung cancer risk (9). Importantly having three or more of these six 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 eight genes from our previous study (9) 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 cohort study of asymptomatic Stage I lung cancer patients from New Mexico (NM). In addition, we determined whether integrating results from genotyping sequence variants identified through genome-wide association studies, 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 IRB approved. Colorado 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 ratio of ≤ 0.75. 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 (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 18 months prior to cancer diagnosis. This time frame was selected based on our previous study that demonstrated that the prevalence of methylation of gene promoters increased with samples collected within that time period compared to > 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.
New Mexico 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 lung cancer patients irrespective of tumor stage and histology. Lung cancer patients were recruited through the Multidisciplinary Chest Clinic at the University of New Mexico. 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 – 75 years old and able to provide a sputum sample. Spontaneous sputum was collected at the time of enrollment (approximately 1 – 2 weeks prior to surgery) at home and in the morning as described for the Colorado 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 five were excluded due to age resulting in a sample size of 40. Overall approximately 65% and 70% of lung cancer cases and smokers from LSC produced cytological adequate sputum, respectively, as defined by Kennedy et al. (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 y 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 intervals) and gender.

**Sputum processing and methylation specific PCR**

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 diagnosis, although confirmation of case status often takes several years post sample collection. 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 eight genes associated with increased lung cancer risk (p16, MGMT, DAPK, RASSF1A, GATA4, GATA5, PAX5α, and PAX5β) 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 1 PCR to simultaneously amplify four gene promoter regions. Stage 1 primers recognize bisulfite-modified template, but do not discriminate between methylated and unmethylated alleles. Stage 1 PCR product (5 μl of a 1:50 dilution of the stage 1) is then subjected to stage 2 PCR in which primers specific for methylated template are used. All stage 2 PCR reactions are 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. To accurately compare the prevalence of methylation in specimens across all persons with a sensitivity of 1 in 10 – 20,000, 100–150 ng of DNA was used for stage 1 PCR following modification with bisulfite. The methylation-specific primers for
stage 2 for each gene promoter are located around the transcription start site where methylation is strongly correlated with gene silencing and are depicted (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. Due to the issue of stochastic sampling (discussed in detail in [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 non-quantitative assessment of methylation in sputum. Gene methylation levels were quantified by standard QMSP, with and without a nested amplification by our commercial collaborator MDxHealth, formally OncoMethylome Sciences. Methylation cutoffs were set by receiver characteristic operator (ROC) curves (18).

**SNP genotyping**

Five SNPs from chromosomes 15q25, 5p15, and 6p21 identified in genome wide association studies 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 Colorado 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 standard deviations for continuous variables. Differences in demographic variables between cases and controls were assessed with Fisher’s 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 odds ratios and their corresponding 95% confidence intervals 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 one 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 above. 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 since 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. Receiver operator classification 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 versus IB) of lung cancer also was assessed among the NM cases. Statistical significance was expressed by p values. All analyses were carried out using Statistical Analysis Software (SAS, version 9.2, SAS Institute, Inc., Cary, NC).
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 Colorado Cohort reflecting the enrollment criterion of a minimum of 30 versus 15 pack years for CO and New Mexico participants, respectively. The incidence of squamous cell and non-squamous cell lung cancer (inclusive of adenocarcinoma, large cell, and carcinoma unspecified) was similar among Colorado cases. In contrast, 68% of cancer diagnosed in NM cases was non-squamous lung cancer. All cases from NM were Stage I with an equal distribution of IA and IB, while stage of disease for Colorado 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 (\textit{p16}, \textit{MGMT}, \textit{DAPK}, \textit{RASSF1A}, \textit{GATA4}, \textit{GATA5}, \textit{PAX5}α, \textit{PAX5}β) originally demonstrating increased odds for methylation in cases compared to controls were evaluated in sputum samples from CO. Investigators were blinded to case status and assays were performed in batches of 30 – 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 0 – 70% in sputum showed no increased odds for methylation in cases (Table S2). For the remaining 17 genes analyzed in all subjects, odds were increased 1.4 – 3.6 in cases relative to controls (Table 2). Significant or borderline significant (p < 0.05 – 0.1) odds ratio were seen for five genes that included \textit{PAX5}β, \textit{Dal1}, \textit{PCDH20}, \textit{Kif1a}, and \textit{Dcr2} (Table 2).
These 17 genes were selected for replication in NM cases and controls. Stage I lung cancer was studied because this patient population is asymptomatic for disease and early detection in conjunction with curative intent resection is most likely to improve survival. There was an increase in ability for detecting methylation in NM specimens that may reflect the better quality of the DNA isolated from sputum stored at -80°C. This was evident by the improved consistency for intensity of the stage 1 PCR products from NM compared to CO. Overall, the direction of the effects seen in the NM cohort for most genes were similar to that seen with CO with the exception of TCF21, Dcr2, Dab2, and GATA4 that showed little to no difference between cases and controls. The largest increases in case discrimination observed between the NM and CO cohorts were for the PAX5α, GATA5 and SULF2 genes that showed a 3.2 – 4.2-fold increase in odds ratio in NM compared to 1.4 – 1.8-fold in CO (Table 2). Overall, for NM subjects, significant differences were seen for PAX5α, GATA5, DAL1, and SULF2 (p < 0.05) and borderline significance for DAPK and PAX5β (p ≤ 0.07). Finally, the methylation status of two genes, CXCL12 and CXCL14 discovered through genome-wide transcriptome profiling (24) with functions distinct from the 31 genes studied above was evaluated in the NM 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). Quantitative MSP (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 using nested QMSP and standard QMSP. Neither approach using methylation cutoffs 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 increased levels of methylated alleles for quantitative differences.

**Association of gene methylation panels with lung cancer risk**

Receiver operator classification curves were generated from the CO and NM cohorts to determine how well the methylation of different gene panels distinguished lung cancer cases from controls. Initially the 11 genes (p16, MGMT, DAPK, PAX5α, PAX5β, GATA5, Dal-1, PCDH20, Jph3, Kif1a, and SULF2) with the highest ORs in common between the two study groups were evaluated; CXCL14 was included for NM. The ROC curves show that gene methylation increased the classification accuracy obtained with only covariates from 62% to 69% for CO (p < 0.08) and 58% to 73% (p < 0.01) for NM (Fig. 1). Methylation of several genes within the 11- and 12-gene panels was highly correlated in sputum; therefore we evaluated the performance of the seven gene panels that provided the most distinction between cases and controls. These panels were comprised of MGMT, DAPK, PAX5β, Dal-1, PCDH20, Jph3, and Kif1a for CO and DAPK, PAX5β, PAX5α, Dal-1, GATA5, SULF2, and CXCL14 for NM. With these seven gene panels classification accuracy was increased to 71% for CO (p < 0.05) and 77% NM (p < 0.001, Fig. 1) compared to only covariates. Thus, with the sensitivity set at 75%, the false-positive rate was approximately 32% for the NM case-control study.
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 seven gene panels. When the number of methylated genes was used as a continuous variable (0 to 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 three or more methylated genes compared to < 3 methylated genes from the 7-gene panels increased to 2.3 and 5.0 for Colorado and NM, respectively (Table 3). Moreover, 38% of NM cases compared to only 3% of controls showed methylation of five or more genes compared to those with fewer than five genes. This equates to a 22.3-fold increased risk for lung cancer.

Assessment of SNPs identified through GWAS for lung cancer risk

Genome-wide association studies have identified three chromosomal regions at 15q25, 5p15, and 6p21 as being associated with risk for lung cancer (19 – 23). We genotyped five 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 (\textit{nAChR} gene cluster), one within 6p21.31 (\textit{HLA-DQA1}), one within 6p21.33 (\textit{BAT3-MSH5}) and two within 5p15.33 (\textit{CLPTM1L-TERT} 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 two of the SNPs (rs1051730 Chr15q25 and rs3117582 Chr6p21) showed protection within the Colorado 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 five 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 two 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 five or more genes in a seven-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 compared to 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, while 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 seven gene panels for CO and NM make it difficult to reduce the 12-gene panel for subsequent validation studies. Six genes, p16, MGMT, DAPK, PAX5α, PAX5β, and GATA5 were identified in our initial study and have shown utility as biomarkers for uncovering genetic determinants for methylation in the aerodigestive tract and identifying dietary factors associated with reduced gene promoter methylation (12, 13, 25). Dal-1, 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 NSCLC 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 pro-
apoptosis genes (24). Overall, this gene panel is largely comprised of genes that code for transcription factors (GATA5, PAX5α, PAX5β) and genes involved in regulating cell proliferation, adhesion, and apoptosis (p16, DAPK, DAL-1, CXCL14, PCDH20, SULF2; (14 – 17, 24, 26).

The challenge is considerable for identifying non-invasive biomarkers for a disease developing over 30 years, and diagnosed at an annual incidence of 1 – 2% in current and former smokers age 45 to over 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 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 utility as one type of biomarker for predicting these tumor types. Integrating our gene methylation panel with other classes of biomarkers such as chromosomal aneusomy detected by fluorescence in situ hybridization in sputum is one avenue currently being pursued (27). Serum-based assays to detect altered expression of microRNAs, that inherently are more stable in the circulation compared to large RNAs, are showing promise as biomarkers for lung cancer detection, as are a panel of autoantibodies recently developed by Hanash and coworkers (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 to disease free controls for establishing “classification cut offs” 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 31). Implementing a multifaceted approach for validating and integrating the best DNA, RNA, and protein-based biomarkers interrogated across common
specimens (blood, sputum, 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.

Grant Support

This work was supported by National Cancer Institute Specialized Program of Research Excellence P50 CA58187 and CA58184, R01 CA097356, a research grant from MDxHealth, formerly Oncomethylome Sciences, Inc. and the State of New Mexico.
References


Figure Legend

**Figure 1.** ROC curve comparing sensitivity and specificity of gene methylation panels in the (A) Colorado (CO) and (B) NM studies for classifying lung cancer cases and controls. The covariates included in the ROC curve were age, gender, smoking status, and pack years.
Table 1. Summary of selected variable by case-control status

<table>
<thead>
<tr>
<th>Variable value</th>
<th>Colorado</th>
<th></th>
<th>p-value</th>
<th>New Mexico</th>
<th></th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
<td>Controls</td>
<td></td>
<td>Cases</td>
<td>Controls</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 64)</td>
<td>(n = 64)</td>
<td></td>
<td>(n = 40)</td>
<td>(n = 90)</td>
<td></td>
</tr>
<tr>
<td>Age (yrs)</td>
<td>68.3</td>
<td>67.6</td>
<td>0.64</td>
<td>63.8</td>
<td>64.8</td>
<td>0.48</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>(8.1)</td>
<td>(7.2)</td>
<td></td>
<td>(7.8)</td>
<td>(7.7)</td>
<td></td>
</tr>
<tr>
<td>Gender (%)</td>
<td>23.4</td>
<td>21.9</td>
<td>&gt;0.99</td>
<td>20.0</td>
<td>20.0</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>Sm. status (%)</td>
<td>39.1</td>
<td>29.7</td>
<td>0.35</td>
<td>40.0</td>
<td>44.4</td>
<td>0.70</td>
</tr>
<tr>
<td>Pk. years Mean (SD)</td>
<td>71.5</td>
<td>68.5</td>
<td>0.20</td>
<td>57.8</td>
<td>49.8</td>
<td>0.38</td>
</tr>
<tr>
<td>Ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHW</td>
<td>90.6</td>
<td>95.3</td>
<td>NA</td>
<td>60.0</td>
<td>62.2</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
<td>27.5</td>
<td>30.0</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>9.4</td>
<td>4.7</td>
<td></td>
<td>12.5</td>
<td>7.8</td>
<td></td>
</tr>
</tbody>
</table>

1P-values are from Fisher’s exact test for categorical variables and the Wilcoxon rank sum test for continuous variables.
2Test excludes missing ethnicity.
Sm, smoking; Pk, pack; NHW, non-Hispanic white.
## Table 2. Prevalence and odd ratios for gene promoter methylation in sputum samples from Colorado and New Mexico case-control studies

<table>
<thead>
<tr>
<th>Gene</th>
<th>Colorado</th>
<th>New Mexico</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases (n = 64)</td>
<td>Controls (n = 64)</td>
</tr>
<tr>
<td></td>
<td>% positive</td>
<td></td>
</tr>
<tr>
<td>P16</td>
<td>35.9</td>
<td>26.6</td>
</tr>
<tr>
<td>MGMT</td>
<td>26.6</td>
<td>17.2</td>
</tr>
<tr>
<td>DAPK</td>
<td>35.9</td>
<td>25.0</td>
</tr>
<tr>
<td>PAX5α</td>
<td>25.0</td>
<td>20.3</td>
</tr>
<tr>
<td>PAX5β</td>
<td>42.2</td>
<td>26.6</td>
</tr>
<tr>
<td>GATA5</td>
<td>32.8</td>
<td>26.6</td>
</tr>
<tr>
<td>Dal-1</td>
<td>26.6</td>
<td>9.4</td>
</tr>
<tr>
<td>PCDH20</td>
<td>45.3</td>
<td>31.2</td>
</tr>
<tr>
<td>Jph3</td>
<td>29.7</td>
<td>17.2</td>
</tr>
<tr>
<td>Kif1a</td>
<td>34.4</td>
<td>21.9</td>
</tr>
<tr>
<td>SULF2</td>
<td>34.4</td>
<td>25.0</td>
</tr>
<tr>
<td>CXCL14</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>CXCL12</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>RASSF1a</td>
<td>12.5</td>
<td>6.2</td>
</tr>
<tr>
<td>GATA4</td>
<td>52.4</td>
<td>40.3</td>
</tr>
<tr>
<td>Dab2</td>
<td>4.8</td>
<td>1.6</td>
</tr>
<tr>
<td>Dcr2</td>
<td>27.4</td>
<td>11.1</td>
</tr>
<tr>
<td>RASSF2</td>
<td>9.7</td>
<td>4.8</td>
</tr>
<tr>
<td>TCF21</td>
<td>24.2</td>
<td>14.1</td>
</tr>
</tbody>
</table>

¹Adjusted for age, gender, smoking status and pack years.
ND, not determined.
Table 3. Association of gene methylation panels and risk for lung cancer

<table>
<thead>
<tr>
<th>Methylation Panel</th>
<th>Colorado (n = 128)</th>
<th></th>
<th>New Mexico (n = 130)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>p-value</td>
<td>OR (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>Count</td>
<td></td>
<td></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>
<td>1.4 (1.2, 1.6)</td>
<td>0.0001</td>
</tr>
<tr>
<td>7 gene panel</td>
<td>1.5 (1.1, 1.9)</td>
<td>0.003</td>
<td>2.0 (1.5, 2.6)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Categorical</td>
<td></td>
<td></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>
<td>3.1 (0.9, 9.9)</td>
<td>0.06</td>
</tr>
<tr>
<td>≥ 3 genes –7 gene panel</td>
<td>2.3 (1.0, 5.2)</td>
<td>0.048</td>
<td>5.0 (2.2, 11.8)</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

1 Adjusted for age, gender, smoking status and pack years.
2 OR indicates the effect associated with a one-gene difference.
3 OR compares high to low methylation groups.
Table 4. Association of sequence variants with risk for lung cancer in Colorado and New Mexico case-control studies

<table>
<thead>
<tr>
<th>SNP</th>
<th>Genotype</th>
<th>Colorado</th>
<th>New Mexico</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>OR (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>rs1051730</td>
<td>G–&gt;A</td>
<td>0.45 (0.2, 0.8)</td>
<td>0.01</td>
</tr>
<tr>
<td>Chr15q25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs17426593</td>
<td>T–&gt;C</td>
<td>1.2 (0.6, 2.6)</td>
<td>0.57</td>
</tr>
<tr>
<td>Chr6p21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3117582</td>
<td>T–&gt;G</td>
<td>0.3 (0.1, 0.8)</td>
<td>0.02</td>
</tr>
<tr>
<td>Chr6p21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2736100</td>
<td>A–&gt;C</td>
<td>1.5 (0.8, 2.6)</td>
<td>0.57</td>
</tr>
<tr>
<td>Chr5p15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs401681</td>
<td>C–&gt;T</td>
<td>1.5 (0.8, 2.6)</td>
<td>0.20</td>
</tr>
<tr>
<td>Chr5p15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1Letters 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).

2With the exception of rs3117582, an additive model was used so the OR is the effect per allele. Since 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.
Clinical Cancer Research

Defining a gene promoter methylation signature in sputum for lung cancer risk assessment

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

Clin Cancer Res Published OnlineFirst April 17, 2012.

Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-11-3049

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2012/06/13/1078-0432.CCR-11-3049.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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.