Research Article
See related commentary by Schneider and Peek, p. 253

Prevention of *Helicobacter pylori*–Induced Gastric Cancers in Gerbils by a DNA Demethylating Agent

Tohru Niwa[1], Takeshi Toyoda[2], Tetsuya Tsukamoto[3], Akiko Mori[1], Masae Tatematsu[4], and Toshikazu Ushijima[1]

Abstract

Suppression of aberrant DNA methylation is a novel approach to cancer prevention, but, so far, the efficacy of the strategy has not been evaluated in cancers associated with chronic inflammation. Gastric cancers induced by *Helicobacter pylori* infection are known to involve aberrant DNA methylation and associated with severe chronic inflammation in their early stages. Here, we aimed to clarify whether suppression of aberrant DNA methylation can prevent *H. pylori*–induced gastric cancers using a Mongolian gerbil model. Administration of a DNA demethylating agent, 5-aza-2′-deoxycytidine (5-aza-dC), to gerbils (0.125 mg/kg for 50–55 weeks) decreased the incidence of gastric cancers induced by *H. pylori* infection and *N*-methyl-*N*-nitrosourea (MNU) treatment from 55.2% to 23.3% ($P < 0.05$). In gastric epithelial cells, DNA methylation levels of six CpG islands (HE6, HG2, SB1, SB5, SF12, and SH6) decreased to 46% to 68% ($P < 0.05$) of gerbils without 5-aza-dC treatment. Also, the global DNA methylation level decreased from 83.0% ± 4.5% to 80.3% ± 4.4% (mean ± SD) by 5-aza-dC treatment ($P < 0.05$). By 5-aza-dC treatment, *Il1b* and *Nos2* were downregulated (42% and 58% of gerbils without, respectively) but *Tnf* was upregulated (187%), suggesting that 5-aza-dC treatment induced dysregulation of inflammatory responses. No obvious adverse effect of 5-aza-dC treatment was observed, besides testicular atrophy. These results showed that 5-aza-dC treatment can prevent *H. pylori*–induced gastric cancers and suggested that removal of induced DNA methylation and/or suppression of DNA methylation induction can become a target for prevention of chronic inflammation–associated cancers. *Cancer Prev Res; 6(4); 263–70. ©2013 AACR.*

Introduction

DNA methylation is an epigenetic mechanism for gene regulation. Methylation of promoter CpG islands (CGIs) consistently suppresses expression of their downstream genes (1), and physiologic methylation of retrotransposons is involved in their transcriptional repression (2). In cancers, tumor suppressor genes are frequently inactivated by aberrant methylation of their promoter CGIs (3, 4). Such aberrant methylation is present not only in cancers but also in noncancerous tissues exposed to chronic inflammation, such as colonic mucosae with ulcerative colitis, liver tissues exposed to hepatitis, and gastric mucosae exposed to chronic gastritis (5–10). In the case of the stomach, *Helicobacter pylori* infection is known to induce severe chronic inflammation (11–13) and aberrant methylation in gastric epithelial cells (GEC; ref. 14). Accumulation levels of aberrant methylation in gastric mucosae correlate with risk of gastric cancers (8–10).

Suppression of aberrant methylation is considered as one of the novel targets for cancer chemoprevention (15, 16). Traditionally, chemoprevention has used substances based on 2 strategies: the anti-initiation and anti-promotion/progression strategies (17–19). In the former strategy, blockage of activity of carcinogens that induce genetic or epigenetic alterations and enhancement of repair systems have been targeted. In the latter strategy, suppression of proliferation of initiated cells and induction of their apoptosis have been targeted. However, neither of these strategies targeted removal of genetic or epigenetic alterations accumulated in the cells, which can be achieved by DNA demethylating agents, such as 5-aza-2′-deoxycytidine (5-aza-dC; refs. 20, 21).

The usefulness of DNA demethylating agents in cancer chemoprevention has been shown in several animal models, including intestinal tumors in $Apc^{min/+}$ mice (22, 23), prostate tumors in transgenic mice harboring probasin promoter-driven SV40 antigen (24), 4-(methyl-nitrosamino)-1-(3-pyridyl)-1-butanone–induced mouse lung tumors (25), and 4-nitroquinoline 1-oxide–induced mouse

Authors' Affiliations: [1]Division of Epigenomics, National Cancer Center Research Institute; [2]Division of Pathology, National Institute of Health Sciences, Tokyo; [3]Department of Diagnostic Pathology, Fujita Health University School of Medicine, Kutsukake-cho, Toyoake, Aichi; and [4]Japan Bioassay Research Center, Hadano, Kanagawa, Japan

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

Corresponding Author: Toshikazu Ushijima, National Cancer Center Research Institute, 5-1-1-Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. Phone: 81-3-3547-5240; Fax: 81-3-5565-1753; E-mail: tushijim@ncc.go.jp

doi: 10.1158/1940-6207.CAPR-12-0369

oral tumors (26). Genetic suppression of a maintenance DNA methyltransferase (*Dnmt1*) also suppressed tumor development in some of these models (22, 27, 28). However, so far, the efficacy of suppression of aberrant DNA methylation was not evaluated in chronic inflammation–associated cancers, in which aberrant DNA methylation is heavily involved (29, 30). From this aspect, gastric cancers induced by *H. pylori* infection of Mongolian gerbils (*Meriones unguiculatus*) have several advantages. In gerbils, *H. pylori* infection induces severe chronic inflammation, as in humans, and promotes gastric cancers initiated by *N*-methyl-*N*-nitrosourea (MNU; refs. 31, 32). Also, 10 CGIs have already been established as markers that can be methylated by *H. pylori* infection, and a critical role of inflammation triggered by *H. pylori* infection, not a direct effect of *H. pylori*, in methylation induction has been shown (14). In contrast, few markers for methylation induction have been isolated in *H. pylori*- or *Helicobacter felis*–infected mice, except *Tff2* promoter (33).

In this study, using the gerbil model, we aimed to clarify whether 5-aza-dC treatment can prevent chronic inflammation–induced gastric cancers and evaluate its effects on methylation induction and inflammation triggered by *H. pylori* infection.

Materials and Methods

Animals and sample preparation

Male Mongolian gerbils (MGS/Sea) were purchased from Kyudo and divided into 10 groups (G1–6 in Fig. 1A and G7–10 in Fig. 2A). Gerbils were inoculated with *H. pylori* [~4×10^8 colony-forming units (CFU)/gerbil, ATCC 43504; American Type Culture Collection] at 5 weeks of age (34). In a carcinogenicity experiment, 10 ppm of MNU (Sigma-Aldrich) was given in drinking water to gerbils. 5-Aza-dC (125 μg/kg body weight in sterilized PBS; Sigma-Aldrich) was administered to gerbils intraperitoneally twice per week. The dose was selected from the 3 doses (125, 250, and 500 μg/kg) tested in a preliminary experiment for lack of toxicity. Timing and duration of the treatments are shown in Figs. 1A and 2A.

The stomach was resected and cut along the greater curvature. In a carcinogenicity experiment (G1–6), the antral region was fixed in formalin for histologic analysis. From the body region, GECs were isolated by the gland isolation technique (35) and stored in 100% ethanol at −80°C until DNA extraction. The testes, small intestine, liver, and kidneys were resected, and half parts were fixed in formalin. The other halves were snap-frozen for DNA extraction. In an experiment to induce *H. pylori*–triggered gastritis (G7–10), the antral region was cut into 2 pieces—one was snap-frozen for RNA extraction and the other half was fixed in formalin.

In both experiments, samples (tissues or GECs) were digested with proteinase K, and gDNA was extracted by the standard phenol/chloroform method. RNA of gastric tissue was isolated using ISOGEN (Nippon Gene). Whole blood was obtained from the inferior vena cava, and gDNA was extracted by a QuickGene DNA Whole Blood Kit (Fujifilm). All the animal experiments were approved by the Committee for Ethics in Animal Experimentation at the National Cancer Center.

Histological analysis

Formalin-fixed tissues were sliced along the longitudinal axis into strips of 5 mm width and embedded in paraffin. Sections of 3 μm thickness were prepared and stained with hematoxylin and eosin. Neoplastic lesions in the stomach were diagnosed as previously described (36). The size of a gastric cancer was determined as the maximum diameter of the gastric cancer in the neighboring sections. The degree of infiltration of mononuclear and polymorphonuclear cells was graded on a 4-point scale (0–3; 0, no or faint; 1, mild; 2, moderate; 3, marked) as described (37).

Luminometric methylation assay

gDNA from whole blood was amplified by an illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare) and used as fully unmethylated DNA. The unmethylated DNA was methylated by *Sss*I methylase (New England Biolabs) and used as fully methylated DNA. A series of standard DNA was prepared by serial mixing of the unmethylated DNA and the methylated DNA.

Luminometric methylation assay (LUMA) was conducted as described (38) with slight modifications. Briefly, 3 μg of DNA was digested with 2 pairs of restriction enzymes (*Hpa*II and *Eco*RI or *Msp*I and *Eco*RI) in independent tubes (all restriction enzymes were purchased from Toyobo). The DNA was purified with a DNA Clean & Concentrator Kit (Zymo Research) and eluted in 40 μL of an annealing buffer (2 mmol/L magnesium acetate and 20 mmol/L Tris-acetate, pH7.6). Using the PSQ 96 Pyrosequencing System (Qiagen), 5′-CG overhang produced by *Hpa*II (or *Msp*I) and 5′-AATT overhang produced by *Eco*RI was sequenced, and an *Hpa*II/*Eco*RI (or *Msp*I/*Eco*RI) signal ratio was determined. An *Hpa*II/*Msp*I value was obtained as (*Hpa*II/EcoRI)/(*Msp*I/*Eco*RI) in each sample. The *Hpa*II/*Msp*I value was compared with those of the standard DNA series, and the global methylation level (GML), which is equivalent to the percentage of methylated DNA in the standard DNA, was determined.

Quantitative methylation-specific PCR

DNA digested with *Bam*HI was treated with sodium bisulfite as described (39) and used as a template for real-time PCR. With primer sets specific to methylated CGIs (HE6, HG2, SA9, SB1, SB5, SC3, SD2, SE3, SF12, and SH6; Supplementary Fig. S1) and a B2 repeat sequence, real-time PCR was carried out as described (14). On the basis of the copy number of sequences measured by real-time PCR, the methylation level was calculated as a percentage of the methylated reference (PMR), which was obtained as [(number of methylated fragments of a target CGI in sample)/(number of the B2 repeat in sample)]/[(number of

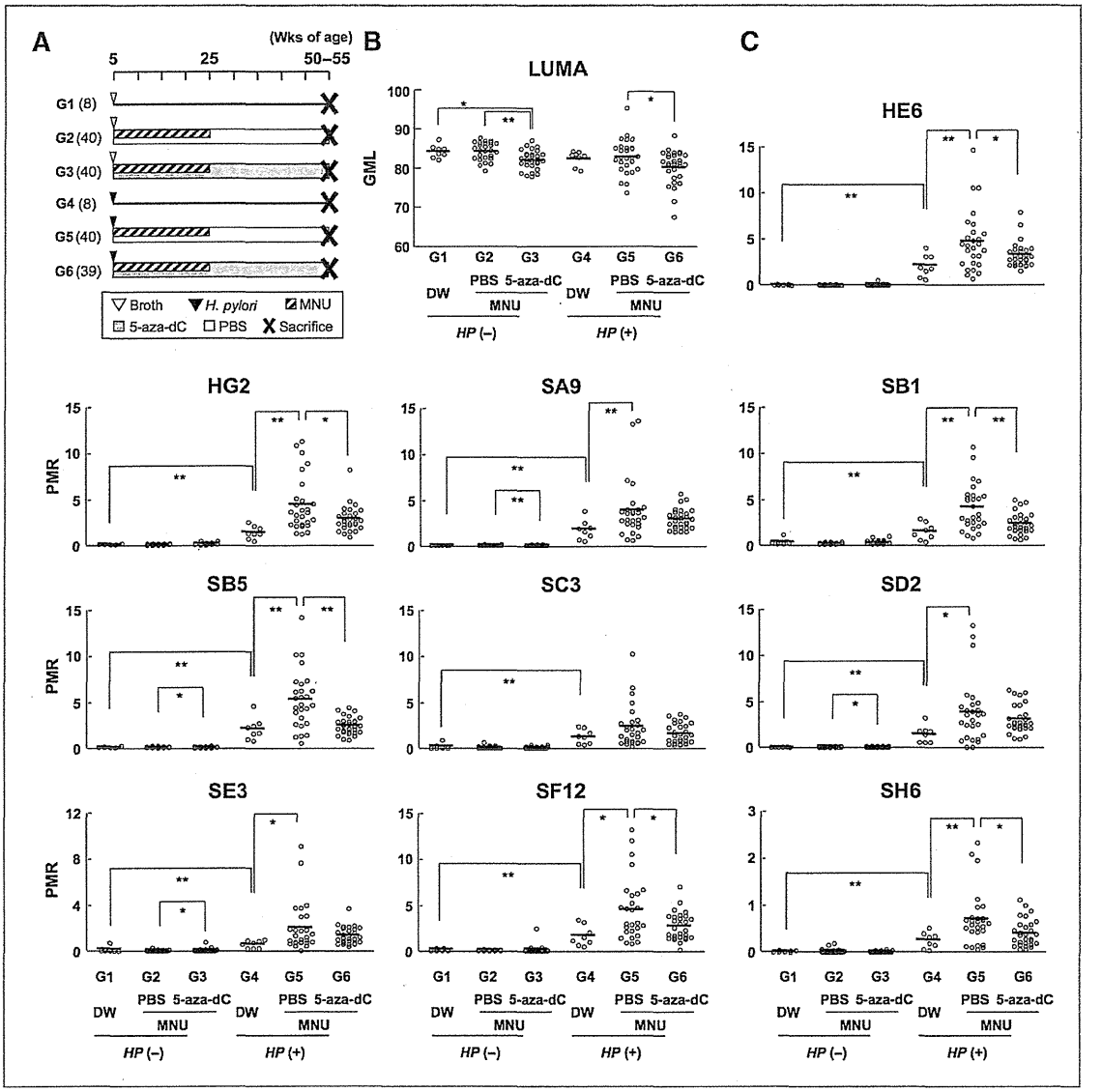

Figure 1. Suppression of DNA methylation by 5-aza-dC treatment. A, design of the carcinogenicity experiment. The number of animals is shown in parentheses. B, global DNA methylation levels analyzed by LUMA in GECs. C, methylation levels of 10 CGIs analyzed by qMSP in GECs. Both preexisting methylation and induction of aberrant methylation were suppressed by 5-aza-dC treatment. Bold horizontal bar, the average. *, $P < 0.05$; **, $P < 0.01$.

methylated fragments of a target CGI in *SssI*-treated DNA)/(number of the B2 repeat in *SssI*-treated DNA)] × 100.

Gene expression analyses

The number of cDNA molecules was quantified by quantitative reverse transcriptase-PCR (qRT-PCR) using gene-specific primers (*Il1b*, *Nos2*, and *Tnf*) as described (14). The number of cDNA molecules of a target gene was normalized to that of *Gapdh*.

Genomic PCR and sequencing

A forward primer (5′-AGATTCCTTGATGCCTGGGTGTC-3′) was designed in a region of the mouse *Tnf* promoter highly conserved with the human corresponding region. A reverse primer (5′-AGATTCCTTGATGCCTGGGTGTC-3′) was designed on the gerbil *Tnf* mRNA sequence (AB177841). The gerbil *Tnf* promoter was amplified using these primers, and the PCR product was directly sequenced with the same primers. The obtained sequence was

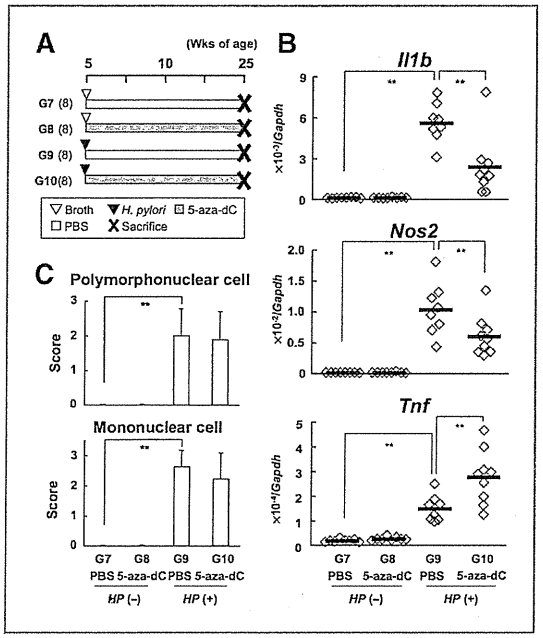

Figure 2. Dysregulation of inflammatory responses by 5-aza-dC treatment. A, design of the experiment for analysis of gastritis. The number of animals is shown in parentheses. B, expression levels of inflammation-related genes in gastric tissues containing submucosal layers. Bold horizontal bar indicates the average. C, infiltration score of polymorphonuclear cells and mononuclear cells in the stomach. Although the degree of *H. pylori*–induced infiltration of polymorphonuclear and mononuclear cells was not affected by 5-aza-dC treatment, expression of *Il1b* and *Nos2* was downregulated and that of *Tnf* was upregulated. Mean and SD are shown. **, $P < 0.01$.

registered in GenBank (AB762083.1). A CGI was searched by a EMBOSS CpG report program (40).

Statistical analysis

Statistical analyses were conducted by SPSS 13.0J (SPSS Japan Inc.). To evaluate significant difference between 2 independent groups of sample data, a Mann–Whitney *U* test was used. The difference of the proportion between 2 groups was evaluated by Fisher exact test.

Results

Suppression of *H. pylori*/MNU-induced gastric carcinogenesis by 5-aza-dC

To evaluate cancer prevention effects of 5-aza-dC, a carcinogenicity experiment was carried out (Fig. 1A). Among the gerbils with MNU treatment and *H. pylori* infection (G5 and G6), 5-aza-dC treatment decreased incidence of gastric cancers from 55.2% (G5) to 23.3% (G6, $P < 0.05$; Table 1). The incidence in G6 was similar to that in MNU-treated gerbils without *H. pylori* infection (G2, 20.7%). There were no significant differences in the tumor multiplicity and size among the groups. These results clearly showed that 5-aza-dC treatment suppressed *H. pylori*/MNU-induced gastric carcinogenesis in gerbils and suggested that it might have completely abrogated the cancer promotion effects of *H. pylori* infection. Among the MNU-treated gerbils without *H. pylori* infection (G2 and G3), 5-aza-dC treatment tended to decrease incidence of gastric cancers but it was not statistically significant.

Regardless of 5-aza-dC treatment, gerbils with MNU treatment (G2, G3, G5, and G6) showed low body weight than the gerbils without MNU treatment (G1 and G4), showing that the body weight loss was dependent upon MNU treatment, not upon 5-aza-dC treatment (Supplementary Fig. S2). Survival rates started to decrease from 25 weeks of age, and the decrease was dependent upon MNU treatment, not upon 5-aza-dC treatment (Supplementary Fig. S3). This showed that the dose of 5-aza-dC used in this study (125 μg/kg body weight) had no obvious effects on body weight and survival of gerbils.

Reduction of DNA methylation levels in GECs by 5-aza-dC

To confirm the demethylating effects of 5-aza-dC *in vivo*, methylation analyses were conducted in GECs. First, the global DNA methylation level was measured by LUMA, in which global CCGG methylation was measured by using

Table 1. Suppression of gastric cancers by 5-aza-dC

		Adenocarcinoma						
Group	Effective number	Well-differentiated	Poorly differentiated	Multiplicity (mean ± SD)	Diameter (mean ± SD), mm	Incidence (%)	Adenoma	Sarcoma
G1. *HP*(–) + DW	8	0	0	0	—	0/8 (0)	0	0
G2. *HP*(–) + MNU + PBS	29	3	3	1.0 ± 0	6.3 ± 1.6	6/29 (20.7)	1	2
G3. *HP*(–) + MNU + 5-aza-dC	33	0	2	1.0 ± 0	3.3 ± 3.3	2/33 (6.1)	1	0
G4. *HP*(+) + DW	8	0	0	0	—	0/8 (0)	0	0
G5. *HP*(+) + MNU + PBS	29	15	1	1.1 ± 0.3	5.7 ± 2.7	16/29 (55.2)[a]	0	2
G6. *HP*(+) + MNU + 5-aza-dC	30	4	3	1.0 ± 0	4.8 ± 1.7	7/30 (23.3)[b]	0	2

NOTE: 5-Aza-dC treatment decreased incidence of gastric cancers (adenocarcinomas) from 55.2% (G5) to 23.3% (G6).

[a]$P < 0.05$ compared with G3.

[b]$P < 0.05$ compared with G5.

a combination of methylation-sensitive and -insensitive restriction enzymes and pyrosequencing. Among the gerbils with MNU treatment and *H. pylori* infection (G5 and G6), 5-aza-dC treatment decreased the global methylation level from 83.0% ± 4.5% (mean ± SD)% (G5) to 80.3% ± 4.4% (G6, $P < 0.05$; Fig. 1B). Among the MNU-treated gerbils without *H. pylori* infection (G2 and G3), the methylation level was decreased by 5-aza-dC treatment (from 84.4% ± 2.3% in G2 to 82.2% ± 2.4% in G3, $P < 0.05$). No significant influence of *H. pylori* infection or MNU treatment was observed. These results indicated that 5-aza-dC worked as a DNA demethylating agent *in vivo* and decreased global methylation levels in GECs.

Next, methylation of 10 CGIs, where *H. pylori* infection was previously shown to induce aberrant methylation (Supplementary Fig. S1; refs. 14, 41), was analyzed by quantitative methylation-specific PCR (qMSP). 5-Aza-dC treatment reduced methylation levels in G6 to 46% to 80% of those in G5 for 6 CGIs (HE6, HG2, SB1, SB5, SF12, and SH6; $P < 0.05$; Fig. 1C). These results showed that 5-aza-dC treatment suppressed methylation induction by *H. pylori* infection and MNU treatment in GECs. The methylation levels in G5 were higher than those in G4 whereas those in G2 were not elevated compared with those in G1, indicating that MNU treatment had an augmenting effect on *H. pylori*–induced aberrant methylation.

Dysregulation of inflammation-related genes by 5-aza-dC

Among 10 inflammation-related genes whose expression was examined in the stomach, expression of 3 genes (*Il1b*, *Nos2*, and *Tnf*) has been shown to be associated with induction of methylation in GECs (14, 41). Therefore, we examined whether 5-aza-dC treatment affected expression of these 3 genes in the stomach after *H. pylori* infection using *H. pylori*–infected and -uninfected gerbils without MNU treatment (Fig. 2A). In the *H. pylori*–infected gerbils with 5-aza-dC treatment (G10), expression levels of *Il1b* and *Nos2* decreased to 42% and 58%, respectively ($P < 0.01$, respectively), of those in *H. pylori*–infected gerbils without 5-aza-dC treatment (G9; Fig. 2B). In contrast, *Tnf* was upregulated to 187% of G9 ($P < 0.01$). These results indicated that 5-aza-dC treatment caused up- and downregulation, namely dysregulation, of inflammation-related genes. As there was a possibility that upregulation of *Tnf* was due to demethylation of its promoter CGI, we sequenced its promoter region but found that there was no CGI (Supplementary Fig. S4). Because methylation of only promoters with CGIs can consistently silence their downstream genes (42), we considered that the upregulation of *Tnf* was unlikely to be due to demethylation by 5-aza-dC.

We also analyzed infiltration of inflammatory cells in the stomach. In gerbils with *H. pylori* infection (G9 and G10), 5-aza-dC treatment did not affect infiltration of mononuclear cells and polymorphonuclear cells (Fig. 2C). In gerbils without *H. pylori* infection (G7 and G8), there was little mononuclear cell and polymorphonuclear cell infiltration, and no effect of 5-aza-dC treatment was observed. These results showed that 5-aza-dC treatment did not affect infiltration of inflammatory cells.

Extra-gastric effects of 5-aza-dC treatment

To evaluate possible adverse effects of 5-aza-dC treatment, we first conducted macroscopic analysis. Although most organs did not show any abnormality, the testes were prominently atrophic in 5-aza-dC–treated gerbils [0.47 ± 0.05 (mean ± SD) g/2 testes in G3 and 0.50 ± 0.05 g/2 testes in G6] compared with gerbils in other groups (1.10 ± 0.11 g/2 testes in G2 and 1.09 ± 0.13 g/2 testes in G5).

We then conducted analysis of histopathologic abnormalities and global DNA methylation levels in the testes, small intestine, liver, and kidneys. In the testes of 5-aza-dC–treated gerbils, the numbers of spermatozoa and spermatids were markedly decreased regardless of *H. pylori*

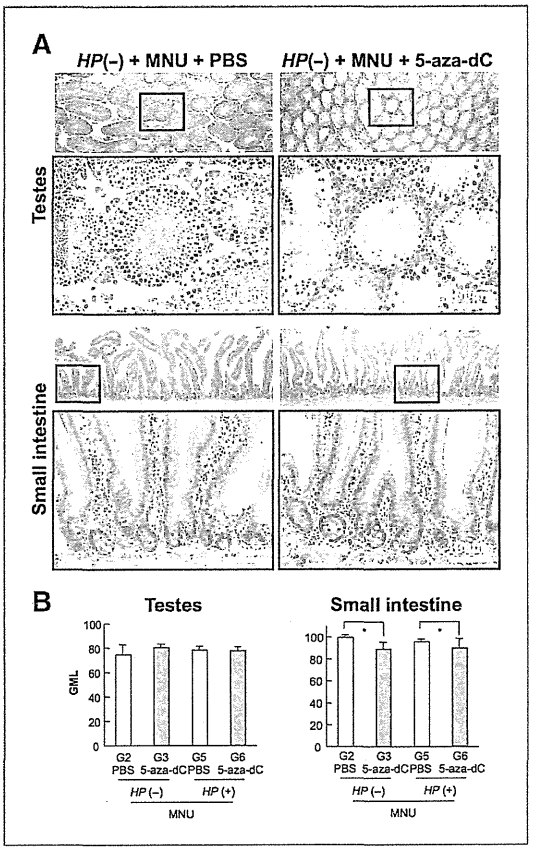

Figure 3. Adverse effects of 5-aza-dC treatment in extra-gastric tissues. A, tissue sections of the testes and small intestine. Bottom, is a magnified view of the region in the black rectangle in the top. Numbers of spermatozoa and spermatids were markedly decreased in the testes of 5-aza-dC–treated gerbil. B, global methylation levels in the testes and small intestine. Five gerbils in each group were randomly selected, and the methylation levels were measured by LUMA. Mean and SD are shown. *, $P < 0.05$. By 5-aza-dC-treatment, the global methylation level did not decrease in the testes but did in the small intestine.

infection status (Fig. 3A). Despite the presence of hypospermatogenesis, there was no significant decrease of the global methylation level (Fig. 3B). In contrast, in the small intestine, the global methylation level was reduced by 10.4% and 5.6% (G3 and G6, respectively Fig. 3B). However, no histologic changes compared with the untreated gerbils were observed (Fig. 3A). As for the liver and kidneys, there were no histologic abnormalities or reduction of global methylation levels in G3 and G6 (Supplementary Fig. S5).

Discussion

Our study using a gerbil model showed that 5-aza-dC treatment suppressed *H. pylori*/MNU-induced gastric cancers. This study showed for the first time that chemoprevention using a DNA demethylating agent is effective for chronic inflammation–associated cancers. As chronic inflammation contributes to about 25% of all cancer cases (43), and aberrant DNA methylation is frequently observed in tissues exposed to chronic inflammation (29), suppression of aberrant methylation might become an effective preventive approach for these types of cancers. This study also showed that induction of aberrant methylation is an important mechanism for gastric carcinogenesis by *H. pylori* infection.

As mechanisms of suppression of gastric cancers by 5-aza-dC, at least 2 modes of action were present. The first one was the DNA demethylating effect. 5-Aza-dC treatment decreased global methylation level in GECs and suppressed hypermethylation of CGIs. These results suggested that 5-aza-dC was capable of removing aberrant DNA methylation induced by *H. pylori* infection and thereby prevented cancer development. The second one was the effect on inflammation. It was previously shown that inflammation triggered by *H. pylori* infection is pivotal for aberrant methylation induction, and expression of inflammation-related genes (*Il1b*, *Nos2*, and *Tnf*) in the stomach is associated with the induction (14, 41). The present study showed that 5-aza-dC treatment dysregulated expression of these genes. Therefore, there is a possibility that altered balances among the related cytokines might have led to the reduced methylation induction.

Recently, in the stomach of hypergastrinemic INS-GAS mice, inhibition of *H. pylori* (*H. felis*)-induced global demethylation by folic acid supplementation was reported to suppress gastric dysplasia (44). The data are seemingly discordant with our data showing a cancer preventive effect by DNA demethylation. However, in the INS-GAS mice, global hypomethylation by *H. felis* was evident, suggesting that the demethylation plays important roles in the carcinogenesis. Hypermethylation by folic acid might exert the cancer preventive effect via suppression of global demethylation. In contrast, in our gerbil study, hypermethylation of CGIs by *H. pylori* infection, rather than global demethylation, was evident, suggesting that hypermethylation of CGIs rather than global demethylation was the major mechanism for the carcinogenesis. Thus, demethylation by 5-aza-dC was considered to have exerted the preventive effect via suppression of hypermethylation of CGIs.

No obvious adverse effect of 5-aza-dC treatment was observed besides hypospermatogenesis in the testes. Hypospermatogenesis due to 5-aza-dC treatment was reported in mice (24). As global hypomethylation was not detected in the testes of the 5-aza-dC–treated gerbils, the effect was speculated to be independent of its DNA demethylating activity. However, we cannot exclude the possibility that decrease of methylation was not detected due to immediate elimination of spermatozoa/spermatids with decreased DNA methylation. Regardless of the mechanism, the presence of this adverse effect precludes 5-aza-dC as a chemoprevention agent for the general population. However, chemoprevention by a DNA demethylating agent itself still might become a promising strategy if a DNA demethylating agent without such toxicity is developed.

Demethylating effects by 5-aza-dC were observed in the stomach and the small intestine, but not in the liver and the kidneys. Specific global demethylation in the stomach and intestine was also observed in the female $Apc^{min/+}$ mice administered a demethylating agent, zebularine (23). Turnover of epithelial cells in these tissues is known to be very rapid, being 3 to 4 days in mice (45, 46). As demethylating effects of 5-aza-dC and zebularine are exerted after their incorporation into gDNA and DNA replication, rapid cell turnover in the stomach and intestine could explain the organ-specific global demethylation.

Individuals with a severe epigenetic field defect, in contrast with the general population, can be considered as a target population for cancer prevention using 5-aza-dC after careful balancing of its preventive and adverse effects. In the case of gastric cancers, eradication of *H. pylori* is the primary strategy for prevention (47), but the incidence of gastric cancers remains high, even after *H. pylori* eradication, especially in persons with intestinal metaplasia and gastric atrophy (48). Notably, aberrant methylation in gastric mucosae decreases by *H. pylori* eradication, but it does not disappear completely (14, 49). The level of the remaining methylation reflects the risk of gastric cancers (8, 9). As DNA demethylating agents are likely to remove such accumulated methylation and suppress gastric cancers development, these individuals with a severe epigenetic field defect may benefit from epigenetic chemoprevention.

In summary, treatment with 5-aza-dC effectively prevented gastric cancers induced by *H. pylori* infection in gerbils, suppressed DNA methylation in GECs, and induced dysregulation of inflammation. Chemoprevention with a DNA demethylating agent is expected to become an effective strategy for prevention of chronic inflammation–associated cancers.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: T. Niwa, M. Tatematsu, T. Ushijima
Development of methodology: T. Niwa, T. Tsukamoto
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): T. Niwa, T. Toyoda
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): T. Niwa, T. Toyoda
Writing, review, and/or revision of the manuscript: T. Niwa, T. Ushijima
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A. Mori
Study supervision: M. Tatematsu

Acknowledgments

The authors thank Ms. R. Shimada (Qiagen K.K.) for her technical support of LUMA.

Grant Support

This study was supported by Grants-in-Aid for The Third-term Comprehensive Cancer Control Strategy from the Ministry of Health, Labour, and Welfare, Japan (T. Ushijima); Grants-in-Aid for Young Scientists (#23701114) from the Japan Society for the Promotion of Science (T. Niwa); and Grants-in-Aid from the Foundation of Promotion of Cancer Research, Japan (T. Niwa).

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

Received August 29, 2012; revised December 19, 2012; accepted January 8, 2013; published online April 4, 2013.

References

1. Taby R, Issa JP. Cancer epigenetics. CA Cancer J Clin 2010;60:376–92.
2. Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene 2008;27:404–8.
3. Esteller M. Epigenetics in cancer. N Engl J Med 2008;358:1148–59.
4. Jones PA, Baylin SB. The epigenomics of cancer. Cell 2007;128:683–92.
5. Hsieh CJ, Klump B, Holzmann K, Borchard F, Gregor M, Porschen R. Hypermethylation of the p16INK4a promoter in colectomy specimens of patients with long-standing and extensive ulcerative colitis. Cancer Res 1998;58:3942–5.
6. Issa JP, Ahuja N, Toyota M, Bronner MP, Brentnall TA. Accelerated age-related CpG island methylation in ulcerative colitis. Cancer Res 2001;61:3573–7.
7. Kondo Y, Kanai Y, Sakamoto M, Mizokami M, Ueda R, Hirohashi S. Genetic instability and aberrant DNA methylation in chronic hepatitis and cirrhosis–A comprehensive study of loss of heterozygosity and microsatellite instability at 39 loci and DNA hypermethylation on 8 CpG islands in microdissected specimens from patients with hepatocellular carcinoma. Hepatology 2000;32:970–9.
8. Maekita T, Nakazawa K, Mihara M, Nakajima T, Yanaoka K, Iguchi M, et al. High levels of aberrant DNA methylation in *Helicobacter pylori*-infected gastric mucosae and its possible association with gastric cancer risk. Clin Cancer Res 2006;12:989–95.
9. Nakajima T, Maekita T, Oda I, Gotoda T, Yamamoto S, Umemura S, et al. Higher methylation levels in gastric mucosae significantly correlate with higher risk of gastric cancers. Cancer Epidemiol Biomarkers Prev 2006;15:2317–21.
10. Perri F, Cotugno R, Piepoli A, Merla A, Quitadamo M, Gentile A, et al. Aberrant DNA methylation in non-neoplastic gastric mucosa of H. Pylori infected patients and effect of eradication. Am J Gastroenterol 2007;102:1361–71.
11. Egan BJ, Holmes K, O'Connor HJ, O'Morain CA. *Helicobacter pylori* gastritis, the unifying concept for gastric diseases. Helicobacter 2007;12 Suppl 2:39–44.
12. Fox JG, Wang TC. Inflammation, atrophy, and gastric cancer. J Clin Invest 2007;117:60–9.
13. Moss SF, Blaser MJ. Mechanisms of disease: inflammation and the origins of cancer. Nat Clin Pract Oncol 2005;2:90–7.
14. Niwa T, Tsukamoto T, Toyoda T, Mori A, Tanaka H, Maekita T, et al. Inflammatory processes triggered by *Helicobacter pylori* infection cause aberrant DNA methylation in gastric epithelial cells. Cancer Res 2010;70:1430–40.
15. Blackburn EH, Tlsty TD, Lippman SM. Unprecedented opportunities and promise for cancer prevention research. Cancer Prev Res 2010;3:394–402.
16. Issa JP. Cancer prevention: epigenetics steps up to the plate. Cancer Prev Res (Phila) 2008;1:219–22.
17. Prevention of cancer in the next millennium: Report of the Chemoprevention Working Group to the American Association for Cancer Research. Cancer Res 1999;59:4743–58.
18. Greenwald P. Cancer chemoprevention. BMJ 2002;324:714–8.
19. Hursting SD, Slaga TJ, Fischer SM, DiGiovanni J, Phang JM. Mechanism-based cancer prevention approaches: targets, examples, and the use of transgenic mice. J Natl Cancer Inst 1999;91:215–25.
20. Issa JP, Kantarjian HM. Targeting DNA methylation. Clin Cancer Res 2009;15:3938–46.
21. Yoo CB, Jones PA. Epigenetic therapy of cancer: past, present and future. Nat Rev Drug Discov 2006;5:37–50.
22. Laird PW, Jackson-Grusby L, Fazeli A, Dickinson SL, Jung WE, Li E, et al. Suppression of intestinal neoplasia by DNA hypomethylation. Cell 1995;81:197–205.
23. Yoo CB, Chuang JC, Byun HM, Egger G, Yang AS, Dubeau L, et al. Long-term epigenetic therapy with oral zebularine has minimal side effects and prevents intestinal tumors in mice. Cancer Prev Res 2008;1:233–40.
24. McCabe MT, Low JA, Daignault S, Imperiale MJ, Wojno KJ, Day ML. Inhibition of DNA methyltransferase activity prevents tumorigenesis in a mouse model of prostate cancer. Cancer Res 2006;66:385–92.
25. Lantry LE, Zhang Z, Crist KA, Wang Y, Kelloff GJ, Lubet RA, et al. 5-Aza-2′-deoxycytidine is chemopreventive in a 4-(methyl-nitrosamino)-1-(3-pyridyl)-1-butanone-induced primary mouse lung tumor model. Carcinogenesis 1999;20:343–6.
26. Tang XH, Albert M, Scognamiglio T, Gudas LJ. A DNA methyltransferase inhibitor and all-trans retinoic acid reduce oral cavity carcinogenesis induced by the carcinogen 4-nitroquinoline 1-oxide. Cancer Prev Res 2009;2:1100–10.
27. Baba S, Yamada Y, Hatano Y, Miyazaki Y, Mori H, Shibata T, et al. Global DNA hypomethylation suppresses squamous carcinogenesis in the tongue and esophagus. Cancer Sci 2009;100:1186–91.
28. Belinsky SA, Klinge DM, Stidley CA, Issa JP, Herman JG, March TH, et al. Inhibition of DNA methylation and histone deacetylation prevents murine lung cancer. Cancer Res 2003;63:7089–93.
29. Niwa T, Ushijima T. Induction of epigenetic alterations by chronic inflammation and its significance on carcinogenesis. Adv Genet 2010;71:41–56.
30. Ushijima T. Epigenetic field for cancerization. J Biochem Mol Biol 2007;40:142–50.
31. Tatematsu M, Tsukamoto T, Mizoshita T. Role of *Helicobacter pylori* in gastric carcinogenesis: the origin of gastric cancers and heterotopic proliferative glands in Mongolian gerbils. Helicobacter 2005;10:97–106.
32. Tsukamoto T, Mizoshita T, Tatematsu M. Animal models of stomach carcinogenesis. Toxicol Pathol 2007;35:636–48.
33. Peterson AJ, Menheniott TR, O'Connor L, Walduck AK, Fox JG, Kawakami K, et al. *Helicobacter pylori* infection promotes methylation and silencing of trefoil factor 2, leading to gastric tumor development in mice and humans. Gastroenterology 2010;139:2005–17.
34. Shimizu N, Ikehara Y, Inada K, Nakanishi H, Tsukamoto T, Nozaki K, et al. Eradication diminishes enhancing effects of *Helicobacter pylori* infection on glandular stomach carcinogenesis in Mongolian gerbils. Cancer Res 2000;60:1512–4.

35. Cheng H, Bjerknes M, Amar J. Methods for the determination of epithelial cell kinetic parameters of human colonic epithelium isolated from surgical and biopsy specimens. Gastroenterology 1984;86:78–85.
36. Tatematsu M, Yamamoto M, Shimizu N, Yoshikawa A, Fukami H, Kaminishi M, et al. Induction of glandular stomach cancers in *Helicobacter pylori* -sensitive Mongolian gerbils treated with N-methyl-N-nitrosourea and N-methyl-N′-nitro-N-nitrosoguanidine in drinking water. Jpn J Cancer Res 1998;89:97–104.
37. Toyoda T, Tsukamoto T, Mizoshita T, Nishibe S, Deyama T, Takenaka Y, et al. Inhibitory effect of nordihydroguaiaretic acid, a plant lignan, on *Helicobacter pylori* -associated gastric carcinogenesis in Mongolian gerbils. Cancer Sci 2007;98:1689–95.
38. Karimi M, Johansson S, Ekstrom TJ. Using LUMA: a Luminometric-based assay for global DNA-methylation. Epigenetics 2006;1:45–8.
39. Yamashita S, Takahashi S, McDonell N, Watanabe N, Niwa T, Hosoya K, et al. Methylation silencing of transforming growth factor-beta receptor type II in rat prostate cancers. Cancer Res 2008;68:2112–21.
40. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 2000;16:276–7.
41. Hur K, Niwa T, Toyoda T, Tsukamoto T, Tatematsu M, Yang HK, et al. Insufficient role of cell proliferation in aberrant DNA methylation induction and involvement of specific types of inflammation. Carcinogenesis 2011;32:35–41.
42. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 2007;39:457–66.
43. Hussain SP, Harris CC. Inflammation and cancer: an ancient link with novel potentials. Int J Cancer 2007;121:2373–80.
44. Gonda TA, Kim YI, Salas MC, Gamble MV, Shibata W, Muthupalani S, et al. Folic acid increases global DNA methylation and reduces inflammation to prevent Helicobacter-associated gastric cancer in mice. Gastroenterology 2012;142:824–33.
45. Gavrieli Y, Sherman Y, Ben-Sasson SA. Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. J Cell Biol 1992;119:493–501.
46. Lee ER. Dynamic histology of the antral epithelium in the mouse stomach: III. Ultrastructure and renewal of pit cells. Am J Anat 1985;172:225–40.
47. Asaka M, Kato M, Takahashi S, Fukuda Y, Sugiyama T, Ota H, et al. Guidelines for the management of *Helicobacter pylori* infection in Japan: 2009 revised edition. Helicobacter 2010;15:1–20.
48. Kabir S. Effect of *Helicobacter pylori* eradication on incidence of gastric cancer in human and animal models: underlying biochemical and molecular events. Helicobacter 2009;14:159–71.
49. Nakajima T, Enomoto S, Yamashita S, Ando T, Nakanishi Y, Nakazawa K, et al. Persistence of a component of DNA methylation in gastric mucosae after *Helicobacter pylori* eradication. J Gastroenterol 2010; 45:37–44.

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.

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 non-cancerous 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

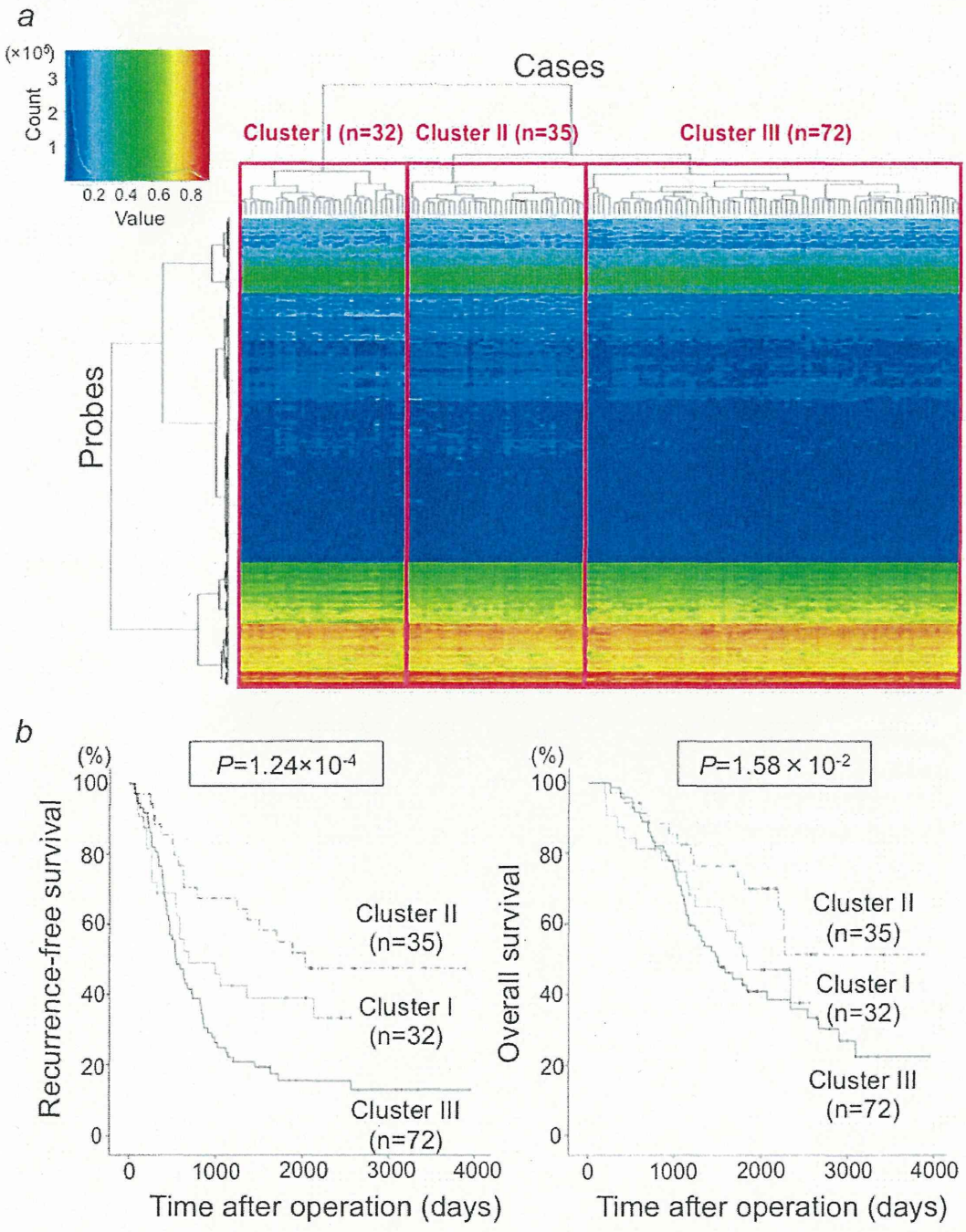

Figure 1. (*a*) Unsupervised hierarchical clustering (Euclidean distance, Ward method) using DNA methylation levels on all 26,447 probes in samples of non-cancerous lung tissue (N) from 139 patients with lung adenocarcinomas in the learning cohort. Based on DNA methylation status in adjacent lung tissue, 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) ≥0.70	24	34	65	9.86×10^{-3} [4]
	FEV_1:FVC <0.70				
	FEV_1 ≥80% of predicted value	4	1	6	
	FEV_1 <80% but ≥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 ($_{average}\ \beta_C > 0.3$) 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).

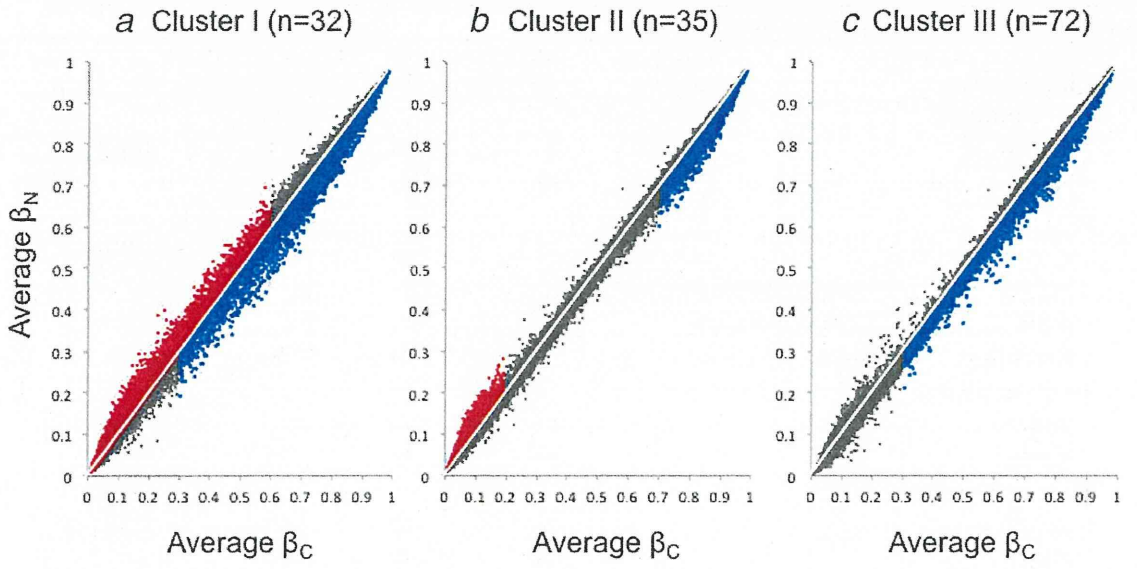

Figure 2. Distribution of average DNA methylation levels on all 26,447 probes of non-cancerous lung tissue (N) samples obtained from patients with lung adenocarcinomas belonging to Clusters I (*a*), II (*b*) and III (*c*) and 36 samples of normal lung tissue (C) obtained from patients without any primary 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)

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)

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

(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			
cg21063899	13	78,109,801	*SCEL*	0.033 ± 0.088	0.013 ± 0.054	−0.081 ± 0.086	3.06×10^{-12}	−0.103	1.47×10^{-9} (Hypo)
cg14074641	16	48,181,753	*ABCC12*	−0.002 ± 0.091	0.025 ± 0.054	−0.109 ± 0.105	1.40×10^{-12}	−0.121	2.44×10^{-1} (Hypo)

[1]Probe ID for the Infinium HumanMethylation27 Bead Array.
[2]Chromosome.
[3]National Center for Biotechnology Information (NCBI) Database (Genome Build 37).
[4]$\Delta\beta_{N\text{-}averageC}$.
[5]Average β_{N-C} in Cluster I *versus* average β_{N-C} in Clusters II and III. Such p values were calculated to reveal the hallmark genes of Cluster I that showed DNA methylation statuses significantly different in their N samples in comparison with N samples from other clusters (Clusters II and III).
[6]Average β_{N-C} in Cluster I minus average β_{N-C} in Clusters II and III. If Δβ (I-II and III) was more than 0.1, N samples in Cluster I were considered to show DNA hypermethylation relative to N samples in other clusters, and if Δβ (I-II and III) was less than −0.1, N samples in Cluster I were considered to show DNA hypomethylation relative to N samples in other clusters.
[7]Stepwise DNA hypermethylation (Hyper) and hypomethylation (Hypo) from normal lung tissue samples to N samples, and then to tumorous tissue samples in Cluster I.
[8]Probe ID for the Infinium HumanMethylation27 Bead Array.
[9]Chromosome.
[10]National Center for Biotechnology Information (NCBI) Database (Genome Build 37).
[11]$\Delta\beta_{N\text{-}averageC}$.
[12]Average β_{N-C} in Cluster II *versus* average β_{N-C} in Clusters I and III. Such *p* value was calculated to reveal the hallmark gene of Cluster II that showed DNA methylation status significantly different in their N samples in comparison with N samples from other clusters (Clusters I and III).
[13]Average β_{N-C} in Cluster II minus average β_{N-C} in Clusters I and III. If Δβ (II−I and III) was more than 0.1, N samples in Cluster II were considered to show DNA hypermethylation relative to N samples in other clusters.
[14]Stepwise DNA hypermethylation (Hyper) and hypomethylation (Hypo) from normal lung tissue samples to N samples, and then to tumorous tissue samples in Cluster II.
[15]Probe ID for the Infinium HumanMethylation27 Bead Array.
[16]Chromosome.
[17]National Center for Biotechnology Information (NCBI) Database (Genome Build 37).
[18]$\Delta\beta_{N\text{-}averageC}$.
[19]Average β_{N-C} in Cluster III *versus* average β_{N-C} in Clusters I and II. Such *p* values were calculated to reveal the hallmark gene of Cluster III that showed DNA methylation statuses significantly different in their N samples in comparison with N samples from other clusters (Clusters I and II).
[20]Average β_{N-C} in Cluster III minus average β_{N-C} in Clusters I and II. If Δβ (III-I and II) was more than 0.1, N samples in Cluster III were considered to show DNA hypermethylation relative to N samples in other clusters and if Δβ (III-I and II) was less than −0.1, N samples in Cluster III were considered to show DNA hypomethylation relative to N samples in other clusters.
[21]Stepwise DNA hypermethylation (Hyper) and hypomethylation (Hypo) from normal lung tissue samples to N samples, and then to tumorous tissue samples in Cluster III.

Table 4. Correlation between DNA methylation levels of hallmark genes for Clusters I, II and III and the clinicopathological parameters in the validation cohort

(A) Hallmark genes for Cluster I

		DNA methylation level in non-cancerous lung tissue (N) samples[2] (mean ± SD)											
		Pleural anthracosis			Emphysematic change			Atypical adenomatous hyperplasia			Tumor anthracosis		
Target ID [1]	Gene symbol	G1	G2-3	*p*-Value[3]	Negative	Positive	*p*-Value[3]	Absence	Presence	*p*-Value[3]	Negative	Positive	*p*-Value[3]
cg20249919	*PCSK6*	−0.126 ± 0.049	−0.049 ± 0.102	1.83×10^{-2}	−0.069 ± 0.067	−0.049 ± 0.131	4.99×10^{-1}	−0.056 ± 0.101	−0.096 ± 0.082	3.57×10^{-1}	−0.077 ± 0.085	−0.047 ± 0.105	3.18×10^{-1}
cg23349790	*IGSF21*	−0.044 ± 0.101	−0.002 ± 0.100	4.13×10^{-1}	−0.035 ± 0.072	0.028 ± 0.118	3.50×10^{-2}	−0.005 ± 0.101	−0.029 ± 0.086	5.79×10^{-1}	−0.054 ± 0.079	0.017 ± 0.098	1.68×10^{-2}
cg22285621	*SSH3*	−0.073 ± 0.043	0.002 ± 0.101	1.25×10^{-2}	−0.001 ± 0.077	−0.014 ± 0.121	6.52×10^{-1}	−0.004 ± 0.101	−0.035 ± 0.059	3.41×10^{-1}	−0.015 ± 0.077	0.001 ± 0.106	5.69×10^{-1}
cg15433631	*IRX2*	−0.041 ± 0.061	0.034 ± 0.074	4.73×10^{-2}	0.025 ± 0.068	0.029 ± 0.083	8.52×10^{-1}	0.026 ± 0.077	0.028 ± 0.056	9.46×10^{-1}	0.010 ± 0.074	0.037 ± 0.072	2.77×10^{-1}
cg21949305	ADORA2A, CYTSA	0.025 ± 0.091	0.026 ± 0.060	9.73×10^{-1}	0.015 ± 0.054	0.039 ± 0.069	1.91×10^{-1}	0.029 ± 0.063	−0.003 ± 0.036	1.28×10^{-1}	−0.004 ± 0.058	0.039 ± 0.061	3.48×10^{-2}
cg10942056	*DISP1*	0.014 ± 0.088	0.015 ± 0.068	9.71×10^{-1}	0.009 ± 0.062	0.023 ± 0.077	5.00×10^{-1}	0.019 ± 0.071	−0.022 ± 0.027	2.48×10^{-2}	−0.007 ± 0.056	0.026 ± 0.072	1.08×10^{-1}
cg15149645	NUPR1	−0.007 ± 0.124	0.013 ± 0.073	7.37×10^{-1}	0.006 ± 0.070	0.015 ± 0.085	7.06×10^{-1}	0.015 ± 0.079	−0.036 ± 0.022	2.81×10^{-3}	−0.015 ± 0.081	0.021 ± 0.077	1.77×10^{-1}
cg06954481	*GBX2*	−0.044 ± 0.031	0.013 ± 0.075	9.57×10^{-3}	0.008 ± 0.062	0.003 ± 0.085	7.95×10^{-1}	0.012 ± 0.073	−0.047 ± 0.040	2.27×10^{-2}	−0.012 ± 0.058	0.016 ± 0.078	1.90×10^{-1}
cg21250978	*PRKAR2B*	−0.013 ± 0.092	0.002 ± 0.058	7.44×10^{-1}	−0.010 ± 0.050	0.014 ± 0.070	1.95×10^{-1}	0.005 ± 0.061	−0.037 ± 0.033	4.81×10^{-2}	−0.032 ± 0.058	0.013 ± 0.059	2.63×10^{-2}
cg22418909	SFRP1	−0.043 ± 0.076	0.002 ± 0.058	2.55×10^{-1}	0.000 ± 0.065	−0.003 ± 0.053	8.32×10^{-1}	0.003 ± 0.061	−0.041 ± 0.022	4.86×10^{-3}	−0.020 ± 0.057	0.007 ± 0.060	1.64×10^{-1}
cg26200585	*PRX*	0.020 ± 0.079	0.015 ± 0.066	8.81×10^{-1}	0.013 ± 0.069	0.016 ± 0.063	8.89×10^{-1}	0.019 ± 0.067	−0.029 ± 0.035	3.48×10^{-2}	−0.006 ± 0.054	0.025 ± 0.070	1.19×10^{-1}
cg24396745	*HCN4*	−0.056 ± 0.035	0.020 ± 0.072	3.53×10^{-3}	0.017 ± 0.066	0.003 ± 0.079	4.87×10^{-1}	0.015 ± 0.073	−0.025 ± 0.054	1.86×10^{-1}	−0.005 ± 0.074	0.022 ± 0.071	2.60×10^{-1}
cg04330449	*NEUROG1*	−0.040 ± 0.033	0.010 ± 0.073	2.33×10^{-2}	0.010 ± 0.065	−0.005 ± 0.078	4.52×10^{-1}	0.006 ± 0.072	−0.020 ± 0.056	3.82×10^{-1}	−0.006 ± 0.044	0.012 ± 0.078	3.40×10^{-1}
cg19589427	*TNFSF18*	−0.008 ± 0.078	−0.010 ± 0.070	9.65×10^{-1}	−0.010 ± 0.078	−0.012 ± 0.057	9.26×10^{-1}	−0.007 ± 0.071	−0.042 ± 0.036	1.10×10^{-1}	−0.040 ± 0.057	0.005 ± 0.069	3.13×10^{-2}
cg16731240	*ZNF577*	−0.042 ± 0.037	0.014 ± 0.087	2.50×10^{-2}	0.012 ± 0.1000	0.003 ± 0.060	7.02×10^{-1}	0.007 ± 0.086	0.014 ± 0.078	8.48×10^{-1}	−0.015 ± 0.074	0.020 ± 0.087	1.79×10^{-1}
cg03544320	*CRMP1*	−0.097 ± 0.083	0.018 ± 0.094	3.14×10^{-2}	0.010 ± 0.109	−0.005 ± 0.084	5.67×10^{-1}	0.008 ± 0.102	−0.040 ± 0.036	4.56×10^{-2}	−0.014 ± 0.095	0.019 ± 0.097	3.04×10^{-1}
cg12864235	*CDH9*	0.032 ± 0.054	0.027 ± 0.057	8.51×10^{-1}	0.025 ± 0.045	0.030 ± 0.066	7.86×10^{-1}	0.031 ± 0.057	−0.004 ± 0.019	1.29×10^{-2}	0.012 ± 0.043	0.034 ± 0.060	1.77×10^{-1}
cg15898840	IGFBP3	−0.043 ± 0.030	0.001 ± 0.057	2.53×10^{-2}	−0.004 ± 0.052	−0.007 ± 0.061	8.16×10^{-1}	−0.002 ± 0.056	−0.036 ± 0.037	1.14×10^{-1}	−0.003 ± 0.044	−0.002 ± 0.060	9.36×10^{-1}
cg08044694	*BRD4*	−0.067 ± 0.036	−0.020 ± 0.049	3.84×10^{-2}	−0.021 ± 0.046	−0.028 ± 0.053	6.30×10^{-1}	−0.022 ± 0.050	−0.039 ± 0.035	3.55×10^{-1}	−0.041 ± 0.041	−0.017 ± 0.052	1.12×10^{-1}
cg03734874	*TMEM179*	−0.032 ± 0.034	0.023 ± 0.074	1.62×10^{-2}	0.021 ± 0.076	0.015 ± 0.068	7.56×10^{-1}	0.024 ± 0.072	−0.037 ± 0.039	1.89×10^{-2}	−0.008 ± 0.052	0.030 ± 0.075	5.63×10^{-2}
cg10599444	*MMP14*	−0.063 ± 0.026	−0.010 ± 0.060	4.85×10^{-3}	−0.008 ± 0.057	−0.023 ± 0.060	3.87×10^{-1}	−0.013 ± 0.060	−0.027 ± 0.041	5.25×10^{-1}	−0.026 ± 0.047	−0.007 ± 0.059	2.66×10^{-1}
cg24133115	*PDE10A*	−0.020 ± 0.025	0.022 ± 0.060	1.58×10^{-2}	0.014 ± 0.059	0.020 ± 0.057	7.02×10^{-1}	0.018 ± 0.060	0.004 ± 0.027	3.88×10^{-1}	0.009 ± 0.037	0.024 ± 0.063	3.25×10^{-1}
cg12594641	LYPD6	−0.024 ± 0.041	0.036 ± 0.068	2.45×10^{-2}	0.021 ± 0.062	0.036 ± 0.075	4.61×10^{-1}	0.029 ± 0.071	0.013 ± 0.038	4.43×10^{-1}	0.025 ± 0.045	0.035 ± 0.075	5.94×10^{-1}
cg05724065	PHKG1	0.016 ± 0.100	0.011 ± 0.057	9.10×10^{-1}	0.014 ± 0.055	0.008 ± 0.067	7.39×10^{-1}	0.016 ± 0.061	−0.032 ± 0.027	1.05×10^{-2}	−0.002 ± 0.066	0.019 ± 0.059	3.33×10^{-1}
cg19466563	SPARCL1	0.018 ± 0.081	0.012 ± 0.055	8.86×10^{-1}	0.007 ± 0.052	0.021 ± 0.060	3.98×10^{-1}	0.019 ± 0.056	−0.035 ± 0.015	4.46×10^{-5}	−0.015 ± 0.047	0.025 ± 0.057	2.24×10^{-2}
cg24433189	SSTR5	0.024 ± 0.075	0.034 ± 0.060	7.96×10^{-1}	0.031 ± 0.052	0.032 ± 0.071	9.63×10^{-1}	0.035 ± 0.062	−0.003 ± 0.027	2.95×10^{-2}	0.026 ± 0.051	0.036 ± 0.065	5.79×10^{-1}
cg24453664	CD59	−0.057 ± 0.039	0.000 ± 0.053	2.44×10^{-2}	−0.002 ± 0.054	−0.009 ± 0.055	6.40×10^{-1}	−0.004 ± 0.054	−0.021 ± 0.050	5.00×10^{-1}	−0.018 ± 0.050	0.001 ± 0.055	2.53×10^{-1}
cg26609631	GSX1	−0.051 ± 0.058	0.006 ± 0.067	9.21×10^{-2}	0.002 ± 0.066	−0.005 ± 0.069	7.29×10^{-1}	0.004 ± 0.068	−0.050 ± 0.025	3.65×10^{-3}	−0.020 ± 0.054	0.010 ± 0.071	1.35×10^{-1}
cg10604646	RGS5	0.038 ± 0.039	0.033 ± 0.058	8.08×10^{-1}	0.013 ± 0.062	0.056 ± 0.041	5.70×10^{-3}	0.037 ± 0.056	−0.014 ± 0.057	1.16×10^{-1}	−0.006 ± 0.063	0.049 ± 0.047	1.07×10^{-2}
cg03355526	ZNF454	−0.061 ± 0.044	0.001 ± 0.075	2.98×10^{-2}	0.003 ± 0.077	−0.012 ± 0.070	4.81×10^{-1}	0.002 ± 0.074	−0.052 ± 0.053	8.77×10^{-2}	−0.007 ± 0.058	0.000 ± 0.078	7.53×10^{-1}
cg27096144	MSX2	−0.066 ± 0.049	0.001 ± 0.063	3.29×10^{-2}	−0.006 ± 0.057	−0.006 ± 0.072	9.95×10^{-1}	−0.003 ± 0.065	−0.026 ± 0.040	2.94×10^{-1}	−0.023 ± 0.055	0.002 ± 0.068	2.05×10^{-1}
cg15520279	HOXD8	−0.021 ± 0.040	0.016 ± 0.070	1.15×10^{-1}	0.015 ± 0.075	0.009 ± 0.056	7.16×10^{-1}	0.014 ± 0.070	−0.006 ± 0.025	2.05×10^{-1}	−0.010 ± 0.038	0.024 ± 0.075	4.76×10^{-2}