Journal of Molecular Biomarkers & Diagnosis
ISSN: 2155-9929

Zhao et al., J Mol Biomark Diagn 2015, 6:6
DOI: 10.4172/2155-9929.1000250

Research Article Open Access

Selection of Sensitive Methylation Markers for the Detection of Non-small Cell Lung Cancer

Xia Zhao[1]*, Jin Jen[1], Tobias Peikert[2], Eric Edell[2], Shulan Tian[3], Ping Yang[4], Yajue Huang[5] and Hongzhi Zou[1]*

[1]*Division of Experiemental Pathology and Laboratory Medicine, Mayo Clinic, Minnesota, USA*
[2]*Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Minnesota, USA*
[3]*Division of Biomedical Statistics and Informatics, Mayo Clinic, Minnesota, USA*
[4]*Division of Epidemiology, Mayo Clinic, Minnesota, USA*
[5]*Division of Anatomic Pathology, Mayo Clinic, Rochester, Minnesota 55905, USA*

Abstract

Introduction: While early-stage lung cancer is curable by surgical resection, most patients are diagnosed with advanced- stage disease. Annual low-dose computed tomography screening decreases lung cancer mortality, however effective biomarkers to address the high false positive rate and to better define high risk individuals are lacking. This study was designed to identify potential DNA methylation markers for the detection of non-small cell lung cancer, the most common type of lung cancer.

Methods: 152 candidate methylation genes were first investigated in lung cancer cell lines and a pilot set of lung tissues. Five promising methylated genes, DMRTA, HOXA9, ZIC4, HOXA7, and SIX3, were selected and further validated in 150 non-small cell lung cancers and 142 tumor-free surrounding lung tissues using the quantitative methylation-specific PCR.

Results: Methylation levels of DMRTA2, HOXA9, ZIC4, HOXA7, and SIX3 were significantly higher in tumors compared to tumor-free surrounding lung tissues ($P<2.2e-16$ for all). Receiver operation curve analysis showed that methylation of DMRTA2, HOXA9, ZIC4, HOXA7, and SIX3 identified 93%, 91%, 89%, 81%, and 59% of non-small cell lung cancers (n=150) with a specificity of 95%. Comparing tumors to tumor-free surrounding lung tissues, area under the curve values were 0.967, 0.955, 0.950, 0.904, and 0.819, respectively. The predicted area under the curve value after combining DMRTA2 and HOXA9 was 0.971. Methylation levels of these genes were not correlated to cancer stages ($P>0.05$).

Conclusion: We identified a group of highly sensitive and specific methylation markers in non-small cell lung cancer. These markers are potential valuable candidates to improve the performance of lung cancer screening.

Keywords: DNA methylation; Biomarker; Non-small cell lung cancer; DMRTA2; HOXA9

Introduction

Lung cancer is the leading cause of cancer-related death worldwide [1]. In the United States, more than 221,000 new cases of lung cancer will be diagnosed in 2015, and over 158,000 individuals will die of this disease. This account for 26.8% of total deaths caused by all cancers [2]. Based on histological characteristics, lung cancer is divided into two major subtypes: non-small cell lung cancer (NSCLC, 85% of cases) and small cell lung carcinoma (SCLC, 15% of cases) [3]. NSCLC is further sub-classified into adenocarcinoma (AD), squamous cell carcinoma (SCC), and large cell carcinoma (LC). While the survival of patients diagnosed with SCLC is almost universally poor, surgical resection can be curative in patients with early-stage NSCLC [4,5]. However, in the absence of effective early detection, the majority of NSCLC patients are diagnosed with advanced-stage disease.

Chest radiography and sputum cytology are ineffective for lung cancer screening [6]. In 2011, the National Lung Cancer Screening Trial demonstrated that screening with annual low-dose computed tomography (LDCT) reduces lung cancer mortality by 20% in high-risk individuals [6,7]. LDCT lung cancer screening programs are being implemented throughout the United States. While this represents a breakthrough for lung cancer screening, several challenges remain. LDCT is expensive, associated with cumulative radiation exposure and a very high false-positive rate (>96%) [8]. Additional diagnostic tests and unnecessary invasive procedures to determine the nature of over-abundant lung nodules would result in potentially increased mortality, morbidity, and healthcare costs. In addition the currently recommended definition for the high-risk individuals to be offered lung cancer screening only includes a subgroup of individuals who will ultimately develop lung cancer [9]. Therefore, there is an urgent need to identify specific, cost-effective non-invasive biomarkers for lung cancer screening. These tests could be applied independently or in combination with LDCT to help to identify a high-risk screening population and reduce the high false-positive rate of LDCT screening.

DNA methylation has been proposed to be a valuable source of biomarkers for disease diagnosis and risk stratification [10-12]. The clinical application of methylation markers has been successful in screening for different malignancies. For instance, methylated BMP3 (bone morphogenetic protein 3) and NDRG4 (NDRG family member 4) have been used as the major markers in the ColoGuard™ stool DNA testing recently approved by Food and Drug Administration (FDA) for

***Corresponding authors:** Hongzhi Zou and Xia Zhao, Department of Laboratory Medicine and Pathology, Mayo Clinic, 200 First Street SW, Rochester, Minnesota 55905, USA, Tel: +01 507-293-2464, +01 507-538-7547; E-mail: zou.hongzhi@mayo.edu, zhao.xia@mayo.edu

Received August 06, 2015; **Accepted** October 02, 2015; **Published** October 04, 2015

Citation: Zhao X, Jen J, Peikert T, Edell E, Tian S, et al. (2015) Selection of Sensitive Methylation Markers for the Detection of Non-small Cell Lung Cancer. J Mol Biomark Diagn 6: 250. doi:10.4172/2155-9929.1000250

colorectal cancer screening [13,14]. Previous studies have demonstrated that methylation markers are of diagnostic value when measured in bodily fluids from patients with lung cancer, including sputum, plasma/serum, and bronchial lavage fluid [15-20]. Gene methylation is associated with elevated cancer risk and can be detected in sputum up to three years before diagnosis of lung cancer [21]. These studies proved that methylate genes could serve as valuable biomarkers to improve lung cancer detection, however the sensitivity and specificity of previously reported methylation markers were suboptimal.

In this study, we selected discriminant methylation markers from 152 candidate genes reported to be methylated in various cancers using lung cancer cell lines and tissues, and further validated them in a large cohort of lung tissues. These candidate genes included biomarkers, transcriptional factors, tumor suppressors, as well as genes with unspecific functions. They represented the majority of reported methylation genes. We identified a group of highly sensitive and specific potential methylation markers to be further investigated with the goal to develop non-invasive tests to improve early diagnosis of NSCLC.

Materials and Methods

Approval of this study was obtained from the Institutional Review Board of Mayo Foundation.

Lung cancer cell lines

Four human non-small cell lung cancer cell lines from different origins, including NCI-H358 (lung/bronchiole epithelial cancer), NCI-H460 (large cell carcinoma), NCI-H520 (squamous cell carcinoma), and NCI-H1299 (non-small cell lung carcinoma), were used for the primary selection of methylated markers. Cell lines were authenticated (Genetica DNA Laboratories, Burlington, NC).

Study subjects

A total of 318 lung tissues collected from 195 NSCLC patients were examined and split into two sets. The pilot set of 26 samples was used for the secondary selection of methylation markers, including 9 pairs of NSCLC tumors/ tumor-free surrounding lung tissues (TFSLTs), 4 unpaired NSCLC tumor tissues, and 4 unpaired TFSLTs. The validation set of 292 samples was used to validate the top methylation markers from the secondary selection, including 114 NSCLC tumor/TFSLT pairs, 38 unpaired NSCLC tumors, and 28 unpaired TFSLTs. All tissues were optimal cutting temperature (OCT) compound-embedded fresh frozen samples. Cancer stages were determined according to the TNM American Joint Committee on Cancer (AJCC) staging system (the 7th edition). Detailed demographic and clinical characteristics were only available for patient subjects in the validation set (Table 1).

Characteristic	Matched NSCLC[a]/ TFSLT[b] (n=114)	NSCLC only (n=36)	TFSLT only (n=28)
Age-yr Median (range)	70 (32-88)	71 (37-86)	67.5 (28-83)
Sex-no. (%)			
Male	68 (59.6)	16 (44.4)	15 (53.6)
Female	46 (40.4)	20 (55.6)	13 (46.4)
Stage-no. (%)			
I/II	85 (74.6)	22 (61.1)	23 (82.1)
III/IV	29 (25.4)	14 (38.9)	5 (17.9)
Histologic characteristics-no. (%)			
Adenocarcinoma	35 (30.7)	17 (47.2)	14 (50.0)
Squamous-cell	24 (21.1)	10 (27.8)	7 (25.0)
Large-cell	55 (48.2)	9 (25.0)	7 (25.0)
Tumor size-no. (%)			
<3 cm	53 (46.5)	18 (50.0)	17 (60.7)
≥ 3 cm	61 (53.5)	18 (50.0)	11 (39.3)
Smoking status-no. (%)			
Never	7 (6.1)	2 (5.6)	4 (14.3)
Light (< 20 pky[c])	13 (11.4)	2 (5.6)	2 (7.1)
Heavy (≥ 20 pky)	92 (80.7)	32 (88.9)	22 (78.6)
Unknown	2 (1.8)	0 (0.0)	0 (0.0)

[a]non-small cell lung cancer; [b]tumor free surrounding lung tissue; [c]pack-year (a pack of cigarettes a day for a year), a unit for measuring the amount a person has smoked over a long period of time.

Table 1: Clinic characteristics of subjects.

DNA extraction and bisulfite treatment

Genomic DNA from cultured cell line was extracted with DNAzol (Invitrogen, Grand Island, NY). Tissue sections were examined by an experienced pathologist who circled out histologically distinct lesions to direct careful microdissection. Genomic DNA from tissue sections was extracted using QIAamp DNA Mini Kit (Qiagen, Valencia, CA). DNA (~500 ng) was bisulfite treated using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA) and eluted into 30 µl of elution buffer.

Primary selection of methylated markers with lung cancer cell lines

A total of 152 candidate genes (Supplemental Table 1) reported to be methylated in various cancers were evaluated in the four lung cancer cell lines, including NCI-H358, NCI-H460, NCI-H520, and NCI-H1299, with the conventional methylation-specific PCR (MSP). Genes methylated in at least three of the four lung cancer cell lines were selected for further testing in tissue samples. Methylated primers for each candidate gene were either from literature, or designed with at least three CpGs on each primer to discriminate methylated DNA sequences from unmethylated ones. Primer sequences and annealing temperatures were listed in Supplemental Table 1. One microliter of bisulfite-modified DNA was amplified in a total volume of 20 µl containing 1x iQ SYBR® Green Supermix (Bio-Rad, Hercules, CA) and 250 nM of each primer. The amplification process included hot start at 95°C for 3 min, denaturing at 95°C for 30 sec, annealing at optimized temperatures for 30 sec, and extension at 72°C for 30 sec for 35 cycles, and a final 5 min extension step at 72°C. Bisulfite-treated human genomic DNA and CpGenome Universal Methylated DNA (EMD Millipore, Billerica, MA) were used as positive controls for unmethylation and methylation, respectively. MSP products were verified by 2% agarose gel electrophoresis.

Secondary selection of tumor-specific methylated markers with lung tissues

Methylated genes selected using cancer cell lines were further assessed in the pilot set of lung tissues, including 26 NSCLC tumors and TFSLTs, with fluorescence-based quantitative real-time methylation-specific PCR (qMSP). Quantification of a region without CpG sites in β-actin (ACTB) was considered as a reference of bisulfite treatment and DNA input. qMSP reaction was run in a 20 µl volume containing 10 µl of 2x LightCycler 480 SYBR Green Master (Roche, Indianapolis, IN), 0.125 µM each primer, and 1 µl of bisulfite-treated DNA as template. Bisulfite-treated CpGenome Universal Methylated DNA (EMD Millipore) was used as a positive control and 5-fold serially diluted to create standard curves for all plates. DNA copy numbers of the standards were approximately calculated according to the equation published by Applied Biosystems (http://www6.appliedbiosystems.com/support/tutorials/pdf/quant_pcr.pdf):

Gene copy number=(gDNA (pg))/(3.3 pg (the approximated human genome mass, haploid))

The copy number of methylated gene in each sample was derived through the standard curve. A simplified percent of methylated reference (sPMR) can be defined to normalize gene methylation level. It was represented as the ratio of the copy number of target gene to that of β-actin multiplied by 100 [13]. A cutoff sPMR value was set to calculate sensitivity for each gene at a specificity of 100% in the pilot set of tissue samples. Genes methylated in more than 80% (11/13) of the tumors, but not in any of the TFSLTs (0/13), or compensatory to the pre-selected top methylation markers were chosen for further study.

Validation of tumor-specific methylated markers

Five methylated genes, including DMRT-like family A2 (DMRTA2), homeobox A9 (HOXA9), zic family member 4 (ZIC4), homeobox A7 (HOXA7), and SIX homeobox 3 (SIX3) were tested in the validation set of lung tissues, including 150 NSCLC tumors and 142 TFSLTs. Primers and probes were re-designed to target the bisulfite-modified methylated sequences in gene promoter regions (Supplemental Table 2 and Supplemental Figure 1). Optimized qMSP reaction was done in a volume of 25 μl consisting of 0.5 μM of each primer, 0.2 μM of probe, 5 mM of $MgCl_2$, 0.4 mM of each of deoxynucleotide triphosphate, 5 μl of 5x GoTaq buffer, and 2.5 units of GoTaq R Hot Start Polymerase (Promega, Madison, WI). One microliter of bisulfite- treated DNA was used as template. Bisulfite-converted CpGenome Universal Methylated DNA (EMD Millipore) was used as positive control and 5-fold serially diluted to create standard curves for all plates. Quantification of each marker was denoted by sPMR values and illustrated in boxplots to show different methylation levels in tumors and TFSLTs.

Gene symbol	Specificity (%)	Sensitivity (%)	AUC*
DMRTA2	95	93	0.967
HOXA9	95	91	0.955
ZIC4	95	89	0.950
HOXA7	95	81	0.904
SIX3	95	59	0.819
DMRTA2+HOXA9	95	95	0.971
All five together	95	95	0.971

*AUC: area under the curve value calculated through receiver operation curve analysis (ROC)

Table 2: Performance of the five markers in the validation set

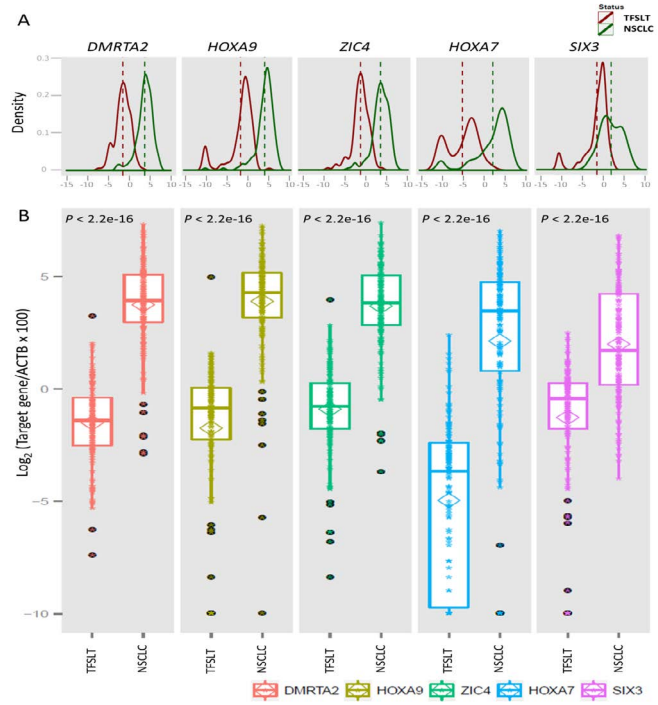

Figure 1: Methylation levels of DMRTA2, HOXA9, ZIC4, HOXA7, and SIX3 genes in 150 NSCLC tumors and 142 TFSLTs. (A) Distribution of log transformed methylation levels in NSCLC tumors and TFSLTs. (B) Boxplots showing the difference of methylation levels between NSCLC tumors and TFSLTs. Methylation levels were log transformed.

Statistical analysis

Wilcoxon rank sum test or signed-rank tests was performed to compare methylation levels between tissue groups. Data distribution was plotted for each of the five methylation markers after log transformation to reduce skewness in the validation set. Generalized linear model was applied to evaluate the correlation of methylation levels with demographic and clinical characteristics, such as patient gender, patient age, smoking history, cancer cell type, cancer stage, tumor size, and tumor location. Clinical characteristics correlated to methylation levels were further tested by unpaired two sample t-test within subgroups. Receiver operation curve (ROC) was constructed to compare gene methylation levels in cancers versus normal controls for each of the five markers and their combinations. The associated area under the curve (AUC) value for each curve was calculated. Comparisons of different ROC curves were performed in order to select the most discriminant markers for use in the future. To evaluate the prevalence of CpG island methylator phenotype (CIMP), the indicator for comethylation, methylation levels of multiple markers in the validation set were dichotomized. The dichotomization threshold was chosen at a methylation level of sPMR =8 to obtain a high discriminant rate by removing background noise. Fisher's exact test or Chi-square test was performed to analyze the association of gene co-methylation frequencies with clinical characteristics in of tumor patients. Statistical analysis was conducted with R package (version 3.0.3).

Results

Selection of methylation markers with lung cancer cell lines and tissue samples

Methylation status of 152 candidate genes was assessed in four lung cancer cell lines with MSP (Supplemental Table 3). DNA methylation was detected in all four cell lines, for 23 genes, in three cell lines for 10 genes, in two cell lines for 30 genes, in one cell line for 39 genes, and in none of cell lines for 50 genes, respectively. A total of 33 genes methylated in at least three of the four cell lines were selected for further testing in tissue samples. We excluded genes methylated in less than three of four cell lines for further study because of their likely low detection rates for NSCLC patients.

Methylation levels of the 33 genes were further quantified in the pilot set of 26 tissue samples with qMSP. At a specificity cutoff of 100%, 20 and 13 genes were respectively methylated in more than 30% (4/13) and 50% (7/13) of the cancers (Supplemental Table 4). Four genes, DMRTA2, HOXA9, ZIC4, and HOXA7, were methylated in more than 80% (11/13) of the tumors, but not in any of the TFSLTs (0/13). At a specificity of 100%, methylation of DMRTA2, HOXA9, ZIC4, and HOXA7 was detected in 12,12,13,11 of 13 NSCLCs, respectively. Gene SIX3 was methylated in 61.5% of the tumors, but was compensatory to methylated HOXA9 for tumor detection in this set of tissues: one

			CIMP (≥ 2 loci)			CIMP (≥ 3 loci)			CIMP (≥ 4 loci)			CIMP (= 5 loci)		
Tissue	**Characteristics**	**Category**	**+**	**-**	***P***	**+**	**-**	***P***	**+**	**-**	***P***	**+**	**-**	***P***
NSCLC	Total		123	27		109	41		77	73		36	114	
	Gender	Male (n=84)	69	15		66	18		43	41		18	66	
		Female (n=66)	54	12	1.00	43	23	0.10	34	32	1.00	18	48	0.41
	Age	< 60 y (n=19)	11	8		9	10		9	10		3	16	
		≥60 y (n=131)	112	19	0.008**	100	31	0.01*	68	63	0.81	33	98	0.54
	Histological type	Adenocarcinoma (n=52)	40	12		33	19		19	33		14	38	
		Squamous cell (n=65)	55	10		50	15		35	30		14	51	
		Large cell (n=33)	28	5	0.50	26	7	0.20	23	10	0.01*	8	25	0.79
	Tumor size	< 3 cm (n=71)	56	15		52	19		34	37		14	57	
		≥3 cm (n=79)	67	12	0.40	57	22	1.09	43	36	0.51	22	57	0.33
	Stage	I/II (n =107)	87	20		75	32		53	54		25	82	
		III/IV (n=43)	36	7	0.82	34	9	0.36	24	19	0.59	11	32	0.83
	Smoking history	Never (n=9)	8	1		7	2		5	4		3	6	
		Light (< 20 pky, n=15)	15	0		13	2		11	4		6	9	
		Heavy (≥20 pky, n=124)	98	26		87	37		60	64		27	97	
		Unknown (n=2)	2	0	0.18	2	0	0.54	1	1	0.26	0	2	0.32
	Tumor location	Left lobe (n=65)	55	10		48	17		36	29		13	52	
		Right lobe (n=85)	68	17	0.53	61	24	0.85	41	44	0.41	23	62	0.34
TFSLT	Total		1	141		1	141		0	142		0	142	
	Gender	Male (n=83)	1	84		1	84		0	85		0	85	
		Female (n=59)	0	59	1.00	0	59	1.00	0	59	1.00	0	59	1.00
	Age	<60 y (n=21)	0	21		0	21		0	21		0	21	
		≥60 y (n=121)	1	120	1.00	1	120	1.00	0	121	1.00	0	121	1.00

*P<0.05; ** P<0.01; a CpG island methylator phenotype.

Table 3: The association analyses of clinical characteristics with gene co-methylation status (CIMPa).

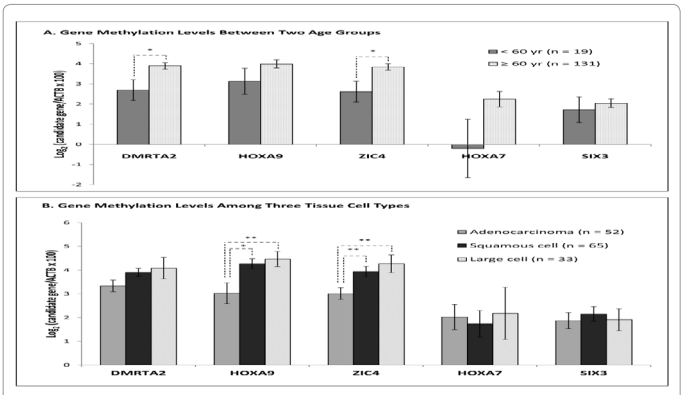

Figure 2: Association of methylation levels with patient age and cancer cell type. (A) Histogram showing methylation levels in two age groups. (B) Histogram showing methylation levels in three subtypes of NSCLCs.

NSCLC tumor sample could not be detected by HOXA9 but by SIX3. Therefore, DMRTA2, HOXA9, ZIC4, HOXA7, and SIX3 were selected for further assessment with the validation set of tissue samples.

Performance of methylation markers DMRTA2, HOXA9, ZIC4, HOXA7, and SIX3 in NSCLCs

The five tumor-specific methylated genes, including DMRTA2, HOXA9, ZIC4, HOXA7, and SIX3, were tested in the validation set of 292 tissue samples. Median methylation levels (sPMRs) in TFSLTs and lung tumor tissues were respectively 0.38 (95% CI: 0.29–3.44) and 15.25 (95% CI: 0.42–114.5) for DMRTA2, 0.56 (95% CI: 0–2.79) and 19.54 (95% CI: 0.37–102.89) for HOXA9, 0.59 (95% CI: 0.02–4.19) and 14.27 (95% CI: 0.25–86.94) for ZIC4, 0.08 (95% CI: 0–2.02) and 11.13 (95% CI: 0–101.11) for HOXA7, and 0.75 (95% CI: 0–3.48) and 3.3 (95% CI: 0.15–88.34) for SIX3.. The distribution of log- transformed methylation levels of DMRTA2, HOXA9, and ZIC4 in NSCLC tumors showed clear separations between two status groups in Figure 1A. Methylation levels were significantly higher in NSCLCs than in TFSLTs for each of the five genes (P<2.2e-16 for each gene, Figure 1B). When only the 114 paired samples were included into data analysis, the methylation levels were still significantly higher in NSCLC tumors than in TFSLTs for each of the all five genes (P =8.9e-16 for SIX3 and P<2.2e-16 for the other 4 genes, Supplemental Figure 2). Methylation was not associated with gender, cancer stage, smoking history, tumor size, and tumor location for each of the five genes (P>0.05 for each). However, methylation level of NSCLCs correlated with older age for DMRTA2 and ZIC4 (P<0.05 for each, Figure 2A). Methylation levels of HOXA9 and ZIC4 were significantly higher in large and squamous cell carcinomas than in adenocarcinomas (P<0.05 for HOXA9 and P<0.01 for ZIC4, Figure 2B).

Receiver operating curves were constructed for each tumor-specific marker and their most discriminant combinations. At a specificity of 95%, methylation of DMRTA2, HOXA9, ZIC4, HOXA7, and SIX3 was detected in 93% (140/150), 91% (136/150), 89% (133/150), 81% (122/150) and 59% (89/150) of NSCLC tissues. Comparing 150 tumors to 142 TFSLTs, AUC values were 0.967, 0.955, 0.950, 0.904 and 0.819 for DMRTA2, HOXA9, ZIC4, HOXA7 and SIX3, respectively. Combining all five markers, the predicted AUC value was 0.971 (Figure 3A and Table 2). At a specificity of 95%, 95% (142/150) of NSCLC tissues could be detected by these five markers. Remarkably, the performance of DMRTA2 or HOXA9 could not be significantly improved by adding anyone of the other three markers including ZIC4, HOXA7, or SIX3 (P>0.05). The predicted AUC value of the combination of DMRTA2 and HOXA9 was also 0.971 (Figure 3B and Table 2), which is similar to that of the combination of all five markers.

Comethylation in NSCLCs

Genes DMRTA2, HOXA9, ZIC4, HOXA7, and SIX3 were

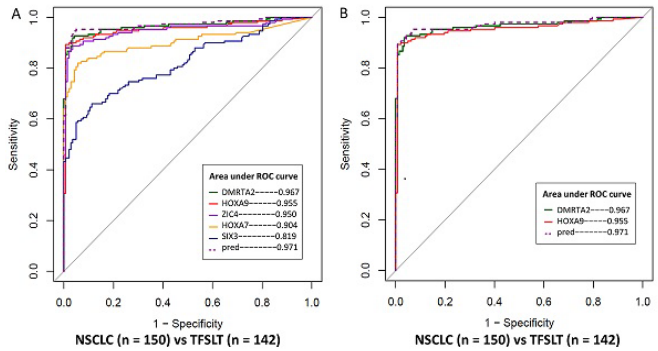

Figure 3: ROC curves for gene methylation levels in NSCLC tumors versus TFSLTs. (A) ROC curves for methylation levels of five genes in NSCLC tumors versus TFSLTs. AUC values were 0.967, 0.955, 0.950, 0.904 and 0.819 for single marker DMRTA2, HOXA9, ZIC4, HOXA7 and SIX3, respectively. AUC was 0.971 for the predicted combination of five methylated markers. (B) ROC curves of methylated markers DMRTA2 and HOXA9, as well as their combinations; in NSCLCs versus TFSLTs. AUC was 0.971 for the predicted combination of DMRTA2 and HOXA9.

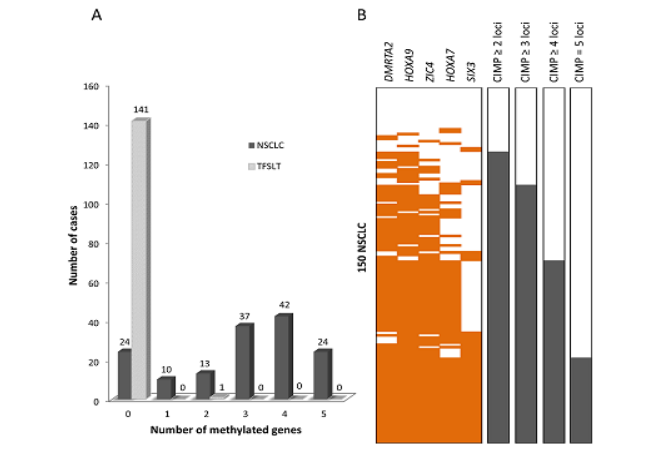

Figure 4: Comethylation of the five tumor-specific methylated genes and their distributions in NSCLCs. (A) Histogram showing numbers of cases with different numbers of co-methylated genes in 150 tumors and 142 TFSLTs. (B) Dichotomous heat map demonstrating CIMP phenomenon in NSCLCs using different numbers of methylation markers. On the heat map, orange bars indicate NSCLC samples with sPMR ≥ 8, and white bars indicate NSCLC samples with sPMR < 8. To the right of the heat map, grey bars indicate "CIMP +" samples, white bars indicate "CIMP –" samples.

commonly comethylated in NSCLCs, and the subset of subjects showing comethylation shared certain characteristics. Methylation levels in the validation set were dichotomized to simplify panel assembly and to allow easier translation of quantitative to qualitative panels [22]. The dichotomization thresholds of methylation levels were chosen at a point (sPMR =8) sufficiently above background levels but well below the much higher levels in NSCLCs. Methylation of at least two markers (≥2 loci), at least three markers (≥3 loci), at least four markers (≥4 loci), and five out of five markers (=5 loci) was detected in 116 (77%), 103 (69%), 66 (44%) and 24 (16%) of 150 NSCLC tissues (Figure 4A and 4B). Comethylation was only detected in 0.7% (≥2 loci) and 0% (≥3 or ≥4 or =5 loci) of TFSLTs (Figure 4A). This again confirms that comethylation is much more common in tumors compared to TFSLTs. Statistical analyses showed that CIMPs ≥2 loci and ≥3 loci were significantly associated with older patient age (P<0.01 and P<0.05, Table 3). Comethylation of CIMP (≥4 loci) correlated with cancer cell subtypes (P<0.05, Table 3). No association was observed for CIMPs with other clinical characteristics. Notably, comethylation was not associated with age in TFSLTs. Therefore, none of the gene markers in CIMPs was excluded due to non-specificity (Table 3).

Discussion

DNA methylation patterns are tissue-specific and maintained during cell divisions. However, they can be reprogrammed in cancer through de novo methylation which may take place in an instructive manner by interaction between cis-acting sequences on the DNA and trans-acting protein complexes capable of recruiting DNA methyltransferases [12]. Comparing methylation status of genes between paired tumor and tumor-free surrounding lung tissues could facilitate marker identification by uncovering de novo methylation loci. Alternated DNA methylation patterns, especially hypermethylation of CpG islands nearby gene promoter regions, are frequent events in cancer development and may provide sensitive markers for cancer detection [23].

Methylation status of CpG islands in a large number of genes has been assayed in tissue, plasma/serum, and sputum from lung cancer patients [18-20,24]. In this study, methylation-specific primers were designed to amplify CpG islands in gene promoter regions to find genes specifically methylated within the lung tumors rather than TFSLTs. A total of 152 candidate genes reported to be methylated in various cancers were evaluated first with lung cancer cell lines and then with a pilot set of lung tissues. Cell lines are widely used experimental tool to understand the behavior of primary tumors. However, limitations exist in applying this tool, such as cross-contamination, representativeness of cell lines, as well as the influence of cell culture environment on cell properties [25]. To minimize the impact of these limitations on marker identification using lung cancer cell lines, we authenticated the cell lines to confirm their purity and picked cell lines from different origins to allow broad coverage of NSCLCs. Moreover, methylation markers primarily selected with cell lines were secondarily selected with a pilot set of tissue samples. Five discriminant methylated genes, DMRTA2, HOXA9, ZIC4, HOXA7, and SIX3, were extensively analyzed in a large validation set of tissue samples. Three single genes, DMRTA2, HOXA9, and ZIC4, were found to be methylated in about 90% of NSCLC tumors, but rarely in adjacent tumor-free lung tissues. HOXA7 and SIX3 were also specifically methylated in the majority of NSCLC tumors. Therefore, each of them would serve as a potentially valuable marker for the detection of lung cancer. Although methylation of these genes was previously reported [26-28], their methylation profiles and associations with clinical characteristics in NSCLC were not well characterized. Methylation of DMRTA2 and ZIC4 was identified in lung cancer by microarray technology [27], but their potential value as diagnostic markers was never explored. Methylated HOXA9 was suggested as an early biomarker for lung cancer detection by other investigators [28], which is consistent with our findings in this study. Of note, some well-known methylation genes in lung cancer, such as Ras association (RalGDS/AF-6) domain family member 1 (RASSF1A), deleted in lung and esophageal cancer 1 (DLEC1), paired box 5 (PAX5α), and retinoic acid receptor, beta (RARβ) [11], did not show sufficient discriminant value in the present study, which might be due to different study design, sample size, or ethnic background [29].

Comethylation (CIMP) was first proposed as a distinct epigenotype for colorectal cancer [30], and was also observed in lung cancer [31,32]. However, the definition of CIMP was inconsistent due to different marker panels used [33]. For example, CIMP+ with four

or more genes methylated in a panel of nine genes, including APC, CDH13, KLK10, DLEC1, RASSF1A, EFEMP1, SFRP1, RARβ, and p16, was observed in 65.38% of lung tumor tissues [34]. CIMP+ with at least three synchronously genes methylated in a group of six genes, including hOGG1, RARβ, SEMA3B, RASSF1A, BLU, and FHIT, on chromosome 3p was detected in 43.8% of peripheral blood mononuclear cell specimens from patients with NSCLC [32]. In our study, the five methylation markers evaluated were frequently co-methylated in same NSCLC tumors, and were associated with distinct clinical features. Methylation of at least two markers (≥2 loci), at least three markers (≥3 loci), at least four markers (≥4 loci), and five out of five markers (=5 loci) was respectively detected in 77%, 69%, 44% and 24% of 150 NSCLCs. Since CIMP- was observed in a subset of NSCLCs, the improvement of marker sensitivity by adding more methylation markers was minimal [22]. Therefore, implementing markers that are positive in CIMP- samples, such as genetic alterations [35], protein [36], or imaging biomarkers, will be necessary to provide a completely informative marker panel for the detection of NSCLC tumors.

As our ultimate goal is to develop DNA tests for the detection of NSCLCs utilizing non/mimimally invasively obtained biospecimes, such as sputum, bronchial aspirate, and blood, an informative panel of methylated markers will be essential. Certain clinical characteristics may impact the performance of methylation markers. In this study, except for age and cancer cell type, the performance of methylated DMRTA2, HOXA9, and ZIC4 for the detection of NSCLC was not affected by key clinical characteristics, such as gender, smoking history, tumor size and location, and disease stage. Therefore, because of their broad spectrum of coverage and early onset, methylated DMRTA2, HOXA9, and ZIC4 are valuable markers for the early detection of NSCLC.

Despite their frequent methylation in NSCLCs, the carcinogenic roles of these genes are not well understood. Since inactivation of tumor suppressor genes by aberrant promoter methylation play a potential role during tumorigenesis [37], it is possible that these genes function as tumor suppressors. For example, polycomb complexes were proposed to contribute to silencing of homeobox genes, and then initiates tumorigenesis [11,38]. The carcinogenic roles of these genes could be related to their mechanistic importance in embryo development. All of these five genes can act as transcription factors to affect early embryo development. For instances, Dmrta2 in mice is essential in the early development of the telencephalon via the formation of the cortical hem and maintaining of neural progenitors [39]; homeobox genes HOXA9 and HOXA7 spatially and temporally regulate morphogenesis and differentiation during embryonic development [40]; homeobox gene SIX3 provides necessary instructions for the formation of the forebrain and eye development [41]; ZIC4 has been proposed to regulate late-emerging characteristics in the dorsal surface with ZIC1 [42]. Due to their important roles in embryo development, methylation of these genes could affect important cellular functions and eventually promote tumorigenesis. Therefore, Further investigation to characterize their mechanisms of action is needed.

In conclusion, we have identified three highly tumor-specific methylated genes, DMRTA2, HOXA9, and ZIC4, in NSCLC tumors compared to TFSLTs. Our findings will help the development of non-invasive molecular diagnostic tests for the early detection of lung cancer.

Acknowledgement

This study was supported by funding provided by Center for Individualized Medicine at the Mayo Clinic (Hongzhi Zou, Jin Jen), K23CA159391 (Tobias Peikert).

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, et al. (2015) Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer136: E359-386.
2. Siegel RL, Miller KD, Jemal A (2015) Cancer statistics, 2015. CA Cancer J Clin 65: 5-29.
3. Virmani AK, Tsou JA, Siegmund KD, Shen LY, Long TI, et al. (2002) Hierarchical clustering of lung cancer cell lines using DNA methylation markers. Cancer Epidemiol Biomarkers Prev 11: 291-297.
4. Roth JA, Billings P, Ramsey SD, Dumanois R, Carlson JJ (2014) Cost-effectiveness of a 14-gene risk score assay to target adjuvant chemotherapy in early stage non-squamous non-small cell lung cancer. Oncologist 19: 466-476.
5. Khakwani A, Rich AL, Tata LJ, Powell HA, Stanley RA, et al. (2014) Small-cell lung cancer in England: trends in survival and chemotherapy using the National Lung Cancer Audit. PLoS One 9:e89426.
6. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 365: 395-409.
7. Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, et al. (2011) The National Lung Screening Trial: overview and study design. Radiology 258: 243-253.
8. Greenberg AK, Lu F, Goldberg JD, Eylers E, Tsay JC, et al. (2012) CT scan screening for lung cancer: risk factors for nodules and malignancy in a high-risk urban cohort. PLoS One. 7: e39403.
9. Wang Y, Midthun DE, Wampfler JA, Deng B, Stoddard SM, et al. (2015) Trends in the proportion of patients with lung cancer meeting screening criteria. JAMA 313: 853-855.
10. Schubeler D (2015) Function and information content of DNA methylation. Nature 517: 321-326.
11. Rauch T, Wang Z, Zhang X, Zhong X, Wu X, et al. (2007) Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. Proc Natl Acad Sci U S A 104: 5527-5532.
12. Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, et al. (2006) Evidence for an instructive mechanism of de novo methylation in cancer cells. Nat Genet 38: 149-153.
13. Zou H, Harrington JJ, Shire AM, Rego RL, Wang L, et al. (2007) Highly methylated genes in colorectal neoplasia: implications for screening. Cancer Epidemiol Biomarkers Prev 16: 2686-2696.
14. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, et al. (2014) Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med 370: 1287-1297.
15. Leng S, Liu Y, Weissfeld JL, Thomas CL, Han Y, et al. (2015) 15q12 Variants, Sputum Gene Promoter Hypermethylation, and Lung Cancer Risk: A GWAS in Smokers. J Natl Cancer Inst 107.
16. Ponomaryova AA, Rykova EY, Cherdyntseva NV, Skvortsova TE, Dobrodeev AY, et al. (2011) RARbeta2 gene methylation level in the circulating DNA from blood of patients with lung cancer. Eur J Cancer Prev 20: 453-455.
17. Topaloglu O, Hoque MO, Tokumaru Y, Lee J, Ratovitski E, et al. (2004) Detection of promoter hypermethylation of multiple genes in the tumor and bronchoalveolar lavage of patients with lung cancer. Clin Cancer Res 10: 2284-2288.
18. Belinsky SA, Grimes MJ, Casas E, Stidley CA, Franklin WA, et al. (2007) Predicting gene promoter methylation in non-small-cell lung cancer by evaluating sputum and serum. Br J Cancer 96: 1278-1283.
19. Belinsky SA, Klinge DM, Dekker JD, Smith MW, Bocklage TJ, et al. (2005) Gene promoter methylation in plasma and sputum increases with lung cancer risk. Clin Cancer Res 11: 6505-6511.
20. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, et al. (2002) Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer Res 62: 2370-2377.
21. Palmisano WA, Divine KK, Saccomanno G, Gilliland FD, Baylin SB, et al. (2000) Predicting lung cancer by detecting aberrant promoter methylation in sputum. Cancer Res 60: 5954-5958.
22. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, et al. (2006) CpG island methylator phenotype underlies sporadic microsatellite instability

and is tightly associated with BRAF mutation in colorectal cancer. Nat Genet 38: 787-793.

23. Esteller M (2002) CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. Oncogene 21: 5427-5440.

24. Tsou JA, Hagen JA, Carpenter CL, Laird-Offringa IA (2002) DNA methylation analysis: a powerful new tool for lung cancer diagnosis. Oncogene 21: 5450-5461.

25. Paz MF, Fraga MF, Avila S, Guo M, Pollan M, et al. (2003) A systematic profile of DNA methylation in human cancer cell lines. Cancer Res 63: 1114-1121.

26. Kalari S, Jung M, Kernstine KH, Takahashi T, Pfeifer GP (2013) The DNA methylation landscape of small cell lung cancer suggests a differentiation defect of neuroendocrine cells. Oncogene 32: 3559-3568.

27. Carvalho RH, Haberle V, Hou J, van Gent T, Thongjuea S, et al. (2012) Genome-wide DNA methylation profiling of non-small cell lung carcinomas. Epigenetics Chromatin 5: 9.

28. Hwang SH, Kim KU, Kim JE, Kim HH, Lee MK, et al. (2011) Detection of HOXA9 gene methylation in tumor tissues and induced sputum samples from primary lung cancer patients. Clin Chem Lab Med 49: 699-704.

29. Enokida H, Shiina H, Urakami S, Igawa M, Ogishima T, et al. (2005) Ethnic group-related differences in CpG hypermethylation of the GSTP1 gene promoter among African-American, Caucasian and Asian patients with prostate cancer. Int J Cancer 116: 174-181.

30. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, et al. (1999) CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A 96: 8681-8686.

31. Shinjo K, Okamoto Y, An B, Yokoyama T, Takeuchi I, et al. (2012) Integrated analysis of genetic and epigenetic alterations reveals CpG island methylator phenotype associated with distinct clinical characters of lung adenocarcinoma. Carcinogenesis 33: 1277-1285.

32. Liu Z, Li W, Lei Z, Zhao J, Chen XF, et al. (2010) CpG island methylator phenotype involving chromosome 3p confers an increased risk of non-small cell lung cancer. J Thorac Oncol 5: 790-797.

33. Issa JP (2004) CpG island methylator phenotype in cancer. Nat Rev Cancer 4: 988-993.

34. Zhang Y, Wang R, Song H, Huang G, Yi J, et al. (2011) Methylation of multiple genes as a candidate biomarker in non-small cell lung cancer. Cancer Lett 303: 21-28.

35. Lam DC, Luo SY, Deng W, Kwan J, Rodriguez-Canales J, et al. (2015) Oncogenic mutation profiling in new lung cancer and mesothelioma cell lines. Onco Targets Ther 8: 195-209.

36. Jakobsen KR, Paulsen BS, Baek R, Varming K, Sorensen BS, et al. (2015) Exosomal proteins as potential diagnostic markers in advanced non-small cell lung carcinoma. J Extracell Vesicles 4: 26659.

37. Momparler RL (2003) Cancer epigenetics. Oncogene 22: 6479-6483.

38. Kim SY, Paylor SW, Magnuson T, Schumacher A (2006) Juxtaposed Polycomb complexes co-regulate vertebral identity. Development 133: 4957-4968.

39. Konno D, Iwashita M, Satoh Y, Momiyama A, Abe T, et al. (2012) The mammalian DM domain transcription factor Dmrta2 is required for early embryonic development of the cerebral cortex. PLoS One 7: e46577.

40. Di Vinci A, Brigati C, Casciano I, Banelli B, Borzì L, et al. (2012) HOXA7, 9, and 10 are methylation targets associated with aggressive behavior in meningiomas. Transl Res 160: 355-362.

41. Lagutin OV, Zhu CC, Kobayashi D, Topczewski J, Shimamura K, et al. (2003) "Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development". Genes Dev 17: 368-379.

42. Kawanishi T, Kaneko T, Moriyama Y, Kinoshita M, Yokoi H, et al. (2013) Modular development of the teleost trunk along the dorsoventral axis and zic1/zic4 as selector genes in the dorsal module. Development 140: 1486-1496.