

Cancer Research

Differentiation of Lung Adenocarcinoma, Pleural Mesothelioma, and Nonmalignant Pulmonary Tissues Using DNA Methylation Profiles

Brock C. Christensen, Carmen J. Marsit, E. Andres Houseman, et al.

Cancer Res 2009;69:6315-6321. Published OnlineFirst July 28, 2009.

Updated Version

Access the most recent version of this article at:
doi:[10.1158/0008-5472.CAN-09-1073](https://doi.org/10.1158/0008-5472.CAN-09-1073)

Supplementary Material

Access the most recent supplemental material at:
<http://cancerres.aacrjournals.org/content/suppl/2009/07/20/0008-5472.CAN-09-1073.DC1.html>

Cited Articles

This article cites 30 articles, 13 of which you can access for free at:
<http://cancerres.aacrjournals.org/content/69/15/6315.full.html#ref-list-1>

Citing Articles

This article has been cited by 3 HighWire-hosted articles. Access the articles at:
<http://cancerres.aacrjournals.org/content/69/15/6315.full.html#related-urls>

E-mail alerts

[Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions

To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.

Differentiation of Lung Adenocarcinoma, Pleural Mesothelioma, and Nonmalignant Pulmonary Tissues Using DNA Methylation Profiles

Brock C. Christensen,^{1,2} Carmen J. Marsit,² E. Andres Houseman,^{1,3} John J. Godleski,⁴ Jennifer L. Longacker,⁵ Shichun Zheng,⁷ Ru-Fang Yeh,⁸ Margaret R. Wrensch,⁷ Joseph L. Wiemels,⁸ Margaret R. Karagas,⁹ Raphael Bueno,⁶ David J. Sugarbaker,⁶ Heather H. Nelson,¹⁰ John K. Wiencke,⁷ and Karl T. Kelsey^{1,2}

¹Department of Community Health, Center for Environmental Health and Technology; and ²Department of Pathology and Laboratory Medicine, Brown University, Providence, Rhode Island; Departments of ³Biostatistics and ⁴Environmental Health, Harvard School of Public Health; ⁵Department of Environmental Health, Boston University School of Public Health; and ⁶Division of Thoracic Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; Departments of ⁷Neurological Surgery, and ⁸Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California; ⁹Section of Biostatistics and Epidemiology, Department of Community and Family Medicine, Dartmouth Medical School, Lebanon, New Hampshire; and ¹⁰Division of Epidemiology and Community Health, Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota

Abstract

Pathologic differentiation of tissue of origin in tumors found in the lung can be challenging, with differentiation of mesothelioma and lung adenocarcinoma emblematic of this problem. Indeed, proper classification is essential for determination of treatment regimen for these diseases, making accurate and early diagnosis critical. Here, we investigate the potential of epigenetic profiles of lung adenocarcinoma, mesothelioma, and nonmalignant pulmonary tissues ($n = 285$) as differentiation markers in an analysis of DNA methylation at 1413 autosomal CpG loci associated with 773 cancer-related genes. Using an unsupervised recursively partitioned mixture modeling technique for all samples, the derived methylation profile classes were significantly associated with sample type ($P < 0.0001$). In a similar analysis restricted to tumors, methylation profile classes significantly predicted tumor type ($P < 0.0001$). Random forests classification of CpG methylation of tumors—which splits the data into training and test sets—accurately differentiated mesothelioma from lung adenocarcinoma over 99% of the time ($P < 0.0001$). In a locus-by-locus comparison of CpG methylation between tumor types, 1266 CpG loci had significantly different methylation between tumors following correction for multiple comparisons ($Q < 0.05$); 61% had higher methylation in adenocarcinoma. Using the CpG loci with significant differential methylation in a pathway analysis revealed significant enrichment of methylated gene-loci in Cell Cycle Regulation, DNA Damage Response, PTEN Signaling, and Apoptosis Signaling pathways in lung adenocarcinoma when compared with mesothelioma. Methylation profile-based differentiation of lung adenocarcinoma and mesothelioma is highly accurate, informs on the distinct etiologies of these diseases, and holds promise for clinical application. [Cancer Res 2009;69(15):6315–21]

Introduction

Malignant pleural mesothelioma is a rapidly fatal neoplasm with a clinical presentation that can mimic adenocarcinoma of the lung, complicating diagnosis (1, 2). These malignancies likely have distinct cellular origins, although this remains unclear. Shared signs and symptoms of these diseases include malignant pleural effusion, dyspnea, chest pain, and fatigue (3, 4). An enhanced description of the character of the underlying somatic alterations, and thereby a proper diagnosis, is of paramount importance, especially considering the disparate prognoses and treatment regimens for lung adenocarcinoma and mesothelioma (5, 6).

Several techniques have been used or proposed for differential diagnosis. Cytologic approaches to differential diagnosis have historically had a wide margin of variability in sensitivity depending on sample preparation methods and feature sets analyzed (7, 8). Currently, the most common method uses an immunohistochemical panel containing both epithelial and mesothelial markers (9). Despite recent improvements in antibody panels for differential diagnosis, there is no consensus immunohistochemical panel or evidence-based guidelines for panel selection (9, 10). Another method, using mRNA expression gene ratios, has reported differential diagnosis accuracy of 95% and 99% for mesothelioma and adenocarcinoma, respectively (11). The instability of mRNA, though, may make wide-scale implementation of this technology challenging, particularly outside of major academic surgical centers.

It is well recognized that promoter DNA hypermethylation is a mechanism of stable control of transcription, and an important contributor to carcinogenesis. When certain cytosines in specific clustered regions primarily located in gene promoters are hypermethylated, aberrant, stable gene silencing can occur. Regulatory CpG clusters are common, often occur in tumor suppressor genes, and are thought to remain largely unmethylated in noncancerous cells. In fact, about half of all human genes contain CpG islands and are potentially subject to aberrant methylation silencing (12, 13). 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 (14). Our previous work has shown hundreds of differentially methylated CpG loci in pleural mesothelioma compared with nondiseased pleura (15). Other reports, using a small number of candidate loci, have shown significant differences

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Requests for reprints: Karl T. Kelsey, Brown University, Box GE-5, 70 Ship Street, Providence, RI, 02903. Phone: 401-863-6420; Fax: 401-863-9008; E-mail: Karl_Kelsey@Brown.edu.

©2009 American Association for Cancer Research.
doi:10.1158/0008-5472.CAN-09-1073

Table 1. Patient demographics, exposures, and tissue characteristics

| Covariate | Lung | | Pleura | |
|--------------------------|----------------------------|--|--|---|
| | Nontumor (<i>n</i> = 52)* | Adenocarcinoma (<i>n</i> = 57) [†] | Nontumor (<i>n</i> = 18) [‡] | Mesothelioma (<i>n</i> = 158) [§] |
| Age | | | | |
| Range | 47–89 | 35–89 | 38–77 | 30–84 |
| Mean (SD) | 68.8 (9.2) | 68.2 (11.4) | 58.3 (11.3) | 61.7 (9.8) |
| Gender (<i>n</i>) % | | | | |
| Male | 26 (55.4) | 23 (40.4) | 14 (77.8) | 120 (75.9) |
| Female | 21 (44.6) | 34 (59.6) | 4 (22.2) | 38 (24.1) |
| Histology (<i>n</i>) % | | | | |
| Adenocarcinoma | — | 57 (100) | — | — |
| Epithelioid | — | — | — | 109 (69.0) |
| Biphasic | — | — | — | 44 (27.8) |
| Sarcomatoid | — | — | — | 5 (3.2) |
| Smoking status | | | | |
| Current | 15 (28.8) | 18 (31.6) | — | 34 (27.2) |
| Former | 27 (51.9) | 27 (47.3) | — | 43 (34.4) |
| Never | 5 (9.6) | 12 (21.1) | — | 48 (38.4) |
| Asbestos [¶] | | | | |
| No | 41 (89.1) | 55 (98.2) | 5 (27.8) | 39 (25.9) |
| Yes | 5 (10.9) | 1 (1.8) | 13 (72.2) | 112 (74.1) |

*Five samples missing age and gender data, six samples missing exposure data.

†One sample missing asbestos exposure data.

‡No smoking data available.

§Thirty-three missing smoking data, seven missing asbestos exposure data.

||Excluded from tumor-only analysis.

¶Occupational exposure (lung), known exposure (pleura).

in gene-promoter methylation prevalences between lung adenocarcinoma and mesothelioma (16, 17).

In this study, we exploited the stability of the aberrant cytosine methylation mark and new array-based technology for high throughput measurement of DNA CpG methylation to investigate the methylation status of 1413 autosomal CpG loci associated with 773 cancer-related genes on Illumina's GoldenGate methylation bead-array platform. Using one of the largest case series studies of these diseases and focusing on epigenetic alteration, we show that methylation profiling can differentiate lung adenocarcinoma, mesothelioma, and nonmalignant tissues.

Materials and Methods

Study samples. Mesotheliomas (*n* = 158) and grossly nontumorigenic parietal pleura (*n* = 18) were obtained following surgical resection at Brigham and Women's Hospital through the International Mesothelioma Program from a pilot study conducted in 2002 (*n* = 70) and an incident case series beginning in 2005 (*n* = 88) with a participation rate of 85%. We used biopsy specimens from patients treated for non-small cell lung cancer at the Massachusetts General Hospital from 1992 to 1996 (18) including lung adenocarcinomas (*n* = 57) and nonmalignant pulmonary tissues [*n* = 48; of which 22 (39%) were taken from the adenocarcinoma patients; ref. 18]. Additional normal lung tissues were obtained from the National Disease Research Interchange from donors free of lung malignancy (*n* = 4). All patients provided informed consent under the approval of the appropriate Institutional Review Boards. Clinical information, including histologic diagnosis, was obtained from pathology reports. The study pathologist confirmed the histologic diagnoses and further assessed the percent tumor from resected specimens (mean, >60% for mesotheliomas; >50% for lung adenocarcinomas).

Methylation analysis. DNA from fresh frozen tissue was isolated with QIAamp DNA mini kit (Qiagen), and sodium bisulfite modified using the EZ DNA Methylation kit (Zymo Research). Illumina GoldenGate methylation bead arrays interrogated 1505 CpG loci associated with 803 cancer-related genes processed at the University of California San Francisco Institute for Human Genetics, Genomics Core Facility as described by Bibikova and colleagues (14). Methylation array data are available on the Gene Expression Omnibus archive (accession GSE16559).

Statistical analysis. Illumina BeadStudio Methylation software was used for data set assembly. Fluorescent signals for methylated (Cy5) and unmethylated (Cy3) alleles give methylation level: $\beta = (\max(\text{Cy5}, 0)) / (|\text{Cy3}| + |\text{Cy5}| + 100)$ with ~30 replicate bead measurements per locus. Detection *P* values determined poor performing samples (*n* = 2) and CpG loci (*n* = 8), which were removed from analysis. X chromosome loci were also removed, leaving 1413 CpG loci associated with 773 genes.

Subsequent analyses were conducted in R (19). Hierarchical clustering was performed with the *hclust* function: Manhattan metric and average linkage for CpG loci with the highest variance. For inference, data were clustered using a recursively partitioned mixture model (RPMM; ref. 20). Associations between covariates and methylation at individual CpG loci were tested with generalized linear models, accounting for the beta-distribution of average β as in Hsuing and colleagues (21). False discovery rate correction via *Q*-values were computed by the *qvalue* package (22).

Recognizing the importance of using a *training set* to build a classifier, and a *test set* upon which to test the validity of the classification scheme, we have used the Random Forests approach (RF), R package version 4.5–25 by Liaw and Wiener. RF builds classifiers by repeatedly sampling with replacement from the original data (i.e., bootstrap sampling), sampling from the predictors, and building a classification tree with the resulting samples (23). Upon every iteration, approximately a third of the original data are not sampled; the unsampled, or “out of the bag” observations are

used as a test set against which the tree is assessed with respect to classification error. The out of the bag error rate—the average classification error over all iterations—is thus an unbiased estimate of the fraction of times the RF prediction is incorrect.

Canonical pathway analysis was conducted with the use of Ingenuity Pathway Analysis (Ingenuity Systems; ref. 24). CpG gene-loci associated with the Ingenuity Pathway Knowledge Base were considered for analysis and differentially methylated loci from locus-by-locus analysis were compared. The significance of gene-locus enrichment within canonical pathways was measured with a Fisher's exact test ($P < 0.05$).

Results

Incident cases of mesothelioma ($n = 158$), lung adenocarcinoma ($n = 57$) and associated nonmalignant pleural ($n = 18$), and pulmonary tissues ($n = 52$) were assessed for methylation (total $n = 285$). Demographic and tumor characteristic data for these samples are presented in Table 1. Mean age and gender distributions were similar between tumor and their nontumor samples of origin. Lung adenocarcinomas and nontumor lung samples were from individuals with similar exposures to smoking, and their asbestos exposure histories did not differ. Mesotheliomas and nontumor pleural samples were from individuals with similar exposures to asbestos.

Unsupervised hierarchical clustering of the 500 most methylation-variable autosomal CpG loci revealed readily apparent differences in the epigenetic profiles among lung adenocarcinoma, mesothelioma, and nonmalignant tissues (Fig. 1A). However, nonmalignant pleural and pulmonary tissues did not seem to segregate from each other. Unsupervised hierarchical clustering of tumors only is shown in Fig. 1B. We next applied a modified model-

based form of unsupervised clustering known as RPM (20). The RPM returned 17 methylation classes whose average methylation profiles are shown in Fig. 2; 11 of these classes (68%) perfectly captured a single sample type, and methylation profiles were a significant predictor of tissue sample type ($P < 0.0001$). The 50 CpG loci whose methylation status most effectively discriminates among methylation classes are listed in Supplementary Table S1.

A supervised RF classification of methylation data in all samples returned a confusion matrix showing which samples are correctly classified, those that are misclassified, and the misclassification error (ME) rate for each sample type (Table 2). The overall ME rate of 7.0% was significantly lower than the expected error rate under the null hypothesis ($P < 0.0001$).

Consistent with the patterns observed from unsupervised clustering, nonmalignant tissues had a higher misclassification error (ME, 24.3%), than tumors (ME, 1.4%). Of 52 nonmalignant pulmonary tissues, 4 were confused as lung adenocarcinoma, and 1 as a mesothelioma (ME, 9.6%). Among 18 nonmalignant pleural tissues, 7 were confused as nontumor lung, and 5 as mesothelioma (ME, 66.6%). On the other hand, only one lung adenocarcinoma was misclassified, as a nontumor lung (ME, 1.8%); and only two mesotheliomas were misclassified, both as lung adenocarcinoma (ME, 1.3%). The 50 most discriminatory CpG loci from this RF analysis are given in Supplementary Table S2.

We next restricted our analysis to lung adenocarcinoma and nonsarcomatoid mesotheliomas ($n = 210$) and applied the RPM approach (Fig. 3). In this model, 14 methylation classes resulted, and 12 (86%) perfectly capture a single tumor type. Methylation classes significantly predicted tumor type ($P < 0.0001$). The 50 most critical loci for differentiating the methylation classes in this model

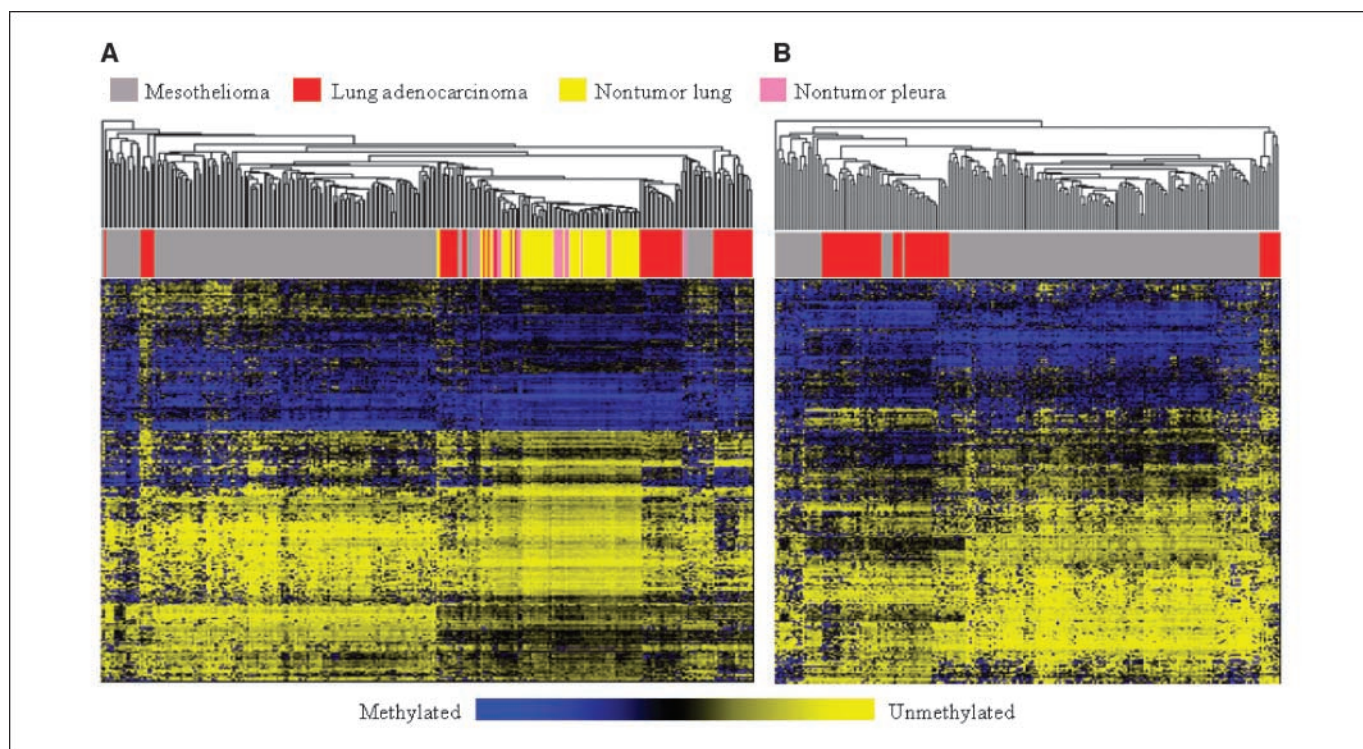


Figure 1. Unsupervised clustering heatmap of CpG loci in all samples and tumors only. Unsupervised hierarchical clustering heat map based on Manhattan distance and average linkage of the 500 autosomal CpG loci with the highest variance. Columns, samples; rows, CpG loci. Blue, methylated; yellow, unmethylated; A, all samples; gray bars, mesotheliomas; red bars, lung adenocarcinomas; yellow bars, nontumor lung samples; pink bars, nontumor pleura samples. B, tumor samples only.

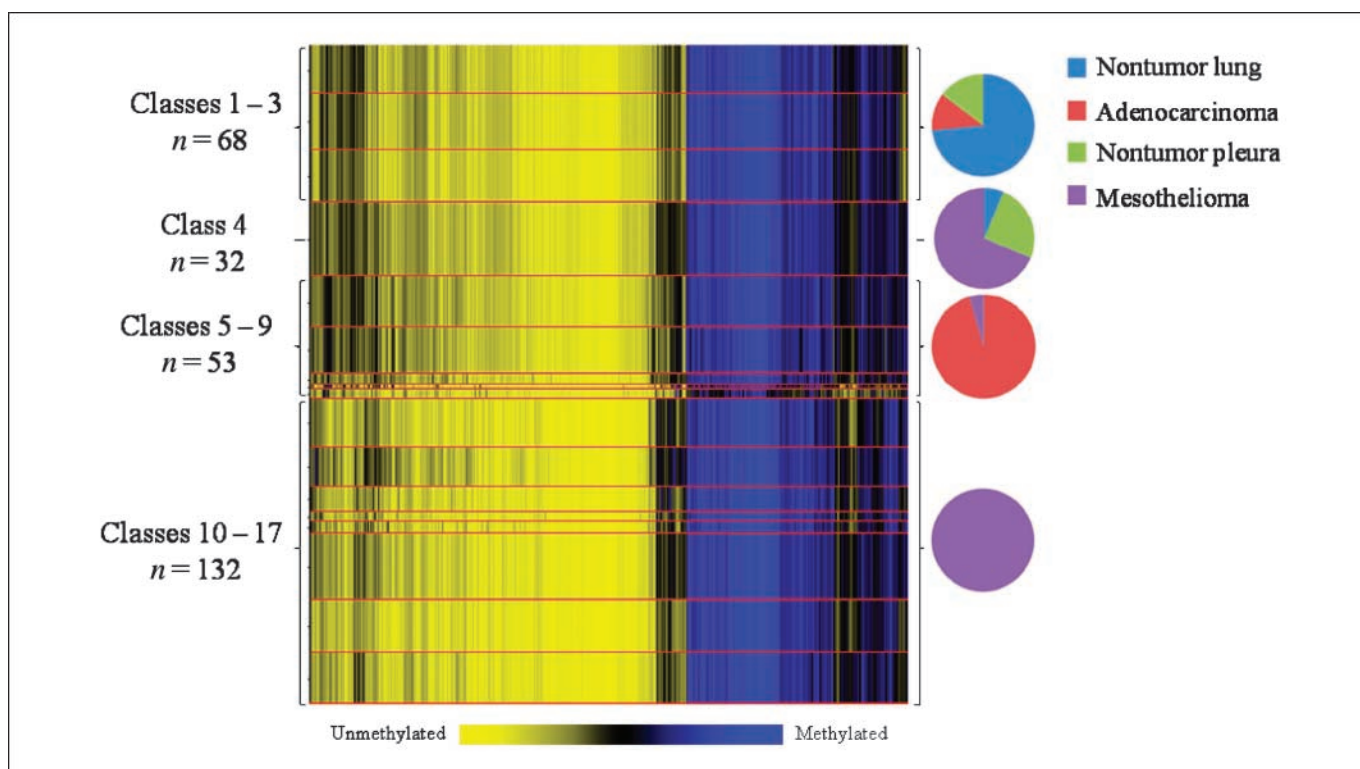


Figure 2. RPMM of CpG methylation for lung adenocarcinomas, mesotheliomas, and nonmalignant pulmonary tissues. The figure depicts the results of RPMM. Columns, CpG sites; rows, methylation classes. The height of each row is proportional to the number of observations residing in the class, and the color of the columns within the row represents the average methylation of the CpG for that class. Blue, methylated; yellow, unmethylated. Pie charts represent the composition of the group of classes indicated with respect to tissue type. Methylation profile classes differentiate sample types (permutation test $P < 0.0001$).

are listed in Supplementary Table S3. Results were again followed up with RF classification resulting in a confusion matrix with an overall ME of $<1\%$, ($P < 0.0001$; Table 2). The 50 most discriminatory CpG loci for RF classification of tumors are given in Supplementary Table S4.

In a univariate approach, we tested all CpG loci individually for an association between methylation and tumor type with

generalized linear models followed by correction for multiple comparisons. In this manner, 1266 CpG loci had methylation levels that differed between lung adenocarcinoma and mesothelioma ($Q < 0.05$; Supplementary Table S5). Among these 1266 CpG loci, 61% had higher methylation in lung adenocarcinoma compared with mesothelioma. In addition, epithelioid and sarcomatoid mesotheliomas had differential methylation ($Q < 0.05$) at 87 CpG

Table 2. Random forests analysis confusion matrices

| | Lung | | Pleura | | Classification error |
|---|----------------|----------------|----------------|--------------|----------------------|
| | Nontumor | Adenocarcinoma | Nontumor | Mesothelioma | |
| Lung | | | | | |
| Nontumor | 47 | 4 | 1 | — | 9.6% |
| Adenocarcinoma | 1 | 56 | — | — | 1.8% |
| Pleura | | | | | |
| Nontumor | 7 | — | 5 | 6 | 66.7% |
| Mesothelioma | — | 2 | — | 156 | 1.3% |
| Overall error estimate = 7.0% $P < 0.0001$ | | | | | |
| | Adenocarcinoma | | Mesothelioma | | Classification error |
| | Adenocarcinoma | Mesothelioma | Adenocarcinoma | Mesothelioma | |
| Adenocarcinoma | 56 | 1 | 1 | 152 | 1.75% |
| Mesothelioma | 1 | 152 | 1 | 152 | 0.65% |
| Overall error estimate = 0.95% $P < 0.0001$ | | | | | |

loci including 15 gene-loci (e.g., *SLC22A18*, *RARA*, and *SEPT9*) with >1 CpG displaying differential methylation (Supplementary Table S6).

Lastly, using the locus-by-locus data, we performed a pathway analysis comparing methylation profiles between lung adenocarcinoma and mesothelioma. Among mesotheliomas, *Fc Epsilon RI Signaling*, and *Calcium Signaling* pathways were significantly enriched (Fisher's $P < 0.05$) for methylation versus lung adenocarcinoma (Table 3). Lung adenocarcinomas had six pathways with significant enrichment (Fisher's $P < 0.05$) of methylated gene-loci versus mesothelioma including *Cell Cycle Regulation*, *DNA Damage Response*, *PTEN Signaling*, and *Apoptosis Signaling*.

Discussion

The microscopic assessment of adenocarcinoma of the lung can resemble malignant pleural mesothelioma. There is no absolute standardized approach to differential diagnosis of these diseases, which can be challenging. As is the case with any disease, proper diagnosis is paramount; a rapid, accurate diagnosis has the potential to improve patient outcome. Using DNA methylation profiling, we successfully differentiated these tumors, suggesting that this approach may be a useful adjunct in diagnosis.

All somatic cells in a given individual are genetically identical (excluding T and B cells). However, different cell types form distinct anatomic structures and carry out a wide range of physiologic functions. This is made possible largely via control of gene

expression. One approach for differentiating pleural mesothelioma and lung adenocarcinoma relies on the differential gene expression profiles of these tumors (11). Although this approach is sound, and has been reproduced in malignant pleural effusions (1), the instability of mRNA transcripts makes methods relying upon RNA measures difficult to standardize and implement. DNA methylation profiles reflect phenotypically important differences in gene transcription and the molecular structure of DNA is inherently more stable than RNA, making assessment of DNA methylation profiles attractive as a highly accurate and reproducible diagnostic test.

Unsupervised clustering achieved excellent segregation of tumor tissues from each other and from nontumor tissues, although there was indistinct clustering of nontumorigenic lung and pleural samples. Similarly, some RPMM methylation classes contained a mixture of both nontumor lung and nontumor pleura samples, and in RF classification, nontumor pleura samples had the highest misclassification error. The most likely reason for pleura being misclassified as lung tissue is the potential contamination of the pleural sample with adjacent lung tissue. In addition, in this and other RF classifications of methylation data from our group, we found a significant correlation between sample size and classification error. Therefore, some of the ME for pleural samples may be attributable to small sample size. In the future, arrays with larger panels of CpG methylation markers may further increase the accuracy with which these tissue types can be differentiated.

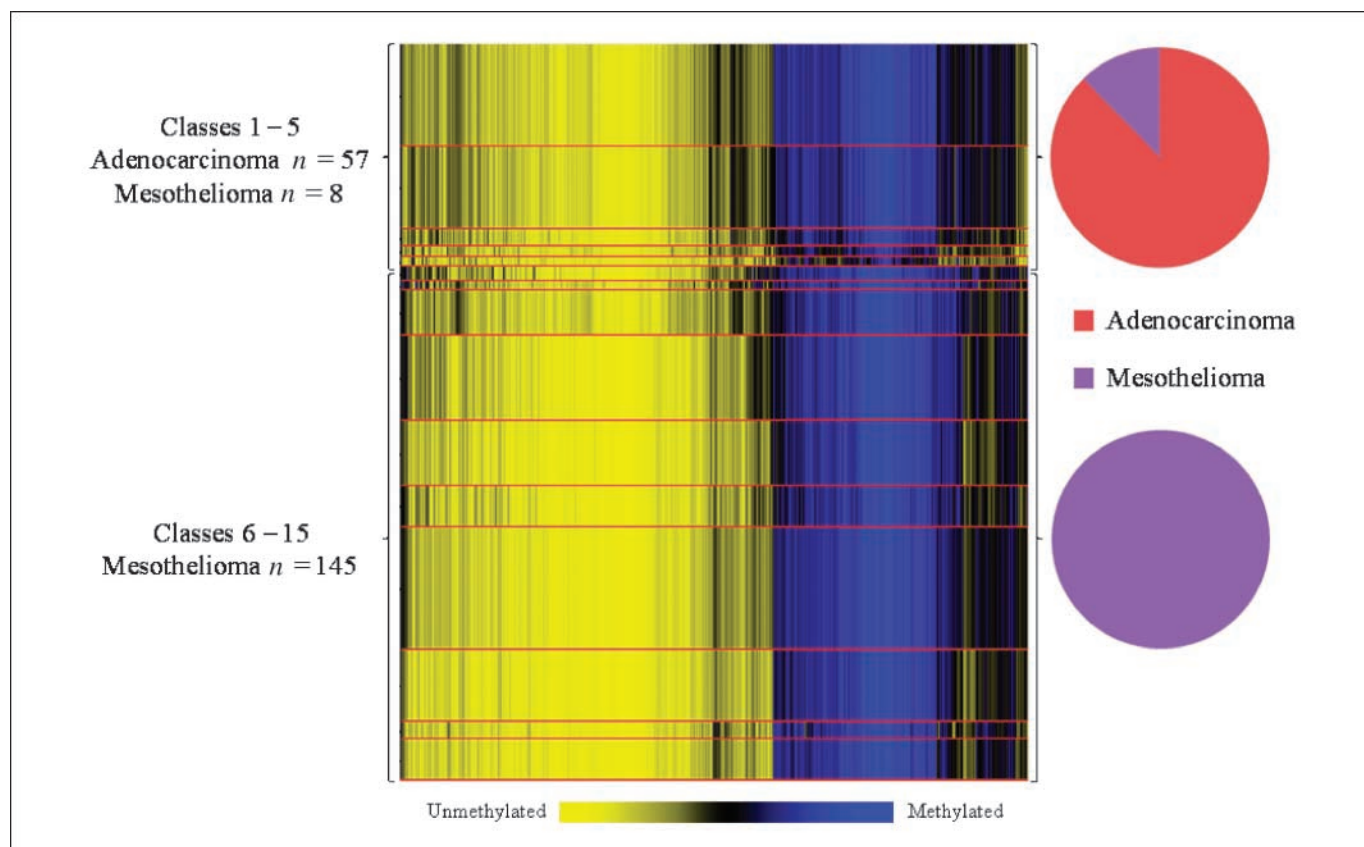


Figure 3. RPMM of CpG methylation for lung adenocarcinomas and mesotheliomas. The figure depicts the results of RPMM. Columns, CpG sites; rows, methylation classes. The height of each row is proportional to the number of observations residing in the class, and the color of the columns within the row represents the average methylation of the CpG for that class. Blue, methylated; yellow, unmethylated. Pie charts represent the composition of the group of classes indicated with respect to tissue type. Methylation profile classes significantly differentiate tumor types (permutation test $P < 0.0001$).

Table 3. Pathways with gene-loci enriched for methylation in mesothelioma versus lung adenocarcinoma

| Pathway | Increased methylation | P | Genes |
|---|-----------------------|------|---|
| Cell cycle: G ₁ -S checkpoint regulation | Adenocarcinoma | 0.01 | <i>CDKN2A, RBL2, NRG1, CDK6, ABL1, E2F3, CDKN2B, CCND1, HDAC5, CCNE1, CCND3, CCND2, HDAC11, CDKN1A, TGFB3, E2F5, TGFB2, SMAD4, CDK2</i> |
| Fc Epsilon RI signaling | Mesothelioma | 0.01 | <i>VAV2, PIK3R1, MAPK9, PLA2G2A, IL13, MAPK14, SYK, LAT, LYN, VAV1, IL3, CSF2, TNF, IL4</i> |
| DNA damage response | Adenocarcinoma | 0.02 | <i>RBL2, GADD45A, FANCG, CDKN1A, E2F5, FANCF, E2F3, HLTf, RAD50, SMARCA4, MLH1</i> |
| PTEN signaling | Adenocarcinoma | 0.02 | <i>ITGB1, RAF1, NRAS, CASP3, RRAS, ITGA2, KRAS, NFKB2, NFKB1, CCND1, PTEN, AKT1, BMPR1A, NGFR, CDKN1A, PDGFRA, INSR, PDGFRB, EGFR</i> |
| Apoptosis signaling | Adenocarcinoma | 0.03 | <i>RAF1, NRAS, CASP3, RRAS, KRAS, BCL3, NFKB2, BAX, NFKB1, FAS, CASP6, CASP2, TNFRSF1B, CASP8, CASP10</i> |
| IL-6 signaling | Adenocarcinoma | 0.04 | <i>MAP2K6, ABCB1, RAF1, IL8, NRAS, RRAS, KRAS, BCL3, IL6, NFKB2, JAK2, MAPK12, NFKB1, COL1A1, IL1RN, NGFR, MAPK10, IL1B, TNFRSF1B</i> |
| Calcium signaling | Mesothelioma | 0.04 | <i>HDAC7, ITPR3, CREB1, HDAC1, CREBBP, HDAC9, ACTG2, PRKARIA</i> |
| Erythropoietin signaling | Adenocarcinoma | 0.04 | <i>EPO, RAF1, STAT5A, PTPN6, NRAS, RRAS, BCL3, KRAS, NFKB2, JAK2, MAPK12, NFKB1, AKT1</i> |

In an analysis restricted to tumors, we showed the great extent to which CpG methylation varies between mesothelioma and lung adenocarcinoma. Disparate CpG methylation profiles between these tumor types can be attributed in part to differential methylation profiles in the tissues of origin. Although there has been a general consensus that normal cells maintain CpG islands in an unmethylated state permissive to transcription (13), tissue-specific methylation of CpG islands has been described in nondiseased cells (25). In fact, data from the Human Epigenome Project have shown that there is tissue-specific methylation among 90 genes associated with the human major histocompatibility complex (26), and others have reported tissue-specific promoter-region methylation of monocytes, testis, and brain tissues (27). Consistent with these findings, our data show that, in general, normal lung and pleura have different basal methylation profiles.

The different etiologic factors associated with the induction of these tumors likely contribute to their differential methylation. Although the majority of lung adenocarcinomas are related to smoking, smoking is not a risk factor for mesothelioma; rather, the vast majority of mesotheliomas are linked to asbestos exposure. Although asbestos is also a risk factor for lung adenocarcinoma, in our study population, only one lung adenocarcinoma patient had a known occupational asbestos exposure, and this individual was also a smoker. Significant smoking-related and asbestos-related methylation-induced gene inactivation events have been described in lung adenocarcinoma and mesothelioma, respectively (28, 29). It is possible that differences in carcinogen exposure result in differences in methylation profiles within and between tumor types.

In a locus-by-locus analysis of tumor samples, over 1,000 CpG loci were differentially methylated between tumor types. Previously, with a combined sample of over 100 mesotheliomas and lung adenocarcinomas, Toyooka and colleagues (16) reported significantly increased methylation in lung adenocarcinoma at *APC*, *CDH13*, *CDKN2A*, *MGMT*, and *RARB*. Consistent with these results, in our study, all 12 CpG loci examined among these five genes had significantly higher methylation in adenocarcinomas after correcting for multiple comparisons. In another study, methylation of *CDH1*, *ESR1*, *PTGS2*, and *RASSF1* had significantly different methylation among normal lung, mesothelioma and adenocarcino-

ma (total $n = 24$), with all gene-loci exhibiting higher methylation in lung adenocarcinoma versus mesothelioma (17). Similarly, in our results, at least one of the two CpG loci investigated in each of these genes had significantly higher methylation in lung adenocarcinoma and none of the CpG loci we examined in these genes had higher methylation in mesothelioma.

Pathway analysis of differentially methylated CpG loci suggested that there is significant, tumor-type-specific enrichment for methylation-based silencing of genes in specific pathways. As tumorigenesis requires somatic inactivation of several pathways, our observations suggest that either the differing etiologic factors or the differential response of the target cells to these factors is driving the mode of pathway inactivation (i.e., epigenetic versus genetic). For example, the enrichment for methylation inactivation of differential cytokine signaling pathway genes (*IL-6 Signaling* in lung adenocarcinoma and *Fc Epsilon Signaling* in mesothelioma) could represent a differential immune-regulated inflammatory response to the primary carcinogens of tobacco smoke and asbestos for these tumors. Furthermore, our group and others have shown that there is an increasing prevalence of DNA methylation of *CDKN2A* with greater smoking duration in lung cancers (30, 31), whereas this gene is often inactivated through homozygous deletion in malignant mesothelioma (32, 33). These results suggest that a preferential mode of inactivation may not be occurring in a gene-specific pattern, but instead represents a broader selection of inactivation by exposure and/or target tissue. Alternatively, but not mutually exclusively, the epigenetic status of the genes in these pathways in the stem cells that give rise to these tissues could differ, contributing to the observed differences between these tumors. More complete detailing of the somatic alterations, including profiles of both genetic and epigenetic alterations would assist in characterizing the relationship between exposures and differential pathway inactivation in these cancers.

Future studies that include treatment and survival data for these patients in their respective diseases may identify specific markers of therapeutic value. Epigenetic alterations associated with overall prognosis could potentially contribute to treatment decisions.

In summary, using CpG methylation profiles, we accurately differentiated mesothelioma from lung adenocarcinoma. This

approach is DNA based, inexpensive, commercially available, and individual samples can be classified by simply comparing to existing RPMM data with an empirical Bayes estimator. Furthermore, random forest is a prediction-based algorithm and can, in principle, be used as the basis for diagnostic software. In addition to characterizing the methylation profiles of these tumors for potential diagnostic use, these data and those of the pathway analysis could aid in understanding variation in patients' response to treatment, and/or the identification of novel, critical therapeutic targets. Finally, beyond the classification of lung adenocarcinoma and mesothelioma, this method may be useful for a range of other clinical scenarios.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 3/22/09; revised 5/12/09; accepted 5/27/09; published OnlineFirst 7/28/09.

Grant support: National Cancer Institute (R01CA126939, R01CA105274, R01CA126831, R01CA52689, P50CA097257), National Institutes of Environmental Health Sciences (T32ES007155, P42ES05947, R01ES006717, P30ES00002), NIEHS/NCI (ES/CA06409), International Mesothelioma Program at Brigham and Women's Hospital (Research grant), Mesothelioma Applied Research Foundation (Research grant).

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

References

- Holloway AJ, Diyagama DS, Opeskin K, et al. A molecular diagnostic test for distinguishing lung adenocarcinoma from malignant mesothelioma using cells collected from pleural effusions. *Clin Cancer Res* 2006;12:5129–35.
- Robinson BW, Lake RA. Advances in malignant mesothelioma. *The New England Journal of Medicine* 2005;353:1591–603.
- Nguyen GK, Akin MR, Villanueva RR, Slatnik J. Cytopathology of malignant mesothelioma of the pleura in fine-needle aspiration biopsy. *Diagn Cytopathol* 1999; 21:253–9.
- Antman KH. Current concepts: malignant mesothelioma. *N Engl J Med* 1980;303:200–2.
- Chang MY, Sugarbaker DJ. Extrapleural pneumonectomy for diffuse malignant pleural mesothelioma: techniques and complications. *Thorac Surg Clin* 2004; 14:523–30.
- Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* 2008;83: 584–94.
- Renshaw AA, Dean BR, Antman KH, Sugarbaker DJ, Cibas ES. The role of cytologic evaluation of pleural fluid in the diagnosis of malignant mesothelioma. *Chest* 1997; 111:106–9.
- Ylagan LR, Zhai J. The value of ThinPrep and cytospin preparation in pleural effusion cytological diagnosis of mesothelioma and adenocarcinoma. *Diagnostic Cytopathology* 2005;32:137–44.
- Marchevsky AM. Application of immunohistochemistry to the diagnosis of malignant mesothelioma. *Archiv Pathol Laboratory Med* 2008;132:397–401.
- Marchevsky AM, Wick MR. Evidence-based guidelines for the utilization of immunostains in diagnostic pathology: pulmonary adenocarcinoma versus mesothelioma. *Appl Immunohistochem Mol Morphol* 2007; 15:140–4.
- Gordon GJ, Jensen RV, Hsiao LL, et al. Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer Res* 2002;62:4963–7.
- Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002;16:6–21.
- Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev* 2002;3:415–28.
- Bibikova M, Lin Z, Zhou L, et al. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* 2006;16:383–93.
- Christensen BC, Houseman EA, Godleski JJ, et al. Epigenetic profiles distinguish pleural mesothelioma from normal pleura and predict lung asbestos burden and clinical outcome. *Cancer Res* 2009;69:227–34.
- Toyooka S, Pass HI, Shivapurkar N, et al. Aberrant methylation and simian virus 40 tag sequences in malignant mesothelioma. *Cancer Res* 2001;61:5727–30.
- Tsou JA, Shen LY, Siegmund KD, et al. Distinct DNA methylation profiles in malignant mesothelioma, lung adenocarcinoma, and non-tumor lung. *Lung Cancer* 2005;47:193–204.
- Wiencke JK, Kelsey KT, Varkonyi A, et al. Correlation of DNA adducts in blood mononuclear cells with tobacco carcinogen-induced damage in human lung. *Cancer Res* 1995;55:4910–4.
- R Development CT. R: A Language and Environment for Statistical Computing. Vienna (Austria): R Foundation for Statistical Computing; 2007.
- Houseman EA, Christensen BC, Marsit CJ, et al. Model-based clustering of DNA methylation array data: a recursive-partitioning algorithm for high-dimensional data arising as a mixture of β distributions. *BMC Bioinformatics* 2008;9:365.
- Hsiung DT, Marsit CJ, Houseman EA, et al. Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* 2007;16:108–14.
- Storey J, Taylor J, Siegmund D. Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *J Royal Stat Soc* 2004;Series B:187–205.
- Brieman L. Random Forests. *Machine Learning* 2001; 45:5–32.
- Ingenuity Pathways Analysis application. 6.3, build 54960 ed: Ingenuity Systems; 2008.
- Shiota K. DNA methylation profiles of CpG islands for cellular differentiation and development in mammals. *Cytogenet Genome Res* 2004;105:325–34.
- Rakyan VK, Hildmann T, Novik KL, et al. DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol* 2004;2:e405.
- Schilling E, Rehli M. Global, comparative analysis of tissue-specific promoter CpG methylation. *Genomics* 2007;90:314–23.
- Christensen BC, Godleski JJ, Marsit CJ, et al. Asbestos exposure predicts cell cycle control gene promoter methylation in pleural mesothelioma. *Carcinogenesis* 2008;29:1555–9.
- Marsit CJ, Kim DH, Liu M, 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.
- Kim DH, Nelson HH, Wiencke JK, 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.
- Toyooka S, Maruyama R, Toyooka KO, et al. Smoke exposure, histologic type and geography-related differences in the methylation profiles of non-small cell lung cancer. *Int J Cancer* 2003;103:153–60.
- Cheng JQ, Jhanwar SC, Klein WM, et al. p16 alterations and deletion mapping of 9p21–22 in malignant mesothelioma. *Cancer Res* 1994;54:5547–51.
- Hirao T, Bueno R, Chen CJ, Gordon GJ, Heilig E, Kelsey KT. Alterations of the p16(INK4) locus in human malignant mesothelial tumors. *Carcinogenesis* 2002;23: 1127–30.