Clinical
Cancer
Research

Early Detection of Lung Cancer Using DNA Promoter Hypermethylation in Plasma and Sputum

Alicia Hulbert[1,2], Ignacio Jusue-Torres[3], Alejandro Stark[4], Chen Chen[1,5], Kristen Rodgers[2], Beverly Lee[2], Candace Griffin[2], Andrew Yang[2], Peng Huang[1,6], John Wrangle[7], Steven A. Belinsky[8], Tza-Huei Wang[1,4,9], Stephen C. Yang[2], Stephen B. Baylin[1], Malcolm V. Brock[1,2], and James G. Herman[1,10]

Abstract

Purpose: CT screening can reduce death from lung cancer. We sought to improve the diagnostic accuracy of lung cancer screening using ultrasensitive methods and a lung cancer–specific gene panel to detect DNA methylation in sputum and plasma.

Experimental Design: This is a case–control study of subjects with suspicious nodules on CT imaging. Plasma and sputum were obtained preoperatively. Cases ($n = 150$) had pathologic confirmation of node-negative (stages I and IIA) non–small cell lung cancer. Controls ($n = 60$) had non-cancer diagnoses. We detected promoter methylation using quantitative methylation-specific real-time PCR and methylation-on-beads for cancer-specific genes (*SOX17, TAC1, HOXA7, CDO1, HOXA9,* and *ZFP42*).

Results: DNA methylation was detected in plasma and sputum more frequently in people with cancer compared with controls ($P < 0.001$) for five of six genes. The sensitivity and

specificity for lung cancer diagnosis using the best individual genes was 63% to 86% and 75% to 92% in sputum, respectively, and 65% to 76% and 74% to 84% in plasma, respectively. A three-gene combination of the best individual genes has sensitivity and specificity of 98% and 71% using sputum and 93% and 62% using plasma. Area under the receiver operating curve for this panel was 0.89 [95% confidence interval (CI), 0.80–0.98] in sputum and 0.77 (95% CI, 0.68–0.86) in plasma. Independent blinded random forest prediction models combining gene methylation with clinical information correctly predicted lung cancer in 91% of subjects using sputum detection and 85% of subjects using plasma detection.

Conclusions: High diagnostic accuracy for early-stage lung cancer can be obtained using methylated promoter detection in sputum or plasma. *Clin Cancer Res; 23(8); 1998–2005. ©2016 AACR.*

[1]Sidney Kimmel Cancer Center, Department of Oncology, The Johns Hopkins University School of Medicine, Baltimore, Maryland. [2]Department of Surgery, The Johns Hopkins University School of Medicine, Baltimore, Maryland. [3]Department of Neurosurgery, The Johns Hopkins University School of Medicine, Baltimore, Maryland. [4]Department of Mechanical Engineering, The Johns Hopkins University, Baltimore, Maryland. [5]Department of Thoracic Surgery, Second Xiangya Hospital of Central South University, Changsha, Hunan, PR China. [6]Department of Biostatistics, The Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland. [7]Department of Medicine, Medical University of South Carolina, Charleston, South Carolina. [8]Lung Cancer Program, Lovelace Respiratory Research Institute, Albuquerque, New Mexico. [9]Department of Biomedical Engineering and Institute for NanoBioTechnology, The Johns Hopkins University, School of Medicine, Baltimore, Maryland. [10]Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania.

Introduction

The National Lung Screening Trial (NLST) demonstrated a 20% reduction in lung cancer mortality using low-dose CT screening (1). This survival benefit comes at the price of detecting many indeterminate pulmonary nodules with an overall false-positive rate of 96.4% (1, 2). The likelihood that a nodule is malignant increases with size (3), with a challenge in management for the indeterminate nodules from 7 to 29 mm, with a risk of malignancy between 1.7% and 22% (3). This has led to a cautious adoption of CT screening, because complications, and even deaths, result from further diagnostic procedures (4). One approach to improve the specificity of CT screening involves the use of cancer-specific biomarkers from sputum and plasma. Previous studies have examined DNA methylation as a biomarker for cancer risk, but limited sensitivity and/or specificity were insufficient for lung cancer screening (5–16).

Reduced sensitivity of methylation detection may occur from technical limitations. Extraction methods for DNA have been inefficient for small amounts of DNA (17, 18), a particular problem for bodily fluids. We recently developed methylation-on-beads (MOB), which reduces sample loss thereby potentially increasing sensitivity (19, 20). Another issue for detection is the use of loci with low frequencies of altered DNA methylation, leading to an inability to detect changes in biofluids. We recently identified 6 genes (*SOX17, TAC1, HOXA7, CDO1, HOXA9, ZFP42*) using The Cancer Genome Atlas (TCGA;

AACR

Translational Relevance

The National Lung Screening Trial demonstrated a 20% reduction in lung cancer mortality using low-dose CT screening. Diagnostic accuracy of screening could be improved using cancer-specific biomarkers from sputum and plasma. We developed methylation-on-beads (MOB), reducing sample loss with potentially increased sensitivity. We used MOB and real-time quantitative methylation-specific PCR (qMSP) to detect the promoter methylation using genes frequently methylated in *SOX17*, *TAC1*, *HOXA7*, *CDO1*, *HOXA9*, and *ZFP42*. This study demonstrates that high diagnostic accuracy of early-stage NSCLC can be obtained using a panel of methylated promoter genes in plasma and sputum and that the methylation level of these genes is associated with a high lung cancer risk independent of age, pack-year, and nodule size. This panel could be used to identify patients at high risk for lung cancer, reducing false-positive results, unnecessary tests, and improving the diagnosis of lung cancer at an earlier stage.

ref. 21) with highly prevalent DNA methylation in lung squamous and adenocarcinoma, but not in normal lung tissue (22, 23) one of which (CDO1) has been described elsewhere (22, 23). These were chosen solely on the basis of high frequency cancer-specific methylation and developed into assays using MOB and real-time methylation-specific PCR (qMSP) to determine the diagnostic accuracy for lung cancer detection in sputum and plasma.

Materials and Methods

Study population

The study population consists of a prospective, observational cohort of 651 participants, initiated in 2007 within the Johns Hopkins Lung Cancer Specialized Program of Research Excellence (SPORE). From this cohort, 210 study patients had early-stage node-negative tumors (T1–T2N0) and samples adequate for analysis. Institutional Review Board approval was obtained prior to study initiation (NA_00005998), and all patients signed informed consent. Surgical resection with curative intent and pathologic analyses of suspected lung cancer lesions were completed in all patients and staged according to revised TNM guidelines classification criteria (24). Cases had pathologically confirmed lung cancer. Controls were defined as patients histologically confirmed not to have cancer. Pack-years of cigarette smoking was defined as the average number of packs smoked per day times the number of years smoked. Nodule size was obtained from the pathologic report, and nodule volume was calculated using the ellipsoid volume formula (Volume = 4/3 × π × radius A × radius B × radius C).

Plasma and sputum collection

Prior to surgery, 20 mL of plasma was collected in tubes containing sodium heparin (Becton Dickinson) and then stored at −80°C. For sputum collection, 2 cups containing Saccomanno's fixative solution were used for each patient as previously described (8, 11, 25). Subjects were asked to provide an early morning spontaneous sputum at home in 2 cups for 3 consecutive days within 1 week prior to pulmonary resection (11, 26). Five milliliters of sputum was collected, washed with Saccomanno's solution, vortexed, centrifuged, and stored at −80°C (8).

DNA isolation and bisulfite conversion

DNA extraction from tumor, plasma, and sputum was performed using MOB, a process that allows DNA extraction and bisulfite conversion in a single tube via the use of silica super magnetic beads (20). This approach yields a 1.5- to 5-fold improvement in extraction efficiency compared with traditional conventional techniques (27). We optimized the protocol previously described for plasma (27), using 1.5 mL of plasma and 375 μL (800 units/mL, NEBL p8107s) of proteinase K. For DNA extraction from sputum, we modified the protocol used for plasma by adding 200 μL of sample to 300 μL of Buffer AL and 40 μL of proteinase K and by incubating them together at the same temperature (50°C for 2 hours). After digestion, 300 μL of isopropyl alcohol (IPA) and 150 μL of beads were added. The lysate was incubated and rotated for 10 minutes before adding 5 μL of carrier RNA and incubating for an additional 5 minutes (27).

DNA methylation analysis

The genomic sequence for the genes and 1,000 bases upstream was obtained from the UCSC genomic browser website (28). The primers and hybridization probes for methylation analysis were designed on the basis of this sequence by using Primer3 (v.0.4.0; refs. 29, 30). All primer and probe sequences are listed in Supplementary Table S1. The analysis was performed using real-time qMSP and normalized to a control β-actin assay (18). Each reaction was performed in a 25-μL PCR mixture consisting of 2 μL of bisulfite-converted DNA, 300 nmol/L R-sense primer, 300 nmol/L F-anti-sense primer, 100 nmol/L probe, 100 nmol/L of fluorescein reference dye (Life Technologies), 1.67 mmol/L dNTPs (VWRQuotation), and 1 μL of Platinum Taq DNA Polymerase (Invitrogen). Master mix contained 16.6 mmol/L (NH4)2SO4, 67 mmol/L Tris, pH 8.8, 6.7 mmol/L MgCl$_2$, and 10 mmol/L β-mercaptoethanol in a nuclease-free deionized water solution. Amplification reactions were performed using 96-well plates (MicroAmp) in triplicate. Thermocycling conditions were: 95°C for 5 minutes, 50 cycles at 95°C for 15 seconds, and 65°C for 1 minute and 72°C for 1 minute. An ABI StepOnePlus Real-Time PCR system was used (Applied BioSystems, examples shown in Supplementary Fig. S1).

With the extremely low levels of DNA methylation in plasma and sputum, replicates for some samples produced no detectable methylation as expected. To incorporate this information into the final quantification of methylation, we calculated the $2^{-\Delta Ct}$ for each methylation detection replicate comparing it to the mean C_t for β-actin (ACTB). For replicates which were not detected (ND), a C_T of 100 was used, creating a near zero value for $2^{-\Delta Ct}$. The mean $2^{-\Delta Ct}$ value was calculated with the formula:

$$\mu \ 2^{-\Delta Ct} = \frac{\left(2^{-\Delta Ct\_replicate\_1} + 2^{-\Delta Ct\_replicate\_2} + 2^{-\Delta Ct\_replicate\_3}\right)}{3}$$

Statistical analysis

Quantitative data are expressed as median (interquartile range) for continuous, nonparametric variables and frequency

(percentage) for categorical variables. For intergroup comparison, the Wilcoxon rank-sum test was used for continuous data and the Fisher exact test for categorical data.

Data were analyzed using 2 approaches. The first approach is the receiver operating curve (ROC) analysis using the $2^{-\Delta Ct}$ values for individual genes to determine the performance of each individual marker (R statistic software, version 3.0.2; ref. 31). The area under the curve was reported with 95% confidence intervals (CI). The 3 best-performing genes were selected for diagnostic accuracy for lung cancer detection, on the basis of ROC curves and were used for combined detection. Sensitivity and specificity values were obtained from the presence or absence of detectable methylation as a cutoff.

The second approach utilized a nonparametric machine learning method, random forest, to estimate the prediction accuracy in an independent validation dataset by combining the methylation data and clinical risk factors: nodule size, age, pack-year, chronic obstructive pulmonary disease (COPD) status, and forced vital capacity (FVC) values. Subjects were randomly selected as a training set (67%) and a test set (33%). A statistician (P. Huang), blinded to the diagnoses of the test set, used the training set to build 3 random forest prediction models: (i) used sputum, clinical, and demographic variables, (ii) used plasma and clinical variables, and (iii) used only clinical and demographic variables. The random forest model consisted of 5,000 trees, each using a random sample of the training data. The remaining training data were used for internal cross-validation. Each random forest model provides 2 predictions: the cancer status (a binary prediction) and the probability of cancer (a continuous prediction). The 2 random forest models were then applied to the test set data. Prediction accuracy was reported as the proportion of test set subjects correctly predicted by the random forest classification models, allowing calculation of sensitivity, specificity, and ROC analysis.

Results

Characteristics of the patients

Two hundred and ten patients fulfilled inclusion criteria, with 150 node-negative early-stage lung cancer subjects and 60 controls with non-cancerous lung lesions (Table 1). Clinical and demographic variables were similar in cases and controls with the exception of age, number of pack-year, and nodule size (cm) as well as volume (cm^3). Subjects with lung cancer were older than controls (67 vs. 73 years, $P = 0.007$), smoked more (30 vs. 19.5 pack-years, $P = 0.01$), and had larger nodules (2.0 vs. 1.5 cm, $P = 0.01$). The proportion of smokers, former smokers, and never smokers was not different between cases and controls.

Detection of DNA methylation

We first measured DNA methylation for these genes in tumor tissue, confirming our previous study suggesting these genes were methylated in the majority of lung tumors (Fig. 1). Methylation in sputum was detected more frequently in all 6 genes in patients with cancer compared with controls (Fig. 1), which for some patients was quantitatively similar to lung tumor tissues, but in some cases was at levels previously below conventional methods of detection. For 5 of the 6 genes (SOX17, TAC1, HOXA7, CDO1, and ZFP42), this was statistically significant ($P < 0.001$), whereas HOXA9 showed a lack of

Table 1. Baseline characteristics of the 210 subjects

Patient characteristics	Cancer (N = 150)	Control (N = 60)	P
Age at surgery (IQR), y	68 (62–75)	63 (55–73)	0.007
Gender			
Male (%)	63 (42%)	33 (55%)	0.094
Female (%)	87 (58%)	27 (45%)	
Race			
White (%)	120 (80%)	51 (85%)	0.087
Black (%)	19 (13%)	3 (5%)	
Other (%)	11 (7%)	6 (10%)	
Stage			
IA–IB (%)	136 (91%)	NA	NA
IIA (%)	14 (9%)	NA	
Histology			
Adenocarcinoma (%)	121 (81%)	NA	NA
Squamous cell (%)	26 (17%)	NA	
Adenosquamous (%)	3 (2%)	NA	
Smoking status			
Current (%)	27 (18%)	7 (12%)	0.176
Former (%)	87 (58%)	34 (57%)	
Never (%)	31 (21%)	19 (32%)	
Pack-year (IQR)	30 (10–50)	20 (0–35)	0.010
COPD (%)	41 (27%)	12 (20%)	0.370
FEV1 % predicted (IQR)	84 (70–99)	85 (70–100)	0.861
FVC % predicted (IQR)	92 (80–103)	87 (80–110)	0.682
FEV1/FVC % ratio (IQR)	73 (68–78)	77 (70–79)	0.080
Nodule size, cm	2 (1.5–3)	1.5 (1.1–3)	0.01
<1	6 (4%)	13 (22%)	0.001
1–2	52 (35%)	19 (32%)	
>2	92 (61%)	28 (47%)	
Nodule volume, cm^3	4.19 (1.77–14–14)	1.6 (0.52–18.12)	0.001

NOTE: Nodule size % <1, 1–2, >2 cm.
Abbreviations: FEV1, forced expiratory volume in 1 second; IQR, interquartile range.

specificity. Methylation of all 6 genes was detected more frequently in plasma in cases compared with controls ($P < 0.001$). The worst performing gene was HOXA9 in plasma, which showed a lack of specificity as was seen in the sputum. We determined the sensitivity and specificity in this cohort using the presence or absence of detectable methylation as a cutoff, without considering the quantitation of methylation. This resulted in good sensitivity and specificities (Table 2), showing that the sensitivity and specificity for lung cancer diagnosis using individual genes from sputum ranged from 63% to 93% and 42% to 92%, respectively, and from plasma from 33% to 91% and 52% to 94%, respectively.

Gene methylation and lung cancer diagnostic accuracy

ROC curves for lung cancer detection were obtained for each single gene; using the normalized methylation ΔC_t values calculated as described in Materials and Methods (Supplementary Table S2, ROC curves in Supplementary Figs. S2 and S3). By determining the best quantitative cutoff, the sensitivity and specificity for lung cancer diagnosis from single methylated genes in sputum ranged 63% to 93% and 42% to 92%, respectively, and in plasma from 33% to 91% and 52% to 94%, respectively, and was very similar to that obtained reported in Table 2, with the exception of HOXA9 where quantitative cutoffs improved performance. The AUC values were 0.56 to 0.89 in sputum samples and 0.60 to 0.78 in plasma samples.

The genes with the largest AUC in sputum were: TAC1: AUC, 0.84; 95% CI, 0.74–0.94; HOXA7: AUC, 0.77; 95% CI, 0.67–0.86;
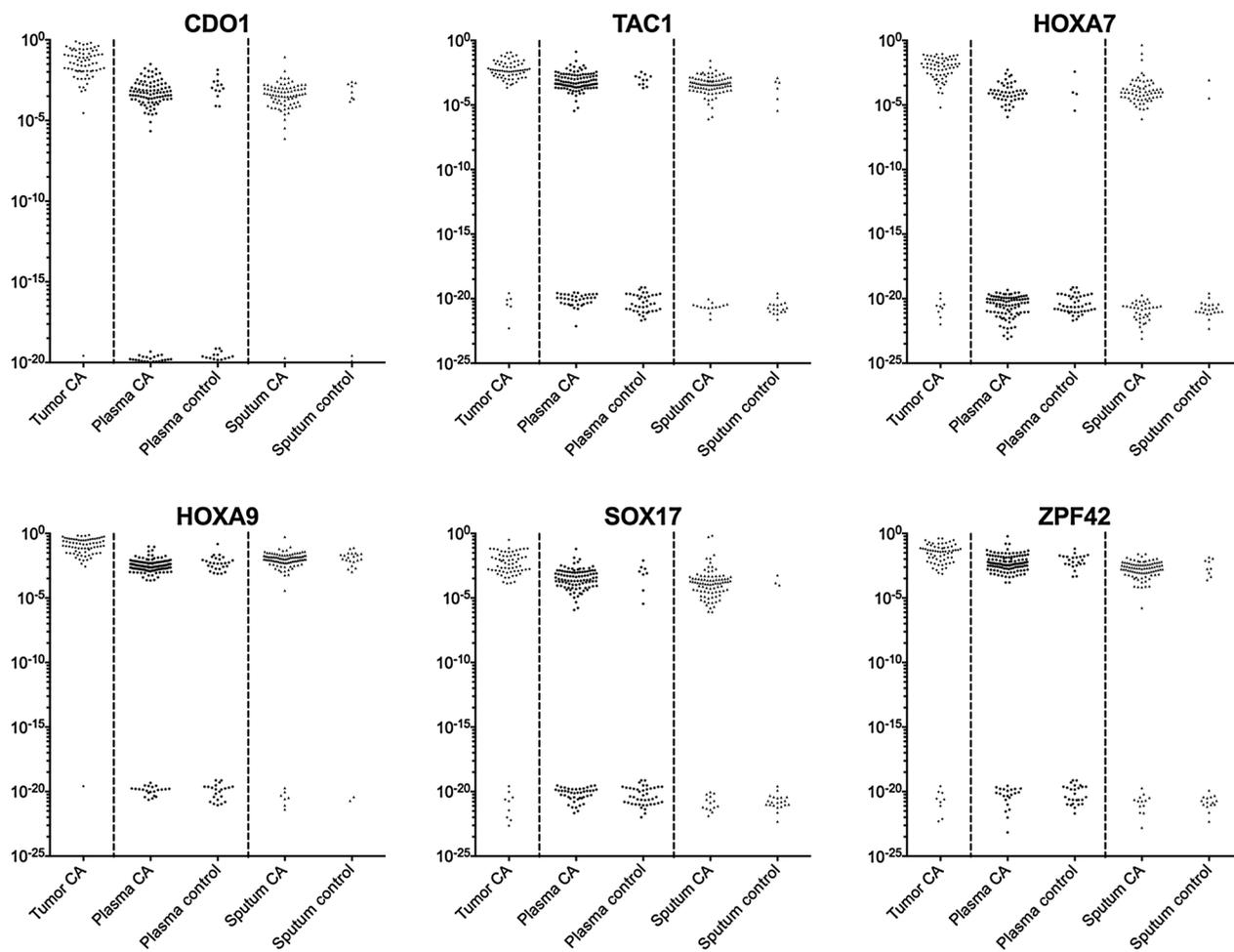
Figure 1.
Methylation detection values of the studied genes. This scatter plot shows the converted ΔC_t methylation values in a logarithmic scale. These values show a bimodal distribution with the lower group the values corresponding to those samples with no detectable amplification (ND). The majority of lung tumor samples have high levels of methylation, as expected from the previous study. Plasma and sputum samples from patients with cancer have detectable methylation that varies from levels nearing that of tumor samples to those at the limits of detection (10^{-5} to 10^{-6}), whereas samples from other patients are undetectable. The majority of controls have undetectable methylation at these loci, although some patients do have detectable methylation that is quantitatively similar to patients with cancer. *HOXA9* methylation is detectable in most control patients, especially in the sputum, suggesting that this change is present in the lung epithelium and not as specific for the detection of cancer.

and *SOX17*: AUC, 0.84; 95% CI, 0.75–0.94 (Fig. 2A), with sensitivities and specificities for *TAC1* at 86% and 75%; *HOXA7* at 63% and 92%; and *SOX17* at 84% and 88, respectively. The positive (PPV) and negative (NPV) predictive values for these 3 genes were: for *TAC1*, 93% and 58%; for *HOXA7*, 97% and 40%; and for *SOX17*, 96% and 60%, respectively.

In plasma, the genes with the largest AUC were: *CDO1*: AUC, 0.68; 95% CI, 0.58–0.77; *TAC1*: AUC, 0.78; 95% CI, 0.70–0.86; and *SOX17*: AUC, 0.78; 95% CI, 0.70–0.86 (Fig. 2B), with corresponding sensitivities and specificities for *CDO1* at 65% and 74%; *TAC1* at 76% and 78%; and *SOX17* at 73% and 84%, respectively. The PPV and NPV for these genes were: for *CDO1*, 86% and 46%; for *TAC1*, 90% and 57%; and for *SOX17*, 92% and 55%, respectively.

The sensitivity and specificity obtained from the combination of the 3 best performing markers (*TAC1*, *HOXA17*, and *SOX17*) in sputum was 98% and 71%, respectively, with a corresponding

ROC AUC of 0.89 (95% CI, 0.80–0.98; Fig. 2C). In plasma, the combination of *CDO1*, *TAC1*, and *SOX17* showed a sensitivity, specificity, and AUC of 93%, 62%, and 0.77 (95% CI, 0.68–0.86), respectively (Fig. 2D).

Smokers' subset analysis

As CT screening for lung cancer is currently recommended for current and ex-smokers, we explored the diagnostic accuracy when only smokers were considered ($n = 155$; 114 with cancer and 41 without cancer; Supplementary Table S4). The results in only smokers were similar to the entire study population for the prevalence of methylated patients, sensitivity, specificity, and AUC (Supplementary Table S5). AUC in smokers only was 0.89 (95% CI, 0.79–0.99) for the combination of the methylation status of the best 3 genes from sputum and AUC was 0.85 (95% CI, 0.76–0.94) from the best 3 genes from plasma (Supplementary Table S5).

Table 2. Gene methylation detection in sputum and plasma

Sputum	Cancer (N = 90)		Control (N = 24)		PPV	NPV
	n	Sensitivity	n	Specificity		
SOX17	76	84%	3	88%	96%	60%
TAC1	77	86%	6	75%	93%	58%
HOXA7	57	63%	2	92%	97%	40%
CDO1	70	78%	8	67%	90%	44%
HOXA9	84	93%	22	8%	79%	25%
ZFP42	78	87%	9	63%	90%	56%
TAC1, HOXA7, SOX17	88	98%	7	71%	93%	89%
Plasma	**Cancer (N = 125)**		**Control (N = 50)**		**PPV**	**NPV**
	n	Sensitivity	n	Specificity		
SOX17	91	73%	8	84%	92%	55%
TAC1	95	76%	11	78%	90%	57%
HOXA7	42	34%	4	92%	91%	36%
CDO1	81	65%	13	74%	86%	46%
HOXA9	108	86%	27	46%	80%	58%
ZFP42	105	84%	23	54%	82%	57%
CDO1, TAC1, SOX17	116	93%	19	62%	86%	78%

Independent prediction accuracy performance

While the above analysis looked at individual gene methylation in cases and controls to detect cancer, independent blinded random forest prediction models were used to consider all DNA methylation biomarkers in combination with clinical risk factors. Risk factors included in the first 2 random forest prediction models were methylation C_t values from all 6 genes, age, pack-year, COPD status, and FVC values. The methylation C_t values were not included in the last prediction model. The randomly selected training dataset has 140 subjects with 99 (70.7%) cancers and 41 (29.3%) controls. The independent test set has 70 subjects with 51 (72.9%) cancers and 19 (27.1%) controls. In the variable of importance output of the first 2 random forest prediction models, methylation C_t values were ranked as more important variables than demographic and clinical variables (Supplementary Fig. S4). Supplementary Table S3 summarizes the prediction accuracies of these 3 models when they were applied to the independent test set patients. With sputum samples, the random forest model correctly predicted lung cancer in 91% of subjects in the test subset. The corresponding AUC was 0.85 (95% CI, 0.59–1.0; Fig. 3). The sensitivity and specificity of the prediction in the testing subset from the ROC curve were 0.93 and 0.86, respectively. Using plasma samples, the random forest model correctly predicted lung cancer in 85% of subjects in the testing subset. The corresponding AUC was 0.89 (95% CI, 0.79–0.99; Fig. 3). The sensitivity and specificity of the prediction in the testing subset from the ROC curve were 0.93 and 0.67, respectively. Using clinical and demographic risk factors alone, the accuracies were lower than the first 2 models with a diagnostic accuracy of 68%, AUC of 0.64, PPV of 75%, and NPV of 38% (Fig. 3; Supplementary Table S3).

Discussion

High diagnostic accuracy for early-stage lung cancer can be obtained using a panel of methylated promoter genes and an ultrasensitive detection strategy on the basis of MOB in sputum or plasma., This assay has several characteristics which make it clinically useful: (i) the sensitivity and specificity in sputum and plasma exceeds the diagnostic accuracy required by most clinical standards (32, 33); (ii) it can be performed with minute quantities of DNA from sputum or plasma; and (iii) it can help distinguish malignant versus benign nodules, addressing the current problem of high false-positive scans in lung cancer screening. This discrimination is independent of age, pack-year, and even nodule size, which allows the detection of early-stage lung cancer in smokers. Finally, as a PCR-based assay, it is simple and relatively inexpensive.

Previous studies have sought to improve lung cancer risk assessment by the use of molecular biomarkers obtained from plasma and sputum (8, 10, 11, 25, 26, 34, 35). However none of these tests have been used clinically because their achieved sensitivities and specificities were usually not high enough for clinical decision making (8, 10, 11, 25, 26, 34–38). With improvements in DNA extraction methods and processing for methylation detection, along with the use of highly prevalent cancer-specific methylation targets, we have overcome these limitations. Direct comparisons between serum and plasma for detection of DNA methylation have not been conducted in this study, but the use of plasma may reduce the amount of lymphocyte DNA present for analysis.

Despite the improved sensitivity of this approach, there are some patients with undetectable DNA methylation in either blood or sputum. In examining these nondetectable patients, this is not related to clinical characteristics, including smoking status (see similar detection in only smokers, Supplementary Data). This does not appear to be related to PCR failure or assay efficiency, which was assessed for each assay using appropriate controls (Materials and Methods). We also examined whether tumor size and therefore tumor burden affected our ability to detect DNA methylation in the plasma or sputum. There was no statistical difference in tumor size between patients with cancer with or without detectable DNA methylation. (Supplementary Table S6), and notably nodules less than 2 cm were readily detected.

In this study, detection of methylation in sputum samples was slightly better than the detection of these same genes in plasma. The access of early cancers to the airways may be one explanation for this difference. Indeed, changes in
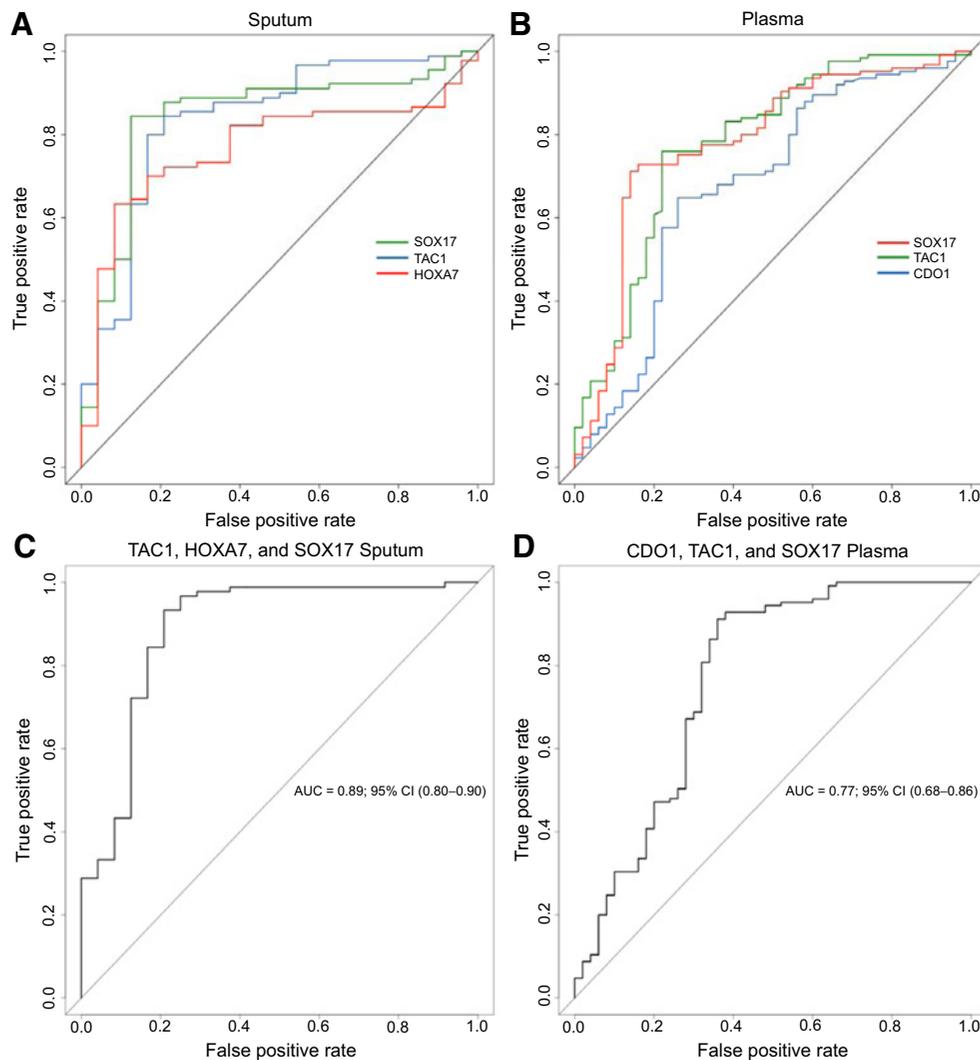
Figure 2.
ROC curves for lung cancer detection. **A,** ROC curves comparing the 3 genes with the largest areas under the curve for sputum. **B,** ROC curves comparing the 3 genes with the largest areas under the curve for plasma. **C,** ROC of the combined methylation status of the genes from sputum with the largest area under the curve. **D,** ROC of the combined methylation status of the genes from plasma with the largest area under the curve.

the airways form the basis for the AEGIS Study, which reported an improved diagnostic yield of bronchoscopy using gene expression classifiers from epithelial cells collected during bronchoscopy (38). The AUC, sensitivities, and specificities reported in the AEGIS Study were lower than we report here. In our model where methylation markers from plasma were considered simultaneously with age and number of pack-years, we observed a predictive accuracy close to that of sputum. This suggests that plasma could substitute for sputum in lung cancer detection in those cases where sputum cannot be obtained.

According to the NLST, the chances of having lung cancer with a positive CT screening are less than 5% (1, 2). This is because lung cancer with CT screening in the NLST study yielded a 71% sensitivity but a 63% specificity with a 96.4% false-positive rate (1, 2). Our current findings suggest that methylation detection with a few genes from either plasma or sputum could potentially

guide management of positive CT screening results. Although our study included some patients who would not meet current lung cancer screening guidelines (non-smokers), we observed similar detection rates when only smokers were analyzed. Replication and external validation of our findings in a large, prospective, multicenter case–control trial are essential before this approach can be adopted.

This study shows that high sensitivity and specificity detection of early-stage NSCLC can be obtained using a panel of methylated promoter genes in plasma and sputum and that the methylation level of these genes is associated with a high lung cancer risk independent of age, pack-year, and nodule size. If confirmed in a validation study, this panel could be used as an adjunct to CT screening, identifying patients at high risk for lung cancer, reducing false-positive results, unnecessary tests, and improving the diagnosis of lung cancer at an earlier stage.
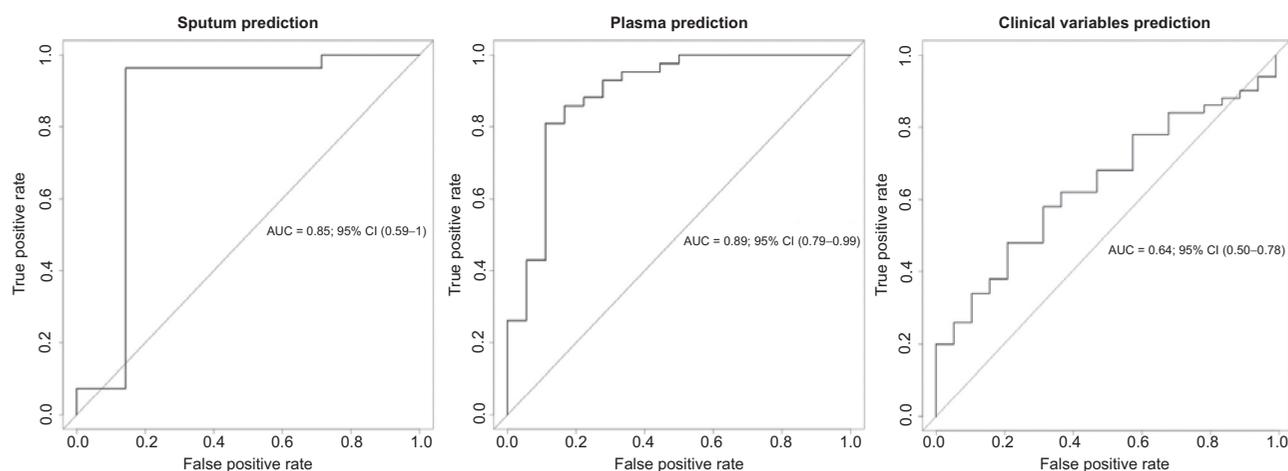
Figure 3.
ROC curves for cancer predictions. ROC curves assessing the accuracy of the predictions for lung cancer performed on the testing subset by using as predictors the ΔC_t values for all 6 genes, age, pack-year, COPD status, and FVC values. The left plot is obtained using sputum samples, the middle one using plasma samples, and the right one using the ROC curve for the clinical predictors alone.

Disclosure of Potential Conflicts of Interest

A. Hulbert holds ownership interest (including patents) in Johns Hopkins Technology Ventures. M.V. Brock holds ownership interest (including patents) in and is a consultant/advisory board member for Cepheid Corporation. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: A. Hulbert, I. Jusue-Torres, A. Stark, M.V. Brock, J.G. Herman

Development of methodology: A. Hulbert, I. Jusue-Torres, A. Stark, K. Rodgers, B. Lee, J. Wrangle, T.-H. Wang, J.G. Herman

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): C. Chen, K. Rodgers, B. Lee, C. Griffin, A. Yang, S.C. Yang

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): I. Jusue-Torres, A. Stark, C. Chen, K. Rodgers, A. Yang, P. Huang, J. Wrangle, S.C. Yang, M.V. Brock, J.G. Herman

Writing, review, and/or revision of the manuscript: A. Hulbert, I. Jusue-Torres, A. Stark, C. Griffin, P. Huang, S.A. Belinsky, T.-H. Wang, S.C. Yang, S.B. Baylin, M.V. Brock, J.G. Herman

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A. Stark, K. Rodgers, B. Lee, C. Griffin, S.C. Yang, J.G. Herman

Study supervision: A. Hulbert, I. Jusue-Torres, K. Rodgers, B. Lee, C. Griffin, S.C. Yang, M.V. Brock, J.G. Herman

References

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365: 395–409.
2. Tammemagi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med 2013;368:728–36.
3. Gierada DS, Pinsky P, Nath H, Chiles C, Duan F, Aberle DR. Projected outcomes using different nodule sizes to define a positive CT lung cancer screening examination. J Natl Cancer Inst 2014;106. pii: dju284.
4. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, et al. Benefits and harms of CT screening for lung cancer: a systematic review. JAMA 2012;307:2418–29.
5. Esteller M, Sanchez-Cespedes M, Rosell R, Sidransky D, Baylin SB, Herman JG. Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. Cancer Res 1999;59:67–70.
6. Wong IH, Lo YM, Zhang J, Liew CT, Ng MH, Wong N, et al. Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. Cancer Res 1999;59:71–3.
7. Palmisano WA, Divine KK, Saccomanno G, Gilliland FD, Baylin SB, Herman JG, et al. Predicting lung cancer by detecting aberrant promoter methylation in sputum. Cancer Res 2000;60:5954–8.
8. Belinsky SA, Liechty KC, Gentry FD, Wolf HJ, Rogers J, Vu K, et al. Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. Cancer Res 2006;66:3338–44.
9. Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, et al. DNA methylation markers and early recurrence in stage I lung cancer. N Engl J Med 2008;358:1118–28.
10. Ostrow KL, Hoque MO, Loyo M, Brait M, Greenberg A, Siegfried JM, et al. Molecular analysis of plasma DNA for the early detection of lung cancer by quantitative methylation-specific PCR. Clin Cancer Res 2010;16:3463–72.
11. Leng S, Do K, Yingling CM, Picchi MA, Wolf HJ, Kennedy TC, et al. Defining a gene promoter methylation signature in sputum for lung cancer risk assessment. Clin Cancer Res 2012;18:3387–95.
12. Li L, Shen Y, Wang M, Tang D, Luo Y, Jiao W, et al. Identification of the methylation of p14ARF promoter as a novel non-invasive biomarker for early detection of lung cancer. Clin Transl Oncol 2013;16:581–9.

13. Sandoval J, Mendez-Gonzalez J, Nadal E, Chen G, Carmona FJ, Sayols S, et al. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. J Clin Oncol 2013;31:4140–7.
14. Kim Y, Kim D-H. CpG island hypermethylation as a biomarker for the early detection of lung cancer. New York, NY: Springer New York; 2014. p.141–71.
15. Nawaz I, Qiu X, Wu H, Li Y, Fan Y, Hu L-F, et al. Development of a multiplex methylation specific PCR suitable for (early) detection of non-small cell lung cancer. Epigenetics 2014;9:1138–48.
16. Yang X, Dai W, Kwong DL-w, Szeto CYY, Wong EH-w, Ng WT, et al. Epigenetic markers for noninvasive early detection of nasopharyngeal carcinoma by methylation-sensitive high resolution melting. Int J Cancer 2014;136:E127–35.
17. Herman JG, Graff JR, Myohanen 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;93:9821–6.
18. Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D, et al. MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res 2000;28:E32.
19. Bailey VJ, Keeley BP, Razavi CR, Griffiths E, Carraway HE, Wang TH. DNA methylation detection using MS-qFRET, a quantum dot-based nanoassay. Methods 2010;52:237–41.
20. Bailey VJ, Zhang Y, Keeley BP, Yin C, Pelosky KL, Brock M, et al. Single-tube analysis of DNA methylation with silica superparamagnetic beads. Clin Chem 2010;56:1022–5.
21. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature 2012;489:519–25.
22. Wrangle J, Machida EO, Danilova L, Hulbert A, Franco N, Zhang W, et al. Functional identification of cancer-specific methylation of CDO1, HOXA9, and TAC1 for the diagnosis of lung cancer. Clin Cancer Res 2014;20:1856–64.
23. Diaz-Lagares A, Mendez-Gonzalez J, Hervas D, Saigi M, Pajares MJ, Garcia D, et al. A novel epigenetic signature for early diagnosis in lung cancer. Clin Cancer Res 2016;22:3361–71.
24. Ettinger DS, Wood DE, Akerley W, Bazhenova LA, Borghaei H, Camidge DR, et al. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) Non-small cell lung cancer, version 1.2015. J Natl Compr Canc Netw 2014;12:1738–61.
25. Belinsky SA, Klinge DM, Dekker JD, Smith MW, Bocklage TJ, Gilliland FD, et al. Gene promoter methylation in plasma and sputum increases with lung cancer risk. Clin Cancer Res 2005;11:6505–11.
26. Prindiville SA, Byers T, Hirsch FR, Franklin WA, Miller YE, Vu KO, 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;12:987–93.
27. Keeley B, Stark A, Pisanic TR2nd, Kwak R, Zhang Y, Wrangle J, et al. Extraction and processing of circulating DNA from large sample volumes using methylation on beads for the detection of rare epigenetic events. Clin Chim Acta 2013;425:169–75.
28. Genome Bioinformatics Group of UC Santa Cruz. UCSC genome bioinformatics. 2015. [cited 2015 Jun 30]. Available from: http://genome.ucsc.edu/.
29. Brandes JC, Carraway H, Herman JG. Optimal primer design using the novel primer design program: MSPprimer provides accurate methylation analysis of the ATM promoter. Oncogene 2007;26:6229–37.
30. Untergrasser A CI, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3web. 2012. [cited 2015 Jun 30]. Available from: http://primer3.ut.ee/.
31. Team RC. R: A language and environment for statistical computing. In: R-Project Org, Version 3.0.2 ed. Vienna, Austria: R Foundation for Statistical Computing; 2013.
32. Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, Reid B, et al. The case for early detection. Nat Rev Cancer 2003;3:243–52.
33. Belinsky SA. Gene-promoter hypermethylation as a biomarker in lung cancer. Nat Rev Cancer 2004;4:707–17.
34. Kennedy TC, Proudfoot SP, Franklin WA, Merrick TA, Saccomanno G, Corkill ME, et al. Cytopathological analysis of sputum in patients with airflow obstruction and significant smoking histories. Cancer Res 1996;56:4673–8.
35. Kennedy TC, Proudfoot SP, Piantadosi S, Wu L, Saccomanno G, Petty TL, et al. Efficacy of two sputum collection techniques in patients with air flow obstruction. Acta Cytol 1999;43:630–6.
36. Ahrendt SA, Chow JT, Xu LH, Yang SC, Eisenberger CF, Esteller M, et al. Molecular detection of tumor cells in bronchoalveolar lavage fluid from patients with early stage lung cancer. J Natl Cancer Inst 1999;91:332–9.
37. Sanchez-Cespedes M, Esteller M, Wu L, Nawroz-Danish H, Yoo GH, Koch WM, et al. Gene promoter hypermethylation in tumors and serum of head and neck cancer patients. Cancer Res 2000;60:892–5.
38. Silvestri GA, Vachani A, Whitney D, Elashoff M, Porta Smith K, Ferguson JS, et al. A bronchial genomic classifier for the diagnostic evaluation of lung cancer. N Engl J Med 2015;373:243–51.