

# Hierarchical Clustering of Lung Cancer Cell Lines Using DNA Methylation Markers<sup>1</sup>

Arvind K. Virmani, Jeffrey A. Tsou,  
Kimberly D. Siegmund, Linda Y. C. Shen,  
Tiffany I. Long, Peter W. Laird, Adi F. Gazdar, and  
Ite A. Laird-Offringa<sup>2</sup>

Hamon Center for Therapeutic Oncology Research and Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9593 [A. K. V., A. F. G.], and Norris Cancer Center and Departments of Surgery and Biochemistry & Molecular Biology [J. A. T., L. Y. C. S., T. I. L., P. W. L., I. A. L.-O.], Preventive Medicine [K. D. S.], and Medicine [L. Y. C. S.], Keck School of Medicine, University of Southern California, Los Angeles, California 90089-9176

## Abstract

Recent analyses of global and gene-specific methylation patterns in cancer cells have suggested that cancers from different organs demonstrate distinct patterns of CpG island hypermethylation. Although certain CpG islands are frequently methylated in many different kinds of cancer, others are methylated only in specific tumor types. Because distinct patterns of CpG island hypermethylation can be seen in tumors from different organs, it seems likely that histological subtypes of cancer within a given organ may exhibit distinct methylation patterns as well. The goal of our study was to determine whether the patterns of CpG island hypermethylation could be used to distinguish between different histological subtypes of lung cancer. We analyzed the methylation status of 23 loci in 91 lung cancer cell lines using the quantitative real-time PCR method MethyLight. Genes *PTGS2* (*COX2*), *CALCA*, *MTHFR*, *ESR1*, *MGMT*, *MYOD1*, and *APC* showed statistically significant differences in the level of CpG island methylation between small cell lung cancer (SCLC) and non-small cell lung cancer cell lines (NSCLC). Hierarchical clustering using a panel consisting of these seven loci yielded two major groups, one of which contained 78% of the SCLC lines. Within this group, a large cluster consisted almost exclusively of SCLC cell lines. Our results show that DNA methylation patterns differ between NSCLC and SCLC cell lines and suggest that these patterns could be

developed into a powerful molecular marker to achieve accurate diagnosis of lung cancer.

## Introduction

The four most frequently occurring types of lung cancer are adenocarcinoma, squamous cell carcinoma, large cell lung cancer, and SCLC<sup>3</sup> (1). They are often grouped as NSCLC<sup>3</sup> and SCLC because of major differences in the therapeutic approach to these patients (2). SCLC, which exhibits many neuroendocrine properties, is generally treated with chemotherapy and has an aggressive clinical course with early widespread metastases, resulting in a mean 5-year survival of ~5% (3, 4). In contrast, NSCLC is much more commonly resected (dependent on tumor stage), with a resulting mean 5-year survival of 15% (3). Because of the important differences in biology, treatment, and prognosis between SCLC and NSCLC, it is imperative that a correct diagnosis be made. However, morphological distinction between these two categories using standard histopathological techniques can be difficult, because a subset of NSCLC can exhibit neuroendocrine features (5). Therefore, it would be of great benefit to obtain molecular markers that could unequivocally distinguish between NSCLC and SCLC.

One potential cancer-specific marker that has recently garnered considerable attention is DNA hypermethylation, which is commonly seen in cancer cells. Methylation of cytosines at CpG dinucleotides is a natural epigenetic modification of DNA that is essential for proper development of mammalian organisms (6). CpG islands are G:C- and CpG-rich regions of ≥200 bp that are generally unmethylated in normal somatic tissues (7). However, in cancer cells, hypermethylation of CpG islands is commonly seen, and it frequently coincides with gene silencing (8–10). Because of this, DNA hypermethylation is an important mechanism by which tumor suppressor genes can be inactivated in cancer (8–10).

Recent analyses of global and gene-specific methylation patterns in cancer cells have suggested that cancers from different organs demonstrate distinct patterns of CpG island hypermethylation (11, 12). Although certain CpG islands are frequently methylated in many different kinds of cancer, others are methylated only in specific tumor types (12, 13). Because distinct patterns of CpG island hypermethylation can be seen in tumors from different organs, it seems likely that histological subtypes of cancer within a given organ may exhibit distinct methylation patterns as well. The goal of our study was to determine whether the patterns of CpG island hypermethylation could be used to distinguish between different histological subtypes of lung cancer. Such tumor type- and tumor subtype-specific signatures could be invaluable tools for accurate cancer diagnosis and early detection and might also shed light on the

Received 8/24/01; revised 11/28/01; accepted 1/10/02.

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.

<sup>1</sup> Supported by Public Health Service Specialized Program of Research Excellence (SPORE) and Developmental Grant 4P50CA7097-0452 from the National Cancer Institute, NIH, Department of Health and Human Services (to A. K. V.), and in part by American Lung Association Grant RG-027-N (to I. A. L.-O.) and a generous donation from Mary Lou and Eri Mettler (to I. A. L.-O. and J. A. T.).

<sup>2</sup> To whom requests for reprints should be addressed, at USC/Norris Cancer Center, Room NOR6420, MC# 9176, 1441 Eastlake Avenue, Los Angeles, CA 90089-9176. Phone: (323) 865-0655; Fax: (323) 865-0158; E-mail: ilaird@hsc.usc.edu.

<sup>3</sup> The abbreviations used are: SCLC, small cell lung cancer; NSCLC, non-small cell lung cancer; PMR, percentage of methylated reference.

possible developmental relationships between tumor subtypes (14). Because SCLC tumor tissues are seldom available for laboratory studies, we limited our study to a large panel of well-characterized cell lines.

## Materials and Methods

**Tumor Cell Lines.** Cell lines were initiated by Gazdar *et al.* (15) at the National Cancer Institute and Hamon Cancer Center, and information regarding the tumor type and characteristics of patients from which they were derived has been described for most of these lines. NSCLC lines (47) and 44 SCLC cell lines were analyzed.

**MethyLight.** DNA was isolated according to standard procedures. Sodium bisulfite conversion, which converts all unmethylated Cs to Us, resulting in their replacement by T in the final PCR product, was performed as described previously (14, 16). Methylation analysis was performed by the fluorescence-based, real-time PCR assay MethyLight (17). Primers and probes (designated by their Human Genome Organization name followed by a probe identifier) were *APC-M1*, *ARF-M1*, *CALCA-M1*, *CHD1-M1*, *CDKN2A-M2*, *CDKN2B-M1*, *CTNNB1-M1*, *ESR1-M1*, *GSTP1-M1*, *HIC1-M1*, *MGMT-M1*, *MLH1-M1*, *MTHFR-M1*, *MYOD1-M1*, *PTGS2-M1*, *RB1-M1*, *TGFBR2-M1*, *TIMP3-M1*, and *TYMS-M1* as described previously (14). Sequences for primers and probes for *AR1-M1*, *ESR2-M1*, *MGMT-M2*, *PGR-M2*, and *THBS1-M1* were as follows (given in 5' to 3' direction): *AR1-M1*: forward primer: GCGTTTTTTCGAGATTTTCGG probe: ACCGTC-CCGCTCTCCCAACAACTA reverse primer: GCCTC-CTCTACCTATAACTACTCCG; *ESR2-M1*: forward primer: TTTGAAATTTGTAGGGCGAAGAGTAG probe: CCGACC-CAACGCTCGCCG reverse primer: ACCCGTCGCAACTC-GAATAA; *MGMT-M2*: forward primer: GCGTTTCGACGT-TCGTAGGT probe: CGCAAACGATACGCACCCGCA reverse primer: CACTCTTCCGAAAACGAAACG (*MGMT-M1* and *-M2* refer to sites sampling different locations of a large CpG island spanning the *MGMT* promoter, exon 1 and intron 1; Ref. 18); *PGR-M2*: forward primer: TTATAATTCGAGGCGGTTA-GTGTTT probe: ATCATCTCCGAAAATCTCAAATCCCA-ATAATACG reverse primer: TCGAATTCTACTAACTCCG-TACTACGA; *THBS1-M1*: forward primer: CGACGCACCA ACCTACCG probe: ACGCGCGCTCACCTCCCT reverse primer: GTTTTGAGTTGGTTTTACGTTTCGTT. Three sets of primers and probes, designed specifically for bisulfite-converted DNA, were used concurrently: a methylation-specific set for the gene of interest and two reference sets to normalize for input DNA (one for  $\beta$ -actin (*ACTB-C1*; Ref. 14) and one for type 2 collagen (*COL2A1-C1*; forward primer: TCTAACAATTATAAACTC-CAACCACCA probe: CCTTCATTCTAACCCAATACCTA-TCCCACCTCTAAA reverse primer: GGGAAAGATGGGATA-GAAGGGAATAT). The PMR<sup>3</sup> for each locus was calculated by dividing the *GENE*:reference ratio of a sample by the *GENE*:reference ratio of highly methylated *SssI*-treated human sperm DNA and multiplying by 100 (14). *GENE* methylation levels were normalized independently using each of the two reference reactions, and the mean of the resulting PMR was used as the final PMR value.

**Statistics.** The PMR values were used as continuous variables, and median methylation levels in NSCLC and SCLC lines were compared using the Mann-Whitney *U* test (SAS Statview software), because the data were not normally distributed. A subset consisting of the seven loci showing the best discrimination between SCLC and NSCLC were selected for the cluster analysis: *PTGS2*, *CALCA*, *MTHFR*, *ESR1*, *MGMT-M1*, *MYOD1*,

and *APC*. Because of the large proportion of cell lines with undetectable methylation, we categorized the PMR values into three groups: (a) no detectable methylation; (b) low levels of methylation; and (c) high levels. At each of the seven loci, the median of the positive PMR values was used to classify methylation values; cell lines with methylation levels less than or equal to the median were classified as low, and cell lines with methylation levels above the median were classified as high. The median was used because it provides an unbiased method for ranking methylation levels for the different genes. The trichotomized data were used for an agglomerative hierarchical cluster analysis performed with S-PLUS 2000 (Insightful Corp.). The distance between two samples was measured by the sum of absolute deviations across the seven loci, commonly known as Manhattan distance. To reflect the biological importance of high methylation values, which may be correlated with gene silencing, greater weight was given to differences between high and low PMR values compared with low and undetectable PMR. This was done by coding categories as follows: (a) 0 = no detectable methylation; (b) 1 = low methylation; and (c) 3 = high methylation. The distance between clusters was measured using the group average method (average link). This corresponds to the average distance across pairs of samples drawn such that no two samples are drawn from the same cluster.

## Results

**Characteristics of NSCLC and SCLC Cell Lines.** We studied a total of 91 lung cancer cell lines, 47 NSCLC and 44 SCLC. The majority of cell lines were derived from advanced stage disease (87% stage 3 or stage 4 in NSCLC and 85% extensive in SCLC). The average age of the patients was 54 years (SD = 10); 87% were Caucasian, and 13% were African-American. There was a ~2:1 ratio of males to females and 3:1 ratio of smokers to nonsmokers. The distribution of these characteristics did not differ by cell type (NSCLC or SCLC). A greater number of SCLC cell lines than NSCLC cell lines were derived from metastatic tumors (93 versus 69%;  $\chi^2 P = 0.01$ ), and more SCLC patients showed either a complete or partial response to therapy (82 versus 35%;  $\chi^2 P = 0.001$ ).

**CpG Island Hypermethylation.** Methylation analysis was performed by the fluorescence-based, real-time PCR assay MethyLight (17). Each set of two primers and the corresponding probe is designed to sample a total of 6–12 CpG dinucleotides located in the area to which the primers and probe hybridize. On average, 8 CpGs are sampled by each primer/probe set. Primers and probes are designed in such a way that they will hybridize only when all CpGs occurring in the area are methylated (17). Thus, the primer/probe sets are designed for the highest level of stringency. This means that a positive result indicates methylation at all CpGs present in the sampled region. A negative result could indicate partial methylation in the sampled area, no methylation in the sampled area, or in rare cases, absence of the sampled area (*e.g.*, through homozygous deletion). Because some heterogeneity can occur (even in cell lines) in the methylation of CpGs within a given tract (*e.g.*, not all 8 CpGs in a sampled region may be methylated in each individual cell), an observation of <100% methylation can be common, although most molecules may be methylated (17). Although this may lead to an underestimation of methylation, it provides a good reflection of the level of extensively methylated alleles. This is important, because extensive methylation is most likely to result in gene silencing and, thus, tends to affect genes whose expression would antagonize tumor growth (8, 10). To increase our odds of detecting hypermethylated loci, we

have chosen the genes analyzed based on their previously shown involvement in cancer, demonstrated potential for tumor suppression, and/or demonstrated propensity to become hypermethylated in cancer. However, it should be noted that for the purpose of developing methylation markers, any methylation occurring consistently and specifically in a given tumor type would be highly relevant, irrespective of its biological consequences.

We used DNA from the 91 cell lines to analyze the methylation status of loci in 23 genes. The PMR<sup>3</sup> (14) values obtained (>2000 data points) were trichotomized and organized to emphasize genes that showed higher methylation in SCLC (Fig. 1A), those that showed higher methylation in NSCLC (Fig. 1B), those that showed substantial methylation but were indistinguishable in both kinds of cells (Fig. 1C), and those showing sporadic (Fig. 1D) or undetectable methylation (Fig. 1E). PMR values were used to determine whether statistically significant differences in methylation levels and percentage of cell lines showing methylation were seen between SCLC and NSCLC cell lines (Table 1). *PTGS2*, *CALCA*, *MTHFR*, *ESR1*, *MGMT-M1*, *MYOD1*, and *APC* showed statistically significant differences in the median methylation levels in NSCLC and SCLC, and *TIMP3*, *TGFBR2*, and *CDKN2A* approached statistical significance. The highest significant difference was seen for *PTGS2* ( $P < 0.001$ ), which showed a significantly higher proportion of SCLC cell lines methylated ( $P = 0.03$  using Fisher's exact test) and significantly higher median methylation levels in SCLC cell lines. The higher median methylation levels were observable when all cell lines were taken into account but also when only cell lines showing positive methylation were considered. Thus, *PTGS2* (*COX2*) appears to be a good indicator of SCLC, showing both qualitative and quantitative differences in methylation between SCLC and NSCLC cell lines. Qualitative and quantitative differences were also seen for *MYOD1*, whereas only quantitative differences were seen for *CALCA*, *MTHFR*, *ESR1*, *MGMT-I*, and *APC*. The methylation patterns of the rest of the loci were similar in both SCLC and NSCLC lines. *HIC1* showed very frequent methylation in both cell types. *TGFBR2*, *THBS1*, and *TYMS* showed sporadic methylation in various lines. *ARF*, *CTNNB1*, and *RB1* genes showed no detectable methylation in any of the lung cancer cell lines. These results indicate that there can be considerable differences in methylation levels of CpG islands in cell lines, varying from very highly methylated to undetectable methylation. In addition, differences that are highly statistically significant can be observed.

**Hierarchical Clustering Using a Panel of Seven Loci.** Although a number of the loci analyzed might be used individually to discern differences between tumor subtypes, correct identification of tumor samples is more likely to occur when multiple informative markers are used. For this reason, we wished to determine whether a combination of loci might serve better as a diagnostic panel. On the basis of the fact that 7 of the 23 loci showed significant ( $P < 0.05$ ) differences in methylation levels (Table 1), we used these loci as a panel in an agglomerative hierarchical cluster analysis (Fig. 2). The cell lines fell into two major groups, one of which contained 78% of the SCLC cell lines. Within the major SCLC group, one of the two subgroups consisted almost entirely of SCLC (91%), containing only two cell lines that were not SCLC lines (an atypical carcinoid line and a bronchioalveolar carcinoma cell line). Among NSCLC cell lines, grouping of subtypes was less obvious. Some clustering of adenocarcinomas was seen, but the current set of markers does not allow this group to be clearly

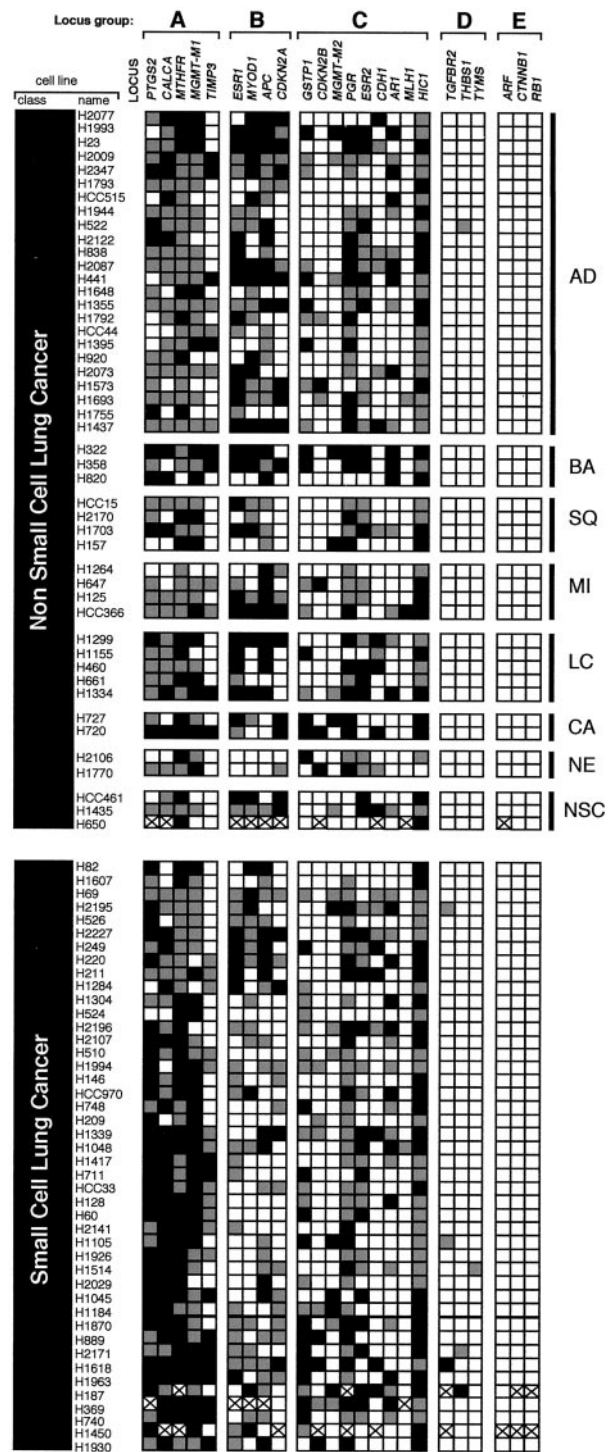


Fig. 1. Methylation analysis of a panel of 23 genes in 47 NSCLC and 44 SCLC cell lines. Cell line names are indicated at left. Cell lines with the prefix HCC- originated at Hammond Cancer Center; all others are from NIH (the prefix NIH- has been omitted for brevity). Loci are listed at the top. *MGMT-M1* and *-M2* refer to sites sampling different locations of a large CpG island spanning the *MGMT* promoter, exon 1, and intron 1 (18). NSCLC cell lines were derived from samples classified as adenocarcinoma (AD), bronchioalveolar carcinoma (BA), squamous cell lung cancer (SQ), mixed histology (MI), large cell lung cancer (LC), carcinoids (CA), neuroendocrine NSCLC (NE), and unclassified NSCLC (NSC). Methylation values are marked as □ (undetectable), gray boxes (PMR below the median), or ■ (PMR above the median). Samples for which no PMR was obtained because of failure of one or both references are indicated by a crossed box.



Table 1 Comparison of methylation levels in NSCLC and SCLC cell lines

Gene (HUGO <sup>a</sup> name)	Non-small cell lung cancer (n = 47) <sup>b</sup>			Small cell lung cancer (n = 44) <sup>d</sup>			P <sup>e</sup>
	% pos.	Median	IQR <sup>c</sup>	% pos.	Median	IQR <sup>c</sup>	
<i>PTGS2</i>	78	1.2	0.2–5.3	95	29.5	1.6–55.7	<0.0001*
<i>CALCA</i>	76	19.7	6.8–65.6	86	87.9	48.7–129.0	0.0004*
<i>MTHFR</i>	98	75.8	32.2–111.5	100	117.4	89.3–148.6	0.0001*
<i>ESR1</i>	74	40.5	2.5–79.0	56	1.4	0.5–17.3	0.0024*
<i>MGMT-M1</i>	81	3.7	1.0–9.5	91	7.8	2.6–19.4	0.0145*
<i>MYOD1</i>	65	12.2	0.8–61.2	42	3.8	0.5–13.2	0.0133
<i>APC</i>	74	79.7	65.6–107.1	58	36.0	1.2–103.7	0.0441
<i>TIMP3</i>	30	68.3	0.6–108.0	50	40.3	10.2–94.0	0.0571
<i>TGFBR2</i>	0			7	0.2	0.1–1.8	0.0658
<i>CDKN2A</i>	48	34.4	0.5–47.9	37	0.4	0.0–5.0	0.1019
<i>GSTP1</i>	40	112.3	79.3–227.7	59	72.4	14.1–173.6	0.2894
<i>TYMS</i>	0			2	0.1	0.1–0.1	0.3118
<i>PGR</i>	81	12.9	2.8–63.8	88	2.5	0.4–42.9	0.3312
<i>CDH1</i>	28	5.9	1.6–13.5	21	9.6	2.6–35.6	0.4980
<i>CDKN2B</i>	13	4.45	2.8–6.5	19	1.6	0.5–45.8	0.5046
<i>THBS1</i>	2	1.4	1.4–1.4	5	23.7	22.0–25.4	0.5127
<i>MGMT-M2</i>	17	31.5	8.2–105.6	23	28.7	3.2–79.4	0.5489
<i>HIC1</i>	98	89.9	63.4–117.1	100	95.7	77.0–112.7	0.5827
<i>MLH1</i>	4	33.7	16.9–50.5	2	63.7	63.7–63.7	0.6086
<i>ESR2</i>	62	3.5	0.1–19.4	52	13.4	0.4–51.3	0.6461
<i>ARI</i>	40	7.4	0.8–47.8	45	3.0	0.1–25.5	0.8568
<i>ARF</i>	0			0			1
<i>CTNNB1</i>	0			0			1
<i>RBI</i>	0			0			1

<sup>a</sup> Human Genome Organization nomenclature (*MGMT-M1* and *-M2* refer to sites sampling different locations of a large CpG island spanning the *MGMT* promoter, exon 1, and intron 1, Ref. 18).

<sup>b</sup> NSCLC lines were derived from 24 adenocarcinomas, three bronchioalveolar carcinomas, four squamous cell lung cancers, four mixed subtype cancers, five large cell carcinomas, two carcinoids, three unspecified NSCLC, and two NSCLC with neuroendocrine features. Ten loci lacked data on one NSCLC cell line (*PTGS2*, *CALCA*, *ESR1*, *MYOD1*, *APC*, *CDKN2A*, *CDKN2B*, *CDH1*, *MLH1*, and *ARF*).

<sup>c</sup> IQR, interquartile range (25<sup>th</sup>–75<sup>th</sup> percentile).

<sup>d</sup> Ten loci lacked data on one SCLC cell line (*PTGS2*, *CALCA*, *ESR1*, *MYOD1*, *APC*, *CDKN2A*, *CDKN2B*, *CDH1*, *MLH1*, and *ARF*); 5 loci lacked data on two SCLC cell lines (*MTHFR*, *TGFBR2*, *PGR*, *CTNNB1*, and *RBI*).

<sup>e</sup> Mann-Whitney *U* test. Loci showing significant ( $P < 0.05$ ) differences in methylation levels are highlighted in *bold type*. Of these, loci for which a significant difference ( $P < 0.05$ ) was found using only the median methylation levels of positive cell lines are marked by an *asterisk*.

separated from other NSCLCs. This is supported by the fact that only 2 of the 23 loci show significant differences between adenocarcinoma and all other NSCLCs: (a) *MTHFR* (Mann-Whitney  $P = 0.01$ ); and (b) *ARI* (Mann-Whitney  $P = 0.04$ ). To improve the distinctions within the NSCLC group, expanding the number of loci analyzed would be required.

To test whether the clustering might be due to certain characteristics of the cell lines (stage of tumor from which the line was developed or gender, race, and smoking history of the patient), we tested whether there was a statistically significant association between our derived cluster variable and the patient characteristics. No such correlation was found (all  $P > 0.10$ ), suggesting that the clustering does indeed reflect the histological origin of the cancer rather than any other patient characteristic. This was supported by the fact that the median methylation level did not appear to vary by patient characteristic in any of the seven loci used for the cluster analysis. The association of the cluster variable with cell type described earlier was highly significant ( $P = 4.9 \times 10^{-7}$ ). Response to therapy also showed a statistically significant difference between the clusters ( $P = 0.02$ ), but this can be explained by the fact that the different cell types differed with respect to response to therapy. Using logistic regression, we found that after adjusting for cell type, response to therapy no longer predicted the clustering variable. This further supports our conclusion that clustering reflects the type of cancer rather than any of the patient characteristics.

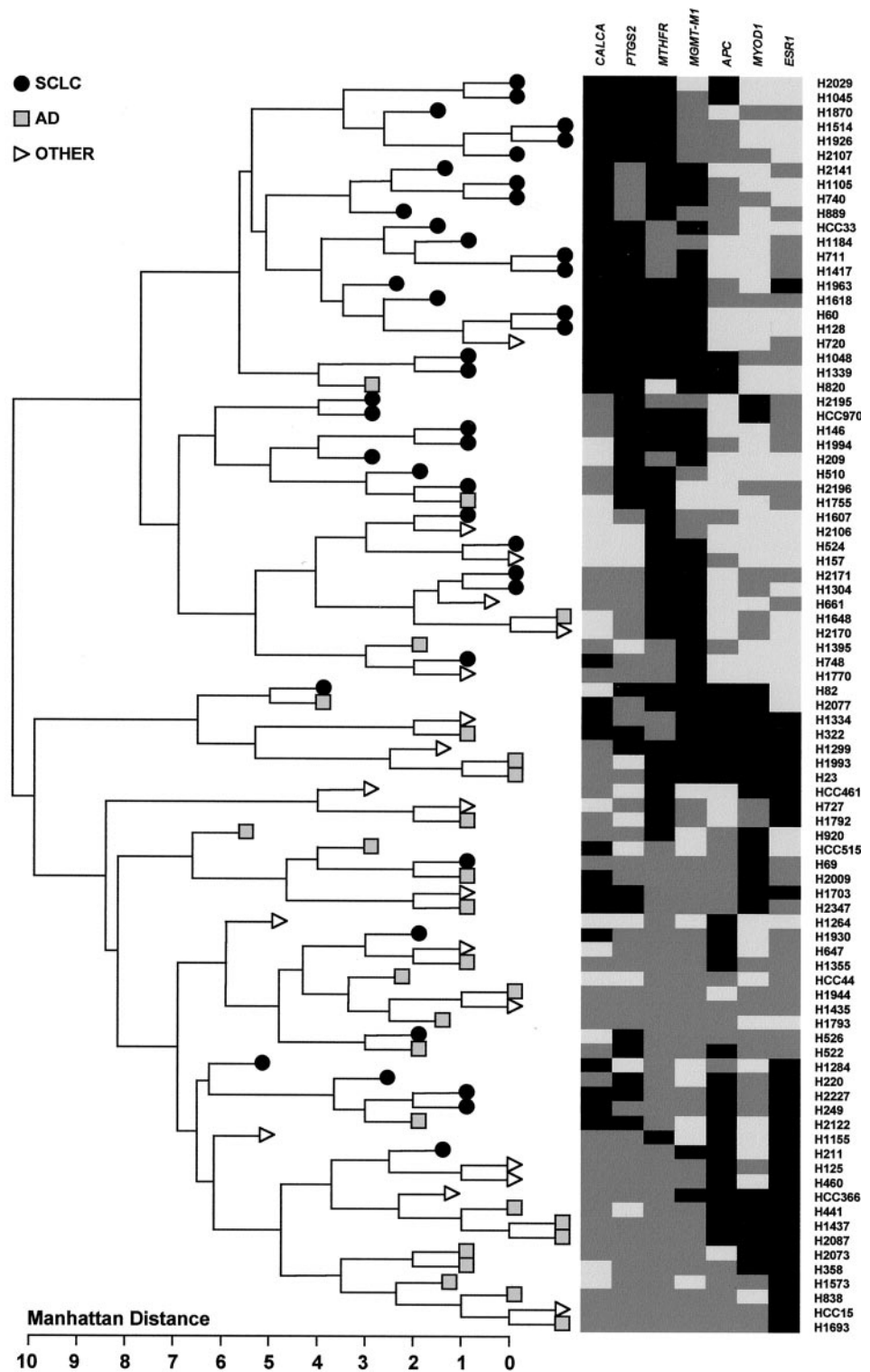
Including additional, less informative loci in the panel

used for the clustering reduced the propensity of the SCLCs to group together, whereas reducing the number of loci led to a loss of information and concomitant loss of clustering. This requirement for an optimal number of genes to most clearly observe relationships between samples has been observed previously, e.g., when using artificial neuronal networks to classify cells based on gene expression levels (19). Although our approach was a selective one, it demonstrates the ability to observe clusters based on methylation data. It will be important to validate the utility of the seven marker loci using an independent set of cell lines and/or tumor samples in the future.

## Discussion

Here we describe studies aimed at determining whether CpG island hypermethylation patterns could be used to distinguish between different histological subtypes of lung cancer. Because SCLC is usually diagnosed by transbronchial biopsy or cytology, and resections are rare, tumor material from SCLC patients is scarce. For this reason, we have used a collection of NSCLC and SCLC cell lines. These cell lines were established from primary or metastatic tumors precisely to address the shortage of available tumor material (15). Extensive analysis has shown that in general, our lung cancer cell lines maintain the original characteristics of the tumor remarkably well (15, 20, 21). Early experimenters had questioned the validity of using cell lines to study DNA methylation, because it was reported that immortalized mouse fibroblasts exhibited hypermethylation (22).

Fig. 2. Clustering analysis of cell lines based on methylation data from seven loci. A dendrogram obtained from agglomerative hierarchical clustering analysis, using data from 89 cell lines for which complete methylation data were available for the seven genes that showed statistically significant differences (Table 1), is shown at *left*. To calculate the Manhattan distance, cell lines were assigned a value of 0, 1, or 3 for each locus, denoting no detectable methylation, a PMR value below the median, or a PMR value above the median, respectively. The higher value assigned to the latter group was used to reflect the biological importance of high methylation levels, which are correlated with gene silencing. ●, the cell lines derived from SCLC. Gray squares, cell lines derived from adenocarcinoma or bronchiolaveolar carcinoma (grouped together because they are histologically indistinguishable). △, other NSCLC cell lines. A diagram depicting the methylation levels corresponding to the clustered cell lines is shown at *right*. Methylation levels for the seven loci used for clustering (listed at *top*) are given, indicating undetectable (light gray), below median (dark gray), or above median methylation levels (black).



However, our analysis shows that a number of loci showed low or no detectable methylation in any of the 91 cell lines studied, arguing against a universal hypermethylation of CpG islands in cell lines. A recent study of bladder cancer tumors and the cell lines derived from them showed relatively stable methylation patterns *in vivo* and *in vitro* over time (23).

Of the 23 loci we have tested, only *APC* has been studied previously in SCLC (24). Using nonquantitative methylation-specific PCR, we previously saw methylation in 26% of SCLC cell lines and 59% of NSCLC cell lines. The somewhat higher percentages found in the current study (58 and 74%, respectively) are likely due to the increased sensitivity of the Meth-

yLight method. Thus far, none of the 23 loci have been studied in SCLC tumor samples. Thus, it is not yet clear whether the values obtained for SCLC cell lines match those that would be found in SCLC tumors. However, the most commonly studied methylation site in lung cancer to date (the *CDKN2A*/p16 promoter, tested in NSCLC in  $\geq 10$  studies, among others, Refs. 25–28) showed methylation in an average of 40% of NSCLC tumor samples analyzed, which agrees well with the 47% of NSCLC cell lines positive in our study. Furthermore, recent studies of methylation of *RARB*, *FHIT*, and *CDH13* showed similar methylation levels in lung cancer tumors and cell lines (29–31). An important advantage of cell lines is that they offer a virtually inexhaustible source of DNA. Thus, the analysis of cell lines might be a very useful first step in the identification of potentially informative methylation markers, allowing the prescreening of very large numbers of DNA methylation loci. Once potentially informative loci have been identified, the utility of these loci as epigenetic markers can be determined using much more scarcely available tumor material. Naturally, it will be important to validate panels of markers identified in this manner using large collections of tumor specimens and normal tissue.

Our observation that 3 of the 23 loci show no detectable methylation is of interest, in particular, because one of those loci lies in the promoter of the *RBI* gene, which is very frequently inactivated in lung cancer (32). This suggests that the gene is primarily inactivated by other mechanisms, such as deletion and mutation. Another interesting observation is the fact that the *M1* site in *MGMT* was informative ( $P = 0.0145$ ), whereas the *M2* site, located  $\sim 300$  nucleotides upstream in the same large CpG island (which spans the promoter, exon 1, and intron 1), showed less frequent methylation and was not statistically different between NSCLC and SCLC cell lines. This supports the observation that methylation patterns within a given CpG island may not be uniform (33) and suggests that when trying to establish epigenetic signatures, it may be wise to sample large CpG islands at different locations.

Estimates for the number of CpG islands vary from 28,876 to 195,706, depending on the exact definition used (34, 35). Assuming the homogeneous presence or absence of methylation at each CpG island, this offers at least  $2^{28,876}$  different combinations of methylation patterns that could be found in cancer cells. However, as indicated above, methylation within a given CpG island may not be uniform, so that the number of informative sites may be larger than the number of CpG islands. In addition, a given tumor may consist of a mixed population of cells, yielding differences in the observed levels of methylation. Such quantitative differences in methylation at a given locus may also be informative, as indicated by the highly statistically significant differences in methylation levels seen for *PTGS2*, *CALCA*, and *MTHFR*. The difference in methylation levels of the latter two genes, detected with the MethyLight assay, would not have been revealed with the routinely used methylation-specific PCR method. (*CALCA* and *MTHFR* showed no significant difference in the percentage of NSCLC versus SCLC cell lines methylated using Fisher's exact test.) Thus, methylation patterns obtained with the MethyLight assay offer an epigenetic readout for cancer cells that approaches gene expression studies in its potential wealth of information. However, three important advantages distinguish the utility of methylation versus expression analyses for studies of cancer cells: (a) methylation patterns are present in DNA molecules, which are much more stable than RNA, increasing their ability to be detected in biologically labile samples, such as archival tissue and serum; (b) methylation analyses are compatible with routine clinical

processing procedures, such as formalin fixation (14); and (c) methylation analyses may be more selective for cancer-specific changes than expression studies, because promoter CpG islands are usually unmethylated in normal tissues, so that observed hypermethylation is frequently significant. This is supported by our ability to observe clustering using 7 loci out of a set of merely 23, which is remarkable, considering the large numbers of genes (thousands) routinely analyzed to define clusters in gene expression assays. Our study is, to our knowledge, the first to use methylation analyses to develop clusters. Our results suggest that with the inclusion of sufficient informative loci, it will become possible to classify lung cancer subtypes with a high degree of accuracy using methylation profiles.

### Acknowledgments

We thank Dr. Jeffrey Hagen and members of the Laird-Offringa lab for helpful criticisms of the manuscript and Binh Trinh for help with the figures.

### References

- WHO. Histological typing of lung tumors, Ed. 2. Geneva: WHO, 1982.
- Travis, W. D., Linder, J., and Mackay, B. Classification, histology, cytology, and electron microscopy. In: H. I. Pass, J. B. Mitchell, D. H. Johnson, and A. T. Turrisi (eds.), Lung Cancer: Principles and Practice, pp. 361–395. Philadelphia: Lippincott-Raven Publishers, 1996.
- NCI. In: L. A. G. Ries, B. A. Miller, B. F. Hankey, C. L. Kosary, A. Harras, and B. K. Edwards (eds.), SEER cancer statistics review, 1973–1991: tables and graphs, Vol. Pub. No. 94–2789. Bethesda, MD: NIH, 1994.
- Viallet, J., and Ihde, D. C. Small cell carcinoma of the lung: clinical and biological aspects. Crit. Rev. Oncol. Hematol., 11: 109–135, 1991.
- Travis, W. D., Gal, A. A., Colby, T. V., Klimstra, D. S., Falk, R., and Koss, M. N. Reproducibility of neuroendocrine lung tumor classification. Hum. Pathol., 29: 272–279, 1998.
- Li, E., Bestor, T. H., and Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell, 69: 915–926, 1992.
- Bird, A. P. CpG-rich islands and the function of DNA methylation. Nature, 321: 209–213, 1986.
- Jones, P. A., and Laird, P. W. Cancer epigenetics coming of age. Nat. Genet., 21: 163–167, 1999.
- Robertson, K., and Wolffe, A. DNA methylation in health and disease. Nat. Rev. Genet., 1: 11–19, 2000.
- Herman, J. G., and Baylin, S. B. Promoter-region hypermethylation and gene silencing in human cancer. Curr. Top. Microbiol. Immunol., 249: 35–54, 2000.
- Costello, J. C., Fruhwald, M. C., Smiraglia, D. J., Rush, L. J., Robertson, G. P., Gao, X., Wright, F. A., Feramisco, J. D., Peltomaki, P., Lang, J. C., Schuller, D. E., Yu, L., Bloomfield, C. D., Caligiuri, M. A., Yates, A., Nishikawa, R., Huang, H. J. S., Petrelli, N. J., Zhang, X., O'Dorisio, M. S., Held, W. A., Cavenee, W. K., and Plass, C. Aberrant CpG-island methylation has non-random and tumor-type-specific patterns. Nat. Genet., 25: 132–138, 2000.
- Esteller, M., Corn, P. G., Baylin, S. B., and Herman, J. G. A gene hypermethylation profile of human cancer. Cancer Res., 61: 3225–3229, 2001.
- Goessl, K., Krause, H., Muller, M., Heicappell, R., Schrader, M., Sachsinger, J., and Miller, K. Fluorescent methylation-specific polymerase chain reaction for DNA-based detection of prostate cancer in bodily fluids. Cancer Res., 60: 5941–5945, 2000.
- Eads, C. A., Lord, R. V., Wickramasinghe, K., Long, T. I., Kurumboor, S. K., Bernstein, L., Peters, J. H., DeMeester, S. R., DeMeester, T. R., Skinner, K. A., and Laird, P. W. Epigenetic patterns in the progression of esophageal adenocarcinoma. Cancer Res., 61: 3410–3418, 2001.
- Phelps, R. M., Johnson, B. E., Ihde, D. C., Gazdar, A. F., Carbone, D. P., McClintock, P. R., Linnoila, R. I., Matthews, M. J., Bunn, P. A., Jr., Carney, D., Minna, J. D., and Mulshine, J. L. NCI-Navy Medical Oncology Branch cell line data base. J. Cell. Biochem. Suppl., 24: 32–91, 1996.
- Olek, A., Oswald, J., and Walter, J. A modified and improved method for bisulphite based cytosine methylation analysis. Nucleic Acids Res., 24: 5064–5066, 1996.
- Eads, C. A., Danenberg, K. D., Kawakami, K., Saltz, L. B., Blake, C., Shibata, D., Danenberg, P. V., and Laird, P. W. MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res., 28: E32, 2000.
- Patel, S. A., Graunke, D. M., and Pieper, R. O. Aberrant silencing of the CpG island-containing human *O6*-methylguanine DNA methyltransferase gene is as-

- sociated with the loss of nucleosome-like positioning. *Mol. Cell. Biol.*, *17*: 5813–5822, 1997.
19. Khan, J., Wei, J. S., Ringer, M., Saal, L. H., Ladanyi, M., Westerman, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C., and Meltzer, P. S. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neuronal networks. *Nat. Med.*, *7*: 673–679, 2001.
  20. Hay, R. J. Human cells and cell cultures: availability, authentication and future prospects. *Hum. Cell.*, *9*: 143–152, 1996.
  21. Wistuba, I. I., Bryant, D., Behrens, C., Milchgrub, S., Virmani, A. K., Ashfaq, R., Minna, J. D., and Gazdar, A. F. Comparison of features of human lung cancer cell lines and their corresponding tumors. *Clin. Cancer Res.*, *5*: 991–1000, 1999.
  22. Antequera, F., Boyes, J., and Bird, A. High levels of *de novo* methylation and altered chromatin structure at CpG islands in cell lines. *Cell*, *62*: 503–514, 1990.
  23. Markl, I. D. C., Cheng, J., Liang, G., Shibata, D., Laird, P. W., and Jones, P. A. Global and gene-specific epigenetic patterns in human bladder cancer genomes are relatively stable *in vivo* and *in vitro* over time. *Cancer Res.*, *61*: 5875–5884, 2001.
  24. Virmani, A. K., Rathi, A., Sathyanarayana, U. G., Padar, A., Huang, C. X., Cunningham, H. T., Farinas, A. J., Milchgrub, S., Euhus, D. M., Gilcrease, M., Herman, J. G., Minna, J. D., and Gazdar, A. F. Aberrant methylation of the Adenomatous Polyposis Coli (APC) gene promoter 1A in breast and lung carcinomas. *Clin. Cancer Res.*, *7*: 1998–2001, 2001.
  25. Belinsky, S. A., Nikula, K. J., Palmisano, W. A., Michels, R., Saccomanno, G., Gabrielson, E., Baylin, S. B., and Herman, J. G. Aberrant methylation of p16 INK4a is an early event in lung cancer and potential biomarker for early diagnosis. *Proc. Natl. Acad. Sci. USA*, *95*: 11891–11896, 1998.
  26. Sanchez-Cespedes, M., Reed, A. L., Buta, M., Wu, L., Westra, W. H., Herman, J. G., Yang, S. C., Jen, J., and Sidransky, D. Inactivation of the INK4a/ARF locus frequently coexists with TP53 mutations in non-small cell lung cancer. *Oncogene*, *18*: 5843–5849, 1999.
  27. Ahrendt, S. A., Chow, J. T., Xu, L. H., Yang, S. C., Eisenberger, C. F., Esteller, M., Herman, J. G., Wu, L., Decker, P. A., Jen, J., and Sidransky, D. Molecular detection of tumor cells in bronchoalveolar lavage fluid from patients with early stage lung cancer. *J. Natl. Cancer Inst.* (Bethesda), *91*: 332–339, 1999.
  28. Zochbauer-Muller, S., Fong, K. M., Virmani, A. K., Geradts, J., Gazdar, A. F., and Minna, J. D. Aberrant promoter methylation of multiple genes in non-small cell lung cancers. *Cancer Res.*, *61*: 249–255, 2001.
  29. Virmani, A. K., Rathi, A., Zochbauer-Muller, S., Sacchi, N., Fukuyama, Y., Bryant, D., Maitra, A., Heda, S., Fong, K. M., Thunnissen, F., Minna, J. D., and Gazdar, A. F. Promoter methylation and silencing of the retinoic acid receptor- $\beta$  gene in lung carcinomas. *J. Natl. Cancer Inst.* (Bethesda), *92*: 1303–1307, 2000.
  30. Zochbauer-Muller, S., Fong, K. M., Maitra, A., Lam, S., Geradts, J., Ashfaq, R., Virmani, A. K., Milchgrub, S., Gazdar, A. F., and Minna, J. D. 5' CpG island methylation of the *FHIT* gene is correlated with loss of gene expression in lung and breast cancer. *Cancer Res.*, *61*: 3581–3585, 2001.
  31. Toyooka, K. O., Toyooka, S., Virmani, A. K., Sathyanarayana, U. G., Euhus, D. M., Gilcrease, M., Minna, J. D., and Gazdar, A. F. Loss of expression and aberrant methylation of the *CDH13* (H-cadherin) gene in breast and lung cancer. *Cancer Res.*, *61*: 4556–4560, 2001.
  32. Weintraub, S. J. Inactivation of tumor suppressor proteins in lung cancer. *Am. J. Respir. Cell Mol. Biol.*, *15*: 150–155, 1996.
  33. Zheng, S., Chen, A., McMillan, A., Lafuente, A., Lafuente, M. J., Ballesta, A., Trias, M., and Wiencke, J. K. Correlations of partial and extensive methylation at the p14(ARF) locus with reduced mRNA expression in colorectal cancer cell lines and clinicopathological features in primary tumors. *Carcinogenesis*, *21*: 2057–2064, 2000.
  34. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., and D., G. J. The sequence of the human genome. *Science*, *291*: 1304–1351, 2001.
  35. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., and Gage, D. Initial sequencing and analysis of the human genome. *Nature*, *409*: 860–921, 2001.