Tian et al. BMC Cancer 2014, 14:772
http://www.biomedcentral.com/1471-2407/14/772

BMC Cancer

RESEARCH ARTICLE Open Access

Prognostication of patients with clear cell renal cell carcinomas based on quantification of DNA methylation levels of CpG island methylator phenotype marker genes

Ying Tian[1], Eri Arai[1*], Masahiro Gotoh[1], Motokiyo Komiyama[2], Hiroyuki Fujimoto[2] and Yae Kanai[1]

Abstract

Background: The CpG island methylator phenotype (CIMP) of clear cell renal cell carcinomas (ccRCCs) is characterized by accumulation of DNA methylation at CpG islands and poorer patient outcome. The aim of this study was to establish criteria for prognostication of patients with ccRCCs using the ccRCC-specific CIMP marker genes.

Methods: DNA methylation levels at 299 CpG sites in the 14 CIMP marker genes were evaluated quantitatively in tissue specimens of 88 CIMP-negative and 14 CIMP-positive ccRCCs in a learning cohort using the MassARRAY system. An additional 100 ccRCCs were also analyzed as a validation cohort.

Results: Receiver operating characteristic curve analysis showed that area under the curve values for the 23 CpG units including the 32 CpG sites in the 7 CIMP-marker genes, i.e. *FAM150A, ZNF540, ZNF671, ZNF154, PRAC, TRH and SLC13A5,* for discrimination of CIMP-positive from CIMP-negative ccRCCs were larger than 0.95. Criteria combining the 23 CpG units discriminated CIMP-positive from CIMP-negative ccRCCs with 100% sensitivity and specificity in the learning cohort. Cancer-free and overall survival rates of patients with CIMP-positive ccRCCs diagnosed using the criteria combining the 23 CpG units in a validation cohort were significantly lower than those of patients with CIMP-negative ccRCCs ($P = 1.41 \times 10^{-5}$ and 2.43×10^{-13}, respectively). Patients with CIMP-positive ccRCCs in the validation cohort had a higher likelihood of disease-related death (hazard ratio, 75.8; 95% confidence interval, 7.81 to 735; $P = 1.89 \times 10^{-4}$) than those with CIMP-negative ccRCCs.

Conclusions: The established criteria are able to reproducibly diagnose CIMP-positive ccRCCs and may be useful for personalized medicine for patients with ccRCCs.

Keywords: DNA methylation, CpG island methylator phenotype (CIMP), Prognostication, MassARRAY system, Clear cell renal cell carcinoma (ccRCC)

Background

Clear cell renal cell carcinoma (ccRCC) is the most common histological subtype of adult kidney cancer [1]. In general, ccRCCs at an early stage are curable by nephrectomy. However, some ccRCCs relapse and metastasize to distant organs, even if the resection has been considered complete [2]. Even though novel targeting agents have been developed for treatment of ccRCC, unless relapsed or metastasized tumors are diagnosed early by close follow-up, the effectiveness of any therapy is restricted [3]. Therefore, reliable prognostic criteria need to be established.

Not only genetic, but also epigenetic events appear to accumulate during carcinogenesis, and DNA methylation alterations are one of the most consistent epigenetic changes in human cancers [4-6]. We and other groups have revealed that DNA methylation alterations participate in renal carcinogenesis and are significantly correlated with the clinicopathological diversity of ccRCCs [7-11]. In addition, a distinct cancer phenotype known

* Correspondence: earai@ncc.go.jp
[1]Division of Molecular Pathology, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
Full list of author information is available at the end of the article

as the CpG island methylator phenotype (CIMP), characterized by accumulation of DNA methylation at CpG islands, has been defined in well-studied cancers [12,13] such as those of the colorectum [14] and stomach [15], and shown to be significantly correlated with clinicopathological parameters. Although the relevance of the CIMP-positive phenotype in the context of ccRCCs has not yet been clearly defined [16], our group very recently identified CIMP-positive ccRCCs based on genome-wide DNA methylation analysis [7]. We also identified 17 genes, i.e. *FAM150A, GRM6, ZNF540, ZFP42, ZNF154, RIMS4, PCDHAC1, KHDRBS2, ASCL2, KCNQ1, PRAC, WNT3A, TRH, FAM78A, ZNF671, SLC13A5* and *NKX6-2*, which are hallmarks of CIMP in ccRCCs [7], using single CpG-resolution Infinium assay [17]. The CIMP-positive ccRCCs in our cohort were clinicopathologically aggressive and associated with poorer patient outcome [7], indicating that CIMP in ccRCCs might be applicable as a prognostic indicator.

However, in our previous study, CIMP-positive ccRCCs were identified using hierarchical clustering analysis based on DNA methylation profiles in the examined cohort [7]. The DNA methylation status of entire promoter CpG islands, other than Infinium probe sites, in the CIMP marker genes has not been evaluated quantitatively. Therefore, to establish criteria for CIMP diagnosis that would be applicable to individual patients, CpG sites having the largest diagnostic impact should be identified in the entire promoter CpG islands of the CIMP marker genes based on quantification of DNA methylation levels. Moreover, appropriate cutoff values of DNA methylation levels need to be established for the identified CpG sites in order to discriminate CIMP-positive from CIMP-negative ccRCCs.

In the present study, we quantitatively evaluated DNA methylation levels at 299 CpG sites throughout the promoter CpG islands of the ccRCC-specific CIMP marker genes in 88 CIMP-negative ccRCCs and 14 CIMP-positive ccRCCs using the MassARRAY system. We then validated the prognostic impact of the established criteria for CIMP diagnosis in a validation cohort of 100 additional ccRCCs.

Methods

Patients and tissue samples

As a learning cohort, 102 samples of cancerous tissue obtained from specimens surgically resected from 102 patients with primary ccRCCs were subjected to the present analysis. These patients did not receive preoperative treatment and underwent nephrectomy at the National Cancer Center Hospital, Tokyo, Japan. There were 71 men and 31 women with a mean (± standard deviation) age of 62.9 ± 10.4 years (range, 36 to 85 years). Histological diagnosis was made in accordance with the World Health Organization classification [18].

In our previous study, unsupervised hierarchical clustering based on genome-wide DNA methylation analysis using single CpG-resolution Infinium assay divided the 102 ccRCCs in the learning cohort into 88 CIMP-negative ccRCCs and 14 CIMP-positive ccRCCs [7]. In the same study, we showed that the CIMP-positive ccRCCs were clinicopathologically more aggressive and associated with a poorer patient outcome than CIMP-negative ccRCCs [7]: the clinicopathological characteristics [19,20] of CIMP-negative and CIMP-positive ccRCCs in the learning cohort are summarized in Additional file 1: Table S1.

As a validation cohort, 100 samples of cancerous tissue were obtained from specimens surgically resected from 100 patients with primary ccRCCs. These patients also did not receive preoperative treatment and underwent nephrectomy at the National Cancer Center Hospital, Tokyo, Japan. The patients comprised 68 men and 32 women with a mean (± standard deviation) age of 62.5 ± 11.4 years (range, 33 to 87 years). The clinicopathological characteristics [19,20] of ccRCCs in the validation cohort are summarized in Additional file 2: Table S2.

Tissue specimens were taken and frozen immediately after surgical removal and have been stored in liquid nitrogen until DNA extraction. ccRCCs are hypervascular tumors with an increased opportunity for infiltration of non-cancerous cells such as lymphocytes [21]: the microscopically examined tumor cell contents (%) of all ccRCC tissue specimens in the learning and validation cohorts are shown in Additional file 3: Table S3. Tissue specimens were provided by the National Cancer Center Biobank, Tokyo, Japan. This study was approved by the Ethics Committee of the National Cancer Center, Tokyo, Japan, and was performed in accordance with the Declaration of Helsinki. All the patients provided written informed consent prior to inclusion in the study.

DNA extraction and bisulfite modification

High-molecular-weight DNA was extracted from fresh-frozen tissue samples using phenol-chloroform followed by dialysis [22]. One microgram of genomic DNA was subjected to bisulfite treatment using an EpiTect Bisulfite Kit (QIAGEN GmbH, Hilden, Germany), in accordance with the manufacturer's protocol. This process converts non-methylated cytosine to uracil, while methylated cytosine remains unchanged [23].

Quantitative DNA methylation analysis with the MassARRAY system

DNA methylation levels at individual CpG sites were evaluated quantitatively using the MassARRAY platform (Sequenom, San Diego, CA). This method utilizes base-specific cleavage and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) [24]. Specific PCR primers for bisulfite-converted

DNA were designed using the EpiDesigner software package (www.epidesigner.com, Sequenom), encompassing all promoter CpG islands of the previously identified ccRCC-specific CIMP marker genes [7]. The sequences of the 16 primer sets are given in Additional file 4: Table S4. A T7-promoter tag (5′-CAGTAATACGACTCACTATAGG GAGAAGGCT-3′) was added to each reverse primer for in vitro transcription, and a 10-mer tag (5′-AGGAAGA GAG-3′) was added to each forward primer to balance the PCR.

To overcome PCR bias in DNA methylation analysis, we optimized the annealing temperature and type of DNA polymerase: 0%, 50% and 100% methylated control DNA (Epitect methylated human control DNA; QIAGEN) was used as template to test the linearity of the protocol. Using HotStar Taq DNA polymerase (QIAGEN) or TaKaRa Taq HS DNA polymerase (Takara Bio, Shiga, Japan), the annealing temperature for each of the 16 primer sets was set to give a correlation coefficient (R^2) of more than 0.9 and to make the slope of the standard curve close to 1 (Additional file 5: Figure S1 and Additional file 4: Table S4). The PCR products were separated electrophoretically on 2% agarose gel and stained with ethidium bromide to confirm that specific products of the appropriate size and no non-specific products were obtained upon amplification.

Then, the PCR products were used as a template for in vitro transcription and the RNase A-mediated cleavage reaction using an EpiTYPER Reagent Kit (Sequenom). The fragmented samples were dispensed onto a SpectroCHIP array, and then detected on a MassARRAY analyzer compact MALDI-TOF MS instrument. The data were visualized using EpiTYPER Analyzer software v1.0 (Sequenom). The DNA methylation level (%) at each CpG site was determined by comparing the signal intensities of methylated and non-methylated templates. A cluster of consecutive CpG sites, each giving one measured value by the MassARRAY system, is defined as a "CpG unit" in the manufacturer's protocol. The DNA methylation levels at the 299 examined CpG sites in the CIMP marker genes were then expressed as data for the 193 CpG units. Experiments were performed in triplicate for each sample-CpG unit, and the mean value for the three experiments was used as the DNA methylation level.

Statistics

Differences in DNA methylation levels at individual CpG units between CIMP-positive ccRCCs and CIMP-negative ccRCCs were analyzed using Mann–Whitney *U* test. The CpG units having the largest diagnostic impact were identified by receiver operating characteristic (ROC) curve analysis [25]: For 23 CpG units showing area under the curve (AUC) values larger than 0.95, appropriate cutoff values were determined in order to discriminate CIMP-positive from CIMP-negative ccRCCs [26]. For discriminating CIMP-positive from CIMP-negative ccRCCs, the Youden index [26] was used as a cutoff value for each CpG unit. Survival curves for patients with ccRCCs were analyzed by the Kaplan-Meier method and the log-rank test. Correlations between DNA methylation levels and recurrence and disease-related death were analyzed using the Cox proportional hazards model. All statistical analyses were performed using SPSS statistics version 20 (IBM Corp., Armonk, NY). Differences at *P* values of less than 0.05 were considered statistically significant.

Results

DNA methylation status of CIMP marker genes in CIMP-negative and CIMP-positive ccRCCs

Previously, we had identified 17 ccRCC-specific CIMP marker genes based on genome-wide DNA methylation analysis using the Infinium HumanMethylation27K BeadChip [7]. Six exact Infinium probe CpG sites in ccRCC-specific CIMP marker genes (Probe ID: cg06274159 for the *ZFP42* gene, cg03975694 for the *ZNF540* gene, cg08668790 for the *ZNF154* gene, cg01009664 for the *TRH* gene, cg22040627 for the *SLC13A5* gene, and cg19246110 for the *ZNF671* gene) were examined using the MassArray system in the learning cohort (Additional file 6; Figure S2). Significant correlations between DNA methylation levels determined by our previous Infinium assay [7] and those determined by the present MassArray analysis were statistically confirmed ($P = 1.25 \times 10^{-35}$, $P = 1.98 \times 10^{-32}$, $P = 1.31 \times 10^{-41}$, $P = 5.30 \times 10^{-34}$, $P = 7.91 \times 10^{-22}$ and $P = 7.61 \times 10^{-44}$, respectively).

In the present study, our primary intention was to evaluate quantitatively the DNA methylation status of not only the Infinium probe sites but also the entire promoter CpG islands in the ccRCC-specific CIMP marker genes using the MassARRAY system [24]. Since the promoter regions of the CIMP marker genes, *KCNQ1, FAM78A* and *NKX6-2*, have a very high GC content, for these three genes we were unable to set optimized PCR conditions. Then, the DNA methylation status of 193 CpG units including 299 CpG sites in the remaining 14 ccRCC-specific CIMP marker genes, i.e. *FAM150A, GRM6, ZNF540, ZFP42, ZNF154, RIMS4, PCDHAC1, KHDRBS2, ASCL2, PRAC, WNT3A, TRH, ZNF671* and *SLC13A5*, was evaluated quantitatively using the MassARRAY system. The average DNA methylation levels of 38 CpG units including 68 CpG sites located within the 1347 bp 5′-region of the representative CIMP marker gene, *SLC13A5*, in CIMP-negative (n = 88) and CIMP-positive (n = 14) ccRCCs in the learning cohort are shown in Figure 1A. Similarly, the average DNA methylation levels of 21 CpG units including 29 CpG sites located within the 428 bp 5′-region of another representative CIMP marker gene, *ZNF671*, in CIMP-negative and CIMP-positive ccRCCs in

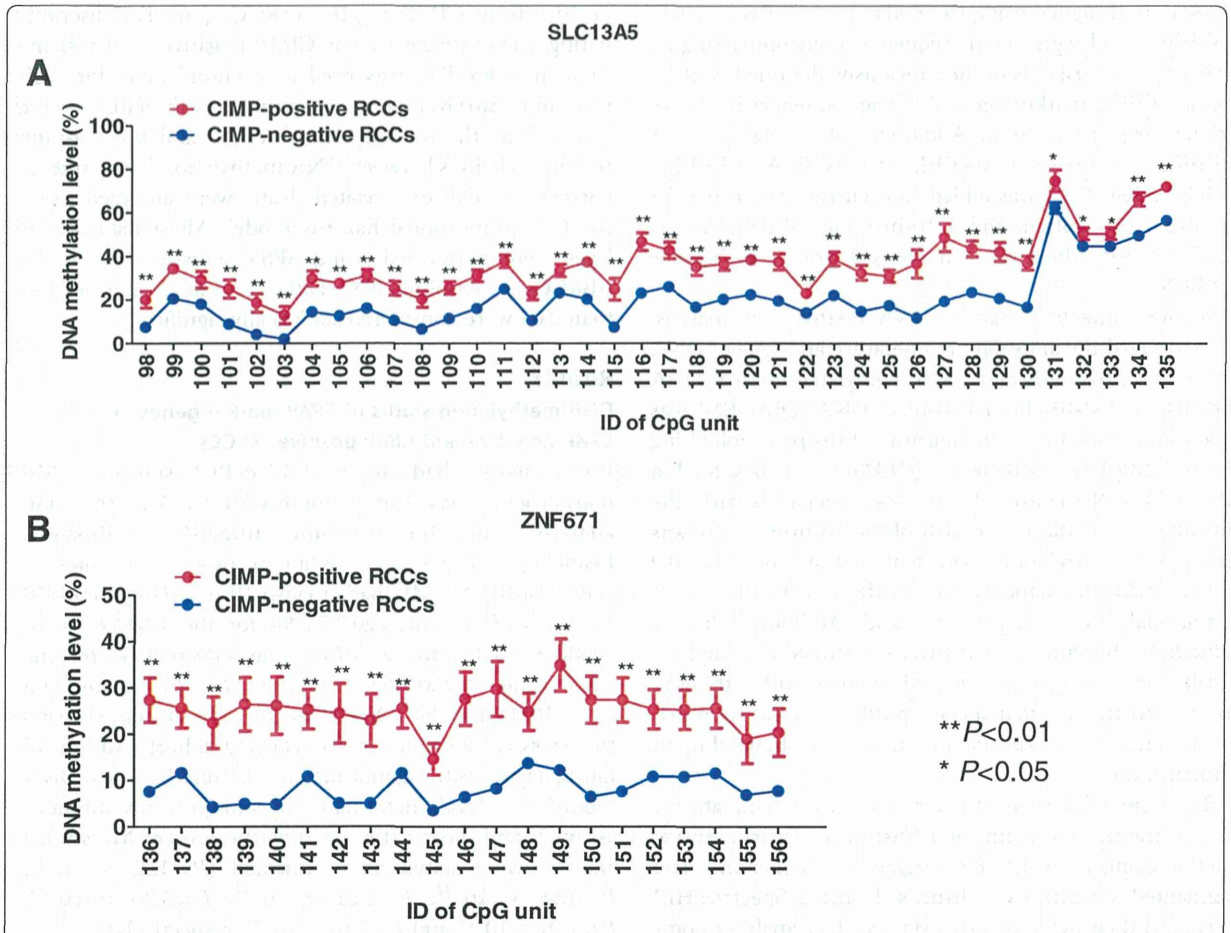

Figure 1 Average DNA methylation levels at promoter CpG islands in the *SLC13A5* (A) and *ZNF671* (B) genes in CIMP-negative (n = 88) and CIMP-positive (n = 14) ccRCCs in the learning cohort. DNA methylation levels of each CpG unit were evaluated quantitatively using the MassARRAY system. **A**. Average DNA methylation levels of all examined 38 CpG units including 68 CpG sites located within 1347 bp 5′ region of the *SLC13A5* gene in CIMP-positive ccRCCs (red line) were significantly higher than those in CIMP-negative ccRCCs (blue line). **B**. Average DNA methylation levels of all examined 21 CpG units including 29 CpG sites located within the 428 bp 5′ region of the *ZNF671* gene in CIMP-positive ccRCCs (red line) were significantly higher than those in CIMP-negative ccRCCs (blue line). $*P < 0.05$ and $**P < 0.01$. Exact *P* values for each CpG unit of the *SLC13A5* and *ZNF671* genes are summarized in Additional file 7: Table S5. Error bar: standard error.

the learning cohort are shown in Figure 1B. The average DNA methylation levels of all the CpG units examined (59 in total) in the *SLC13A5* and *ZNF671* genes in the CIMP-positive ccRCCs were significantly higher than those in CIMP-negative ccRCCs (the *P* values for each CpG unit are shown in Additional file 7: Table S5). Similarly, the average DNA methylation levels of 130 CpG units including 195 CpG sites, out of the 134 CpG units examined including 202 CpG sites in the remaining 12 CIMP marker genes, in the CIMP-positive ccRCCs were significantly higher than those in CIMP-negative ccRCCs (Additional file 7: Table S5). These data indicated that almost the entire promoter CpG islands in all the CIMP marker genes examined were methylated in CIMP-positive ccRCCs.

Establishment of criteria for discriminating CIMP-positive from CIMP-negative ccRCCs in the learning cohort

Since quantitative DNA methylation analysis using the MassARRAY system revealed that many CpG sites showed significant differences in DNA methylation levels between CIMP-negative and CIMP-positive ccRCCs among all the promoter CpG islands of CIMP marker genes (Figure 1 and Additional file 7: Table S5), we attempted to identify CpG sites having the largest diagnostic impact, and to establish criteria for discriminating CIMP-positive from CIMP-negative ccRCCs. ROC curves were constructed for all 193 CpG units examined, including 299 CpG sites in the 14 CIMP marker genes examined, and the corresponding AUC values [25] were calculated. Eighty-six CpG units, including 135 CpG sites, showed AUC values

larger than 0.9 (Additional file 8: Table S6). Among these 86, the top 23 CpG units including 32 CpG sites showing AUC values larger than 0.95 were used to establish the criteria for discriminating CIMP-positive from CIMP-negative ccRCCs (Table 1). For discriminating CIMP-positive from CIMP-negative ccRCCs, the Youden index [26] was used as a cutoff value for each CpG unit (Table 1).

Figure 2A shows scattergrams of the DNA methylation levels of representative CpG units in CIMP-negative and CIMP-positive ccRCCs in the learning cohort along with cutoff values listed in Table 1. The sensitivity and specificity of such discrimination using the cutoff values derived for each CpG unit are shown in Figure 2A and Table 1. A histogram showing the number of CpG units showing DNA methylation levels higher than the cutoff values listed in Table 1 in the learning cohort is shown in Figure 2B. All 14 ccRCCs showing DNA methylation levels higher than the cutoff values listed in Table 1 at 16 or more CpG units based on the present MassARRAY analysis (red bars in Figure 2B) were CIMP-positive ccRCCs identified by our previous hierarchical clustering based on the Infinium assay. All 88 ccRCCs showing DNA methylation levels higher than the cutoff values listed in Table 1 at less than 16 CpG units based on the present MassARRAY analysis (blue bars in Figure 2B) were CIMP-negative ccRCCs identified by our previous hierarchical clustering based on the Infinium assay.

Based on Figure 2B, we established the following criteria: when ccRCC tissue shows DNA methylation levels higher than the cutoff values listed in Table 1 at 16 or more CpG units (green line in Figure 2B), it is judged to be CIMP-positive. When ccRCC tissue shows DNA methylation levels higher than the cutoff values listed in Table 1 at less than 16 CpG units, it is judged to be

Table 1 The 23 CpG units showing area under the curve (AUC) values larger than 0.95 in receiver operating characteristic curve analysis for discrimination of CpG island methylator phenotype (CIMP)-positive clear cell renal cell carcinomas (ccRCCs) from CIMP-negative ccRCCs in the learning cohort

ID of CpG unit[1]	Gene symbol	Chromo-some	Position of CpG site[2]	AUC value	Cutoff value[3](%)	Sensitivity[4](%)	Specificity[4](%)
81	*TRH*	3	129693406, 129693412	0.973	30.8	100.0	88.6
85	*TRH*	3	129693518, 129693521, 129693528	0.950	18.2	100.0	78.4
89	*TRH*	3	129693586	0.952	11.0	92.3	92.0
94	*TRH*	3	129693635	0.967	6.6	100.0	87.5
8	*FAM150A*	8	53478477	0.968	27.2	83.3	94.1
11	*FAM150A*	8	53478511	0.968	27.2	83.3	94.1
78	*PRAC*	17	46799755	0.957	40.7	92.9	89.8
102	*SLC13A5*	17	6616733	0.983	7.5	92.9	96.6
105	*SLC13A5*	17	6616812	0.983	18.5	100.0	94.3
106	*SLC13A5*	17	6616826, 6616828	0.951	23.3	100.0	88.6
107	*SLC13A5*	17	6616851, 6616854, 6616857	0.954	14.8	100.0	87.5
110	*SLC13A5*	17	6616927, 6616929	0.951	23.3	100.0	88.6
30	*ZNF540*	19	38042496	0.983	41.0	100.0	98.3
32	*ZNF540*	19	38042518	0.960	35.7	100.0	93.1
33	*ZNF540*	19	38042530, 38042532	0.991	36.4	100.0	96.6
43	*ZNF154*	19	58220567	0.956	13.3	92.9	90.9
44	*ZNF154*	19	58220627	0.966	14.8	85.7	95.5
45	*ZNF154*	19	58220657, 58220662	0.959	22.2	92.9	95.5
149	*ZNF671*	19	58238780	0.954	15.2	85.7	89.7
158	*ZNF671*	19	58238928	0.965	10.5	100.0	88.5
160	*ZNF671*	19	58238954	0.954	15.2	85.7	89.7
161	*ZNF671*	19	58238987	0.954	15.2	85.7	89.7
163	*ZNF671*	19	58239012	0.951	10.5	85.7	92.0

[1]ID of CpG unit is defined in Additional file 4: Table S4.
[2]National Center for Biotechnology Information (NCBI) Database (Genome Build 37).
[3]The Youden index was used as a cutoff value for discriminating CIMP-positive ccRCCs in the learning cohort from CIMP-negative ccRCCs. When the cancerous tissue shows a DNA methylation level equal to or higher than the cutoff value, the ccRCC is considered to be CIMP-positive; when the cancerous tissue shows a DNA methylation level lower than the cutoff value, the ccRCC is considered to be CIMP-negative.
[4]Sensitivity and specificity for discrimination of CIMP-positive ccRCCs in the learning cohort from CIMP-negative ccRCCs using individual CpG units.

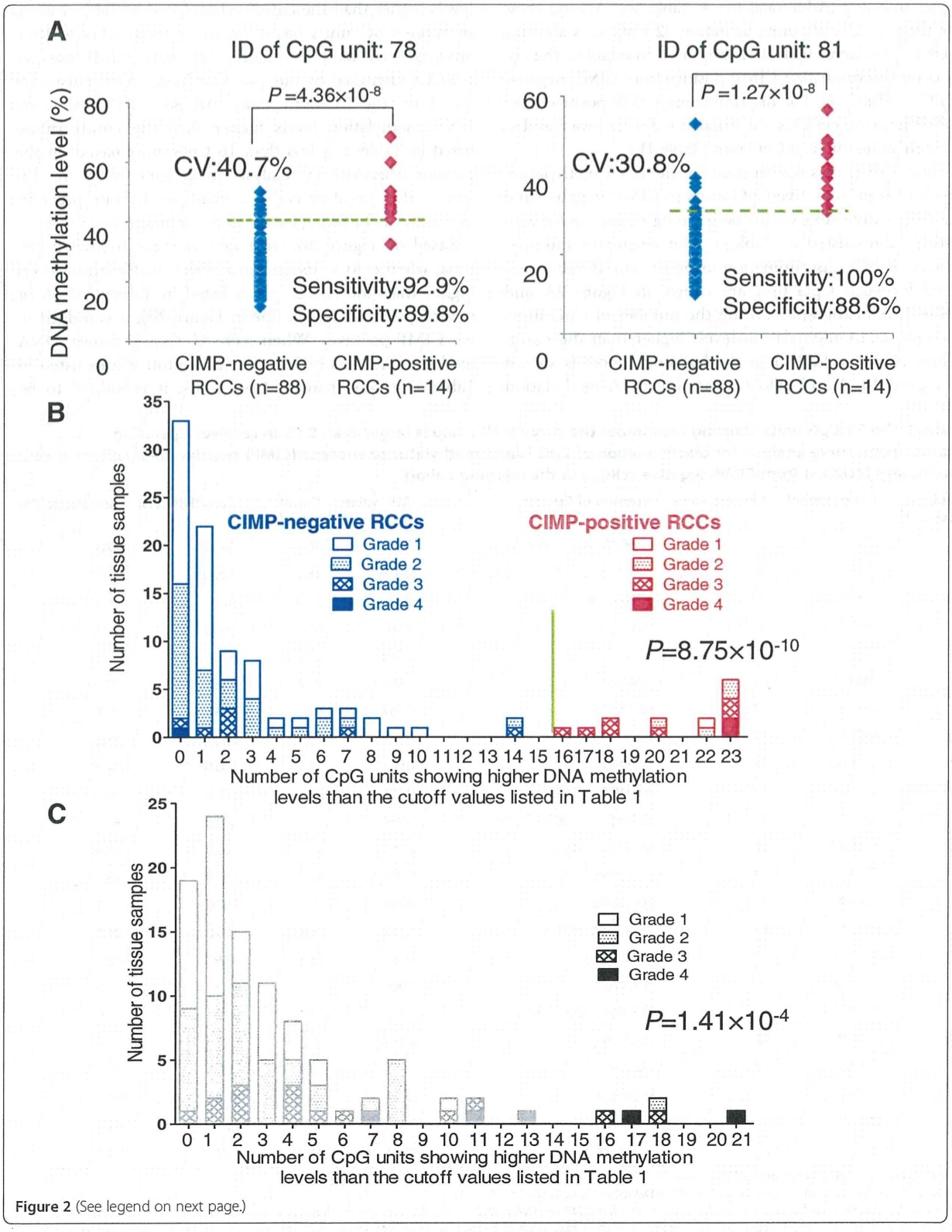

Figure 2 (See legend on next page.)

(See figure on previous page.)
Figure 2 The criteria for CIMP diagnosis discriminating CIMP-positive from CIMP-negative ccRCCs based on the MassARRAY system. **A**. Scattergrams of DNA methylation levels of representative CpG units in the learning cohort. Using each CpG unit and its cutoff value (CV) described in Table 1, CIMP-positive ccRCCs were discriminated from CIMP-negative ccRCCs with sufficient sensitivity and specificity. **B**. Histogram showing the number of CpG units with DNA methylation levels higher than the cutoff values listed in Table 1 in the learning cohort. All 14 ccRCCs (red columns) showing DNA methylation levels higher than the cutoff values at 16 or more CpG units were CIMP-positive ccRCCs, and all 88 ccRCCs (blue columns) showing DNA methylation levels higher than the cutoff values at less than 16 CpG units were CIMP-negative ccRCCs. On the basis of this histogram, we established the following criteria: When the cancerous tissue showed DNA methylation levels higher than the cutoff values at 16 (green bar) or more CpG units, it was judged to be CIMP-positive. The number of CpG units showing higher DNA methylation levels than the cutoff values in CIMP-positive ccRCCs (20.79 ± 0.69) was higher than that of CIMP-negative ccRCCs (2.09 ± 0.32, $P = 8.75 \times 10^{-10}$). **C**. Histogram showing the number of CpG units with DNA methylation levels higher than the cutoff values listed in Table 1 in the additional 100 ccRCCs comprising the validation cohort. Using the criteria established on the basis of panel **B**, 5 ccRCCs (black bars) were diagnosed as CIMP-positive ccRCCs, whereas 95 ccRCCs (gray bars) were diagnosed as CIMP-negative ccRCCs. The number of CpG units showing higher DNA methylation levels than the cutoff values in ccRCCs diagnosed as CIMP-positive (18.00 ± 0.84) was higher than that of ccRCCs diagnosed as CIMP-negative (2.73 ± 0.30, $P = 1.41 \times 10^{-4}$).

CIMP-negative. Using these criteria, CIMP-positive ccRCCs in the learning cohort were discriminated from CIMP-negative ccRCCs with 100% sensitivity and specificity.

Prognostic impact of CIMP diagnosis in the validation cohort

It has previously been revealed that patients with CIMP-positive ccRCCs show a poorer outcome [7]. Therefore, we attempted to validate the prognostic impact of CIMP diagnosis using criteria based on the cutoff values listed in Table 1. Using the additional 100 ccRCCs in the validation cohort, DNA methylation levels at the 23 CpG units including the 32 CpG sites in Table 1 were evaluated quantitatively using the MassARRAY system. The DNA methylation statuses of the 100 ccRCCs in the validation cohort were used to construct a histogram showing the number of CpG units with DNA methylation levels higher than the cutoff values listed in Table 1 (Figure 2C). The distribution of DNA methylation status at the 23 CpG units of the ccRCCs in the validation cohort (Figure 2C) was similar to that in the learning cohort (Figure 2B). Based on the criteria for CIMP diagnosis established using the learning cohort, 5 ccRCCs showing DNA methylation levels higher than the cutoff values listed in Table 1 at 16 or more CpG units were diagnosed as CIMP-positive, whereas 95 ccRCCs showing such higher DNA methylation levels at less than 16 CpG units were diagnosed as CIMP-negative.

Survival curves of the 100 patients belonging to the validation cohort were calculated by the Kaplan-Meier method (Figure 3). The period covered ranged from 27 to 5,031 days (mean: 1,860 days). Cancer-free (Figure 3A) and overall (Figure 3B) survival rates of patients with CIMP-positive ccRCCs diagnosed using the criteria based on the cutoff values listed in Table 1 were significantly lower than those of patients with CIMP-negative ccRCCs ($P = 1.41 \times 10^{-5}$ and 2.43×10^{-13}, respectively, log-rank test).

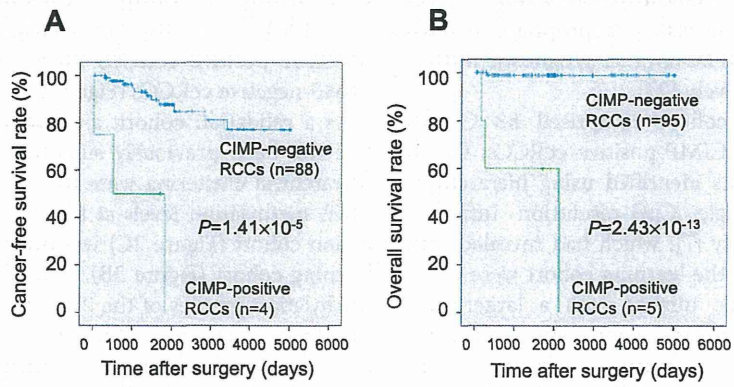

Figure 3 Kaplan–Meier survival curves of patients with CIMP-positive and negative ccRCCs in the validation cohort. Cancer-free (Panel **A**, $P = 1.41 \times 10^{-5}$) and overall (Panel **B**, $P = 2.43 \times 10^{-13}$) survival rates of patients with ccRCCs showing DNA methylation levels higher than the cutoff values listed in Table 1 at 16 or more CpG units (diagnosed as CIMP-positive ccRCCs) were significantly lower than those of patients with ccRCCs showing DNA methylation levels higher than the cutoff values listed in Table 1 at less than 16 CpG units (diagnosed as CIMP-negative ccRCCs). Patients who underwent curative resection were included in panel **A**. The prognostic significance of the criteria for CIMP-diagnosis established in the present study was clearly confirmed in the validation cohort.

Among 5 ccRCCs diagnosed as CIMP-positive in the validation cohort, one tumor was grade 2, two were grade 3, and two were grade 4 (Figure 2C); four were stage III and one was stage IV. Even after adjusting the grades, the cancer-free ($P = 1.01 \times 10^{-3}$) and overall ($P = 7.04 \times 10^{-4}$) survival rates of patients with CIMP-positive high-grade (grades 3 and 4) ccRCCs were significantly lower than those of patients with CIMP-negative high-grade (grades 3 and 4) ccRCCs (log-rank test, Additional file 9: Figure S3). The cancer-free ($P = 7.76 \times 10^{-4}$) and overall ($P = 5.48 \times 10^{-5}$) survival rates of patients with CIMP-positive high-stage (stages III and IV) ccRCCs were significantly lower than those of patients with CIMP-negative high-stage (stages III and IV) ccRCCs in the validation cohort (log-rank test, Additional file 9: Figure S3).

When compared with CIMP-negative ccRCCs, the CIMP-positive ccRCCs in the validation cohort had a significantly higher likelihood of recurrence (hazard ratio, 10.6; 95 percent confidence interval, 2.81 to 40.2; $P = 5.03 \times 10^{-4}$), and of disease-related death (hazard ratio, 75.8; 95 percent confidence interval, 7.81 to 735; $P = 1.89 \times 10^{-4}$) (Cox proportional hazards model). These data indicated that the validation cohort clearly demonstrated the prognostic significance of the criteria for CIMP diagnosis established in the present study.

Discussion

Since the effectiveness of any therapy for relapsed or metastasized ccRCC is restricted unless it is diagnosed early by close follow-up after nephrectomy [3], significant prognostic criteria need to be established. Unlike alterations of mRNA and protein expression, which can be easily affected by the microenvironment of cancer cells, DNA methylation alterations are stably preserved on DNA double strands by covalent bonds [4,5]. Therefore, DNA methylation levels at appropriate marker CpG sites would appear to be optimal prognostic indicators if evaluated quantitatively [27].

The present learning cohort comprised 88 CIMP-negative ccRCCs and 14 CIMP-positive ccRCCs: CIMP in the learning cohort was identified using hierarchical clustering based on single CpG-resolution Infinium assay in our previous study [7], which had revealed that CIMP-positive ccRCCs in the learning cohort were clinicopathologically aggressive tumors with a larger diameter, more frequent vascular involvement, infiltrating growth, and renal pelvis invasion, as well as having higher histological grades and pathological TNM stages than CIMP-negative ccRCCs [7] (Additional file 1: Table S1). During the follow-up period after nephrectomy, the cancer-free and overall survival rates of patients with CIMP-positive ccRCCs in the learning cohort were significantly lower than those of patients with CIMP-negative ccRCCs in our previous study [7], indicating that CIMP in ccRCCs might be applicable as a prognostic indicator.

We previously identified ccRCC-specific CIMP marker genes whose DNA methylation levels differed markedly between CIMP-negative and CIMP-positive ccRCCs based on the Infinium assay [7]. Since hierarchical clustering is not applicable to clinical use, in the present study we attempted to establish criteria for CIMP diagnosis that would be applicable to patients admitted to hospitals on an individual basis. The DNA methylation status of all promoter CpG islands, even CpG sites other than the Infinium probe sites, in the CIMP marker genes was evaluated quantitatively using the MassARRAY system, which is known to be suitable for quantification of multiple CpG sites [24]. Moreover, we carefully optimized the experimental conditions for MassARRAY analysis in order to avoid any PCR bias (Additional file 4: Table S4).

It was revealed that the entire promoter CpG islands in all the CIMP marker genes examined, i.e. *FAM150A, GRM6, ZNF540, ZFP42, ZNF154, RIMS4, PCDHAC1, KHDRBS2, ASCL2, PRAC, WNT3A, TRH, ZNF671* and *SLC13A5*, were methylated in CIMP-positive ccRCCs without exception (Figure 1 and Additional file 7: Table S5). Within such promoter CpG islands, there were many CpG sites where DNA methylation levels were useful for discrimination of CIMP-positive ccRCCs in the learning cohort from CIMP-negative ccRCCs (Additional file 8: Table S6). We identified the top 23 CpG units whose AUC values were larger than 0.95 in ROC analysis, and the Youden index was used as a cutoff value for such discrimination in each CpG unit (Table 1). The sensitivity and specificity of each of the 23 CpG units was sufficient for such discrimination (Table 1 and Figure 2A). Moreover, combination of the 23 CpG units generated criteria with 100% sensitivity and specificity for discrimination of CIMP-positive ccRCCs in the learning cohort from CIMP-negative ccRCCs (Figure 2B).

As a validation cohort, an additional 100 ccRCCs that had not been previously subjected to Infinium assay or hierarchical clustering were analyzed. The distribution of DNA methylation levels at the 23 CpG units in the validation cohort (Figure 2C) was quite similar to that in the learning cohort (Figure 2B), indicating that distinct DNA methylation profiles of the 23 CpG units are reproducible in ccRCCs. In the validation cohort, 5 ccRCCs were diagnosed as CIMP-positive based on the criteria established in the present MassARRAY analysis (Table 1). CIMP-positive ccRCCs diagnosed in the validation cohort had significantly lower cancer-free and overall survival rates than those of CIMP-negative ccRCCs (Figure 3). Even after adjusting the grades and stages, the cancer-free and overall survival rates of patients with high-grade (grade 3/4) and

high-stage (stage III/IV) CIMP-positive ccRCCs were significantly lower than those of patents with high-grade (grade 3/4) and high-stage (stage III/IV) CIMP-negative ccRCCs (Additional file 9: Figure S3). Moreover, CIMP-positive ccRCCs had a higher likelihood of both recurrence and disease-related death (hazard ratios 10.6 and 75.8, respectively). These data indicated that CIMP of ccRCCs can be reproducibly diagnosed using the criteria established in the present study, and that CIMP diagnosis is useful for prognostication of patients with ccRCCs.

Reproducible diagnosis of CIMP using the criteria established in the present study makes it possible to explore the molecular background of CIMP-positive renal carcinogenesis. Since CIMP-positive ccRCCs show clinicopathological aggressiveness and poorer outcome [7], the molecular pathways participating in CIMP-positive renal carcinogenesis should be clarified and the therapeutic targets of CIMP-positive ccRCCs need to be identified. Even though we [28] and another group [29,21] reported the results of multilayer omics analyses in ccRCCs, such reports did not focus on CIMP. Therefore we are now performing multilayer omics (i.e. genome (whole-exome), transcriptome and proteome) analyses of tissue specimens from CIMP-negative and -positive ccRCCs. Frequently affected molecular pathways that might potentially become therapeutic targets are now being identified in more aggressive CIMP-positive ccRCCs (unpublished data).

The criteria for CIMP diagnosis established in the present study may be useful for not only prognostication but also companion diagnostics for personalized medicine [30]. If our CIMP diagnosis reveals CIMP-negativity in samples of tumor tissue obtained by nephrectomy, the risk of recurrence and metastasis would be considered low, and such patients would not require adjuvant therapy. On the other hand, if our CIMP diagnosis reveals CIMP-positivity, then the risk of recurrence and metastasis would be considered high. Therefore, close follow-up and frequent imaging diagnosis are recommended for early diagnosis of recurrence. In addition, inhibitors for frequently affected molecular pathways identified by multilayer omics analysis in CIMP-positive ccRCCs might be effective after recurrence. If further preclinical examinations support the effectiveness of adjuvant therapy using inhibitors for frequently affected molecular pathways in CIMP-positive ccRCCs, such adjuvant therapy may be recommended immediately after nephrectomy in patients with CIMP-positive ccRCCs.

Conclusions

CIMP of ccRCCs is characterized by accumulation of DNA methylation at CpG islands and poorer patient outcome. Based on quantification of DNA methylation levels of the ccRCC-specific CIMP marker genes, the criteria for CIMP diagnosis have been established. CIMP of ccRCCs can be reproducibly diagnosed using the criteria established in the present study. The prognostic significance of the criteria has been clearly validated in the validation cohort. Frequently affected molecular pathways that might potentially become therapeutic targets are now being identified using multilayer omics analyses in more aggressive CIMP-positive ccRCCs. The criteria for CIMP diagnosis may be useful for not only prognostication but also companion diagnostics for personalized medicine.

Additional files

Additional file 1: Table S1. Correlation between CpG island methylator phenotype (CIMP) and clinicopathological parameters of clear cell renal cell carcinomas (ccRCCs) in the learning cohort.

Additional file 2: Table S2. Clinicopathological characteristics of clear cell renal cell carcinomas (ccRCCs) in the validation cohort.

Additional file 3: Table S3. Microscopically examined tumor cell content (%) of specimens of clear cell renal cell carcinoma tissue from the learning and validation cohorts.

Additional file 4: Table S4. Primer sequences and optimal PCR conditions for MassARRAY.

Additional file 5: Figure S1. Standard curves for optimization of PCR conditions for the MassARRAY system on representative CpG units. To test the linearity of the protocol, 0%, 50% and 100% methylated control DNA was used as a template. Experiments were performed in triplicate for each sample-CpG unit, and the mean value for the three experiments was used as the DNA methylation level. Error bar: standard deviation. Optimized PCR conditions (annealing temperature and type of DNA polymerases) are summarized in Additional file 4: Table S4.

Additional file 6: Figure S2. Scattergrams of DNA methylation levels determined by Infinium assay and those determined by MassArray analysis. [a]Probe ID for the Infinium HumanMethylation27 Bead Array. Six exact Infinium probe CpG sites (cg06274159 for the *ZFP42* gene, cg03975694 for the *ZNF540* gene, cg08668790 for the *ZNF154* gene, cg01009664 for the *TRH* gene, cg22040627 for the *SLC13A5* gene, and cg19246110 for the *ZNF671* gene) were examined by the MassArray platform in the learning cohort. Significant correlations between DNA methylation levels determined by our previous Infinium assay [7] and those determined by the present MassArray analysis were confirmed ($P = 1.25 \times 10^{-35}$, $P = 1.98X10^{-32}$, $P = 1.31 \times 10^{-41}$, $P = 5.30 \times 10^{-34}$, $P = 7.91 \times 10^{-22}$ and $P = 7.61 \times 10^{-44}$, respectively).

Additional file 7: Table S5. Differences of DNA methylation levels at all examined 193 CpG units including 299 CpG sites of 14 CpG island methylator phenotype (CIMP) marker genes between CIMP-negative and CIMP-positive clear cell renal cell carcinomas (ccRCCs) in the learning cohort.

Additional file 8: Table S6 Eighty-six CpG units showing area under the curve (AUC) values larger than 0.9 in receiver operating characteristic curve analysis for discrimination of CpG island methylator phenotype (CIMP)-positive clear cell renal cell carcinomas (ccRCCs) from CIMP-negative ccRCCs in the learning cohort.

Additional file 9: Figure S3. Kaplan–Meier survival curves of patients with CIMP-positive and -negative high-grade (grades 3 and 4) and high-stage (stages III and IV) clear cell renal cell carcinomas (ccRCCs) in the validation cohort. The cancer-free (Panel A, $P = 1.01 \times 10^{-3}$) and overall (Panel B, $P = 7.04 \times 10^{-4}$) survival rates of patients with CIMP-positive grade 3/4 ccRCCs were significantly lower than those of patients with CIMP-negative grade 3/4 ccRCCs (log-rank test). The cancer-free (Panel C, $P = 7.76 \times 10^{-4}$) and overall (Panel D, $P = 5.48 \times 10^{-5}$) survival rates of patients with CIMP-positive stage III/IV ccRCCs were significantly lower than those of patients with CIMP-negative stage III/V ccRCCs (log-rank test). Patients who underwent curative resection are included in panels A and C.

Abbreviations
AUC: Area under the curve; CIMP: CpG island methylator phenotype; CV: Cutoff value; NCBI: National Center for Biotechnology Information; ccRCC: Clear cell renal cell carcinoma; ROC: Receiver operating characteristic.

Competing interests
The authors declare that they have no competing interests.

Author's contributions
EA and YK were responsible for the study design, development of the analysis plan and study management. YT, EA, and MG performed MassARRAY and statistical analyses. EA, MK, HF and YK collected tissue samples and performed clinicopathological analysis. YT, EA and YK interpreted the data and prepared the manuscript. All authors read and approved the final manuscript.

Acknowledgments
This work was supported by The Program for Promotion of Fundamental Studies in Health Sciences (10–42) of the National Institute of Biomedical Innovation (NiBio), Japan, A Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control (19140201) from the Ministry of Health, Labor and Welfare of Japan and Grants-in-Aid for Scientific Research (B) (23390090) and (C) (25460487) from the Japan Society for the Promotion of Science (JSPS), Japan. National Cancer Center Biobank is supported by the National Cancer Center Research and Development Fund (23A-1), Japan.

Author details
[1]Division of Molecular Pathology, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. [2]Department of Urology, National Cancer Center Hospital, Tokyo 104-0045, Japan.

Received: 25 June 2014 Accepted: 8 October 2014
Published: 20 October 2014

References
1. Ljungberg B, Campbell SC, Choi HY, Jacqmin D, Lee JE, Weikert S, Kiemeney LA: **The epidemiology of renal cell carcinoma.** *Eur Urol* 2011, **60**:615–621.
2. Donat SM, Diaz M, Bishoff JT, Coleman JA, Dahm P, Derweesh IH, Herrell SD 3rd, Hilton S, Jonasch E, Lin DW, Reuter VE, Chang SS: **Follow-up for clinically localized renal neoplasms: AUA guideline.** *J Urol* 2013, **190**:407–416.
3. Logan JE, Rampersaud EN, Sonn GA, Chamie K, Belldegrun AS, Pantuck AJ, Slamon DJ, Kabbinavar FF: **Systemic therapy for metastatic renal cell carcinoma: a review and update.** *Rev Urol* 2012, **14**:65–78.
4. You JS, Jones PA: **Cancer genetics and epigenetics: two sides of the same coin?** *Cancer Cell* 2012, **22**:9–20.
5. Baylin SB, Jones PA: **A decade of exploring the cancer epigenome - biological and translational implications.** *Nat Rev Cancer* 2011, **11**:726–734.
6. Kanai Y: **Genome-wide DNA methylation profiles in precancerous conditions and cancers.** *Cancer Sci* 2010, **101**:36–45.
7. Arai E, Chiku S, Mori T, Gotoh M, Nakagawa T, Fujimoto H, Kanai Y: **Single-CpG-resolution methylome analysis identifies clinicopathologically aggressive CpG island methylator phenotype clear cell renal cell carcinomas.** *Carcinogenesis* 2012, **33**:1487–1493.
8. Arai E, Wakai-Ushijima S, Fujimoto H, Hosoda F, Shibata T, Kondo T, Yokoi S, Imoto I, Inazawa J, Hirohashi S, Kanai Y: **Genome-wide DNA methylation profiles in renal tumors of various histological subtypes and non-tumorous renal tissues.** *Pathobiology* 2011, **78**:1–9.
9. Arai E, Ushijima S, Fujimoto H, Hosoda F, Shibata T, Kondo T, Yokoi S, Imoto I, Inazawa J, Hirohashi S, Kanai Y: **Genome-wide DNA methylation profiles in both precancerous conditions and clear cell renal cell carcinomas are correlated with malignant potential and patient outcome.** *Carcinogenesis* 2009, **30**:214–221.
10. Arai E, Kanai Y, Ushijima S, Fujimoto H, Mukai K, Hirohashi S: **Regional DNA hypermethylation and DNA methyltransferase (DNMT) 1 protein overexpression in both renal tumors and corresponding nontumorous renal tissues.** *Int J Cancer* 2006, **119**:288–296.
11. Rydzanicz M, Wrzesiński T, Bluyssen HA, Wesoły J: **Genomics and epigenomics of clear cell renal cell carcinoma: Recent developments and potential applications.** *Cancer Lett* 2013, **341**:111–126.
12. Issa JP: **CpG island methylator phenotype in cancer.** *Nat Rev Cancer* 2004, **4**:988–993.
13. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP: **CpG island methylator phenotype in colorectal cancer.** *Proc Natl Acad Sci U S A* 1999, **96**:8681–8686.
14. Shen L, Toyota M, Kondo Y, Lin E, Zhang L, Guo Y, Hernandez NS, Chen X, Ahmed S, Konishi K, Hamilton SR, Issa JP: **Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer.** *Proc Natl Acad Sci U S A* 2007, **104**:18654–18659.
15. Toyota M, Ahuja N, Suzuki H, Itoh F, Ohe-Toyota M, Imai K, Baylin SB, Issa JP: **Aberrant methylation in gastric cancer associated with the CpG island methylator phenotype.** *Cancer Res* 1999, **59**:5438–5442.
16. Morris MR, Maher ER: **Epigenetics of renal cell carcinoma: the path towards new diagnostics and therapeutics.** *Genome Med* 2010, **2**:59.
17. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL: **Genome-wide DNA methylation profiling using Infinium® assay.** *Epigenomics* 2009, **1**:177–200.
18. Eble JN, Sauter G, Epstein JI, Sesterhenn IA: **Renal cell carcinoma.** In *World Health Organization classification of tumours. Pathology and genetics. Tumours of the urinary system and male genital organs.* Lyon: IARC Press; 2004:10–43.
19. Fuhrman SA, Lasky LC, Limas C: **Prognostic significance of morphologic parameters in renal cell carcinoma.** *Am J Surg Pathol* 1982, **6**:655–663.
20. Sobin LH, Wittekind C: *International Union Against Cancer (UICC). TNM classification of malignant tumors.* 6th edition. New York: Wiley-Liss; 2002:193–195.
21. The Cancer Genome Atlas Research Network: **Comprehensive molecular characterization of clear cell renal cell carcinoma.** *Nature* 2013, **499**:43–49.
22. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: A Laboratory Manual.* 3rd edition. New York: Cold Spring Harbor Laboratory Press; 2001:6.14–6.15.
23. Clark SJ, Harrison J, Paul CL, Frommer M: **High sensitivity mapping of methylated cytosines.** *Nucleic Acids Res* 1994, **22**:2990–2997.
24. Jurinke C, Denissenko MF, Oeth P, Ehrich M, van den Boom D, Cantor CR: **A single nucleotide polymorphism based approach for the identification and characterization of gene expression modulation using MassARRAY.** *Mutat Res* 2005, **573**:83–95.
25. Fan J, Upadhye S, Worster A: **Understanding receiver operating characteristic (ROC) curves.** *CJEM* 2006, **8**:19–20.
26. Akobeng AK: **Understanding diagnostic tests 3: Receiver operating characteristic curves.** *Acta Paediatr* 2007, **96**:644–647.
27. Arai E, Kanai Y: **DNA methylation profiles in precancerous tissue and cancers: carcinogenetic risk estimation and prognostication based on DNA methylation status.** *Epigenomics* 2010, **2**:467–481.
28. Arai E, Sakamoto H, Ichikawa H, Totsuka H, Chiku S, Gotoh M, Mori T, Nakatani T, Ohnami S, Nakagawa T, Fujimoto H, Wang L, Aburatani H, Yoshida T, Kanai Y: **Multilayer-omics analysis of renal cell carcinoma, including the whole exome, methylome and transcriptome.** *Int J Cancer* 2014, **135**:1330–1342.
29. Sato Y, Yoshizato T, Shiraishi Y, Maekawa S, Okuno Y, Kamura T, Shimamura T, Sato-Otsubo A, Nagae G, Suzuki H, Nagata Y, Yoshida K, Kon A, Suzuki Y, Chiba K, Tanaka H, Niida A, Fujimoto A, Tsunoda T, Morikawa T, Maeda D, Kume H, Sugano S, Fukayama M, Aburatani H, Sanada M, Miyano S, Homma Y, Ogawa S: **Integrated molecular analysis of clear-cell renal cell carcinoma.** *Nat Genet* 2013, **45**:860–867.
30. Ziegler A, Koch A, Krockenberger K, Grosshennig A: **Personalized medicine using DNA biomarkers: a review.** *Hum Genet* 2012, **131**:1627–1638.

doi:10.1186/1471-2407-14-772
Cite this article as: Tian *et al.*: Prognostication of patients with clear cell renal cell carcinomas based on quantification of DNA methylation levels of CpG island methylator phenotype marker genes. *BMC Cancer* 2014 **14**:772.

UICC global cancer control

IJC International Journal of Cancer

Epigenetic clustering of lung adenocarcinomas based on DNA methylation profiles in adjacent lung tissue: Its correlation with smoking history and chronic obstructive pulmonary disease

Takashi Sato[1,2], Eri Arai[1], Takashi Kohno[3], Yoriko Takahashi[4], Sayaka Miyata[4], Koji Tsuta[5], Shun-ichi Watanabe[6], Kenzo Soejima[2], Tomoko Betsuyaku[2] and Yae Kanai[1]

[1] Division of Molecular Pathology, National Cancer Center Research Institute, Tokyo 104-0045, Japan
[2] Division of Pulmonary Medicine, Department of Medicine, Keio University School of Medicine, Tokyo 160-8582, Japan
[3] Division of Genome Biology, National Cancer Center Research Institute, Tokyo 104-0045, Japan
[4] Bioscience Department, Research and Development Center, Mitsui Knowledge Industry Co., Ltd., Tokyo 105-6215, Japan
[5] Division of Pathology, Department of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo 104-0045, Japan
[6] Division of Thoracic Surgery, Department of Thoracic Oncology, National Cancer Center Hospital, Tokyo 104-0045, Japan

The aim of this study was to clarify the significance of DNA methylation alterations during lung carcinogenesis. Infinium assay was performed using 139 paired samples of non-cancerous lung tissue (N) and tumorous tissue (T) from a learning cohort of patients with lung adenocarcinomas (LADCs). Fifty paired N and T samples from a validation cohort were also analyzed. DNA methylation alterations on 1,928 probes occurred in N samples relative to normal lung tissue from patients without primary lung tumors, and were inherited by, or strengthened in, T samples. Unsupervised hierarchical clustering using DNA methylation levels in N samples on all 26,447 probes subclustered patients into Cluster I (n = 32), Cluster II (n = 35) and Cluster III (n = 72). LADCs in Cluster I developed from the inflammatory background in chronic obstructive pulmonary disease (COPD) in heavy smokers and were locally invasive. Most patients in Cluster II were non-smokers and had a favorable outcome. LADCs in Cluster III developed in light smokers were most aggressive (frequently showing lymphatic and blood vessel invasion, lymph node metastasis and an advanced pathological stage), and had a poor outcome. DNA methylation levels of hallmark genes for each cluster, such as *IRX2*, *HOXD8*, *SPARCL1*, *RGS5* and *EI24*, were again correlated with clinicopathological characteristics in the validation cohort. DNA methylation profiles reflecting carcinogenetic factors such as smoking and COPD appear to be established in non-cancerous lung tissue from patients with LADCs and may determine the aggressiveness of tumors developing in individual patients, and thus patient outcome.

Cancer Genetics

Lung cancer is the leading cause of cancer-related death worldwide,[1] and adenocarcinoma is the most common histological subtype, both in smokers and non-smokers. Differences in the genetic features of lung adenocarcinomas (LADCs) between smokers and non-smokers have been described.[2] LADCs arising in individuals who have never smoked, especially women and those of East Asian ethnicity, have been reported to have *EGFR* mutation and are thus responsive to tyrosine kinase inhibitors, whereas those arising in smokers frequently show oncogenic missense mutations in *KRAS*. *EGFR* and *KRAS* mutations in LADCs are almost entirely mutually exclusive. With regard to *TP53*

Key words: DNA methylation, infinium assay, lung adenocarcinoma, cigarette smoking, chronic obstructive pulmonary disease

Abbreviations: AAH: atypical adenomatous hyperplasia; C: normal lung tissue; COPD: chronic obstructive pulmonary disease; FDR: false discovery rate; LADC: lung adenocarcinomas; N: non-cancerous lung tissue; ROC: receiver operating characteristic curve; T: tumorous tissue; TNM: tumor-node-metastasis

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Cancer Center Research and Development Fund; **Grant number:** 23A-1 (National Cancer Center Biobank); **Grant sponsor:** The National Institute of Biomedical Innovation (NiBio), the Ministry of Health, Labor and Welfare of Japan, and the Japan Society for the Promotion of Science (JSPS)

DOI: 10.1002/ijc.28684

History: Received 26 June 2013; Revised 29 Nov 2013; Accepted 5 Dec 2013; Online 19 Dec 2013

Correspondence to: Eri Arai, Division of Molecular Pathology, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan, Tel.: +81-3-3542-2511, Fax: +81-3-3248-2463, E-mail: earai@ncc.go.jp

What's new?

While genetic abnormalities are well studied in human cancers, epigenetic changes, especially in the early stages of carcinogenesis, remain largely unknown. Here, the authors perform a genome-wide analysis focusing on DNA methylation profiles in "normal" lung tissue adjacent to lung adenocarcinomas. Using single-CpG-resolution Infinium assays, they identify distinct DNA methylation profiles clustering with specific risk factors such as cigarette smoking, inflammation and chronic obstructive pulmonary disease. The authors speculate that these epigenetic profiles detected in the neighboring cells may influence the aggressiveness of tumors developing in individual patients and may thus help predict disease outcome.

mutations, G:C to T:A transversions and A:T to G:C transitions at CpG sites are characteristic of smoking-related lung cancers, whereas G:C to A:T transitions at non-CpG sites are associated with lung cancers in individuals who have never smoked. However, the molecular changes responsible for the development of LADCs in both smokers and nonsmokers, especially at the very early stages, are not yet fully understood.

As well as genetic abnormalities, epigenetic changes have been described in human cancers,[3] one of the most consistent being DNA methylation alterations. In LADCs, silencing of the *RASSF1A, CDKN2A, RARβ, MGMT, APC, DAPK, FHIT* and *CDH13* genes due to DNA hypermethylation around their promoter regions has been frequently reported.[4] Moreover, in various organs, DNA methylation alterations are characteristically observed even at the precancerous stage[5–7]: we and other groups have reported aberrant DNA methylation of specific genes or chromosomal loci in non-cancerous lung tissue from LADC patients, or in lung tissue from cancer-free smokers.[4,8,9] DNA methylation alterations of tumor-related genes have been reported in airway epithelial cells from smokers.[8,10,11] Recently, methylome analysis using single-CpG-resolution Infinium assay has been introduced.[12] Although studies of lung cancers using the Infinium assay by Selamat *et al.*[13] and Lockwood *et al.*[14] did not focus on noncancerous lung tissue obtained from the same patients, our previous study revealed that alterations of DNA methylation status in adjacent lung tissue are not nonsensical, but in fact create alterations in the expression of mRNAs for specific genes in cancerous tissue developing in the same individual patients.[15]

It is known that DNA methylation profiles at the precancerous stage are determined by carcinogenetic factors. For examples, distinct DNA methylation profiles at the chronic hepatitis or liver cirrhosis stage as a precancerous condition for hepatocellular carcinoma[16,17] or those in the stomach mucosa harboring *Helicobacter pylori* infection as a precancerous condition for stomach adenocarcinoma have been reported.[18] In this study, to further understand the significance of DNA methylation alterations during lung carcinogenesis, we examined correlations between epigenetic clustering of patients with LADCs based on DNA methylation profiles in adjacent lung tissue and carcinogenetic factors such as cigarette smoking and chronic obstructive lung disease (COPD).

Material and Methods

Patients and tissue samples

As a learning cohort, 139 paired samples of non-cancerous lung tissue (N) and the corresponding tumorous tissue (T) were obtained from patients with primary LADCs who underwent lung resection at the National Cancer Center Hospital, Japan, between December 2000 and May 2008. None of these patients had received any preoperative treatment. Sixty-nine patients were males and seventy were females with a median age of 60 years (range, 30–76 years). Clinicopathological parameters in the learning cohort are summarized in Supporting Information Table S1. Pleural anthracosis, which mainly reflects the cumulative effects of smoking history, was evaluated macroscopically according to the criteria described previously.[19] Presence or absence of emphysematous change, respiratory bronchiolitis, interstitial fibrosis[20,21] and atypical adenomatous hyperplasia (AAH, a precancerous lesion for LADC)[22,23] was evaluated microscopically on the basis of the criteria described previously. Histological diagnosis and grading were based on the 2004 World Health Organization classification.[24] When, within a tumor, black dusty material[25] is seen to have accumulated in foci of active fibroblast proliferation, reflecting active cancer–stromal interaction associated with a poorer outcome in LADC patients,[26] the tumor is considered to be tumor anthracosis-positive (Supporting Information Fig. S1). All the tumors were classified according to the pathological tumor-node-metastasis (TNM) classification.[27] Recurrence was diagnosed by clinicians on the basis of physical examination and imaging modalities such as computed tomography, magnetic resonance imaging, scintigraphy or positron-emission tomography, and sometimes confirmed histopathologically by biopsy. A proportion of this cohort had also been included in our previous study focusing on recurrence-related genes.[15]

DNA methylation profiles of the 139 N samples and 139 T samples were compared with previously reported DNA methylation profiles of 36 samples of normal lung tissue (C) obtained from specimens surgically resected from 36 patients without any primary lung tumors.[15] Briefly, 22 of these patients were males and 14 were females, with a median age of 63 years (range, 27–83 years). Thirty-five had undergone lung resection for metastatic lesions from primary cancers of the colon, rectum, kidney, urinary bladder, thyroid, breast, pancreas, ampulla of Vater and salivary gland, osteosarcoma, synovial sarcoma, leiomyosarcoma, rhabdomyosarcoma,

liposarcoma, dermatofibrosarcoma and myxofibrosarcoma. The remaining one patient had undergone chest wall resection for lipoma with removal of adjacent lung tissue.

As a validation cohort, 50 paired samples of N and the corresponding T were obtained from patients with primary LADCs who underwent lung resection at the National Cancer Center Hospital, Japan, between December 1997 and May 2000. None of these patients had received any preoperative treatment. Thirty-three patients were males and seventeen were females with a median age of 63 years (range, 40–81 years). Clinicopathological parameters in the validation cohort are summarized in Supporting Information Table S1.

Tissue specimens were provided by the National Cancer Center Biobank, Japan. This study was approved by the Ethics Committee of the National Cancer Center, Japan, and was performed in accordance with the Declaration of Helsinki. All patients included in this study provided written informed consent.

Infinium assay

Genomic DNA was extracted from all tissue samples using a QIAamp DNA Mini kit (Qiagen, Valencia, CA). Five-hundred-nanogram aliquots of DNA were subjected to bisulfite conversion using an EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, CA). Subsequently, DNA methylation status at 27,578 CpG loci was examined at single-CpG resolution using the Infinium HumanMethylation27 Bead Array (Illumina, San Diego, CA). This array contains CpG sites located mainly within the proximal promoter regions of the transcription start sites of 14,475 consensus coding sequences in the National Center for Biotechnology Information Database. An Evo robot (Tecan, Männedorf, Switzerland) was used for automated sample processing. After whole-genome amplification and hybridization, the specifically hybridized DNA was fluorescence-labeled by a single-base extension reaction and detected using a BeadScan reader (Illumina) in accordance with the manufacturer's protocols. The data were then assembled using GenomeStudio methylation software (Illumina). At each CpG site, the ratio of the fluorescence signal was measured using a methylated probe relative to the sum of the methylated and unmethylated probes, that is, the so-called β-value, which ranges from 0.00 to 1.00, reflecting the methylation level of an individual CpG site.

The reliability of DNA methylation levels (β-values) determined by Infinium assay has been verified in our previous studies.[7,15] In addition, DNA methylation levels of the representative genes (*NUPR1, EVI2B, CASP8* and *KRTAP11-1* genes) based on the Infinium assay in representative samples included in this study were verified using the quantitative pyrosequencing method (Supporting Information Fig. S2), thus confirming the reliability of the Infinium assay. Moreover, we compared the DNA methylation levels of 545 representative Infinium probes, whose β values were unrelated to the clinicopathological parameters of the tumors or patient outcome (recurrence or death), between all samples in the learning cohort (obtained between December 2000 and May 2008) and the validation cohort (obtained between December 1997 and May 2000). No significant differences in DNA methylation levels between the learning and validation cohorts were observed in any of the 545 probes examined (Supporting Information Fig. S3). Supporting Information Figure S3 clearly indicates the excellent concordance of DNA methylation status between the two cohorts ($r = 1.000$, $p < 2.20 \times 10^{-16}$), confirming that the epigenetic changes did not degrade over time.

Statistics

In the Infinium assay, all CpG sites on chromosomes X and Y were excluded, to avoid any gender-specific methylation bias. In addition, the call proportions (p-value of <0.01 for detection of signals above the background) for 39 probes (shown in Supporting Information Table S2) in 36 C samples, 139 N samples and 139 corresponding T samples in the learning cohort were less than 90%. As such a low proportion may be attributable to polymorphism at the probe CpG sites, these 39 probes were excluded from the present assay, leaving a final total of 26,447 autosomal CpG sites.

Infinium probes showing significant differences in DNA methylation levels between the 36 C samples and 139 N samples in the learning cohort were identified by the Welch's t-test. Ordered differences from 36 C to 139 N, and then to 139 T samples themselves in the learning cohort were examined by the Jonckheere–Terpstra trend test. A false discovery rate (FDR) of $q = 0.01$ was considered significant. Unsupervised hierarchical clustering (Euclidean distance, Ward method) based on DNA methylation levels of the 139 N samples in the learning cohort was performed. Correlations between clusters of patients and clinicopathological parameters were examined using Kruskal–Wallis test, Fisher's exact test and Kruskal–Wallis exact test at a significance level of $p < 0.05$. Survival curves of patients belonging to each cluster were calculated by the Kaplan–Meier method, and the differences were compared by the Log-rank test. The hallmark genes discriminating the clusters were identified by Welch's t-test. Correlations between DNA methylation levels of such hallmark genes in N samples and clinicopathological parameters of patients in the validation cohort were examined using Welch's t-test and ANOVA test at a significance level of $p < 0.05$. All statistical analyses were performed using programming language R.

Results

DNA methylation alterations during lung carcinogenesis

(*i*) Welch's t-test revealed that DNA methylation levels on the 3,778 probes were already altered in N samples in the learning cohort relative to those in C samples (FDR, $q = 0.01$, Table 1A). (*ii*) The Jonckheere–Terpstra trend test revealed ordered differences in the DNA methylation level from the 39 C samples to the 139 N samples, and then to the 139 T samples themselves in the learning cohort on the 12,368 probes (FDR, $q = 0.01$, Table 1B). (*iii*) Among the probes, 1,928 satisfied

Table 1. DNA methylation alterations during lung carcinogenesis

The number of probes showing DNA hypermethylation and DNA hypomethylation	
(A) The probes on which DNA methylation levels were altered in 139 samples of non-cancerous lung tissue (N) obtained from patients with lung adenocarcinomas (LADCs) in the learning cohort relative to those in 39 samples of normal lung tissue (C) obtained from patients without any primary lung tumors. (Welch's *t*-test, False discovery rate [FDR] $q = 0.01$)	
DNA hypermethylation ($\beta_C < \beta_N$)	1,526
DNA hypomethylation ($\beta_C > \beta_N$)	2,252
Total	3,778
(B) The probes on which DNA methylation levels showed ordered differences from 39 C samples to 139 N samples, and then to 139 tumorous tissue (T) samples in the learning cohort. (Jonckheere–Terpstra trend test, FDR $q = 0.01$)	
DNA hypermethylation ($\beta_C < \beta_N < \beta_T$, $\beta_C < \beta_N \fallingdotseq \beta_T$ or $\beta_C \fallingdotseq \beta_N < \beta_T$)	6,460
DNA hypomethylation ($\beta_C > \beta_N > \beta_T$, $\beta_C > \beta_N \fallingdotseq \beta_T$ or $\beta_C \fallingdotseq \beta_N > \beta_T$)	5,908
Total	12,368
(C) The probes satisfying both of the above criteria (A) and (B): DNA methylation alterations on these probes occurred even in N samples relative to C samples, and such DNA methylation alterations were inherited by, or strengthened in, T samples.	
DNA hypermethylation ($\beta_C < \beta_N < \beta_T$ or $\beta_C < \beta_N \fallingdotseq \beta_T$)	484
DNA hypomethylation ($\beta_C > \beta_N > \beta_T$ or $\beta_C > \beta_N \fallingdotseq \beta_T$)	1,444
Total	1,928

the above criteria (*i*) and (*ii*): DNA methylation alterations on the 1,928 probes occurred even in N samples relative to C samples, and such DNA methylation alterations were inherited by, or strengthened in, the T samples (Table 1C).

Epigenetic clustering of LADCs based on DNA methylation profiles in N samples

As DNA methylation alterations already occurred in Ns, unsupervised hierarchical clustering using DNA methylation levels in N samples (β_N) on all 26,447 probes was performed in 139 patients with LADCs in the learning cohort. Such clustering based on DNA methylation profiles in N samples subclustered 139 patients in the learning cohort into Cluster I ($n = 32$), Cluster II ($n = 35$) and Cluster III ($n = 72$, Fig. 1*a*). The clinicopathological parameters of the patients in these clusters are summarized in Table 2.

Most of the patients in Cluster I were heavy smokers (median number of cigarettes smoked per day × year index: 810) and frequently showed severe pleural anthracosis, which mainly reflects the cumulative effects of smoking.[19] With regard to the non-cancerous lung tissue, patients belonging to Cluster I frequently showed histological findings compatible with emphysema, respiratory bronchiolitis and interstitial fibrosis, and they frequently suffered from obstructive ventilation impairment (Table 2). In Cluster I, LADCs with a large diameter, a progressed T stage, a high histological grade and frequent pleural invasion were accumulated (Table 2). In addition, tumor anthracosis reflecting active cancer–stromal interaction[26] was frequent in Cluster I (Table 2). These data indicated that LADCs in Cluster I were locally invasive tumors.

Most of the patients in Cluster II were non-smokers (median number of cigarettes smoked per day × year index: 0) and less frequently showed emphysematous changes in their adjacent lung tissue (Table 2). The correlation between epigenetic clustering of LADCs and patient age and sex may be attributable to the fact that younger female non-smokers[28] were accumulated in Cluster II. LADCs in Cluster II showed less aggressive clinicopathological features (Table 2).

Most of the patients in Cluster III were light smokers and tended to have a lower incidence of emphysematous changes in their adjacent lung tissue (Table 2). LADCs in Cluster III frequently showed lymphatic vessel invasion, blood vessel invasion, high N stage and high TNM stage (Table 2), indicating that they were the most aggressive tumors.

Figure 1*b* shows the Kaplan–Meier survival curves of patients belonging to Clusters I, II and III. The period covered ranged from 196 to 3,957 days (mean, 1,634 days). The cancer-free and overall survival rates of patients in Cluster III were significantly lower than those of patients in Cluster II ($p = 1.24 \times 10^{-4}$ and $p = 1.58 \times 10^{-2}$, respectively, Fig. 1*b*).

DNA methylation profiles of N samples belonging to each cluster in the learning cohort

Scattergrams of average DNA methylation levels in N samples ($_{average}\beta_N$) of patients belonging to Clusters I, II and III and average DNA methylation levels in C samples ($_{average}\beta_C$) for all 26,447 probes are shown in Figure 2. In Cluster I, DNA methylation levels on probes normally showing a low or medium degree of DNA methylation ($_{average}\ \beta_C < 0.6$) were elevated in N samples relative to C samples, and DNA methylation levels on probes normally showing a high or medium degree of DNA methylation ($_{average}\ \beta_C > 0.3$) were reduced in N samples relative to C samples (Fig. 2*a*). In Cluster II, DNA methylation levels on probes normally showing a low degree of DNA methylation ($_{average}\ \beta_C < 0.2$) were elevated in N samples relative to C samples, and DNA methylation levels on probes normally showing a high degree of DNA methylation ($_{average}\ \beta_C > 0.7$) were reduced in N

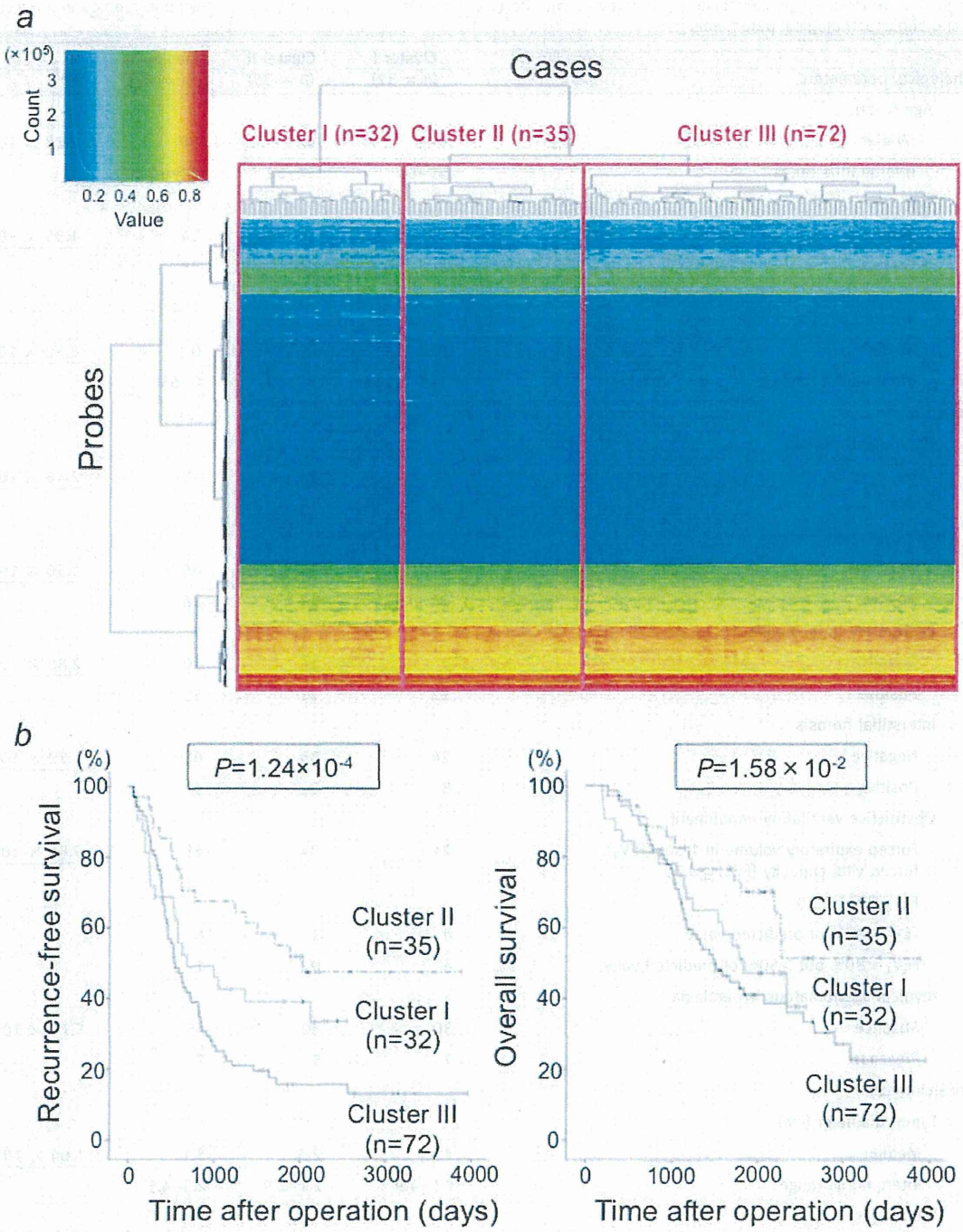

Figure 1. (*a*) Unsupervised hierarchical clustering (Euclidean distance, Ward method) using DNA methylation levels on all 26,447 probes in samples of non-cancerous lung tissue (N) from 139 patients with lung adenocarcinomas in the learning cohort. Based on DNA methylation status in adjacent lung tissue, 139 patients were subclustered into Cluster I ($n = 32$), Cluster II ($n = 35$) and Cluster III ($n = 72$). Correlations between this epigenetic clustering and clinicopathological parameters of the patients are summarized in Table 2. (*b*) Kaplan–Meier survival curves of patients belonging to Clusters I, II and III. The period covered ranged from 196 to 3,957 days (mean, 1,634 days). The cancer-free ($p = 1.24 \times 10^{-4}$) and overall ($p = 1.58 \times 10^{-2}$) survival rates of patients in Cluster III were significantly lower than those of patients in Cluster II (log-rank test).

Table 2. Correlation between epigenetic clustering of patients with lung adenocarcinomas based on DNA methylation profiles in adjacent lung tissue and clinicopathological parameters

Clinicopathological parameters		Cluster I (n = 32)	Cluster II (n = 35)	Cluster III (n = 72)	P[1]
Patients	Age (year)				
	Median	64	57	60	2.03×10^{-2} [2]
	Interquartile range	59–68	54–62	53–64	
	Sex				
	Male	24	11	34	1.35×10^{-3} [3]
	Female	8	24	38	
	Smoking history (number of cigarettes smoked per day × year index)				
	Median	810	0	0	8.80×10^{-6} [2]
	Interquartile range	195–1,113	0–140	0–635	
Adjacent lung tissue					
	Pleural anthracosis				
	G1	13	24	48	2.46×10^{-2} [4]
	G2-3	19	11	24	
	Emphysematic change				
	Negative	8	24	46	2.50×10^{-4} [4]
	Positive	24	11	26	
	Respiratory bronchiolitis				
	Negative	2	14	10	2.80×10^{-3} [4]
	Positive	22	21	58	
	Interstitial fibrosis				
	Negative	24	35	68	5.72×10^{-4} [4]
	Positive	8	0	4	
	Obstructive ventilation impairment				
	Forced expiratory volume in 1 sec (FEV_1): forced vital capacity (FVC) $\geq$0.70	24	34	65	9.86×10^{-3} [4]
	FEV_1:FVC <0.70				
	FEV_1 $\geq$80% of predicted value	4	1	6	
	FEV_1 <80% but $\geq$50% of predicted value	4	0	1	
	Atypical adenomatous hyperplasia				
	Absence	30	30	65	5.72×10^{-1} [4]
	Presence	2	5	7	
Lung adenocarcinomas					
	Tumor diameter (cm)				
	Median	3.4	2.3	3.1	1.64×10^{-4} [4]
	Interquartile range	2.5–4.9	2.1–2.9	2.5–4.5	
	Tumor stage				
	T1a-T1b	6	19	19	1.60×10^{-4} [4]
	T2a-T2b	12	14	39	
	T3-4	14	2	14	
	Histological grades				
	G1	8	20	26	2.37×10^{-3} [4]
	G2	11	12	34	
	G3	13	3	12	

Table 2. Correlation between epigenetic clustering of patients with lung adenocarcinomas based on DNA methylation profiles in adjacent lung tissue and clinicopathological parameters (Continued)

Clinicopathological parameters	Cluster I (n = 32)	Cluster II (n = 35)	Cluster III (n = 72)	P[1]
Tumor anthracosis				
Negative	6	20	39	1.70×10^{-3} [4]
Positive	25	15	33	
Pleural invasion				
Negative	12	22	35	9.62×10^{-3} [4]
Invasion to the visceral pleura beyond the elastic fiber	6	9	17	
Invasion to the surface of the visceral pleura	4	4	15	
Invation to the parietal pleura	10	0	5	
Lymphatic vessel invasion				
Negative	9	18	16	8.54×10^{-3} [4]
Positive	23	17	56	
Blood vessel invasion				
Negative	7	18	15	3.02×10^{-3} [4]
Positive	25	17	57	
Nodal status				
N0	17	26	25	8.72×10^{-5} [4]
N1	10	6	18	
N2-3	5	3	29	
Metastatic status				
M0	31	34	66	4.40×10^{-1} [4]
M1a-1b	1	1	1	
Pathological Tumor-Node-Metastasis stage				
IA-IB	5	24	18	4.36×10^{-6} [4]
IIA-IIB	21	7	19	
IIIA-IV	6	4	35	

[1]*P*values of <0.05 are underlined.
[2]Kruskal-Wallis test.
[3]Fisher's exact test.
[4]Kruskal-Wallis exact test.

samples relative to C samples (Fig. 2*b*). In Cluster III, DNA methylation levels on probes normally showing a high or medium degree of DNA methylation ($_{\text{average}}\ \beta_C > 0.3$) were reduced in N samples relative to C samples (Fig. 2*c*).

Hallmark CpG sites for each cluster in the learning cohort

One hundred sixteen CpG sites were identified as hallmarks of the DNA methylation profile (Fig. 2*a*) of N samples belonging to Cluster I: on these 116 CpG sites, the average β_{N-C} values in Cluster I were significantly different from those in Clusters II and III (Welch's *t*-test, $p < 1 \times 10^{-3}$) and the average β_{N-C} value in Cluster I was 0.1 or more higher or lower than those in Clusters II and III (Table 3A and Supporting Information Table S3). One CpG site was identified as a hallmark for the DNA methylation profile (Fig. 2*b*) of N samples belonging to Cluster II: on the CpG site, the average β_{N-C} value in Cluster II was significantly different from that in Clusters I and III (Welch's *t*-test, $p < 1 \times 10^{-3}$) and the average β_{N-C} value in Cluster II was 0.1 or more higher than those in Clusters I and III (Table 3B). Four CpG sites were identified as a hallmark for the DNA methylation profile (Fig. 2*c*) of N samples belonging to Cluster III: on the four CpG sites, average β_{N-C} values in Cluster III were significantly different from those in Clusters I and II (Welch's *t*-test, $p < 1 \times 10^{-3}$) and average β_{N-C} values in Cluster III were 0.1 or more higher or lower than those in Clusters I and II (Table 3C). In 119 of the 120 CpG sites in Table 3 or Supporting Information Table S3, which were identified based on the DNA methylation profiles in N samples, stepwise DNA methylation alterations from C to N, and then to T samples were revealed by Jonckheere–Terpstra trend test (Table 3 and Supporting Information Table S3).

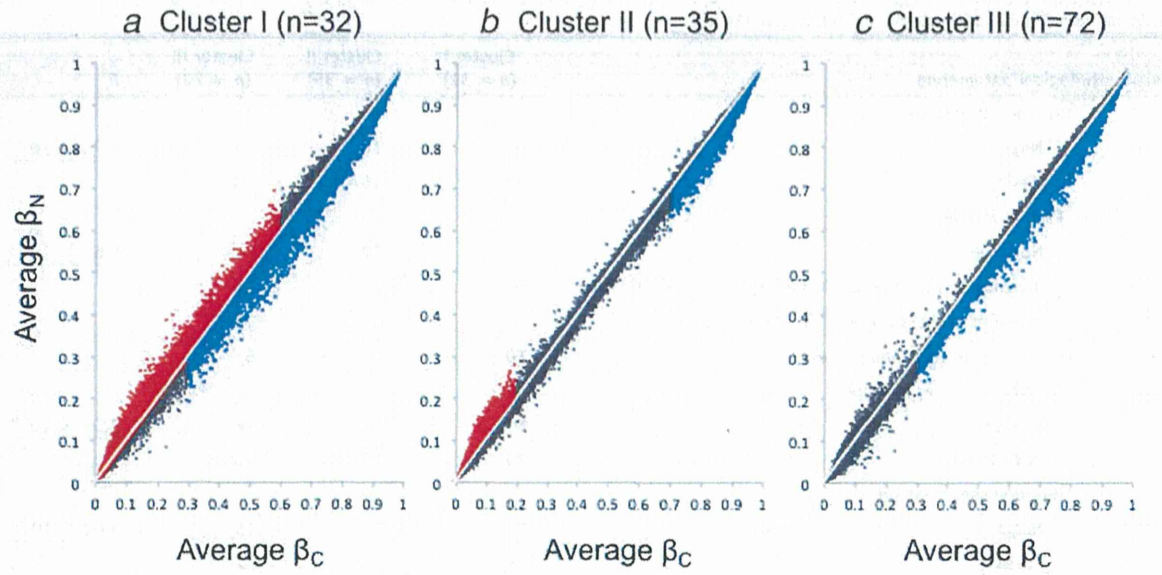

Figure 2. Distribution of average DNA methylation levels on all 26,447 probes of non-cancerous lung tissue (N) samples obtained from patients with lung adenocarcinomas belonging to Clusters I (*a*), II (*b*) and III (*c*) and 36 samples of normal lung tissue (C) obtained from patients without any primary lung tumors. (*a*) In Cluster I, DNA methylation levels on probes normally showing a lower or medium degree of DNA methylation ($_{average}\beta_C < 0.6$, red) were elevated in N samples relative to C samples, and DNA methylation levels on probes normally showing a higher or medium degree of DNA methylation ($_{average}\ \beta_C > 0.3$, blue) were reduced in N samples relative to C samples. (*b*) In Cluster II, DNA methylation levels on probes normally showing a lower degree of DNA methylation ($_{average}\ \beta_C < 0.2$,red) were elevated in N samples relative to C samples, and DNA methylation levels on probes normally showing a higher degree of DNA methylation ($_{average}\ \beta_C > 0.7$,blue) were reduced in N samples relative to C samples. (*c*) In Cluster III, DNA methylation levels on probes normally showing a higher or medium degree of DNA methylation ($_{average}\ \beta_C > 0.3$,blue) were reduced in N samples relative to C samples.

DNA methylation profiles in the validation cohort

The correlations between the DNA methylation status of hallmark CpG sites for Clusters I, II and III in N samples and clinicopathological parameters of patients in the validation cohort were examined. DNA methylation levels on 17 and 2 hallmark CpG sites for Cluster I were significantly correlated with pleural anthracosis and pulmonary emphysema in the adjacent lung tissue in the validation cohort, respectively (Table 4A), whereas hallmark CpG sites for Clusters II and III never showed such a correlation. In addition, in the validation cohort, DNA methylation levels on 18 hallmark CpG sites for Cluster I were significantly correlated with the presence of AAH, a precancerous lesion for LADCs, in the adjacent lung tissue (Table 4A), even though the correlation between the presence of AAH and epigenetic clustering did not reach statistically significant levels (Table 2). DNA methylation levels on 13 hallmark CpG sites for Cluster I were significantly correlated with tumor anthracosis in LADCs in the validation cohort (Table 4A), whereas hallmark genes for Clusters II and III never showed such a correlation. Hallmark genes for Cluster I showing such correlations with pleural anthracosis, emphysema, presence of AAH or tumor anthracosis are described in Table 3A, and hallmark genes not showing such correlations are described in Supporting Information Table S3.

Hallmark gene *ABCC12* was shared between Clusters II and III. The DNA methylation level of *ABCC12* was significantly correlated with N stage and TNM stage in the validation cohort (Table 4B). In the learning cohort, the DNA methylation level of the *ABCC12* gene was high in Cluster II showing low N and TNM stages, and that of the *ABCC12* gene was low in Cluster III showing high N and TNM stages. Therefore, it is feasible that the DNA methylation level of the *ABCC12* gene was significantly higher in patients showing lower N and TNM stages in the validation cohort (Table 4B). DNA methylation levels of two of the three remaining hallmark genes (three hallmark genes other than *ABCC12*) for Cluster III were significantly correlated with lymph vessel invasion in LADCs in the validation cohort, and the DNA methylation levels of all three remaining hallmark genes for Cluster III were significantly correlated with high N and TNM stages (Table 4B). Taken together, correlations between DNA methylation profiles in N samples and clinicopathological characteristics in the adjacent lung tissue or LADCs in the learning cohort were reproduced in the validation cohort.

Discussion

In this study, we focused on DNA methylation profiles in the adjacent non-cancerous lung tissue obtained from patients with LADCs and analyzed the results of methylome analysis of lung tissue samples including 189 N samples at single-CpG resolution. DNA methylation alterations occurred even in N samples relative to C samples, and were inherited by, or

Table 3. Genes for which DNA methylation levels were hallmarks for Clusters I, II and III in the learning cohort

(A) Hallmark genes for Cluster I

Target ID [1]	Chrom[2]	Position[3]	Gene symbol	DNA methylation level in non-cancerous lung tissue (N) samples[4] (mean ± SD)			*p*-Value of Welch's *t*-test (I *vs.* II and III)[5]	Δβ (I-II and III)[6]	*p*-Value of Jonckheere–Terpstra trend test in I[7]
				Cluster I	Cluster II	Cluster III			
cg20249919	15	102,029,706	*PCSK6*	0.091 ± 0.188	−0.047 ± 0.109	−0.070 ± 0.125	9.28×10^{-5}	0.153	6.51×10^{-4} (Hyper)
cg23349790	1	18,434,576	*IGSF21*	0.114 ± 0.133	−0.011 ± 0.111	−0.032 ± 0.108	2.41×10^{-6}	0.139	4.43×10^{-9} (Hyper)
cg22285621	11	67,071,322	*SSH3*	0.103 ± 0.116	−0.031 ± 0.075	−0.033 ± 0.082	2.32×10^{-7}	0.136	3.69×10^{-7} (Hyper)
cg15433631	5	2,751,541	*IRX2*	0.123 ± 0.083	−0.007 ± 0.073	0.000 ± 0.070	8.88×10^{-10}	0.125	6.60×10^{-8} (Hyper)
cg21949305	22	24,828,655	ADORA2A, CYTSA	0.109 ± 0.053	−0.015 ± 0.040	−0.010 ± 0.052	2.91×10^{-15}	0.121	0 (Hyper)
cg10942056	1	223,101,848	*DISP1*	0.095 ± 0.059	−0.027 ± 0.039	−0.026 ± 0.048	1.59×10^{-13}	0.121	4.05×10^{-13} (Hyper)
cg15149645	16	28,550,619	NUPR1	0.090 ± 0.067	−0.023 ± 0.044	−0.033 ± 0.058	7.39×10^{-12}	0.12	1.36×10^{-12} (Hyper)
cg06954481	2	237,076,497	*GBX2*	0.096 ± 0.111	−0.012 ± 0.051	−0.029 ± 0.052	1.02×10^{-6}	0.119	1.25×10^{-7} (Hyper)
cg21250978	7	106,684,541	*PRKAR2B*	0.088 ± 0.060	−0.026 ± 0.044	−0.031 ± 0.056	4.25×10^{-13}	0.118	6.13×10^{-9} (Hyper)
cg22418909	8	41,166,738	SFRP1	0.091 ± 0.082	−0.023 ± 0.055	−0.029 ± 0.052	2.38×10^{-9}	0.118	1.22×10^{-10} (Hyper)
cg26200585	19	40,919,245	*PRX*	0.099 ± 0.059	−0.019 ± 0.040	−0.019 ± 0.054	2.44×10^{-13}	0.118	0 (Hyper)
cg24396745	15	73,660,614	*HCN4*	0.096 ± 0.098	−0.022 ± 0.073	−0.015 ± 0.089	3.31×10^{-7}	0.114	1.96×10^{-8} (Hyper)
cg04330449	5	134,871,166	*NEUROG1*	0.098 ± 0.080	−0.001 ± 0.061	−0.019 ± 0.051	5.89×10^{-9}	0.111	1.08×10^{-13} (Hyper)
cg19589427	1	173,019,720	*TNFSF18*	0.076 ± 0.073	−0.036 ± 0.039	−0.032 ± 0.051	9.08×10^{-10}	0.11	7.78×10^{-10} (Hyper)
cg16731240	19	52,391,250	*ZNF577*	0.090 ± 0.105	−0.015 ± 0.072	−0.022 ± 0.061	1.87×10^{-6}	0.11	0 (Hyper)
cg03544320	4	5,894,691	*CRMP1*	0.088 ± 0.108	−0.016 ± 0.105	−0.022 ± 0.101	7.22×10^{-6}	0.108	1.61×10^{-10} (Hyper)
cg12864235	5	27,038,782	*CDH9*	0.092 ± 0.059	−0.011 ± 0.037	−0.018 ± 0.040	3.56×10^{-12}	0.108	2.37×10^{-13} (Hyper)
cg15898840	7	45,960,834	*IGFBP3*	0.102 ± 0.095	−0.001 ± 0.052	−0.008 ± 0.058	4.67×10^{-7}	0.107	2.02×10^{-8} (Hyper)
cg08044694	19	15,391,927	BRD4	0.068 ± 0.072	−0.029 ± 0.034	−0.044 ± 0.042	1.55×10^{-9}	0.107	1.76×10^{-8} (Hyper)
cg03734874	14	105,071,382	*TMEM179*	0.099 ± 0.068	0.001 ± 0.056	−0.012 ± 0.055	3.06×10^{-10}	0.106	4.39×10^{-13} (Hyper)
cg10599444	14	23,305,941	*MMP14*	0.064 ± 0.065	−0.039 ± 0.040	−0.044 ± 0.056	8.35×10^{-11}	0.106	7.42×10^{-7} (Hyper)
cg24133115	6	166,075,520	*PDE10A*	0.096 ± 0.071	−0.007 ± 0.054	−0.010 ± 0.046	1.50×10^{-9}	0.105	9.66×10^{-10} (Hyper)
cg12594641	2	150,187,223	*LYPD6*	0.111 ± 0.064	0.011 ± 0.071	0.004 ± 0.061	1.05×10^{-10}	0.105	6.56×10^{-7} (Hyper)
cg05724065	7	56,160,528	PHKG1	0.082 ± 0.053	−0.017 ± 0.029	−0.026 ± 0.044	3.01×10^{-13}	0.105	4.43×10^{-11} (Hyper)
cg19466563	4	88,450,506	SPARCL1	0.081 ± 0.053	−0.018 ± 0.027	−0.027 ± 0.042	4.93×10^{-13}	0.104	0 (Hyper)
cg24433189	16	1,128,689	SSTR5	0.092 ± 0.056	−0.005 ± 0.052	−0.015 ± 0.064	2.58×10^{-12}	0.104	9.78×10^{-9} (Hyper)
cg24453664	11	33,758,413	CD59	0.069 ± 0.066	−0.031 ± 0.033	−0.036 ± 0.046	3.23×10^{-10}	0.103	9.78×10^{-9} (Hyper)

Int. J. Cancer: 135, 319–334 (2014)

Table 3. Genes for which DNA methylation levels were hallmarks for Clusters I, II and III in the learning cohort (Continued)

(A) Hallmark genes for Cluster I

Target ID [1]	Chrom[2]	Position[3]	Gene symbol	DNA methylation level in non-cancerous lung tissue (N) samples[4] (mean ± SD)			*p*-Value of Welch's *t*-test (I *vs.* II and III)[5]	Δβ (I-II and III)[6]	*p*-Value of Jonckheere–Terpstra trend test in I[7]
				Cluster I	Cluster II	Cluster III			
cg26609631	13	28,366,814	GSX1	0.077 ± 0.081	−0.025 ± 0.063	−0.026 ± 0.057	4.72×10^{-8}	0.103	1.73×10^{-11} (Hyper)
cg10604646	1	163,172,649	RGS5	0.086 ± 0.041	−0.029 ± 0.059	−0.009 ± 0.060	4.09×10^{-17}	0.102	2.68×10^{-14} (Hyper)
cg03355526	5	178,368,415	ZNF454	0.073 ± 0.070	−0.024 ± 0.043	−0.030 ± 0.061	2.48×10^{-9}	0.101	9.13×10^{-13} (Hyper)
cg27096144	5	174,151,779	MSX2	0.074 ± 0.078	−0.020 ± 0.054	−0.030 ± 0.056	3.60×10^{-8}	0.101	2.11×10^{-7} (Hyper)
cg15520279	2	176,995,088	HOXD8	0.095 ± 0.083	0.008 ± 0.048	−0.013 ± 0.046	1.06×10^{-7}	0.1	1.30×10^{-13} (Hyper)
cg11733245	10	6,104,312	IL2RA	−0.112 ± 0.066	−0.001 ± 0.028	−0.016 ± 0.050	5.58×10^{-10}	−0.101	8.03×10^{-13} (Hypo)
cg22325572	1	111,416,181	CD53	−0.102 ± 0.062	0.013 ± 0.035	−0.007 ± 0.048	9.63×10^{-11}	−0.102	3.52×10^{-12} (Hypo)
cg15691199	14	23,589,419	CEBPE	−0.102 ± 0.061	0.006 ± 0.033	−0.003 ± 0.052	4.72×10^{-11}	−0.102	1.33×10^{-9} (Hypo)
cg16927606	19	36,233,324	U2AF1L4	−0.086 ± 0.048	0.013 ± 0.028	0.018 ± 0.044	3.10×10^{-14}	−0.103	1.79×10^{-8} (Hypo)
cg16240480	1	236,557,473	EDARADD	−0.128 ± 0.064	−0.005 ± 0.039	−0.030 ± 0.049	7.69×10^{-11}	−0.106	1.59×10^{-9} (Hypo)
cg05596756	12	47,610,220	FAM113B	−0.102 ± 0.060	0.009 ± 0.029	0.016 ± 0.047	8.28×10^{-13}	−0.116	1.53×10^{-10} (Hypo)
cg08040471	17	80,407,779	C17orf62	−0.116 ± 0.067	0.008 ± 0.036	0.004 ± 0.047	6.99×10^{-12}	−0.121	5.63×10^{-11} (Hypo)
cg20622019	20	43,279,793	ADA	−0.108 ± 0.072	0.020 ± 0.043	0.012 ± 0.042	3.92×10^{-11}	−0.123	1.56×10^{-13} (Hypo)
cg05109049	17	29,641,333	EVI2B	−0.141 ± 0.081	0.007 ± 0.050	−0.020 ± 0.063	1.98×10^{-10}	−0.13	2.31×10^{-14} (Hypo)
cg07973967	17	62,009,607	CD79B	−0.125 ± 0.061	0.016 ± 0.047	0.002 ± 0.056	2.00×10^{-14}	−0.132	2.81×10^{-11} (Hypo)

(B) Hallmark genes for Cluster II

Target ID[8]	Chrom[9]	Position[10]	Gene symbol	DNA methylation level in non-cancerous lung tissue (N) samples[11] (mean ± SD)			*p*-value of Welch's *t*-test (II *vs.* I and III)[12]	Δβ (II-I and III) [13]	*p*-value of Jonckheere–Terpstra trend test in II[14]
				Cluster I	Cluster II	Cluster III			
cg14074641	16	48,181,753	*ABCC12*	−0.002 ± 0.091	0.025 ± 0.054	−0.109 ± 0.105	1.01×10^{-10}	0.101	7.05×10^{-2} (Hyper)

(C) Hallmark genes for Cluster III

Target ID [15]	Chrom[16]	Position[17]	Gene symbol	DNA methylation level in non-cancerous lung tissue (N) samples[18] (mean ± SD)			*p*-Value of Welch's *t*-test (III *vs.* I and II)[19]	Δβ (III-I and II) [20]	*p*-Value of Jonckheere–Terpstra trend test in III[21]
				Cluster I	Cluster II	Cluster III			
cg26606064	11	125,439,070	*EI24*	0.020 ± 0.083	0.008 ± 0.064	0.115 ± 0.105	8.57×10^{-10}	0.101	2.36×10^{-2} (Hyper)
cg17872476	10	114,205,654	*VTI1A*	−0.034 ± 0.091	−0.035 ± 0.060	−0.137 ± 0.120	1.61×10^{-8}	−0.102	1.51×10^{-2} (Hypo)

Int. J. Cancer: 135, 319–334 (2014) © 2013 The Authors. Published by Wiley Periodicals, Inc. on behalf of UICC