

UCSF

UC San Francisco Previously Published Works

Title

Key epigenetic changes associated with lung cancer development

Permalink

<https://escholarship.org/uc/item/08p692xh>

Journal

Epigenetics, 7(6)

ISSN

1559-2294

Authors

Nelson, Heather H

Marsit, Carmen J

Christensen, Brock C

et al.

Publication Date

2012-06-01

DOI

10.4161/epi.20219

Peer reviewed

# Key epigenetic changes associated with lung cancer development

## Results from dense methylation array profiling

Heather H. Nelson,<sup>1,\*</sup> Carmen J. Marsit,<sup>2</sup> Brock C. Christensen,<sup>2,3</sup> E. Andres Houseman,<sup>3</sup> Milica Kontic,<sup>4</sup> Joseph L. Wiemels,<sup>5</sup> Margaret R. Karagas,<sup>6</sup> Margaret R. Wrensch,<sup>7</sup> Shichun Zheng,<sup>7</sup> John K. Wiencke<sup>7</sup> and Karl T. Kelsey<sup>2,3</sup>

<sup>1</sup>Masonic Cancer Center; Division of Epidemiology and Community Health; University of Minnesota; Minneapolis, MN USA; <sup>2</sup>Department of Pathology and Laboratory Medicine; Brown University; Providence, RI USA; <sup>3</sup>Department of Community Health; Center for Environmental Health and Technology, Brown University; Providence, RI USA;

<sup>4</sup>Department of Oncology; Institute for Lung Diseases and TB; Belgrade University; Belgrade, Serbia; <sup>5</sup>Department of Epidemiology and Biostatistics; University of California at San Francisco; San Francisco, CA USA; <sup>6</sup>Section of Biostatistics and Epidemiology; Department of Community and Family Medicine; Dartmouth-Hitchcock Medical Center; Lebanon, NH USA <sup>7</sup>Department of Neurological Surgery; University of California at San Francisco; San Francisco, CA USA;

**Keywords:** lung cancer, DNA methylation, molecular epidemiology, goldengate, pyrosequencing

Epigenetic alterations are a common event in lung cancer and their identification can serve to inform on the carcinogenic process and provide clinically relevant biomarkers. Using paired tumor and non-tumor lung tissues from 146 individuals from three independent populations we sought to identify common changes in DNA methylation associated with the development of non-small cell lung cancer. Pathologically normal lung tissue taken at the time of cancer resection was matched to tumorous lung tissue and together were probed for methylation using Illumina GoldenGate arrays in the discovery set ( $n = 47$  pairs) followed by bisulfite pyrosequencing for validation sets ( $n = 99$  pairs). For each matched pair the change in methylation at each CpG was calculated (the odds ratio), and these ratios were averaged across individuals and ranked by magnitude to identify the CpGs with the greatest change in methylation associated with tumor development. We identified the top gene-loci representing an increase in methylation (*HOXA9*, 10.3-fold and *SOX1*, 5.9-fold) and decrease in methylation (*DDR1*, 8.1-fold). In replication testing sets, methylation was higher in tumors for *HOXA9* ( $p < 2.2 \times 10^{-16}$ ) and *SOX1* ( $p < 2.2 \times 10^{-16}$ ) and lower for *DDR1* ( $p < 2.2 \times 10^{-16}$ ). The magnitude and strength of these changes were consistent across squamous cell and adenocarcinoma tumors. Our data indicate that the identified genes consistently have altered methylation in lung tumors. Our identified genes should be included in translational studies that aim to develop screening for early disease detection.

### Introduction

Lung cancer remains a significant worldwide public health concern and non-small cell lung cancer (NSCLC) accounts for approximately 70% of lung cancer diagnoses. In the United States in 2012, it is estimated that over 160,000 deaths will be attributed to lung cancer, which represents almost 28% of all cancer-related deaths in the US.<sup>1</sup> The two main forms of NSCLC, adenocarcinoma and squamous cell carcinoma, are both highly linked to tobacco smoking exposure. While we know that smoking causes most lung cancer, it is important to remember that smoking cessation reduces, but does not eliminate, risk. It is therefore essential to continue to identify tumor-specific molecular alterations so that effective screening, chemoprevention and curative therapies may be developed. Alterations to the epigenetics of tumors, including DNA methylation, are particularly appealing as they represent a readily detectable and potentially reversible alteration in malignancy.

Along with genetic alterations, epigenetic alterations are recognized as causal in carcinogenesis. DNA methylation is a

mechanism of stable control of transcription: regulatory CpG clusters are common, often occur in tumor suppressor genes and are thought to remain largely unmethylated in non-cancerous cells. In tumors, the classic example is gene promoter-based hypermethylation of CpGs that is associated with aberrant, stable gene silencing. Approximately half of all human genes contain regulatory CpG islands.<sup>2,3</sup> However, compared with non-tumor cells, tumors may also exhibit losses of methylation at certain CpG loci, which may result in gene activation. Recently, the simultaneous resolution of hundreds of specific, phenotypically defined cancer-related CpG methylation marks has become technologically feasible, allowing for rapid, high-throughput epigenetic profiling of human tissue CpG methylation.<sup>4</sup>

Investigations of lung cancer dominated the early research in aberrant DNA methylation at tumor suppressor loci, particularly investigations of the *CDKN2A* and *RASSF1* genes.<sup>5-8</sup> Panels of candidate genes have been assembled and associated with clinical outcome,<sup>9,10</sup> indicating that molecular profiling of epigenetic alterations in lung cancer will have clinical utility, although

\*Correspondence to: Heather H. Nelson; Email: hhnelson@umn.edu  
Submitted: 12/05/11; Revised: 03/26/12; Accepted: 03/30/12  
<http://dx.doi.org/10.4161/epi.20219>

the translation of these markers to the clinic has been limited. Identifying additional lung cancer specific markers may provide improved utility and further define specific pathways altered in this disease, which can be used as pathways for new personalized and targeted treatment strategies. Therefore, it is critical to utilize high-throughput and genome-wide approaches to define key epigenetic alterations in NSCLC to foster translational studies aiming to develop screenings for early detection of disease and strategies for personalized medicine. Dense methylation profiling studies that include lung tissues are now emerging,<sup>4,11-16</sup> and we hypothesized that there are a limited number of defining epigenetic events that are common to NSCLC tumors. To study this, we used an array-based approach, comparing matched pairs of tumor and adjacent non-tumor tissue to identify crucial epigenetic alterations, and followed these discovery-based approaches with in-depth, quantitative validation of the novel markers in independent tumor series.

## Results

Discovery set tissue pairs ( $n = 47$  pairs) were measured for CpG methylation using the Illumina Goldengate array. Eight CpG loci did not pass QA and were removed, leaving 1,413 autosomal CpG loci associated with 773 cancer-related genes for analysis. Among the 47 NSCLC cases in the discovery set there were 22 adenocarcinomas and 25 squamous cell carcinomas.

CpG loci were ranked according to the magnitude of change in methylation. Given the bounded nature of the  $\beta$  value this was calculated as the Odds Ratio (OR), and is analogous to the fold change in methylation (increase or decrease in methylation in tumors relative to the methylation present in normal lung tissue). There were 107 CpG loci with a greater than 2-fold increase in methylation in tumors, and 43 loci with a greater than 2-fold decrease in methylation in tumors relative to the normal lung epigenome (Fig. 1 and Table 1). More specifically, among adenocarcinomas there were 128 CpG loci with greater than 2-fold increased methylation, and 57 loci with greater than 2-fold decreases in methylation. There were slightly fewer dramatic changes among squamous cell carcinomas with 95 CpG loci demonstrating greater than 2-fold increases in methylation and 40 loci with greater than 2-fold decreases in methylation. Within each histology, rank-ordered lists of ORs for methylation change were generated and there was marked similarity in the top 10 genes for adenocarcinoma and squamous cell carcinoma (Table 2). Among all 47 tissue pairs, there were several genes with CpG having at least a 5-fold increase in methylation, including *homeobox A9* (*HOXA9*, 10.3-fold), *T-cell acute lymphocytic leukemia 1* (*TAL1*, 7.9-fold), *5-hydroxytryptamine (serotonin) receptor 1B* (*HTR1B*, 6.2-fold), *SRY (sex determining region Y)-box 1* (*SOX1*, 5.9-fold), *v-mos Moloney murine sarcoma viral oncogene homolog* (*MOS*, 5.7-fold) and *homeobox A11* (*HOXA11*, 5.0-fold). The CpGs demonstrating the greatest decreases in methylation in tumor compared with non-tumor were associated with genes, including *protein tyrosine phosphatase, non-receptor type 6* (*PTPN6*, 5.3-fold), *nidogen 1* (*NID1*, 3.3-fold), *deleted in liver cancer 1* (*DLC1*, 2.9-fold), *discoidin domain receptor tyrosine kinase*

*1* (*DDR1*, 2.9-fold) and *nitric oxide synthase 3* (*NOS3*, 2.9-fold) (Table 1). Ingenuity Pathways Analysis of genes whose CpGs had at least 2-fold changes in methylation revealed that the top cellular networks for increased methylation were related to development and cell death, and the top cellular networks for decreased methylation were related to cell death and cancer (Table S1).

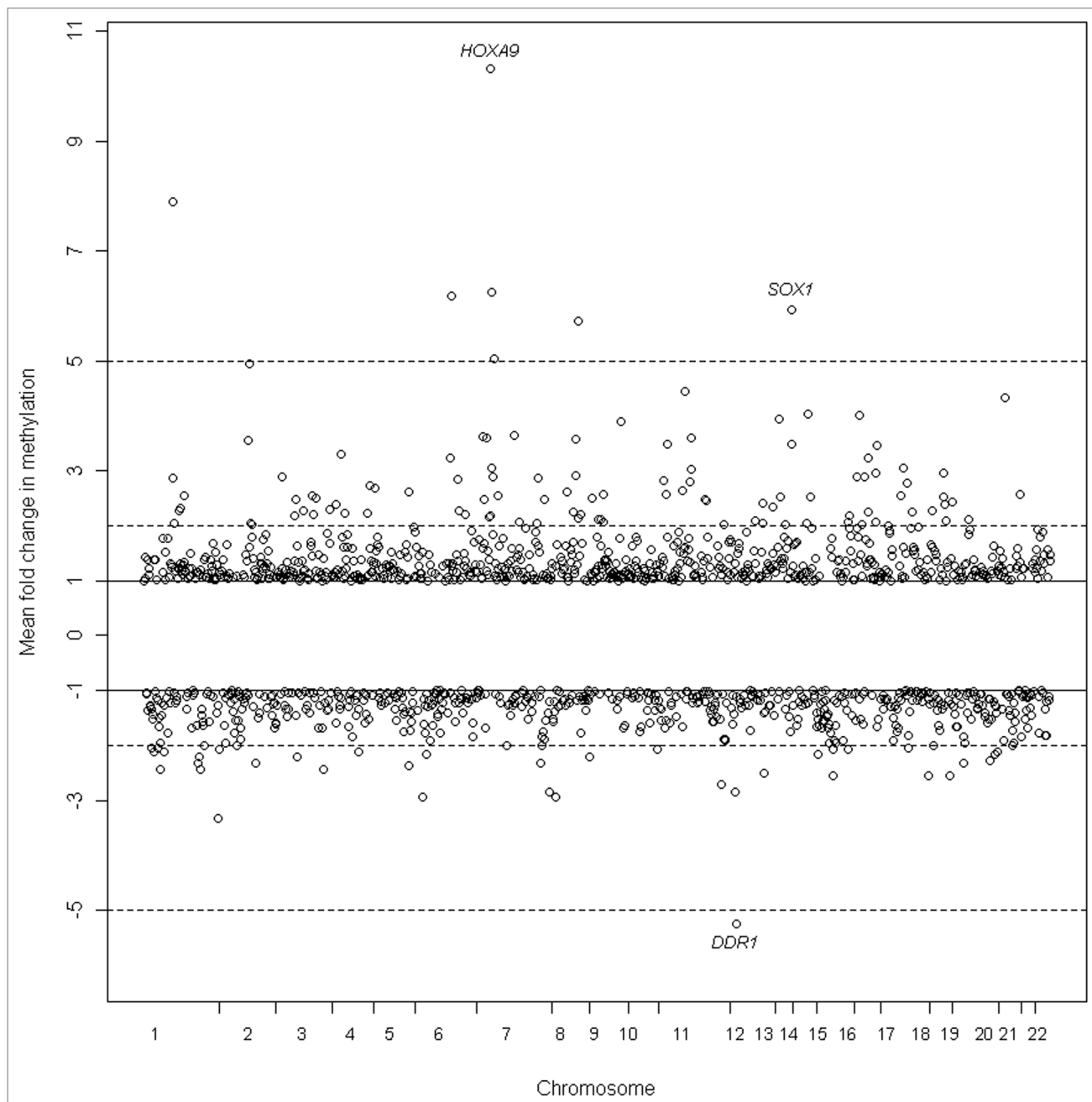
In the replication data sets there were significant differences in methylation comparing the matched normal and tumor tissues. In replication set 1, *HOXA9* had a 2.1-fold methylation increase in tumors, which was comparable to the 3.5-fold methylation increase observed for tumors in replication set 2 ( $p < 7.3 \times 10^{-9}$  for set 1 and  $p < 2.1 \times 10^{-9}$  for set 2, Table 3). Similarly, in tumors compared with matched non-tumor, *SOX1* had a 2.4-fold methylation increase in set 1 and a 3.1-fold methylation increase in replication set 2 ( $p < 5.2 \times 10^{-8}$  and  $p < 1.5 \times 10^{-10}$ , respectively, Table 3). Finally, there were consistent results for decreased methylation at *DDR1* in lung tumors relative to the normal lung genome. In replication set 1, there was a 1.4-fold decrease in methylation of *DDR1* ( $p < 2.6 \times 10^{-11}$ ) and, in replication set 2, there was a 1.4-fold decrease in *DDR1* methylation ( $p < 1.5 \times 10^{-11}$ , Table 3). For all three genes a majority of tumor specimens were abnormally methylated (outside 1 standard deviation of the normal tissue mean): 73% and 64% of tumors were hypermethylated at *SOX1* and *HOXA9*, respectively, and 73% of tumors were hypomethylated at *DDR1*.

Pooling the replication data and examining methylation changes according to histology revealed no significant differences. For squamous cell carcinoma there was a 3.5-fold increase in *HOXA9* methylation, a 3.1-fold increase in *SOX1* methylation, and a 1.5-fold decrease in *DDR1* methylation. For adenocarcinoma these changes were a 2.1-fold increase in *HOXA9*, a 2.3-fold increase in *SOX1* and a 1.4-fold decrease in *DDR1* methylation. There was a higher percentage of squamous cell carcinomas with *SOX1* hypermethylation than adenocarcinomas (83% and 65% of tumors). For *HOXA9*, hypermethylation was relatively constant for adenocarcinoma (68%) and squamous cell carcinoma (72%), whereas *DDR1* hypomethylation was slightly higher in squamous cell carcinoma (78%) than adenocarcinomas (73%). None of these differences by histology were statistically significant.

Finally, for two of our genes, one hypomethylated in tumors (*DDR1*) and one hypermethylated in tumors (*SOX1*), we measured gene expression using RT-PCR (Fig. S1). We observed a higher level of *DDR1* expression among tumors, consistent with the observed gene hypomethylation and decreased *SOX1* expression, again consistent with the hypermethylation observed in tumors.

## Discussion

We used methylation arrays to discover common epigenetic alterations that define NSCLC by comparing patient-matched tumor and non-tumor tissues in three independent populations. By describing tumor-specific epigenetic alterations in NSCLC common to squamous cell and adenocarcinomas we are fostering the development of novel detection and screening strategies. We identified several genes with CpGs that have altered methylation



**Figure 1.** Mean fold change in methylation between tumor and normal for 47 matched tissue pairs and all 1,413 autosomal CpGs by chromosome. Positive fold change values indicate increased methylation in tumor relative to normal tissue and negative fold change values indicate decreased methylation in tumor relative to normal tissue. Horizontal dotted lines indicate 2-fold and 5-fold changes in mean methylation between tumor and normal lung. Genes with CpG loci chosen for replication in independent populations (*HOXA9*, *SOX1*, *DDR1*) are shown.

in tumors compared with non-tumor tissue, and validated three genes; *HOXA9*, *SOX1* and *DDR1* in two replication sets from independent patient populations that have consistently and significantly altered DNA methylation in NSCLC compared with non-tumor lung tissue.

The homeobox (HOX) family of genes encodes transcription factors that are differentially expressed spatially and temporally during embryonic development, and *HOXA9* is part of the *HOXA* family of transcription factors on chromosome 7.

The homeobox genes have been described as being hypermethylated in lung cancer cell lines, and the *HOXA9* gene has been specifically shown to be hypermethylated in primary lung cancers.<sup>15,17</sup> Although we focused on *HOXA9* in our replication sets, among the nine *HOXA* gene CpGs (from *HOXA5*, *HOXA9* and *HOXA11*) measured on the array in the discovery pairs, seven (78%) demonstrated over 2-fold increases in methylation in tumors relative to non-tumors. The strong overrepresentation of *HOXA* gene CpGs with NSCLC-specific increases in methylation

**Table 1.** CpGs with greater than 2-fold increase or decrease in tumor methylation relative to non-tumor lung, discovery set (n = 47 pairs), from illumina golden gate methylation array

| Gene-loci with increased methylation in tumors |       |                    | Gene-loci with decreased methylation in tumors |       |                    |
|--|-------|--------------------|--|-------|--------------------|
| GENE   | CpG*  | OR** (fold change) | GENE   | CpG*  | OR** (fold change) |
| HOXA9  | E252  | 10.3               | PTPN6  | E171  | 5.3                |
| TAL1   | P594  | 7.9                | NID1   | P677  | 3.3                |
| HOXA9  | P1141 | 6.2                | DLC1   | P695  | 2.9                |
| HTR1B  | P222  | 6.2                | DDR1   | P332  | 2.9                |
| SOX1   | P1018 | 5.9                | NOS3   | P38   | 2.9                |
| MOS  | E60   | 5.7                | PTPN6  | P282  | 2.9                |
| HOXA11   | P698  | 5                  | TMPPRS4  | E83   | 2.7                |
| TPEF   | S88   | 5                  | AATK   | E63   | 2.6                |
| MYOD1  | E156  | 4.5                | GABRA5   | P862  | 2.6                |
| HCK  | P858  | 4.3                | SERPINB5                                       | P19   | 2.6                |
| DLK1   | E227  | 4                  | IFNG   | E293  | 2.5                |
| HS3ST2   | E145  | 4                  | SFN  | P248  | 2.4                |
| IPF1   | P750  | 4                  | MUC1   | P191  | 2.4                |
| DBC1   | P351  | 3.9                | TNFSF10  | P2    | 2.4                |
| MDR1   | S300  | 3.6                | ITK  | P114  | 2.4                |
| TWIST1   | E117  | 3.6                | CASP8  | E474  | 2.3                |
| WT1  | E32   | 3.6                | EMR3   | P39   | 2.3                |
| NPY  | P295  | 3.6                | CPA4   | E20   | 2.3                |
| SOX17  | P303  | 3.6                | S100A2   | P1186 | 2.3                |
| FRZB   | E186  | 3.6                | PTPRH  | E173  | 2.3                |
| SOX1   | P294  | 3.5                | ZMYND10  | P329  | 2.2                |
| ASCL2  | P360  | 3.5                | MUC1   | E18   | 2.2                |
| CDH13  | E102  | 3.5                | GML  | P281  | 2.2                |
| PDGFRA   | P1429 | 3.3                | MKRN3  | E144  | 2.2                |
| HTR1B  | E232  | 3.2                | USP29  | E274  | 2.2                |
| MT1A   | P49   | 3.2                | TNF  | P158  | 2.2                |
| RARA   | P1076 | 3.1                | ZIM3   | P718  | 2.1                |
| HOXA9  | P303  | 3                  | HDAC1  | P414  | 2.1                |
| WT1  | P853  | 3                  | PLA2G2A  | P528  | 2.1                |
| DCC  | P471  | 3                  | SPP1   | P647  | 2.1                |
| CDH13  | P88   | 3                  | GABRA5   | P1016 | 2.1                |
| SOX17  | P287  | 2.9                | IL16   | P93   | 2.1                |
| MYH11  | P22   | 2.9                | NID1   | P714  | 2.1                |
| MMP2   | P303  | 2.9                | CYP2E1   | P416  | 2.1                |
| RARB   | E114  | 2.9                | RUNX3  | E27   | 2                  |
| HOXA11   | E35   | 2.9                | BRCA1  | P835  | 2                  |
| IRF5   | E101  | 2.9                | NBL1   | P24   | 2                  |
| TAL1   | E122  | 2.9                | CLDN4  | P1120 | 2                  |
| EYA4   | P794  | 2.9                | PI3  | P1394 | 2                  |
| IGF2AS   | P203  | 2.8                | AIM2   | P624  | 2                  |

\*Location relative to transcription start site in bases, P = promoter, E = exonic. \*\*OR: [(tumor  $\beta$ /1-tumor  $\beta$ )/(normal tissue  $\beta$ /1-normal tissue  $\beta$ )].

suggests that, in general, these genes are excellent targets for the development of lung cancer biomarkers. Other *HOX* family gene members that have been reported to have increased methylation in NSCLC include *HOXC9* and *HOXA1*. Anglim et al. observed a significantly increased prevalence of *HOXC9* methylation in squamous cell lung tumors, though the increase was of moderate magnitude, 66% of tumors compared with 60% of adjacent lung tissue samples.<sup>11</sup> In adenocarcinomas, Tsou et al. reported significantly increased *HOXA1* methylation compared with non-tumor lung.<sup>16</sup> Additional findings that are consistent with ours are included in the original description of the GoldenGate methylation array.<sup>4</sup> In the original description of the array, Bibikova et al. measured methylation in two independent sets (n = 11 and n = 12 pairs) of lung adenocarcinomas and normal lung tissues, identified 55 CpG loci with both statistically significant and a large magnitude of increased methylation in tumors, and five of these CpGs were in *HOXA* family genes.<sup>4</sup>

Similar to the homeobox genes, *SOX1* encodes a transcription factor important in development. Although reports of *SOX1* methylation in NSCLC are lacking, it has been reported to have increased methylation in malignant ovarian tumors compared with benign disease.<sup>18</sup> Additional evidence for the potential utility of *SOX1* as a cancer biomarker comes from Apostolidou et al. who reported a significantly increased prevalence of *SOX1* methylation in high-grade squamous intraepithelial lesions compared with nonspecific cytological alterations in cervical cell suspension samples. Interestingly, in the same report, these authors reported significantly increased *HOXA11* methylation in high-grade lesions and that both *SOX1* and *HOXA11* discriminated high-grade intraepithelial lesions from controls with high sensitivity and specificity. Consistent with the observed increases in methylation of developmentally important *HOXA* family and *SOX1* transcription factors. Ingenuity cellular networks analysis revealed cellular development as a function common to the top three cellular networks with genes having at least 2-fold increases in methylation. In addition, cell death was a common function between both of the top two networks associated with 2-fold increases or decreases in gene methylation, suggesting that there is widespread epigenetic dysregulation of genes that participate in cell death processes.

Most classical investigations of tumor methylation have focused on increased methylation of tumor suppressor genes, though decreased gene-promoter methylation is also common in tumors and may be critically informative in biomarker development. For example, using the GoldenGate array, we have shown that the majority of significant methylation alterations (727 of 969, 75% with  $Q < 0.05$ ) between non-tumor pleura and mesotheliomas are instances of decreased methylation.<sup>20</sup> In the discovery set of lung tumor and non-tumor pairs described here, nearly 30% of CpGs with changes in methylation (of at least 2-fold) were decreases in methylation, including *DDR1*. *DDR1* encodes a receptor tyrosine kinase normally expressed in epithelial cells, including the lung, and it is overexpressed in lung tumors,<sup>21-23</sup> consistent with our finding of DNA methylation loss at this gene.

There is great interest in the development of lung cancer screening biomarkers that can identify early signs of disease,

including tissue-associated biomarkers that could augment differential diagnosis following spiral CT. Examples include identifying tumor-specific genetic or epigenetic alterations in sputum, lung lavage or cell-free DNA in serum. Serum-based detection of epigenetic alterations associated with tumors has been proposed as a viable strategy in lung cancer screening.<sup>24</sup> And, in fact, methylated circulating DNA in cancer patients has been previously associated with disease. For example, lung cancer cases have significantly more serum-derived DNA than controls ( $p < 0.0001$ ).<sup>25</sup> In non-small cell lung cancer patients, Esteller et al. showed that when a patient's tumor was positive for methylation (at one of four investigated genes) 73% of the time methylated DNA was also detectable in the serum.<sup>26</sup> Other work in lung cancer has been similar: in 2002, An et al. showed that 88% of patients ( $n = 64/73$ ) with *CDKN2A* methylation in tumors also had methylation in serum-derived DNA.<sup>27</sup> That same year Usadel and colleagues found that 47% ( $n = 42/89$ ) of lung cancer cases had *APC* methylation in serum DNA.<sup>28</sup> Finally, Ramirez et al. demonstrated a high within-person correlation between lung tumor methylation and sera-derived DNA methylation ( $p < 0.001$ ) at the *DAPK* and *RASSF1* genes,<sup>29</sup> and Hsu et al. reported 75–86% concordance of serum and lung tumor methylation.<sup>30</sup>

Clearly, a major limitation of serum-based epigenetic lung cancer screening biomarkers is that no single methylation markers will capture every tumor. Distinct tumor histologies, different etiologic exposures or differences in underlying susceptibility may all contribute to heterogeneity in the pattern of methylation alterations associated with the tumor phenotype. However, between histologic subtypes and across nearly 50 tumors, our approach was able to identify 150 CpGs with statistically significant, at least 2-fold changes in the magnitude of methylation, and allows for the identification of common tagging marks to maximize biomarker sensitivity. Another potential limitation for epigenetic biomarkers of NSCLC is sub-optimal specificity, and candidate genes are often altered in many malignancies. Both *HOXA9* and *SOX1* are transcription factors known to be important in embryonic development and epigenetic alterations in these genes may not be specific to NSCLC. In light of the cancer stem cell hypothesis and studies in ovarian and cervical cancer discussed above, additional research replicating our other 150 preliminary CpG biomarker candidates may be necessary to proffer epigenetic biomarkers that are highly specific to NSCLC. Nonetheless, our discovery approach allows us to build on the innovative concept that a panel of genes targeted for epigenetic alteration can define a more sensitive and specific screening strategy for this disease.

One possible limitation of our approach is that the matched non-tumor lung tissue may not be entirely epigenetically normal. In an attempt to limit any field cancerization effect, the non-tumorous lung was sampled distant from the tumor mass being resected. However, as this tissue is taken from diseased individuals, most with a smoking history, it is plausible that the normal tissue could have pre-neoplastic change. On the other hand, identifying CpG loci that are altered in tumor tissue relative to non-tumor lung with potential pre-neoplastic epigenetic alterations increases the potential utility of these biomarkers for screening

**Table 1.** CpGs with greater than 2-fold increase or decrease in tumor methylation relative to non-tumor lung, discovery set ( $n = 47$  pairs), from illumina golden gate methylation array (continued)

| Gene-loci with increased methylation in tumors |       |                    | Gene-loci with decreased methylation in tumors |      |                    |
|--|-------|--------------------|--|------|--------------------|
| GENE   | CpG*  | OR** (fold change) | GENE   | CpG* | OR** (fold change) |
| <i>PAX6</i>                                    | P50   | 2.8                | <i>AFF3</i>                                    | P122 | 2                  |
| <i>STAT5A</i>                                  | E42   | 2.8                | <i>MC2R</i>                                    | E455 | 2                  |
| <i>FGF2</i>                                    | P229  | 2.7                | <i>MEST</i>                                    | P4   | 2                  |
| <i>TERT</i>                                    | P360  | 2.7                |  |      |                    |
| <i>CALCA</i>                                   | E174  | 2.6                |  |      |                    |
| <i>SPARC</i>                                   | P195  | 2.6                |  |      |                    |
| <i>NEFL</i>                                    | P209  | 2.6                |  |      |                    |
| <i>ASCL2</i>                                   | E76   | 2.6                |  |      |                    |
| <i>NTRK2</i>                                   | P395  | 2.6                |  |      |                    |
| <i>NTSR1</i>                                   | P318  | 2.6                |  |      |                    |
| <i>ERBB2</i>                                   | P59   | 2.6                |  |      |                    |
| <i>GSTM1</i>                                   | P266  | 2.6                |  |      |                    |
| <i>IGFBP3</i>                                  | P423  | 2.6                |  |      |                    |
| <i>AGTR1</i>                                   | P154  | 2.5                |  |      |                    |
| <i>FLT3</i>                                    | E326  | 2.5                |  |      |                    |
| <i>DCC</i>                                     | P177  | 2.5                |  |      |                    |
| <i>DIO3</i>                                    | P674  | 2.5                |  |      |                    |
| <i>MME</i>                                     | P388  | 2.5                |  |      |                    |
| <i>CDKN2A</i>                                  | S188  | 2.5                |  |      |                    |
| <i>PODXL</i>                                   | P1341 | 2.5                |  |      |                    |
| <i>FGF3</i>                                    | E198  | 2.5                |  |      |                    |
| <i>TWIST1</i>                                  | P44   | 2.5                |  |      |                    |
| <i>RASSF1</i>                                  | P244  | 2.5                |  |      |                    |
| <i>FGF3</i>                                    | P171  | 2.5                |  |      |                    |
| <i>GALR1</i>                                   | E52   | 2.4                |  |      |                    |
| <i>IRAK3</i>                                   | P13   | 2.4                |  |      |                    |
| <i>SLIT2</i>                                   | P208  | 2.4                |  |      |                    |
| <i>DCC</i>                                     | E53   | 2.4                |  |      |                    |
| <i>P2RX7</i>                                   | P119  | 2.4                |  |      |                    |
| <i>GFI1</i>                                    | P45   | 2.3                |  |      |                    |
| <i>FGF12</i>                                   | P210  | 2.3                |  |      |                    |
| <i>ADCYAP1</i>                                 | P398  | 2.3                |  |      |                    |
| <i>PROK2</i>                                   | P390  | 2.3                |  |      |                    |
| <i>EYA4</i>                                    | E277  | 2.3                |  |      |                    |
| <i>GFI1</i>                                    | E136  | 2.3                |  |      |                    |
| <i>MT1A</i>                                    | E13   | 2.3                |  |      |                    |
| <i>HOXB2</i>                                   | P488  | 2.3                |  |      |                    |
| <i>SFRP1</i>                                   | E398  | 2.3                |  |      |                    |
| <i>KDR</i>                                     | P445  | 2.2                |  |      |                    |
| <i>PITX2</i>                                   | E24   | 2.2                |  |      |                    |

\*Location relative to transcription start site in bases, P = promoter, E = exonic. \*\*OR: [(tumor  $\beta$ /1-tumor  $\beta$ )/(normal tissue  $\beta$ /1-normal tissue  $\beta$ )].

**Table 1.** CpGs with greater than 2-fold increase or decrease in tumor methylation relative to non-tumor lung, discovery set (n = 47 pairs), from illumina golden gate methylation array (continued)

| Gene-loci with increased methylation in tumors |       |                    | Gene-loci with decreased methylation in tumors |      |                    |
|--|-------|--------------------|--|------|--------------------|
| GENE   | CpG*  | OR** (fold change) | GENE   | CpG* | OR** (fold change) |
| AGTR1  | P41   | 2.2                |  |      |                    |
| PENK   | P447  | 2.2                |  |      |                    |
| ESR1   | P151  | 2.2                |  |      |                    |
| RASSF1   | E116  | 2.2                |  |      |                    |
| HOXA5  | P1324 | 2.2                |  |      |                    |
| NTRK3  | P752  | 2.2                |  |      |                    |
| HOXA5  | E187  | 2.2                |  |      |                    |
| MOS  | P27   | 2.1                |  |      |                    |
| NPR2   | P618  | 2.1                |  |      |                    |
| JAK3   | P156  | 2.1                |  |      |                    |
| TJP2   | P330  | 2.1                |  |      |                    |
| TCF4   | P317  | 2.1                |  |      |                    |
| HOXC6  | P456  | 2.1                |  |      |                    |
| NTRK3  | E131  | 2.1                |  |      |                    |
| NTRK2  | P10   | 2.1                |  |      |                    |
| MAF  | P826  | 2.1                |  |      |                    |
| COL1A2   | P48   | 2.1                |  |      |                    |
| TPEF   | S36   | 2.1                |  |      |                    |
| IRAK3  | E130  | 2.1                |  |      |                    |
| TAL1   | P817  | 2                  |  |      |                    |
| CFTR   | P115  | 2                  |  |      |                    |
| CHGA   | E52   | 2                  |  |      |                    |
| TMEFF2   | P152  | 2                  |  |      |                    |
| THY1   | P149  | 2                  |  |      |                    |
| CCNA1  | E7    | 2                  |  |      |                    |
| PYCARD   | P150  | 2                  |  |      |                    |
| GUCY2D   | E419  | 2                  |  |      |                    |

\*Location relative to transcription start site in bases, P = promoter, E = exonic. \*\*OR: [(tumor  $\beta$ /1-tumor  $\beta$ )/(normal tissue  $\beta$ /1-normal tissue  $\beta$ )].

applications as they reflect changes that occur in the transition from pre-neoplasia to malignancy.

We identified 150 CpG loci associated with 120 genes that have a significant and at least 2-fold magnitude difference in methylation between tumor and matched non-tumor tissue. We then replicated the results for *HOXA9*, *SOX1* and *DDR1* in 99 additional paired tumor and non-tumor tissues from two independent populations. Our approach provides further proof of concept that strategies in the development of early detection biomarkers for NSCLC that include panels of markers will outperform single-marker approaches; and our results strongly indicate that CpGs in *HOXA9*, *SOX1* and *DDR* are highly attractive as members of such a biomarker panel. Importantly, the discovery of a panel of epigenetic signals that can be combined to develop a sensitive and specific biomarker panel may also provide critical

**Table 2.** Rank order of top differentially methylated CpGs between tumor and normal lung tissue by histology

| GENE_CpG    | Rank    |                 |                          |
|-------------|---------|-----------------|--------------------------|
|             | Overall | Adenocarcinomas | Squamous cell carcinomas |
| HOXA9_E252  | 1       | 1               | 1                        |
| TAL1_P594   | 2       | 2               | 6                        |
| HOXA9_P1141 | 3       | 12              | 2                        |
| HTR1B_P222  | 4       | 7               | 5                        |
| SOX1_P1018  | 5       | 9               | 7                        |
| MOS_E60     | 6       | 5               | 11                       |
| HOXA11_P698 | 7       | 21              | 8                        |
| TPEF_S88    | 8       | 10              | 12                       |
| MYOD1_E156  | 9       | 11              | 17                       |
| HCK_P858    | 10      | 35              | 10                       |

insight into cellular pathways where dysregulation is causal in this disease, further elucidate the mechanisms underlying lung cancer development and enable development of strategies for therapeutic intervention that successfully target the causal pathways.

## Patients and Methods

**Tissues.** Subject demographic, tumor histology and smoking status are in Table 4.

**Discovery set.** Paired tissues were derived from a surgical case series at the Massachusetts General Hospital from 1993 to 1996, as described in Wiencke et al.<sup>31</sup> and Nelson et al.<sup>32</sup> Patients undergoing surgical resection for non-small cell lung cancer consented to tissue collection and a small amount of tumor specimen and normal lung tissue distant to the tumor were collected during surgery and immediately snap frozen for research purposes. From the surgical series, a random subset of 47 patients with matched normal-tumor specimens was used.

**Replication set #1.** A set of paired lung tumor-normal pairs were obtained from the Brown Center for Cancer Research Molecular Pathology Core tissue bank. These were obtained from the pathology surgical suite and snap frozen.

**Replication set #2.** Tissues were obtained from a surgical series at the University of Belgrade Medical School Clinic for Pulmonology. This case-series of NSCLC was initiated in 2008 and is comprised of surgical patients with no pre-surgical chemotherapy or radiotherapy. Demographic data are obtained through patient interview, and clinical information derived from chart review. Tumor and normal tissue are obtained in the operating room and flash-frozen.

**Methylation analysis.** *Preparation of DNA and illumina goldengate array.* DNA from fresh frozen tissue samples was isolated with QIAamp DNA mini kit (Qiagen), and modified with sodium bisulfite using the EZ DNA Methylation kit (Zymo Research). Illumina GoldenGate methylation bead arrays were processed at the University of California San Francisco Institute for Human Genetics, Genomics Core Facility as described by

**Table 3.** Mean CpG methylation and standard deviation at three genes in two independent tissue sets

|              | Replication set 1 (n = 55) |             |                         | Replication set 2 (n = 54) |             |                         | Pooled data (n = 99) |             |                         |
|--------------|----------------------------|-------------|-------------------------|----------------------------|-------------|-------------------------|----------------------|-------------|-------------------------|
|              | Normal                     | Tumor       | p value                 | Normal                     | Tumor       | p value                 | Normal               | Tumor       | p value                 |
| <i>HOXA9</i> | 16.6 ± 3.2                 | 33.9 ± 18.3 | 7.3 × 10 <sup>-9</sup>  | 17.9 ± 8.0                 | 36.5 ± 17.1 | 2.1 × 10 <sup>-9</sup>  | 17.1 ± 8.0           | 35.1 ± 17.6 | 2.2 × 10 <sup>-16</sup> |
| <i>SOX1</i>  | 11.6 ± 3.2                 | 23.9 ± 13.8 | 5.2 × 10 <sup>-8</sup>  | 12.3 ± 4.5                 | 31.3 ± 16.7 | 1.5 × 10 <sup>-10</sup> | 11.9 ± 3.9           | 27.4 ± 15.7 | 2.2 × 10 <sup>-16</sup> |
| <i>DDR1</i>  | 54.6 ± 7.1                 | 37.4 ± 13.1 | 2.6 × 10 <sup>-11</sup> | 51.1 ± 5.6                 | 35.5 ± 12.3 | 1.5 × 10 <sup>-11</sup> | 52.9 ± 6.6           | 36.4 ± 12.7 | 2.2 × 10 <sup>-16</sup> |

Bibikova and colleagues.<sup>4</sup> Illumina GoldenGate array methylation data are publicly available on the GEO archive under accession number GSE27902.

**Bisulfite pyrosequencing.** Three loci were chosen for analysis in the validation tissues using bisulfite pyrosequencing. Assays were designed using Pyromark Assay Design 2.0 (Qiagen), and pyrosequencing data were collected on the Pyromark Q96MD. For the *HOXA9* gene, 1 CpG (the CpG measured by the array) was evaluated, for *SOX1*, 4 CpGs and for *DDR1*, 3 CpGs (both groups included the array CpG) were assessed. The average methylation across CpGs for each gene was calculated.

**mRNA analysis.** RNA was extracted from 30 mg of frozen lung tissue the RNeasy Mini Kit (Qiagen) and quantified with a NanoDrop Spectrophotometer. RNA was converted to cDNA with Invitrogen SuperScript III Reverse Transcriptase according to the manufacturer's protocol and included an Rnase inhibitor (Roche). Expression of the *DDR1* and *SOX1* mRNA transcripts was measured using commercially available Integrated DNA Technologies (IDT) PrimeTime qPCR primer/probe sets [IDT Assay names: *DDR1*-Hs.PT.49a.19433871.g, (exon 6–7), *SOX1*-Hs.PT.49a.2697171.g (exon 1)].

All reactions were run in triplicate on a BioRad CFX Connect Real Time PCR system with GAPDH serving as a referent [*GAPDH*-Hs.PT.49a.20047924.g (exon1)]. A pooled sample of RNAs was run on each plate to allow normalization using CFX Manager Software (version 2.1) and relative expression was then determined using the  $\Delta\Delta C_t$  method.

**Data analysis.** Illumina BeadStudio Methylation software was used for data set assembly. Fluorescent signals for methylated (Cy5) and unmethylated (Cy3) alleles give methylation level:  $\beta = [\max(Cy5, 0)] / (|Cy3| + |Cy5| + 100)$  with  $\sim 30$  replicate bead measurements per locus. For each matched tumor-normal pair the magnitude of methylation change was calculated using the odds ratio (OR: [(tumor  $\beta$ /1-tumor  $\beta$ )/(normal tissue  $\beta$ /1-normal tissue  $\beta$ ])). The OR is analogous to measuring the fold change in methylation, and was applied to these data as  $\beta$  is a restricted value between 0 and 1. For each locus the OR was averaged across all individuals. If the OR was less than one the number was inverted to obtain the relative fold decrease in methylation. Ingenuity Pathways Analysis

## References

- American Cancer Society. Cancer Facts & Figures 2012. Atlanta: American Cancer Society, 2012.
- Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002; 16:6-21; PMID:11782440; <http://dx.doi.org/10.1101/gad.947102>.
- Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 2002; 3:415-28; PMID:12042769.
- Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, et al. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* 2006; 16:383-93; PMID:16449502; <http://dx.doi.org/10.1101/gr.4410706>.
- Kim DH, Nelson HH, Wiencke JK, Zheng S, Christiani DC, Wain JC, et al. p16(INK4a) and histology-specific methylation of CpG islands by exposure to tobacco smoke in non-small cell lung cancer. *Cancer Res* 2001; 61:3419-24; PMID:11309302.
- Marsit CJ, Kim DH, Liu M, Hinds PW, Wiencke JK, Nelson HH, et al. Hypermethylation of RASSF1A and BLU tumor suppressor genes in non-small cell lung cancer: implications for tobacco smoking during adolescence. *Int J Cancer* 2005; 114:219-23; PMID:15540210; <http://dx.doi.org/10.1002/ijc.20714>.

**Table 4.** Subject demographic, tumor histology and smoking status count data

|                         | Discovery set (n = 47) | Replication set 1 (n = 55) | Replication set 2 (n = 54) |
|-------------------------|------------------------|----------------------------|----------------------------|
| <b>Age, mean (SD)</b>   | 68.7 (9.1)             | 72.4 (10.3)                | 59.9 (6.5)                 |
| <b>Sex</b>              |                        |                            |                            |
| Female                  | 21                     | 34                         | 18                         |
| Male                    | 26                     | 21                         | 36                         |
| <b>Histology</b>        |                        |                            |                            |
| Adenocarcinoma          | 22                     | 38                         | 20                         |
| Squamous cell carcinoma | 25                     | 17                         | 34                         |
| <b>Smoking status</b>   |                        |                            |                            |
| Never                   | 2                      |                            | 2                          |
| Former                  | 27                     |                            | 10                         |
| Current                 | 15                     |                            | 41                         |

\*These are counts, not percents.

software was used to determine the top networks enriched with genes whose CpGs had at least 2-fold increases or decreases (separately) in methylation using all autosomal array genes as the referent.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Financial Disclosures

NIH R01ES006717 and R01CA126831 (J.K.W.); NIH P20RR018728 (C.J.M.); NIH R01CA52689 and P50097257 (M.R.W.); NIH R01CA57494 and P42ES007373 (M.R.K.); NIH R01CA078609, R01CA121147, R01CA126939 and R01CA100679 (K.T.K.); NIH P30CA077598 (H.H.N.).

## Supplemental Material

Supplemental materials may be downloaded here: [www.landesbioscience.com/journals/epigenetics/article/20219](http://www.landesbioscience.com/journals/epigenetics/article/20219)



7. Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, et al. Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc Natl Acad Sci USA* 1998; 95:11891-6; PMID:9751761; <http://dx.doi.org/10.1073/pnas.95.20.11891>.
8. Dammann R, Li C, Yoon JH, Chin PL, Bates S, Pfeifer GP. Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. *Nat Genet* 2000; 25:315-9; PMID:10888881; <http://dx.doi.org/10.1038/77083>.
9. Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, et al. DNA methylation markers and prognostic recurrence in stage I lung cancer. *N Engl J Med* 2008; 358:1118-28; PMID:18337602; <http://dx.doi.org/10.1056/NEJMoa0706550>.
10. Buckingham L, Penfield Faber L, Kim A, Liptay M, Barger C, Basu S, et al. PTEN, RASSF1 and DAPK site-specific hypermethylation and outcome in surgically treated stage I and II nonsmall cell lung cancer patients. *Int J Cancer* 2010; 126:1630-9; PMID:19795445.
11. Anglim PP, Galler JS, Koss MN, Hagen JA, Turla S, Campan M, et al. Identification of a panel of sensitive and specific DNA methylation markers for squamous cell lung cancer. *Mol Cancer* 2008; 7:62; PMID:18616821; <http://dx.doi.org/10.1186/1476-4598-7-62>.
12. Ehrich M, Field JK, Liloglou T, Xinarianos G, Oeth P, Nelson MR, et al. Cytosine methylation profiles as a molecular marker in non-small cell lung cancer. *Cancer Res* 2006; 66:10911-8; PMID:17108128; <http://dx.doi.org/10.1158/0008-5472.CAN-06-0400>.
13. Feng Q, Hawes SE, Stern JE, Wiens L, Lu H, Dong ZM, et al. DNA methylation in tumor and matched normal tissues from non-small cell lung cancer patients. *Cancer Epidemiol Biomarkers Prev* 2008; 17:645-54; PMID:18349282; <http://dx.doi.org/10.1158/1055-9965.EPI-07-2518>.
14. Field JK, Liloglou T, Warrak S, Burger M, Becker E, Berlin K, et al. Methylation discriminators in NSCLC identified by a microarray based approach. *Int J Oncol* 2005; 27:105-11; PMID:15942649.
15. Rauch T, Wang Z, Zhang X, Zhong X, Wu X, Lau SK, et al. Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proc Natl Acad Sci USA* 2007; 104:5527-32; PMID:17369352; <http://dx.doi.org/10.1073/pnas.0701059104>.
16. Tsou JA, Galler JS, Siegmund KD, Laird PW, Turla S, Cozen W, et al. Identification of a panel of sensitive and specific DNA methylation markers for lung adenocarcinoma. *Mol Cancer* 2007; 6:70; PMID:17967182; <http://dx.doi.org/10.1186/1476-4598-6-70>.
17. Weisenberger DJ, Trinh BN, Campan M, Sharma S, Long TI, Ananthnarayan S, et al. DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. *Nucleic Acids Res* 2008; 36:4689-98; PMID:18628296; <http://dx.doi.org/10.1093/nar/gkn455>.
18. Su HY, Lai HC, Lin YW, Chou YC, Liu CY, Yu MH. An epigenetic marker panel for screening and prognostic prediction of ovarian cancer. *Int J Cancer* 2009; 124:387-93; PMID:18942711; <http://dx.doi.org/10.1002/ijc.23957>.
19. Apostolidou S, Hadwin R, Burnell M, Jones A, Baff D, Pyndiah N, et al. DNA methylation analysis in liquid-based cytology for cervical cancer screening. *Int J Cancer* 2009; 125:2995-3002; PMID:19609949; <http://dx.doi.org/10.1002/ijc.24745>.
20. Christensen BC, Houseman EA, Godleski JJ, Marsit CJ, Longacker JL, Roelofs CR, et al. Epigenetic profiles distinguish pleural mesothelioma from normal pleura and predict lung asbestos burden and clinical outcome. *Cancer Res* 2009; 69:227-34; PMID:19118007; <http://dx.doi.org/10.1158/0008-5472.CAN-08-2586>.
21. Ford CE, Lau SK, Zhu CQ, Andersson T, Tsao MS, Vogel WF. Expression and mutation analysis of the discoidin domain receptors 1 and 2 in non-small cell lung carcinoma. *Br J Cancer* 2007; 96:808-14; PMID:17299390; <http://dx.doi.org/10.1038/sj.bjc.6603614>.
22. Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* 2007; 131:1190-203; PMID:18083107; <http://dx.doi.org/10.1016/j.cell.2007.11.025>.
23. Yang SH, Baek HA, Lee HJ, Park HS, Jang KY, Kang MJ, et al. Discoidin domain receptor 1 is associated with poor prognosis of non-small cell lung carcinomas. *Oncol Rep* 2010; 24:311-9; PMID:20596615.
24. Heller G, Zielinski CC, Zöchbauer-Müller S. Lung cancer: from single-gene methylation to methylome profiling. *Cancer Metastasis Rev* 2010; 29:95-107; PMID:20099008; <http://dx.doi.org/10.1007/s10555-010-9203-x>.
25. Yoon KA, Park S, Lee SH, Kim JH, Lee JS. Comparison of circulating plasma DNA levels between lung cancer patients and healthy controls. *J Mol Diagn* 2009; 11:182-5; PMID:19324991; <http://dx.doi.org/10.2353/jmoldx.2009.080098>.
26. 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; PMID:9892187.
27. An Q, Liu Y, Gao Y, Huang J, Fong X, Li L, et al. Detection of p16 hypermethylation in circulating plasma DNA of non-small cell lung cancer patients. *Cancer Lett* 2002; 188:109-14; PMID:12406555; [http://dx.doi.org/10.1016/S0304-3835\(02\)00496-2](http://dx.doi.org/10.1016/S0304-3835(02)00496-2).
28. Usadel H, Brabender J, Danenberg KD, Jerónimo C, Harden S, Engles J, et al. Quantitative adenomatous polyposis coli promoter methylation analysis in tumor tissue, serum and plasma DNA of patients with lung cancer. *Cancer Res* 2002; 62:371-5; PMID:11809682.
29. Ramirez JL, Sarries C, de Castro PL, Roig B, Queralt C, Escuin D, et al. Methylation patterns and K-ras mutations in tumor and paired serum of resected non-small-cell lung cancer patients. *Cancer Lett* 2003; 193:207-16; PMID:12706879; [http://dx.doi.org/10.1016/S0304-3835\(02\)00740-1](http://dx.doi.org/10.1016/S0304-3835(02)00740-1).
30. Hsu HS, Chen TP, Hung CH, Wen CK, Lin RK, Lee HC, et al. Characterization of a multiple epigenetic marker panel for lung cancer detection and risk assessment in plasma. *Cancer* 2007; 110:2019-26; PMID:17876837; <http://dx.doi.org/10.1002/cncr.23001>.
31. Wiencke JK, Thurston SW, Kelsey KT, Varkonyi A, Wain JC, Mark EJ, et al. Early age at smoking initiation and tobacco carcinogen DNA damage in the lung. *J Natl Cancer Inst* 1999; 91:614-9; PMID:10203280; <http://dx.doi.org/10.1093/jnci/91.7.614>.
32. Nelson HH, Christiani DC, Mark EJ, Wiencke JK, Wain JC, Kelsey KT. Implications and prognostic value of K-ras mutation for early-stage lung cancer in women. *J Natl Cancer Inst* 1999; 91:2032-8; PMID:10580029; <http://dx.doi.org/10.1093/jnci/91.23.2032>.