Integrated analysis of genetic and epigenetic alterations reveals CpG island methylator phenotype associated with distinct clinical characteristics of lung adenocarcinoma

Keiko Shinjo1,2, Yasuyuki Okamoto1, Byonggu An1, Yoshihiko Yokoyama2, Ichiro Takeuchi2, Makiko Fujii1, Hirotaka Osada1,2, Noriyasu Usami2, Yoshinori Hasegawa3, Hidemi Ho3, Toyoaki Hida7, Nobukazu Fujimoto8, Takumi Kishimoto8, Yoshitaka Sekido1,2 and Yutaka Kondo1,9,

1Division of Molecular Oncology, Aichi Cancer Center Research Institute, 1-1 Kanokoden, Chikusa-ku, Nagoya 464-8681, Japan, 2Department of Cancer Genetics, 3Department of Respiratory Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan, 4Graduate School of Engineering, Nagoya Institute of Technology, Nagoya, Japan, 5Division of General Thoracic Surgery, Nagoya University Graduate School of Medicine, Nagoya, Japan, 6Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan, 7Department of Thoracic Oncology, Aichi Cancer Center Hospital, Nagoya, Japan, 8Department of Respiratory Medicine, Okayama Rosai Hospital, Okayama, Japan, and 9Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency, Saitama, Japan

DNA methylation affects the aggressiveness of human malignancies. Cancers with CpG island methylator phenotype (CIMP), a distinct group with extensive DNA methylation, show characteristic features in several types of tumors. In this study, we initially defined the existence of CIMP in 41 lung adenocarcinomas (AdCas) through genome-wide DNA methylation microarray analysis. DNA methylation status of six CIMP markers newly identified by microarray analysis was further estimated in a total of 128 AdCas by bisulfite pyrosequencing analysis, which revealed that 10 (7.8%), 40 (31.3%) and 78 (60.9%) cases were classified as CIMP-high (CIMP-H), CIMP-low and CIMP-negative (CIMP-N), respectively. Notably, CIMP-H AdCas were strongly associated with wild-type epidermal growth factor receptor (EGFR), males and heavy smokers (P = 0.0089, P = 0.0047 and P = 0.0036, respectively). In addition, CIMP-H was significantly associated with worse prognosis; especially among male smokers, CIMP-H was an independent prognostic factor (hazard ratio 1.7617, 95% confidence interval 1.0030–2.9550, P = 0.0489). Compellingly, the existence of CIMP in AdCas was supported by the available public datasets, such as data from the Cancer Genome Atlas. Intriguingly, analysis of AdCa cell lines revealed that CIMP-positive AdCa cell lines were more sensitive to a DNA methylation inhibitor than CIMP-N ones regardless of EGFR mutation status. Our data demonstrate that CIMP in AdCas appears to be a unique subgroup that has distinct clinical traits from other AdCas. CIMP classification using our six-marker panel has implications for personalized medical strategies for lung cancer patients; in particular, DNA methylation inhibitor might be of therapeutic benefit to patients with CIMP-positive tumors.

Introduction

Lung cancer is the leading cause of human cancer death worldwide (1). Recent targeted therapies have improved the survival of patients with certain types of lung cancer, especially adenocarcinoma (AdCa). EGFR mutations are generally less sensitive to these targeted therapies than tumors with mutations. Given the evidence that the frequency of EGFR mutations accounts for up to 30% of AdCas, and even EML4–ALK fusions are found in AdCas albeit at a lower frequency, elucidating the underlying mechanisms other than such gene alterations in lung carcinogenesis is desirable to facilitate the development of new strategies for lung cancer treatment.

Studies have shown that in addition to genetic alterations, accumulation of epigenetic alterations play an important role in tumorigenesis of lung cancer (4). DNA methylation, an important epigenetic factor, affects the chromatin structure and is closely associated with gene regulation (5). Simultaneous dysregulation of multiple genes, including those involved in cell cycle, cell growth, cell death or cell adhesion, by DNA methylation may be a strong driving force to undergo transformation, sometimes in correlation with potentiated aggressiveness of the tumors (6).

Recent studies in colon cancer have shown that a subset of tumors suffer from a remarkably high rate of aberrant promoter DNA methylation at a large number of loci, referred to as CpG island methylator phenotype (CIMP) (7). CIMP tumors in colon exhibit distinct genetic and clinical features, such as high rates of BRAF and KRAS mutations, low frequency of TP53 mutation, specific histology, proximal location and characteristic clinical outcome, suggesting that CIMP-related cancers may proceed through a unique pathway (8). In lung cancer, some studies have shown the existence of a subgroup of tumors with the characteristic methylation status of CIMP (9–13). However, in comparison with the considerable research of CIMP markers performed in colon cancers (7,14–16), no studies have assessed which DNA methylation markers can predict the most extensively methylated subgroups (i.e. CIMP) in lung AdCas due to the lack of accompanied genome-wide DNA methylation analysis in multiple samples. Since different panels of markers may lead to different classification of lung cancer (9,11), it is important to define markers that can accurately identify CIMP.

To examine whether CIMP exists as a characteristic subgroup in AdCas, we initially performed global screening for genes with aberrant DNA hypermethylation by the methylated CpG island amplification microarray (MCAM) analysis, which provides reproducible results with a high validation rate (16–20). Using the six CIMP markers newly identified by MCAM analysis, we characterized a distinct subgroup of AdCas exhibiting CIMP. In addition, several AdCa cell lines with different CIMP status were treated with a DNA methylation inhibitor, and the relationship between DNA methylation status and drug sensitivity was assessed. Our data provide evidence for a new strategy for lung cancer treatment.

Materials and methods

Cell lines

A549 was purchased from the American Type Culture Collection (Rockville, MD) and PC9 was purchased from Immuno-Biological Laboratories (Fujioka, Gunma, Japan). NCI-H23, NCI-H358, NCI-H920, NCI-H2009, NCI-H1573, NCI-H1650 and HCC4011 were kind gifts from Dr Adi F. Gazdar (University of Texas Southwestern Medical Center, Dallas, TX) and Dr Mitsuo Sato (Nagoya University Graduate School of Medicine, Nagoya, Japan). Y-ML13 (ML13) and ACC-H1 (H1) were established in our institute. All cell lines were maintained in RPMI-1640 medium (Sigma–Aldrich, St Louis, MO) supplemented with 10%
fetal bovine serum (Invitrogen, Carlsbad, CA) and antibiotic–antimycotic reagent (Invitrogen) at 37 °C in a humidified incubator with 5% CO₂.

**5-Aza-dC treatment of cells**

Cells were treated with 50 nM–1 μM 5-aza-2'-deoxycytidine (5-Aza-dC; Sigma–Aldrich) as described previously (17) or the TKI AG1478 (Calbiochem, San Diego, CA) for 72 h. DNA was extracted on the seventh day following treatment. Changes in proliferation were determined by using the TetraColor ONE (Seikagaku, Tokyo, Japan) system, containing 2-(2-methoxy-4-nitropheryl)-3-(4-nitropheryl)-5-(2,4-disulfophenyl)-2H-tetrazolium, monosodium salt and 1-methoxy-5-methylphenazinium methylsulfate as the electron carrier. After 1 h incubation at 37 °C, absorbance was read at 450 nm with a multi-plate reader. Growth inhibition was expressed as a mean ratio of absorbance reading from treated versus untreated cells. Cell numbers were also counted under a light microscope at the same time point.

**Tissue samples**

We collected 128 AdCa samples and 26 normal lung tissues from patients who underwent surgical resection at the Aichi Cancer Center Central Hospital, Okayama Rosai Hospital, Nagoya University Hospital and its affiliated hospitals in Japan, in accordance with the institutional policy. Samples and clinical data were collected after appropriate institutional review board approval was received and written informed consent had been obtained from all the patients. Histological and cytological examination of normal lung tissues, which were obtained from the lung cancer patients, revealed no remarkable findings as malignant tissues. In the normal tissues, no aberrant methylation was detected in 10 genes by pyrosequencing analysis (Supplementary Table 1, available at Carcinogenesis Online). A sample size of 124 patients was calculated to be sufficient to provide a survival rate difference of 25% with a significance level of 0.05 and power of 80%, when the frequency of CIMP-AdCa was estimated to be ~20% as observed in colorectal cancer (8); therefore, we collected 128 AdCa to analyze the significance of CIMP (Table I).

**DNA preparation**

Genomic DNA was extracted using a standard phenol–chloroform method. Fully methylated DNA was prepared by treating genomic DNA with SsI methylase (New England Biolabs, Ipswich, MA), and unmethylated DNA was prepared by treating genomic DNA with phi29 DNA polymerase (Genomi-Phi DNA Amplification Kit; Amersham Biosciences, Uppsala, Sweden) according to the manufacturers’ protocol.

**DNA methylation analysis**

We performed bisulfite treatment as described previously (21,22). The DNA methylation levels were measured using pyrosequencing technology (Pyrosequencing AB, Uppsala, Sweden). Primer sequences and polymerase chain reaction conditions are shown in Supplementary Table 2, available at Carcinogenesis Online. All the primers were designed to examine the methylation status of CpGs within 0.5 kb of the transcription start site. A methylation level >15% was considered methylation positive since lower values could not be easily distinguished from background (17–20).

**Methylated CpG island amplification microarray**

For MCAM analysis, we randomly selected 41 cases from the 128 AdCa without any bias (average age was 61.9 years, ranging from 36 to 83 years old; Table I). A detailed protocol of MCA has been described previously (16–20). Briefly, amplicons from individual AdCa were labeled with Cy3 dye and cohybridized against amplicons from normal controls labeled with Cy5 dye on 15 K custom-promoter microarrays from Agilent Technologies (G4497A; Agilent Technologies, Santa Clara, CA) containing 6157 unique genes, which we had initially validated in a previous study (17). A Cy5/Cy3 signal in excess of 2.0 in MCAM was considered methylation positive (17–19). Comparison of the MCAM signal ratio (Cy5/Cy3 > 2.0 or Cy5/Cy5 ≤ 2.0) with the methylation status (positive or negative) from the pyrosequencing analysis showed a good concordance between the two analyses (sensitivity, 68.0% and specificity, 88.7%; Supplementary Table 3, available at Carcinogenesis Online).

**Hierarchical clustering analysis**

Cluster analysis was performed using an agglomerative hierarchical clustering algorithm (18,23). For specimen clustering, pairwise similarity measures were calculated using the Cluster 3.0 software (http://rana.lbl.gov/EisenSoftware.htm) or Minitab 15 statistical software (http://www.minitab.com) based on the DNA methylation intensity measurements.

### Table I. Clinical and molecular features according to the CIMP status

<table>
<thead>
<tr>
<th></th>
<th>All cases (%)</th>
<th>CIMP-N (%)</th>
<th>CIMP-L (%)</th>
<th>CIMP-H (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>128 (100)</td>
<td>78 (60.9)</td>
<td>40 (31.3)</td>
<td>10 (7.8)</td>
<td>0.4043</td>
</tr>
<tr>
<td>Age (mean ± SD)</td>
<td>64.7 ± 9.8</td>
<td>63.9 ± 10.0</td>
<td>66.4 ± 9.8</td>
<td>63.8 ± 7.0</td>
<td>0.0047</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>71 (55.5)</td>
<td>35 (44.9)</td>
<td>27 (67.5)</td>
<td>9 (90.0)</td>
<td>0.0047</td>
</tr>
<tr>
<td>Female</td>
<td>57 (44.5)</td>
<td>43 (55.1)</td>
<td>13 (32.5)</td>
<td>1 (10.0)</td>
<td></td>
</tr>
<tr>
<td>Stage*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>75 (59.5)</td>
<td>51 (67.1)</td>
<td>21 (52.5)</td>
<td>3 (30.0)</td>
<td>0.2259</td>
</tr>
<tr>
<td>IB</td>
<td>19 (15.1)</td>
<td>7 (9.2)</td>
<td>8 (20.0)</td>
<td>3 (30.0)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>29 (23.0)</td>
<td>16 (21.1)</td>
<td>10 (25.0)</td>
<td>4 (40.0)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>3 (2.4)</td>
<td>2 (2.6)</td>
<td>1 (2.5)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavy smoker</td>
<td>41 (32.3)</td>
<td>21 (27.6)</td>
<td>12 (30.8)</td>
<td>8 (80.0)</td>
<td>0.0036</td>
</tr>
<tr>
<td>Light smoker</td>
<td>29 (22.8)</td>
<td>15 (19.7)</td>
<td>13 (33.3)</td>
<td>1 (10.0)</td>
<td></td>
</tr>
<tr>
<td>Never-smoker</td>
<td>57 (44.9)</td>
<td>42 (55.3)</td>
<td>14 (35.9)</td>
<td>1 (10.0)</td>
<td></td>
</tr>
<tr>
<td>Differentiation*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well</td>
<td>15 (17.4)</td>
<td>11 (21.2)</td>
<td>3 (10.7)</td>
<td>1 (16.7)</td>
<td>0.3552</td>
</tr>
<tr>
<td>Moderate</td>
<td>55 (64.0)</td>
<td>30 (57.6)</td>
<td>22 (78.6)</td>
<td>3 (50.0)</td>
<td></td>
</tr>
<tr>
<td>Poorly</td>
<td>16 (18.6)</td>
<td>11 (21.2)</td>
<td>3 (10.7)</td>
<td>2 (33.3)</td>
<td></td>
</tr>
<tr>
<td>Recurrence*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(–)</td>
<td>41 (42.3)</td>
<td>27 (43.5)</td>
<td>14 (46.7)</td>
<td>0 (0)</td>
<td>0.1394</td>
</tr>
<tr>
<td>(+)</td>
<td>56 (57.7)</td>
<td>35 (56.5)</td>
<td>16 (53.3)</td>
<td>5 (100)</td>
<td></td>
</tr>
<tr>
<td>EGFR mutation*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(–)</td>
<td>80 (62.5)</td>
<td>42 (53.8)</td>
<td>28 (70.0)</td>
<td>10 (100)</td>
<td>0.0089</td>
</tr>
<tr>
<td>(+)</td>
<td>48 (37.5)</td>
<td>36 (46.2)</td>
<td>12 (30.0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>KRAS mutation*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(–)</td>
<td>118 (92.2)</td>
<td>74 (94.9)</td>
<td>36 (90.0)</td>
<td>8 (80.0)</td>
<td>0.2113</td>
</tr>
<tr>
<td>(+)</td>
<td>10 (7.8)</td>
<td>4 (5.1)</td>
<td>4 (10.0)</td>
<td>2 (20.0)</td>
<td></td>
</tr>
<tr>
<td>p53 mutation*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(–)</td>
<td>89 (69.5)</td>
<td>57 (74.0)</td>
<td>28 (73.7)</td>
<td>4 (40.0)</td>
<td>0.16</td>
</tr>
<tr>
<td>(+)</td>
<td>35 (27.3)</td>
<td>20 (26.0)</td>
<td>10 (26.3)</td>
<td>5 (50.0)</td>
<td></td>
</tr>
<tr>
<td>BRAF mutation*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(–)</td>
<td>86 (100)</td>
<td>56 (100)</td>
<td>23 (100)</td>
<td>7 (100)</td>
<td>NA</td>
</tr>
<tr>
<td>(+)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
</tbody>
</table>

*Data were not available in some cases. Recurrence was observed within 5 years after surgery.
across all genes. K-means consensus clustering was performed with the R statistical package. A dendrogram and heat map were constructed using either TreeView (http://rana.lbl.gov/EisenSoftware.htm) or R statistical computing environment (http://cran.r-project.org).

Nearest neighbor classification
Using the DNA methylation status of six CIMP markers (positive or negative by pyrosequencing analysis), nearest neighbor classification was employed to classify the validation set consisting of 87 independent AdCas (24). In this analysis, each validation case was classified into one of the three clusters identified in the training set. The number of nearest neighbors was set as k = 4 because the smallest cluster (cluster 1) was consisted of four cases. The analysis was conducted using R statistical computing environment (http://cran.r-project.org).

Mutation analysis
Mutations in KRAS (codons 12 and 13) were analyzed by direct sequencing and the pyrosequencing method (25,26). EGFR mutations (exons 18–21) and TP53 mutations (exons 5–8) were examined using direct sequencing (19,25). Mutation of BRAF (codon 600) was determined by the pyrosequencing method as previously reported (27). The polymerase chain reaction primer sequences used are listed in Supplementary Table 2, available at Carcinogenesis Online.

The Cancer Genome Atlas data
We obtained the methylation data of AdCa samples from the The Cancer Genome Atlas data (TCGA) web site (http://tcga-data.nci.nih.gov/tcga/tcga-Home2.jsp), and data of 85 AdCa samples (batches 34 and 37) were included in the analysis, which was conducted using the Illumina Infinium Human DNA Methylation 27 platform. The 3835 most variant probes from 27 578 CpG dinucleotides were used for further analysis, and a \( \beta \) value > 0.4 was considered as methylation positive.

Statistical analysis for clinical features
All statistical analyses were performed using the JMP statistical software version 5.1 (SAS Institute, Cary, NC). Fisher’s exact test was used to determine non-random associations between two categorical variables. Kruskal–Wallis analysis was used to evaluate the extent of differences among more than three groups. All reported P values are two sided, with \( P < 0.05 \) taken as statistically significant. Patients were followed until incidence of death or until October 2010, whichever came first. Survival information was available for 118 of the 128 cases. Overall survival was calculated from the date of surgery until the date of death or the date the patient was last known to be alive (censored). The median follow-up time was 42.5 months. Overall survival curves were generated via the Kaplan–Meier method, and the log-rank test was used for statistical analysis. A multivariate analysis using the Cox proportional hazards model was performed to estimate the hazard ratio. All variables for the multivariate analysis were categorical variables (age, stage and CIMP status).

Results
Identification of a distinct subgroup with characteristic DNA methylation profiling in AdCas
First, we evaluated the genome-wide DNA methylation status in a training set of 41 AdCas using MCAM analysis (18–20). Among 6157 genes on the microarray, we selected 1156 genes that are commonly methylated across >10% of AdCas and performed consensus average linkage hierarchical clustering analysis (28). In terms of DNA methylation, AdCas could be divided into three clusters, with clustering stability increasing for \( k = 2 \) to 3 but not for more than \( k = 4 \) (Figure 1A, Supplementary Figure 1, available at Carcinogenesis Online). Intriguingly, all four cases in cluster 1 stably fell into the same cluster regardless of \( k \) values (2–5), whereas 12 cases (92%) and 24 cases (96%) fell into clusters 2 and 3, respectively, indicating a high similarity of their methylation profile among each of the three-cluster member. The number of DNA methylated genes showed bimodal distribution in AdCas; DNA methylation was highly accumulated in two AdCas, both of which fell into cluster 1 (# Figure 1B). Consistently, a majority of the genes were commonly methylated across more than half of the AdCas in cluster 1, whereas common methylation targets were detected in <50% of the AdCas in clusters 2 and 3 ( \( P < 0.001 \), Figure 1C). In addition, the average number of methylated genes was 766.8 ± 70.4, 485.7 ± 40.6 and 319.2 ± 28.1 in clusters 1, 2 and 3 (\( k = 3 \)), respectively ( \( P < 0.001 \) (Figure 1D). These observations indicated that extensively methylated AdCas exist, which appear to be characterized by correlated CpG island DNA methylation of a subset of genes in a subset of tumors, whereas AdCas with less extensive DNA methylation or with rare DNA methylation were classified into discrete subgroups (7,14–16).

Validation analysis of newly identified CIMP markers
It is important to note that this initial selection of the six candidate markers did not introduce a bias for detecting CIMP only in the training set. Therefore, the newly identified panel of six markers was independently confirmed in a validation set of 87 AdCas. Among them, we found 6 (7%) CIMP-H, 30 (34%) CIMP-L and 51 (59%) CIMP-N tumors, which were of a similar frequency as observed in the training set of 41 AdCas (Supplementary Figure 2, available at Carcinogenesis Online). To estimate whether the classification by the six CIMP markers in the validation set was compatible with the CIMP classification in the training set, we performed the nearest four neighborhood prediction analysis (Figure 2C, see Materials and Methods).
analysis defined 6 of 87 AdCas in the validation set as cluster 1, all of which were also categorized as CIMP-H tumors according to our criteria using the six-marker panel. Furthermore, 51 AdCas that were classified as cluster 3 with a probability $\geq 80\%$ by the nearest neighbor classification analysis were also classified as CIMP-N tumors by our six CIMP marker panel. These results indicate that the three clusters are highly reproducible, and our panel of six markers is capable of accurately categorizing cases into these three clusters.

**Identification of AdCa–CIMP in the Cancer Genome Atlas data set**

Next, we examined whether the six CIMP markers could also be applicable in classifying AdCas deposited in TCGA (http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp) (30). DNA methylation data of 85 AdCas were obtained from the TCGA database, which were analyzed by the Infinium BeadChip containing 27,578 probes corresponding to 14,473 genes. Among those probes, we found that some probes corresponding to three of our new CIMP markers (ACAN, CCNA1 and GFRA1) were located close to the regions where we conducted the pyrosequencing analysis. In contrast, no probes corresponding to the other three CIMP markers (CDKN2A, P16 and APC) were located close to the regions where we conducted the pyrosequencing analysis.
Analysis defined 6 of 87 AdCas in the validation set as cluster 1, all of which were also categorized as CIMP-H tumors according to our criteria using the six-marker panel. Furthermore, 51 AdCas that were classified as cluster 3 with a probability of 0.80% by the nearest neighbor classification analysis were also classified as CIMP-N tumors by our six CIMP marker panel. These results indicate that the three clusters are highly reproducible, and our panel of six markers is capable of accurately categorizing cases into these three clusters.

Identification of AdCa–CIMP in the Cancer Genome Atlas data set

Next, we examined whether the six CIMP markers could also be applicable in classifying AdCas deposited in TCGA (http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp) (30). DNA methylation data of 85 AdCas were obtained from the TCGA database, which were analyzed by the Infinium BeadChip containing 27 578 probes corresponding to 14 473 genes. Among those probes, we found that some probes corresponding to three of our new CIMP markers (ACAN, CCNA1, and GFRA1) were located close to the regions where we conducted the pyrosequencing analysis. In contrast, no probes corresponding to the other three CIMP markers (CDKN2A,
MGC45800 and EDARADD) were located within 500 bp of transcription start sites, which was a condition of our pyrosequencing analysis. Therefore, we examined the classification power of the three CIMP markers (ACAN, CCNA1 and GFRAL) in the 85 AdCas obtained from the TCGA. None, 4 (4.7%), 14 (16.5%) and 67 (78.8%) AdCas showed DNA methylation in three, two, one or none of the three CIMP markers, respectively (DNA methylation positive, \( P > 0.4 \), Figure 2D, upper panel). AdCas with DNA methylation in one or two CIMP markers had more methylated probes among the 3833 most variant probes than those who had no methylation in the CIMP markers (average methylated probes: 1430.8 ± 39.5, 2116.4 ± 86.3 and 2288.8 ± 161.5, in AdCas with none, one or two methylated CIMP markers, respectively, \( P < 0.001 \); Figure 2D, lower panel).

These data suggested that the methylation status of the three CIMP markers is also predictive for highly methylated AdCas in the TCGA data set.

Clinical significance of CIMP tumors in lung AdCa

Next, we assessed whether CIMP-positive AdCas form a distinct subgroup with characteristic clinical features. We combined two sets of AdCa cohorts (training set and validation set) for the analysis of CIMP signatures, with a total of 128 cases. We did not access any clinical information before CIMP classification of both sets of AdCas to avoid any bias in the analysis of the clinical significance of CIMP tumors. Of the 128 pooled AdCas, 10 (7.8%) were classified as CIMP-H, 40 (31.3%) as CIMP-L, and 78 (60.9%) as CIMP-N (Figure 3A).

CIMP-H tumors were more prevalent in males (\( P = 0.0047 \)) and associated with frequent exposure to smoking (pack year > 40, \( P = 0.0036 \)). Intriguingly, we found a tight association between CIMP and EGFR status (\( P = 0.0089 \); Table I). None of the 10 CIMP-H AdCas contained any EGFR mutations, whereas 36/78 (46.2%) CIMP-N and 12/40 (30.0%) CIMP-L AdCas had EGFR mutations.

In contrast, no such tendency was observed between CIMP status and TP53, KRAS, or BRAF mutations.

To investigate whether CIMP status had any impact on overall survival, we performed Kaplan–Meier survival analysis and found that CIMP-H was a significantly negative prognostic factor (\( P = 0.0115 \), log-rank test; Figure 3B). Since EGFR mutations have been indicated as a potential positive prognostic factor for survival in advanced non-small cell lung cancer patients treated with chemotherapy with or without TKI (34), we analyzed overall survival according to the EGFR mutation status. Among the AdCas harboring wild-type EGFR, CIMP-H tumors still correlated with a poor survival (\( P = 0.0312 \), log-rank test; Figure 3C). In addition to a worse prognosis in patients with AdCa who were smokers (\( P = 0.0373 \), log-rank test; Figure 3D), we found that CIMP-H was an independent prognostic factor among male smokers (hazard ratio 1.7617, 95% confidence interval 1.0030–2.9550, \( P = 0.0489 \); Figure 3E). Taken together, these findings indicated that CIMP-H tumors have unique clinical features that distinguish them from the other AdCas.

Clinical implication of epigenetic therapy for lung AdCa

To evaluate the relationship between CIMP status and effects of the DNA methylation inhibitor, 5-Aza-dC, as an antitumor agent, we first analyzed DNA methylation status of the six CIMP markers in 14 AdCa cell lines, including one CIMP-H (H358), seven CIMP-L (H23, H1, P90, H2009, H3255, H1975 and H1650) and six CIMP-N cell lines (H920, A549, HCC827, ML13, H1573 and HCC4011) (Supplementary Table 4, available at Carcinogenesis Online). CIMP-H cells harbor wild-type EGFR, whereas the CIMP-L and CIMP-N cells harbor either wild-type (H23, H1, H2009, H1975, H920, A549, HCC827, ML13 and H1573) or mutated (PC9, H3255, H1650 and HCC4011) EGFR. Regardless of the CIMP status, cells with wild-type EGFR showed resistance to the TKI AG1478 (Figure 4A). Intriguingly, antitumor activity of 5-Aza-dC appeared to be associated with CIMP status. Each cell line showed different IC50, which were significantly lower in the CIMP-positive cells (average, CIMP-H and CIMP-L) than in the CIMP-N cells (\( P = 0.02 \), average IC50 was 68, 229 and 982 nM, respectively, Figure 4B). To determine a more accurate relationship between DNA demethylation and antitumor activity, we used level of LINE-1 demethylation, which represents the global level of methylation, as a surrogate marker of 5-Aza-dC treatment. We examined the power of growth inhibition at a concentration of ~20% of LINE-1 demethylation. Cell growth of CIMP-H and CIMP-L cells was significantly inhibited at each concentration of 5-Aza-dC, in contrast to CIMP-N cells, the majority of which did not respond to the treatment (Figure 4B). These data suggest that in addition to CIMP-H AdCas, tumorigenesis of CIMP-L AdCas may also depend on DNA methylation silencing pathway to some extent.

Discussion

In the current study, we performed a comprehensive genome-wide DNA methylation analysis and identified a distinct molecular subgroup (CIMP-H) in human AdCas. This subgroup showed a remarkably high rate of DNA methylation in correlated cancer-specific CpG island hypermethylation of a subset of genes, indicating the existence of CIMP in AdCas (7).

Previously, studies suggested the existence of CIMP in lung cancer. The first study defined a CIMP-positive case as having a tumor with aberrant methylation in either CDH13 or CRBP1 and found that CIMP-positive cases showed poorer prognosis than the CIMP-N ones (9). Although a consistent clinical feature, poor prognosis of CIMP-positive cases, was observed between our study and the previous ones, frequencies of CIMP and the other clinicopathological features associated with CIMP were varied, probably due to the different panels of CIMP marker examined (9,11–13). Thus, it is still unclear which DNA methylation markers can define the most extensively methylated subgroups (CIMP) due to the lack of accompanied genome-wide analysis in the previous AdCa studies. If CIMP affects only a subset of CpG islands in a subgroup of AdCas, collection of data for a large number of markers from numerous tumor samples is required to identify CIMP in AdCas. Indeed, we found in the current study that CIMP-positive tumors diagnosed by the original CIMP markers defined in the colon cancer study (7) were not consistent with the extensively methylated AdCas, suggesting that these markers are not always applicable for diagnosis of CIMP tumors other than colon cancers. Thus, the existence of CIMP in AdCas from the global point of view has remained elusive before the current study. Our genome-wide MCAM analysis successfully identified six practical and representative markers for the prediction of CIMP in AdCas.

Integrated analysis of the DNA methylation status of the six CIMP markers with several cancer-associated gene mutations, including EGFR, TP53, KRAS and BRAF, revealed that CIMP-H tumors in both the training and the validation sets did not harbor any EGFR mutations, suggesting that the two events are mutually exclusive, whereas mutations in the other three genes did not show such strong associations with CIMP status. Thus, our six novel markers enabled us to decipher the negative association between CIMP and EGFR mutations, which had been only suggested by previous studies (35). Indeed, CIMP-H tumors are significantly associated with males, frequent exposure to smoking and high relapse rate of disease, which clearly differs from the typical features of EGFR-mutant AdCas, suggesting that these markers are involved in activation of an alternative pathway, in which tumorigenesis may be minimally dependent on EGFR mutation.

The association of better clinical outcome with CIMP-positive tumors has been reported in breast cancer, colon cancer and glioma. Fang et al. (31) showed that there was significant overlap of CIMP targets in those different types of cancers. Among the 33 overlapped CIMP targets between the cancers, we found that the DNA methylation status of 19 genes was available from our MCAM analysis. Using
the methylation status of these 19 genes, we found that AdCas in our training set were divided into three clusters; the three most extensively methylated AdCas were consistent with the CIMP-H AdCas, suggesting that a panel of 19 genes may also be a potent predictor for CIMP-H in AdCas (Supplementary Figure 3, available at Carcinogenesis Online). However, given the worse prognosis of patients with CIMP-H AdCas, the impact of CIMP to survival in lung cancer might be different from the CIMP in colon cancer, glioma and breast cancer. Indeed, a study in myelodysplastic syndromes also showed that the presence of CIMP was significantly associated with poor prognosis and risk of leukemia transformation (38). The contrasting impacts of CIMP to clinical outcome might be due to the distinct DNA methylation profiles specific to each tumor type; CIMP confers poor prognosis in lung AdCas and myelodysplastic syndromes via inactivation...
of genes critical for tumor progression and for response to chemotherapy.

We found another epigenomic subgroup (CIMP-L) of AdCAs within the major subclasses classified by DNA methylation status. This subtype showed moderate accumulation of DNA methylation. Indeed, our cell line study showed that CIMP-L cells were sensitive to 5-Aza-dC treatment. These data suggest that tumorigenesis pathway of CIMP-L AdCas might also be affected by DNA methylation to a certain extent. However, we could not find any specific features of this subtype. This might be due to the lack of suitable markers to further classify CIMP-L, resulting in a mixture of subpopulations as was found in the colon cancer study (39,40). Sensitive and specific markers for CIMP-L in AdCAs are needed to further characterize CIMP-L. Additional studies will be required to address this problem.

CIMP-positive lung AdCa cell lines appeared to be more sensitive to 5-Aza-dC treatment, in which demethylation effectively occurred even at low doses of 5-Aza-dC, regardless of EGFR mutation status. Epigenetic drugs targeting DNA methylation, such as 5-Aza-dC and 5-azacytidine, have shown clinical effectiveness in cancer treatment, especially for hematological malignancies (41,42). For the treatment of thoracic malignancies, studies showed that a certain population of patients with AdCas clinically benefit from 5-Aza-dC treatment (43,44). One of the important issues of research is the identification of biomarkers predictive of response to DNA methylation inhibitors (45). Our cell line analysis showed that CIMP status appeared to be associated with response to 5-Aza-dC, suggesting that epigenetic therapy might be a useful approach, especially for those individuals who have been diagnosed with CIMP. If this possibility was validated, our findings would be significant for the use of DNA methylation inhibitors in lung tumors.

In conclusion, we demonstrated here that six newly identified CIMP markers may be useful in the accurate and practical epigenomic
classifications of lung cancer. Our findings may enable the development of new molecular diagnostics tools for personalized medicine for lung cancers and confer a new paradigm for cancer treatment.

Supplementary material
Supplementary Figures 1–3 and Tables 1–4 can be found at http://carcin.oxfordjournals.org/.

Funding
This work is supported by grants-in-aid for Cancer Research from the Ministry of Health, Labor and Welfare (to Y.K.) and a grant from the Japan Society for the Promotion of Science (to Y.K.).

Conflict of Interest Statement: None declared.

References

Received November 20, 2011; revised April 9, 2012; accepted April 18, 2012