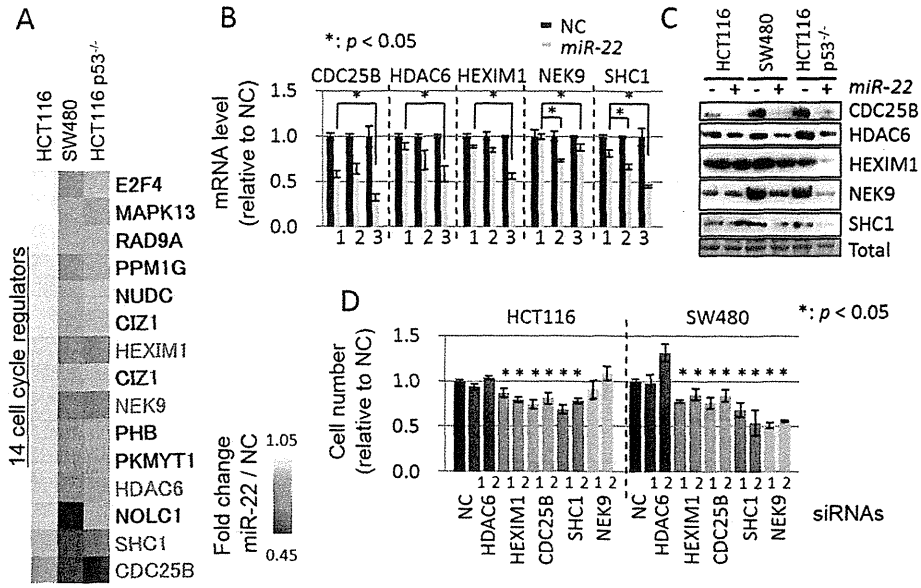

Figure 1 | Identification of NEK9 by miRNA target screen. (A) Fold changes in expression (*miR-22*/NC) of 14 cell-cycle regulators, shown as a heat map. Red letters indicate potential *miR-22* target genes identified by *in silico* analysis. (B) Relative expression levels of five mRNAs identified by this screen were quantitated using TaqMan qRT-PCR assays. *GAPDH* was used as an internal standard. (Lane 1: HCT116; Lane 2: SW480; Lane 3: HCT116 $p53^{-/-}$). (C) Protein levels of candidate genes in the presence of *miR-22* were determined by immunoblot analysis. A 70 kDa protein that stained with amido black was used as a loading control. The original full scan images are available in Supplementary Figure 7. (D) Cell proliferation assay. Cells were counted 72 h after transfection with specified siRNAs, indicated below the graph. Two siRNAs, designated 1 and 2, were used to knock down each gene. Error bars indicate standard deviation (SD, $n = 3$).

cells according to their p53 mutation status. In this study, using *miR-22* as a screening tool, we tried to identify factors that play important roles in proliferation and/or cell-cycle progression of p53-mutant cancer cells. The screen identified NIMA-related kinase 9 (NEK9) as such a factor. Both *in vitro* and *in vivo*, NEK9 depletion inhibited cell growth only in p53-mutant cancer cells, and expression of the NEK9 open reading frame (ORF) in NEK9-knockdown (KD) cells restored cell proliferation. Furthermore, NEK9 KD in cancer cells induced senescence-like changes ascribed to prolonged G1 arrest or slowdown of cell-cycle progression. In addition, human lung adenocarcinoma patients with expression of both NEK9 and mutant p53 proteins exhibited significantly poorer prognoses, suggesting that NEK9 expression confers growth advantages on p53 mutant tumors. Taken together, our findings strongly suggest that the NEK9 network, which is selectively activated in p53-deficient contexts, is a crucial cellular pathway responsible for the growth capacity of cancer cells lacking functional p53.

Results

Identification of NEK9 by miRNA target screen. To identify targets selectively repressed in p53 mutant (MUT) and knockout (KO) cells, we compared changes in mRNA levels resulting from *miR-22* expression between cancer cell lines with p53 KO, MUT, or wild-type (WT). As depicted in Supplementary Figure 1, we first chose the genes down-regulated in both p53 MUT and KO cells by *miR-22*, and compared their expression levels in p53 WT cells, and then selected genes exhibiting a differential response to *miR-22*. Furthermore, we narrowed-down the candidate genes classified as encoding cell-cycle regulators in the Gene Ontology database. The resultant 14 genes were down-regulated by *miR-22* expression to greater extents in p53 MUT and KO cell lines than in WT cells (Fig. 1A, Supplementary Table 1). Finally, we searched for *miR-22* targets among these 14 genes by *in silico* analysis using miRWalk database (http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/), and found that five of these genes are direct targets of *miR-22* (red genes in Fig. 1A).

Quantitative RT-PCR analysis confirmed that these five genes were significantly down-regulated by *miR-22* in both p53 KO and MUT cells, relative to WT cells (Fig. 1B). *MiR-22* expression significantly decreased the level of NEK9 and SHC1 proteins in both p53 KO and MUT cells, but did not affect the levels in WT cells (Fig. 1C). Furthermore, siRNA-mediated knockdown (KD) of the five predicted *miR-22* targets revealed that only NEK9 depletion significantly suppressed proliferation in MUT cells; however, NEK9 KD did not affect proliferation in WT cells (Fig. 1D). Using a reporter-gene assay, we verified that *NEK9* mRNA is a direct target of *miR-22* (Supplementary Figure 2). Based on these results, we concluded that NEK9 is an *miR-22* target gene whose repression specifically suppresses growth of p53-mutant cancer cells.

NEK9 expression is required for proliferation of p53-inactivated cancer cells. To confirm that NEK9 depletion specifically suppresses growth of p53-mutant cells, we knocked down this protein in a set of cancer cell lines with different p53 mutation status (Fig. 2A, Supplementary Figure 3A). NEK9 KD significantly suppressed the growth of three cancer cell lines with p53 mutations, but did not affect the growth of three cancer cell lines with WT p53. In four subsequent experiments, we verified that the effect of NEK9 KD depended on p53 status. First, we stably knocked down p53 in HCT116 cells, which normally express WT p53, via expression of a specific shRNA. NEK9 KD significantly decreased growth in p53 KD cells, but not in cells expressing a control shRNA (Fig. 2B, Supplementary Figure 3B). Second, we established H1299 cell lines (a p53-null lung cancer cell line) that exogenously expressed either WT p53 or one of four p53 mutants (Supplementary Figure 3C). Three of the mutants, R175H, R273H, and R249S, are frequently observed in a variety of human cancers (IARC p53 database, http://p53.iarc.fr/); the fourth, R280K, is less common. In H1299 cells expressing mutant p53, but not in those expressing WT p53, siRNA-mediated NEK9 KD significantly suppressed growth (Fig. 2C). NEK9 KD also suppressed colony formation (Fig. 2D)

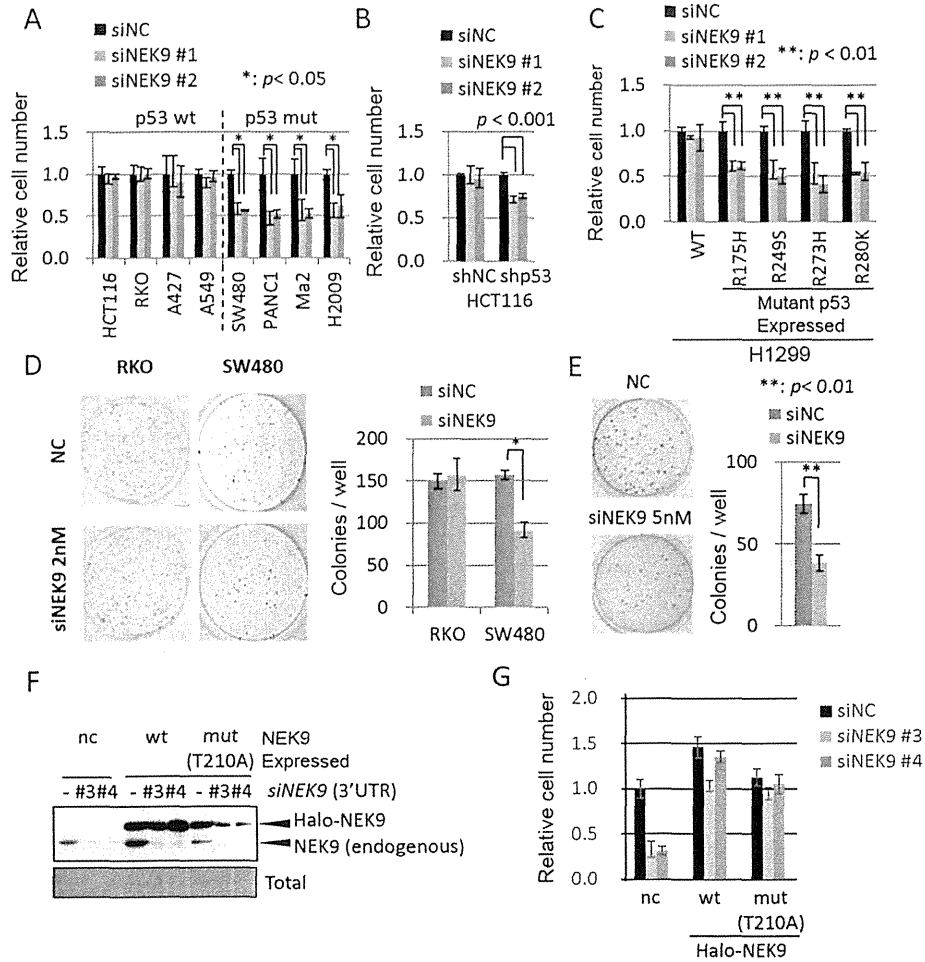

Figure 2 | Depletion of NEK9 inhibits proliferation of p53-inactivated cancer cells. (A) Cells were counted 3 days (or 4 days, for Ma2 and H2009 cells) after transfection with either 2 nM NEK9 siRNA (#1 or #2) or non-targeting control (siNC). Error bars indicate SD ($n = 3$). (B) HCT116 cells in which p53 was stably knocked down were transfected with 5 nM siRNAs, and the cell number was determined. (C) H1299 transfectants expressing WT p53 or one of four p53 mutants were depleted of NEK9, and then subjected to proliferation assays. Error bars indicate SD ($n = 3$). (D) Colony-formation assay on culture dishes was conducted using RKO (p53 WT) and SW480 (p53 MUT) cells, transfected with 2 nM NEK9 siRNAs or NC siRNA, and then allowed to grow for 9 days. Representative images of the colonies, stained with crystal violet, are shown on the left; colony counts are shown in the graph on the right. Three independent experiments yielded similar results; error bars indicate SD ($n = 3$). (E) Images of colonies of NEK9 KD cancer cells are shown at left, and the graph indicates the number of colonies. Error bars indicate SD ($n = 3$). (F) Expression of NEK9 protein after knockdown of endogenous NEK9. SW480 cells were co-transfected with NEK9 siRNAs (#3 and #4) targeting the 3′ UTR region of its mRNA and a siRNA-resistant NEK9 ORF expression vector. KD or expression of Halo-tagged NEK9 was determined by IB. WT and MUT (T210A) indicate wild-type and T210 mutant (kinase-inactivated) of NEK9, respectively. The original full scan images are available in Supplementary Figure 7. (G) Cell proliferation assay for NEK9 ORF–expressing cells. Cell counts are expressed relative to the count in the negative control (i.e., co-transfected with NC siRNA and vector control), which was normalized to 1.

and anchorage-independent growth of MUT cells (Fig. 2E). Third, we asked whether NEK9 protein expression would restore cell proliferation in p53-mutant cells. Endogenous NEK9 was knocked down by siRNA targeting the 3′UTR of *NEK9* mRNA, followed by expression of a cDNA encoding an siRNA-resistant *NEK9* ORF (Fig. 2F), and then the cells were counted. As shown in Fig. 2G, NEK9 re-expression restored cell growth after KD of endogenous NEK9. Expression of an amino-acid–substituted form of NEK9, T210A, in which an autophosphorylation site required for kinase activity of NEK9 is altered[16], also complemented the inhibition of cell proliferation after NEK9 KD, suggesting the possibility that kinase activity is not required for the growth-promoting function (Fig. 2F and G). Fourth, we evaluated the effect of a NEK9 siRNA/drug-delivery system (DDS) carrier complex on cell proliferation in p53-mutant xenograft tumors consisting of SW480 cells, derived from a p53-mutant human colon cancer, in mice. As indicated in Fig. 3A, the siRNA/DDS carrier complex was administered into transplanted tumors five times at 2-day intervals, and tumor size was measured at every 2 days. In these animals, NEK9 KD markedly inhibited *in vivo* growth of p53 mutant cancer cells, but not of p53 WT cells (Fig. 3B, C, and D), providing further evidence that NEK9 contributes to cell proliferation in p53-mutant cancer cells. Taken together, these results clearly indicate that p53 deficiency makes cells dependent on NEK9 for proliferation, both *in vitro* and *in vivo*.

Inhibition of G1-S progression and induction of senescence-like features by NEK9 KD. Because NEK9 regulates proper positioning of bipolar spindles[17–20], we hypothesized that NEK9 KD prevents G2/M progression. In flow-cytometry analysis, however, NEK9 KD had

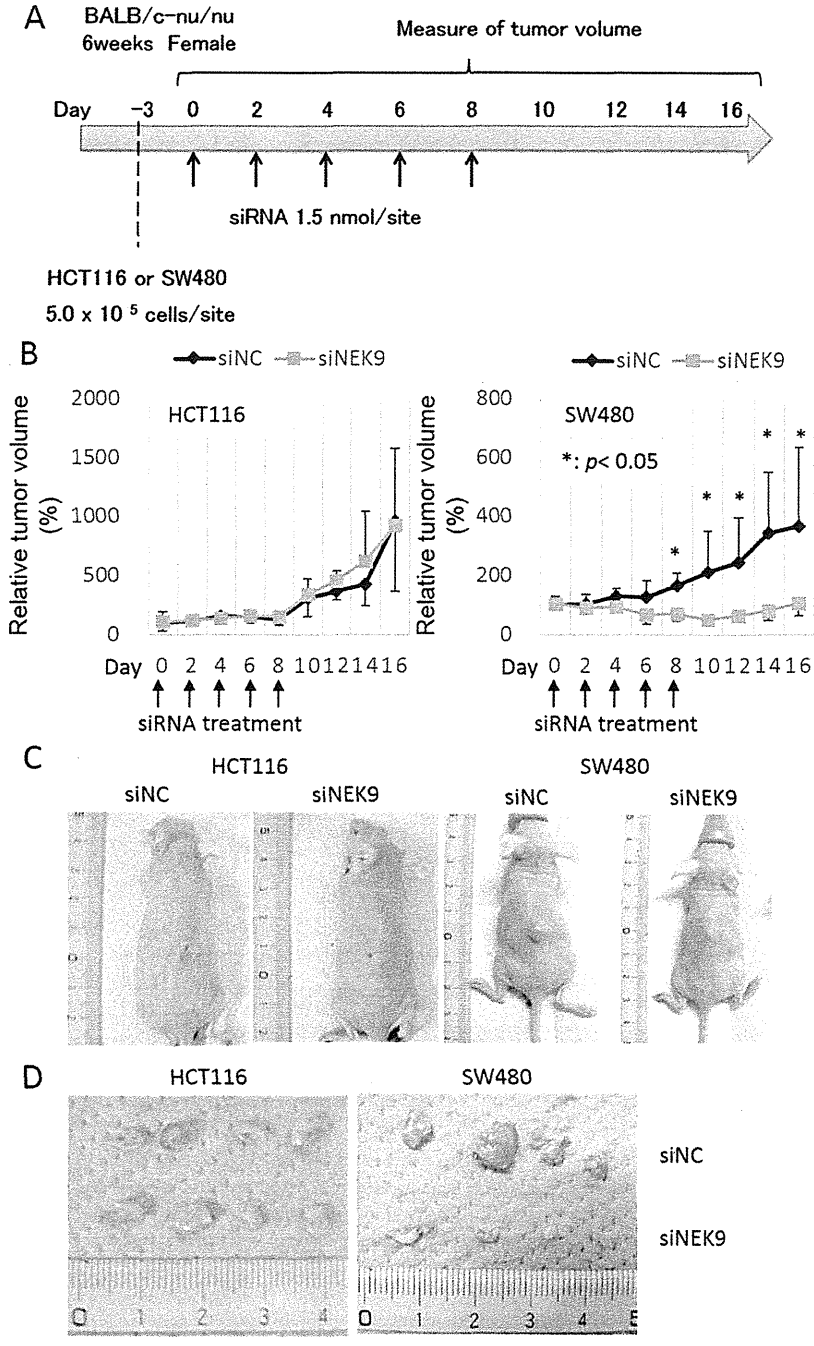

Figure 3 | Growth inhibition of p53-mutant xenograft by NEK9 siRNA. (A) Schedule of administration of NEK9 siRNA into SW480 or HCT 116 xenografts. Arrows show the days on which injections of siRNA/*in vivo*–jetPEI complex were administered. (B) SW480 and HCT 116 cells were inoculated into the backs of nude mice (n = 4), and siRNAs (NEK9 or negative control [NC]) in complex with drug-delivery carrier were injected on the specified days (arrows at the bottom of the graph). Tumor volume is expressed as a percentage (%) relative to the volume at the time of the first siRNA injection (day 0). Error bars indicate SD (n = 4). A left graph indicates tumor growth of p53 WT HCT 116 cells. A right graph is p53 mutant SW480. (C) Representative images of mice with xenograft tumors are shown on the left (HCT 116) and right (SW480). (D) Comparison of xenograft tumors injected with NC (upper panel) and NEK9 siRNA (lower panel).

no significant effect on the fraction of cells in G2/M phase; rather, in the two p53-mutant cell lines SW480 and PANC1, NEK9 KD significantly increased the G1-phase population while significantly decreasing the S-phase population (Fig. 4A). We obtained similar results in H1299 transfectants expressing p53 mutants, as well as in p53 KD HCT116 cells (Supplementary Figure 4). These data indicated that NEK9 is necessary for the G1–S transition in p53-deficient cancer cells.

The growth arrest resulting from NEK9 KD was associated with flattened and enlarged cellular morphologies typical of cellular senescence. Consistent with this, the cells stained positively for senescence-associated β-galactosidase (SA-β-gal) (Fig. 4B). Flow-

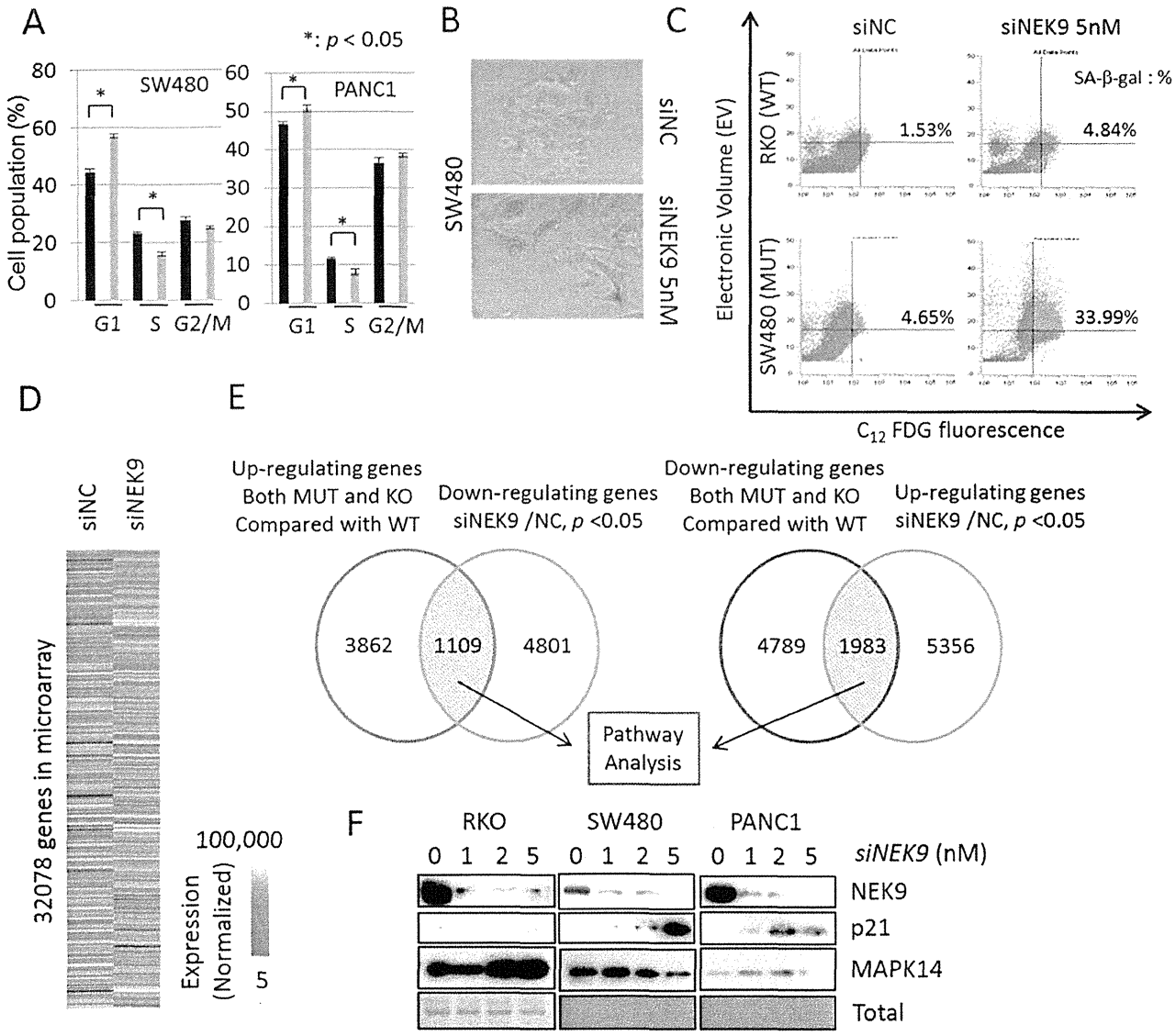

Figure 4 | NEK9 KD induces G1 arrest with senescence-like features and broad changes in gene expression profile. (A) Seventy-two hours after transfection with 2 nM siRNA, cells were subjected to flow-cytometry analysis. Error bars indicate SD of three independent experiments. (B) SW480 cells were transfected with 5 nM siRNAs and incubated for 7 days. (C) SA-β-gal–positive cells were quantified by flow cytometry using C_{12}FDG as a fluorogenic substrate. The upper panels show data from RKO (p53 WT), and the lower panels show data from SW480 (p53 MUT). Electronic volume (EV) indicates cell size. (D) Heat map of changes in mRNA levels 72 h after transfection of SW480 cells with 5 nM siRNAs. (E) Venn diagrams indicate the number of genes up- and down-regulated in p53 MUT/KO and NEK9 KD cells. The blue circle represents genes up-regulated in p53 mutant (MUT) and knockout (KO) cell lines relative to WT cells, and the orange circle represents genes down-regulated genes in NEK9 KD cells. The purple circle represents genes down-regulated in p53 MUT and KO cells relative to WT cells, and the silver circle represents genes up-regulated in NEK9 KD cells. (F) Cells were transfected with the specified concentrations of siRNAs, incubated for 72 h, and then subjected to immunoblot analysis with the indicated antibodies. The original full scan images are available in Supplementary Figure 7.

cytometry analysis[21] confirmed that SA-β-gal staining increased upon NEK9 depletion to a much greater extent in p53-mutant cells than in p53 WT cells (Fig. 4C). These results strongly suggest that depletion of NEK9 causes senescence-like phenotypic changes in p53-deficient cancer cells, because such cells require NEK9 function to progress from G1 to S phase.

Induction of broad changes in the gene expression profile upon NEK9 KD. To further elucidate the molecular connection between NEK9 depletion and cell-cycle arrest/senescence, we analyzed changes in mRNA expression profiles following NEK9 KD in p53-MUT cells. NEK9 KD caused broad changes in gene expression in p53-deficient cells (Fig. 4D): about 25% of genes up-regulated by NEK9 KD were expressed at significantly lower levels in p53 MUT and KO than in WT, suggesting that these genes were down-regulated due to loss of functional p53 (Fig. 4E). On the other hand, about 20% of genes down-regulated by NEK9 KD were expressed at significantly higher levels in MUT and KO than in WT, suggesting that their expression was increased by loss of functional p53 (Fig. 4E). Thus, NEK9 appears to be involved in regulation of a wide range of signaling pathways required for proliferation in p53-deficient cells.

Next, we performed a gene network analysis, using sets of genes affected in opposite directions by NEK9 KD and p53 deficiency

(Fig. 4E). This analysis revealed that NEK9 repression modulates multiple pathways, including receptor signaling cascades, mRNA processing, and cell-cycle control (Supplementary Tables 2 and 3). Upon NEK9 KD in p53-deficient cells, but not p53 WT cells, two crucial regulators of the cell cycle were differentially regulated: p21 was up-regulated, and MAPK14 (also known as p38α) was down-regulated (Fig. 4F). These two factors control the cell cycle regardless of *TP53* status[22–26]. However, in cancer cells with p53 mutations, NEK9 KD modulated the expression of both factors, leading to cell-cycle arrest with characteristics of senescence. These findings reveal a novel role of NEK9: regulation of the cell cycle, via control of multiple genes, in cancer cells carrying p53 mutations.

NEK9 expression in human lung cancer specimens expressing mutant p53. To elucidate whether NEK9 dependency occurs in human tumors, we next analyzed the expression of NEK9 and p53 proteins in cancerous and non-cancerous cells *in vivo* by an immunohistochemical staining of tissue microarray. In this method, positive staining of p53 indicates the expression of mutant p53 protein[27]. NEK9 signal was absent or faint in non-cancerous lung epithelial cells (Supplementary Figure 5A and B). By contrast, NEK9 was frequently expressed in several types of lung cancer: 33/57 (57.9%) of adenocarcinomas (ADC), 54/56 (96.4%) of squamous cell carcinomas (SQC), 23/23 (100%) of large cell neuroendocrine carcinomas (LCNEC), and 22/22 (100%) of small cell carcinomas (SCLC) (Fig. 5A). NEK9 protein was detectable in cancer cells that stained either positively or negatively with an anti-p53 antibody, i.e., again cells harboring either mutant or WT *TP53*, respectively. In the ADC and SQC samples, NEK9 staining did not correlate with either histological differentiation or p53 status. Notably, however, half of ADCs (13/26), the vast majority of SQCs (28/30), and all LCNECs (20/20) and SCLCs (14/14) that were positive for p53 staining also stained positively for NEK9 (Fig. 5B), indicating that NEK9 dependency occurs in lung SQCs, LCNECs, and SCLCs with p53 mutations. In addition, Kaplan–Meier analysis demonstrated that NEK9/p53 double-positive cases had significantly poorer prognoses than other cases (Fig. 5C and Supplementary Figure 5C), suggesting that NEK9 confers a growth advantage on ADCs harboring p53 mutations, resulting in aggressive tumors with poor clinical outcomes. To extend and validate these results, we used microarrays to measure *NEK9* mRNA levels in a different set of 76 lung ADC samples, in which *TP53* mutations were defined by exome sequencing (Fig. 5D, Supplementary Table 4). In addition, we used The Cancer Genome Atlas (TCGA)[28] data to analyze NEK9 expression levels in various cancers. As indicated in Fig. 5D and Supplementary Figure 5D, *NEK9* mRNA was expressed at similar levels in ADCs and SQCs of the lung, colon, stomach, and serous ovarian cancers, regardless of TP53 status. Judging from these results, NEK9 dependency makes a considerable contribution to growth and/or progression of tumors harboring *TP53* mutations.

Discussion

Through *miR-22* target screening, we identified NEK9 as a crucial regulator of cell-cycle progression in cancer cells lacking functional p53. The miRNA target screen is based on unique features of miRNAs, which target multiple mRNAs and induce phenotypes in a cell type–dependent manner. In a previous study, we showed that *miR-22* induces apoptosis in p53 WT cancer cells, but causes cell-cycle arrest in p53 MUT and KO cells that express CDKs and HDACs, critical regulators of the cell cycle that are potential targets of *miR-22*[15]. Therefore, we hypothesized that specific factor(s) repressed by *miR-22* are required for cell-cycle progression in cancer cells with p53 deficiencies. Consistent with this idea, our screen successfully identified NEK9 as a crucial factor involved in cell-cycle progression in p53-deficient cancer cells.

NEK9 regulates mitotic entry by phosphorylating two other NEK family proteins, NEK6 and 7[17–20,29,30], and it is also activated sequentially by PLK1 and CDK1 to regulate formation of the mitotic spindle[31–33]. PLK1 was previously identified as a factor required for survival of isogenic p53 KO cell lines[34]. The results described here demonstrate that NEK9 inhibition does not affect the proportion of cells in G2/M phase, indicating that NEK9 is involved in molecular mechanism(s) distinct from those dependent on PLK1, at least in p53-deficient cancer cells. In fact, proliferation in p53 mutant cells was not specifically inhibited following knockdown of NEK6 or 7 (Supplementary Figure 6A and B).

NEK9 interacts with the FACT (facilitates chromosome transcription) complex in HeLa cells[35], in which functional p53 is disrupted by a viral protein. Knockdown of FACT components inhibits progression from G1 to S phase[35]. These findings are consistent with our own observation that NEK9 KD caused cell-cycle arrest at G1 phase and modulates the global gene expression profile in p53-deficient cells. Furthermore, phosphorylation of NEK9 at T210, which is required for activation of kinase activity[16], is important for the interaction between NEK9 and FACT[35]. Our findings, however, demonstrated that expression of NEK9 T210A restored proliferation in p53 mutant cells, suggesting that the role of NEK9 in regulating G1–S progression is kinase-independent (Fig. 2G). Consistent with this, a previous report showed that phosphorylation at T210 and NEK9 kinase activity are only detectable during G2/M phase, and are required for G2 phase progression[16]. However, the possibility that an uncharacterized phosphorylation site(s) and/or an interaction of protein partner(s) is required for the kinase activity of NEK9 to regulate the progression of the cell cycle from G1 to S phase in the p53-inactivated cells can not be excluded. Further studies are needed to clarify this important issue.

A recent study demonstrated that p53 mutations exert tumor-promoting functions that affect cell-cycle progression and metastatic activity in cancer cells[36]. Therefore, we hypothesized that NEK9 promotes these oncogenic functions of mutant p53. However, we did not detect an interaction between NEK9 and mutant p53 proteins (data not shown). Furthermore, NEK9 KD inhibited the proliferation of p53 KD cells (Fig. 2B). These observations suggest that NEK9 is associated with proliferation and cell-cycle progression in a manner that does not involve a direct interaction with p53. Currently, the relationships between NEK9, mutant p53, and the downstream targets of NEK9 remain unclear due to the unique functions of NEK9, including its kinase independent activity. Therefore, a more detailed analysis of NEK9 function will be required to identify the network(s) controlling the proliferation of p53-inactivated human cancer cells.

Mutant p53 proteins have attracted a great deal of attention as potential targets for cancer therapy. Based on this idea, several approaches have been attempted, including allele-specific activation of WT p53 and restoration of WT function to mutant p53 using small molecules[37,38]. Inhibition of NEK9 scaffolding (or NEK9 complex formation) and kinase functions may provide a means for selectively targeting cancer cells that lack functional p53; thus, our findings might lead to the development of novel cancer therapies.

Methods

Cell lines. Cell lines used for this study were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA). HCT116 $p53^{-/-}$ cells were kindly provided by Dr. Bert Vogelstein (The Johns Hopkins University, Baltimore, MD, USA). Colon and pancreatic cancer cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% heat inactivated fetal bovine serum (FBS), and lung cancer cell lines were maintained in RPMI-1640 (Gibco) supplemented with 10% FBS.

Microarray analyses. Microarray analysis was carried out as described previously[15]. In brief, SW480 or HCT116 $p53^{-/-}$ cells, seeded at 5.0×10^4 cells/ml, were transfected for 48 h with either *miR-22* or miR-negative control (Ambion) at a final concentration of 5 nM using HiPerFect Reagent (Qiagen). Total RNA was prepared using RNeasy columns (Qiagen) and labeled with Cy3 using the Low Input Quick Amp labeling kit (Agilent). Microarray analysis was performed using Whole Human

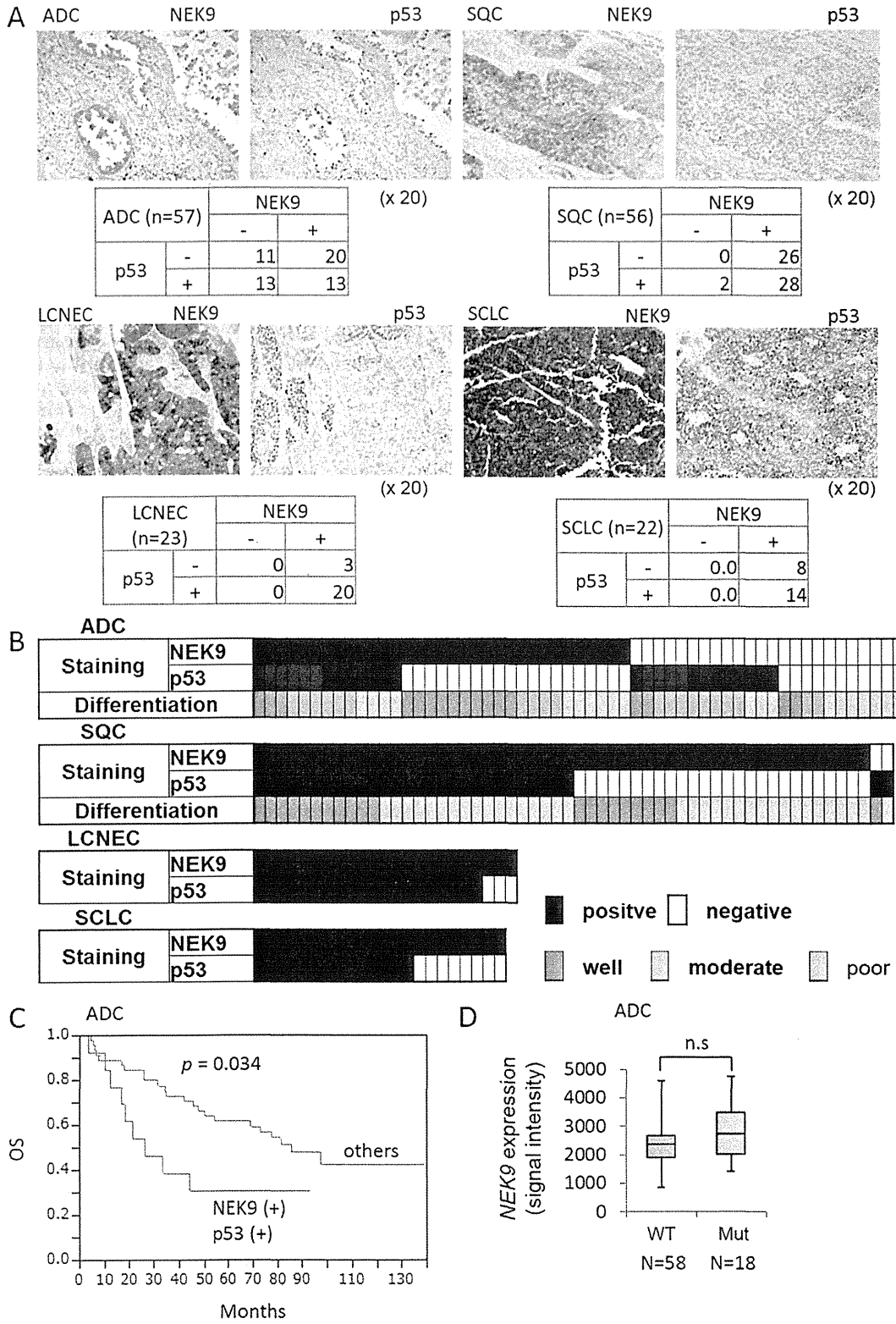

Figure 5 | Expression of NEK9 protein in lung cancer. (A) Representative images of NEK9 and p53 double-positive lung adenocarcinoma (ADC, upper left pair of images), squamous cell carcinoma (SQC, upper right pair), large cell neuroendocrine carcinoma (LCNEC, bottom left pair), and small cell lung carcinoma (SCLC, bottom right pair). The tables below each image pair show the number of samples of each type of cancer that contained positively stained cells. (B) NEK9- and p53-positive cases of ADC and SQC, with histological differentiation type indicated for each case. (C) Kaplan–Meier survival curves comparing NEK9 and p53 double-positive ADCs versus other types of ADCs. Vertical axis indicates overall survival (OS); horizontal axis indicates duration (months). Statistical analysis was carried out by Wilcoxon test. (D) Levels of *NEK9* mRNA in human lung ADCs were determined by microarray analysis. *TP53* mutation status was determined by whole-exome sequencing. Boxes represent the upper and lower quartiles and the median, and whiskers show the maximum and minimum values.

Genome Oligo Microarrays (4 × 44K v2, Agilent), and the data were analyzed using the GeneSpring GX11.5 software (Agilent).

For the generation of mRNA expression profiles in NEK9 KD cells, SW480 cells were transfected with NEK9 siRNAs, total RNAs were prepared, and microarray analysis was carried out as described above. Network analysis was performed using WikiPathways (http://wikipathways.org/).

Human tumor xenograft model. SW480 cells (derived from a p53-mutant human colon cancer) were inoculated at 5.0×10^5 cells per site on the backs of 6-week-old female nude mice (CLEA Japan, Inc.). in vivo–jetPEI (Polyplus Transfection)/siRNA complexes were prepared in a volume of 100 μl (1.5 nmol/site) according to the manufacturer's instructions, and then injected intratumorally into each animal five times at 2-day intervals. Tumor size was monitored every 2 days by measuring the length (L) and width (W) with calipers, and volume (V) was calculated according to the formula $V = (L \times W^2) \times 0.5$. Tumor volume is expressed as percentage of the volume on the day of the first siRNA injection. The animal study protocol was approved by the Committee for Ethics in Animal Experimentation of the National Cancer Center (NCC) Research Institute, and the experiments were carried out according to the NCC guidelines.

Analysis of human lung cancer samples. Surgical specimens of primary human lung cancer were obtained from patients treated at the National Cancer Center (NCC) Hospital, with documented informed consent obtained in each case. Whole-exome sequence analysis was carried out using the Agilent SureSelect Human All-Exon kit, and sequencing was performed on the Illumina HiSeq 2000 platform. *TP53* mutations were evaluated using sequencing data. Gene expression profiles were generated using Agilent Whole Human Genome DNA Microarrays. NEK9 expression levels in lung cancer were then determined.

Immunohistochemical analysis was conducted on tissue microarray sections derived from anonymized samples from patients with ADC, SQC, LCNEC, or SCLC of the lung who were treated in the NCC Hospital. Antibodies against NEK9 (rabbit monoclonal antibody, EP7361, from Epitomics) and p53 (mouse monoclonal antibody, DO-7, from Dako) were used for these analyses. Institutional review board approval for the use of clinical samples was obtained from the NCC.

Supplementary information. Supplementary information incudes extended Methods, eight figures and 5 tables.

1. Vousden, K. H. & Prives, C. Blinded by the Light: The Growing Complexity of p53. *Cell* **137**, 413–431 (2009).
2. Harris, S. L. & Levine, A. J. The p53 pathway: positive and negative feedback loops. *Oncogene* **24**, 2899–2908 (2005).
3. Watson, I. R., Takahashi, K., Futreal, P. A. & Chin, L. Emerging patterns of somatic mutations in cancer. *Nat Rev Genet* **14**, 703–718 (2013).
4. Muller, P. A. & Vousden, K. H. p53 mutations in cancer. *Nat Cell Biol* **15**, 2–8 (2013).
5. Vousden, K. H. & Prives, C. P53 and prognosis: new insights and further complexity. *Cell* **120**, 7–10 (2005).
6. Freed-Pastor, W. A. & Prives, C. Mutant p53: one name, many proteins. *Genes Dev* **26**, 1268–1286 (2012).
7. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
8. Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. *Nat Rev Cancer* **6**, 857–866 (2006).
9. van Kouwenhove, M. Kedde, M. & Agami, R. MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nat Rev Cancer* **11**, 644–656 (2011).
10. Hermeking, H. MicroRNAs in the p53 network: micromanagement of tumour suppression. *Nat Rev Cancer* **12**, 613–626 (2012).
11. Lujambio, A. & Lowe, S. W. The microcosmos of cancer. *Nature* **482**, 347–355 (2012).
12. He, L. *et al.* A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130–1134 (2007).
13. Raver-Shapira, N. *et al.* Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Mol Cell* **26**, 731–743 (2007).
14. Tazawa, H., Tsuchiya, N., Izumiya, M. & Nakagama, H. Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proc Natl Acad Sci U S A* **104**, 15472–15477 (2007).
15. Tsuchiya, N. *et al.* Tumor suppressor miR-22 determines p53-dependent cellular fate through post-transcriptional regulation of p21. *Cancer Res* **71**, 4628–4639 (2011).
16. Tan, B. C. & Lee, S. C. Nek9, a novel FACT-associated protein, modulates interphase progression. *J Biol Chem* **279**, 9321–9330 (2004).
17. Belham, C. *et al.* A mitotic cascade of NIMA family kinases. Nercc1/Nek9 activates the Nek6 and Nek7 kinases. *J Biol Chem* **278**, 34897–34909 (2003).
18. Richards, M. W. *et al.* An autoinhibitory tyrosine motif in the cell-cycle-regulated Nek7 kinase is released through binding of Nek9. *Mol Cell* **36**, 560–570 (2009).
19. Sdelci, S., Bertran, M. T. & Roig, J. Nek9, Nek6, Nek7 and the separation of centrosomes. *Cell cycle* **10**, 3816–3817 (2011).
20. Yang, S. W. *et al.* Nek9 regulates spindle organization and cell cycle progression during mouse oocyte meiosis and its location in early embryo mitosis. *Cell cycle* **11**, 4366–4377 (2012).
21. Debacq-Chainiaux, F., Erusalimsky, J. D., Campisi, J. & Toussaint, O. Protocols to detect senescence-associated beta-galactosidase (SA-betagal) activity, a biomarker of senescent cells in culture and in vivo. *Nature Protoc* **4**, 1798–1806 (2009).
22. McDonald, E. R. 3rd., Wu, G. S., Waldman, T. & El-Deiry, W. S. Repair Defect in p21 WAF1/CIP1 -/- human cancer cells. *Cancer Res* **56**, 2250–2255 (1996).
23. Abbas, T. & Dutta, A. p21 in cancer: intricate networks and multiple activities. *Nat Rev Cancer* **9**, 400–414 (2009).
24. Comes, F. *et al.* A novel cell type-specific role of p38alpha in the control of autophagy and cell death in colorectal cancer cells. *Cell Death and differentiation* **14**, 693–702 (2007).
25. Wagner, E. F. & Nebreda, A. R. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat Rev Cancer* **9**, 537–549 (2009).
26. Paillas, S. *et al.* MAPK14/p38alpha confers irinotecan resistance to TP53-defective cells by inducing survival autophagy. *Autophagy* **8**, 1098–1112 (2012).
27. Iggo, R., Gatter, K., Bartek, J., Lane, D. & Harris, A. L. Increased expression of mutant forms of p53 oncogene in primary lung cancer. *Lancet* **335**, 675–679 (1990).
28. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
29. Roig, J., Mikhailov, A., Belham, C. & Avruch, J. Nercc1, a mammalian NIMA-family kinase, binds the Ran GTPase and regulates mitotic progression. *Genes Dev* **16**, 1640–1658 (2002).
30. O'Regan, L. & Fry, A. M. The Nek6 and Nek7 protein kinases are required for robust mitotic spindle formation and cytokinesis. *Mol Cell Biol* **29**, 3975–3990 (2009).
31. Rapley, J. *et al.* The NIMA-family kinase Nek6 phosphorylates the kinesin Eg5 at a novel site necessary for mitotic spindle formation. *J Cell Sci* **121**, 3912–3921 (2008).
32. Bertran, M. T. *et al.* Nek9 is a Plk1-activated kinase that controls early centrosome separation through Nek6/7 and Eg5. *EMBO J* **30**, 2634–2647 (2011).
33. Sdelci, S., Bertran, M. T. & Roig, J. Nek9, Nek6, Nek7 and the separation of centrosomes. *Cell cycle* **10**, 3816–3817 (2011).
34. Sur, S. *et al.* A panel of isogenic human cancer cells suggests a therapeutic approach for cancers with inactivated p53. *Proc Natl Acad Sci U S A* **106**, 3964–3969 (2009).
35. Roig, J., Groen, A., Caldwell, J. & Avruch, J. Active Nercc1 protein kinase concentrates at centrosomes early in mitosis and is necessary for proper spindle assembly. *Mol Biol Cell* **16**, 4827–4840 (2005).
36. Muller, P. A. & Vousden, K. H. Mutant p53 in Cancer: New Functions and Therapeutic Opportunities. *Cancer Cell* **25**, 304–317 (2014).
37. Brown, C. J., Lain, S., Verma, C. S., Fersht, A. R. & Lane, D. P. Awakening guardian angels: drugging the p53 pathway. *Nat Rev Cancer* **9**, 862–873 (2009).
38. Cheok, C. F., Verma, C. S., Baselga, J. & Lane, D. P. Translating p53 into the clinic. *Nat Rev Clin Oncol* **8**, 25–37 (2011).

Acknowledgments

This work was supported by the Advanced Research for Medical Products Mining Program of the National Institute of Biomedical Innovation (NIBIO, 08-02, 12-01), and by Grants-in-Aid from the Ministry of Health, Labor, and Welfare of Japan; the Ministry of Education, Culture, Sports, Technology of Japan; and the National Cancer Center (NCC) Research and Development Fund (to N. T.). Whole-exome sequencing and expression analysis of clinical samples were supported by NIBIO (10-41, 10-43). We also thank the core facility of NCCRI for microarray analysis and animal experiments, which were supported by the NCC Research and Development Fund (23-A-7).

Author contributions

D.K. and N.T. designed all research, performed experiments, analyzed data, and wrote the paper. F.T. and T.O. carried out some experiments. M.W. and H.N. analyzed some data. H.S., K.M., J.Y. and T.K. performed expression and sequencing analyses of clinical samples. K.T. and S.W. performed clinical and pathological analyses.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Kurioka, D. *et al.* NEK9-dependent proliferation of cancer cells lacking functional p53. *Sci. Rep.* 4, 6111; DOI:10.1038/srep06111 (2014).

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 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

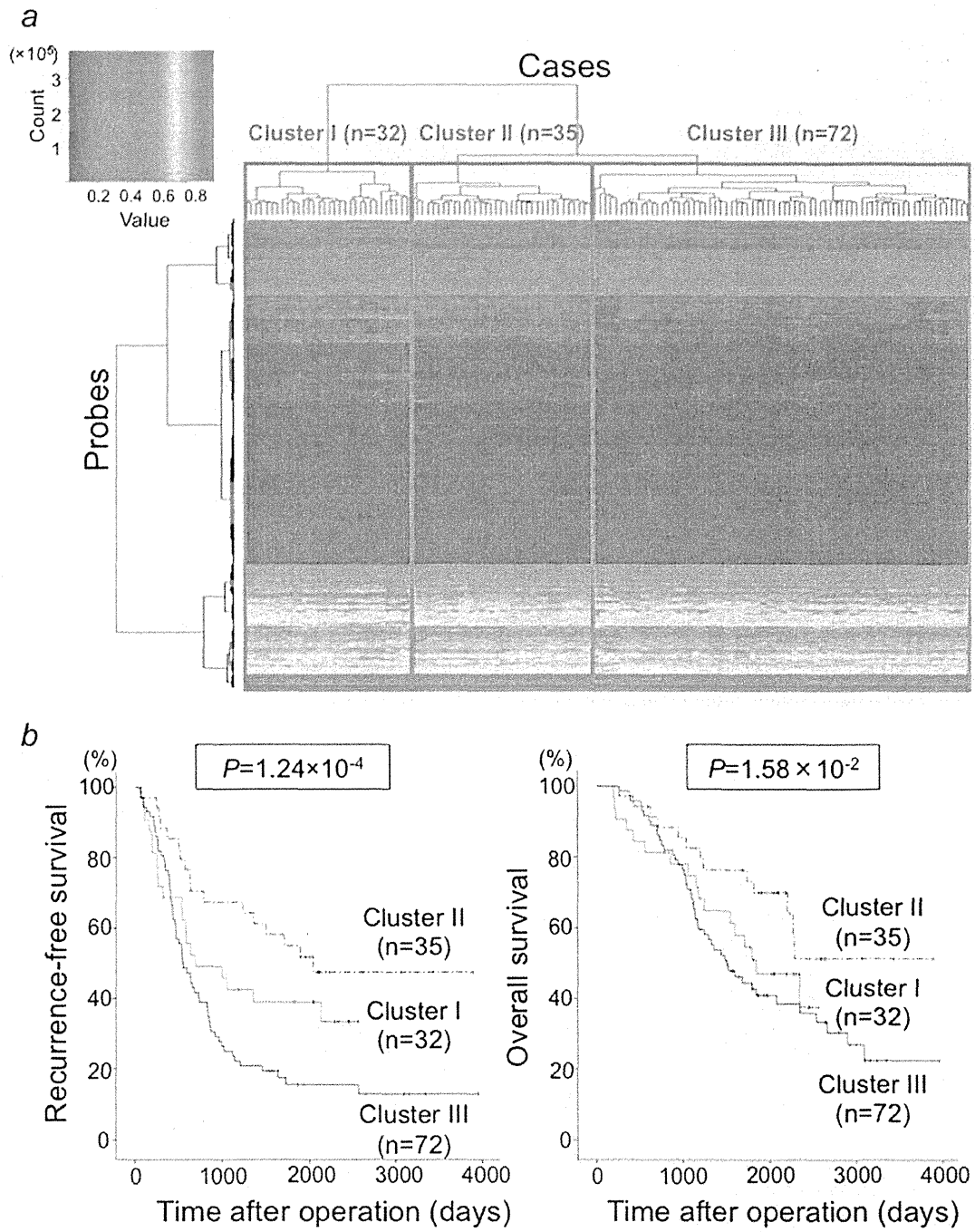

Figure 1. (*a*) Unsupervised hierarchical clustering (Euclidean distance, Ward method) using DNA methylation levels on all 26,447 probes in samples of non-cancerous lung tissue (N) from 139 patients with lung adenocarcinomas in the learning cohort. Based on DNA methylation status in adjacent lung tissue, 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) $\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).

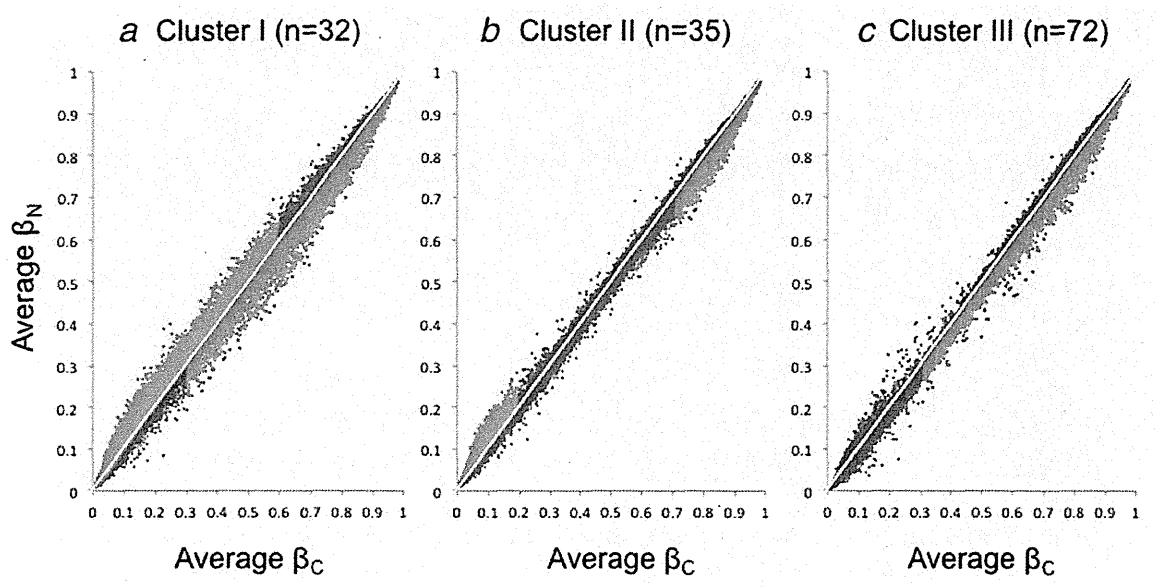

Figure 2. Distribution of average DNA methylation levels on all 26,447 probes of non-cancerous lung tissue (N) samples obtained from patients with lung adenocarcinomas belonging to Clusters I (*a*), II (*b*) and III (*c*) and 36 samples of normal lung tissue (C) obtained from patients without any primary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

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

Target ID [1]	Gene symbol	DNA methylation level in non-cancerous lung tissue (N) samples[2] (mean ± SD)											
		Pleural anthracosis			Emphysematic change			Atypical adenomatous hyperplasia			Tumor anthracosis		
		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}

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

(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]
cg11733245	IL2RA	−0.037 ± 0.042	−0.035 ± 0.051	9.29×10^{-1}	−0.023 ± 0.038	−0.047 ± 0.060	1.10×10^{-1}	−0.035 ± 0.051	−0.013 ± 0.023	1.09×10^{-1}	−0.009 ± 0.035	−0.047 ± 0.050	$\underline{7.07\times10^{-3}}$
cg22325572	CD53	−0.024 ± 0.083	−0.028 ± 0.058	9.23×10^{-1}	−0.017 ± 0.051	−0.040 ± 0.067	1.90×10^{-1}	−0.029 ± 0.060	−0.008 ± 0.057	4.68×10^{-1}	0.003 ± 0.057	−0.041 ± 0.057	$\underline{2.96\times10^{-2}}$
cg15691199	CEBPE	−0.029 ± 0.080	−0.018 ± 0.064	7.94×10^{-1}	−0.016 ± 0.050	−0.022 ± 0.079	7.79×10^{-1}	−0.022 ± 0.066	0.015 ± 0.021	$\underline{1.52\times10^{-2}}$	−0.002 ± 0.061	−0.028 ± 0.067	2.28×10^{-1}
cg16927606	U2AF1L4	−0.012 ± 0.079	−0.004 ± 0.054	8.25×10^{-1}	−0.005 ± 0.048	−0.006 ± 0.065	9.54×10^{-1}	−0.010 ± 0.057	0.032 ± 0.025	$\underline{1.48\times10^{-2}}$	0.009 ± 0.051	−0.012 ± 0.057	2.31×10^{-1}
cg16240480	EDARADD	−0.040 ± 0.103	−0.047 ± 0.072	8.91×10^{-1}	−0.032 ± 0.071	−0.062 ± 0.075	1.61×10^{-1}	−0.048 ± 0.074	−0.017 ± 0.070	3.86×10^{-1}	0.001 ± 0.049	−0.066 ± 0.076	$\underline{1.24\times10^{-3}}$
cg05596756	FAM113B	−0.010 ± 0.085	−0.010 ± 0.062	9.92×10^{-1}	0.000 ± 0.052	−0.022 ± 0.073	2.22×10^{-1}	−0.014 ± 0.064	0.026 ± 0.025	$\underline{2.14\times10^{-2}}$	0.011 ± 0.057	−0.020 ± 0.065	1.30×10^{-1}
cg08040471	C17orf62	0.000 ± 0.073	−0.014 ± 0.066	7.11×10^{-1}	−0.002 ± 0.051	−0.024 ± 0.080	2.80×10^{-1}	−0.015 ± 0.068	0.013 ± 0.029	1.21×10^{-1}	0.019 ± 0.053	−0.027 ± 0.066	$\underline{2.03\times10^{-2}}$
cg20622019	ADA	−0.021 ± 0.073	−0.033 ± 0.068	7.39×10^{-1}	−0.028 ± 0.069	−0.035 ± 0.066	7.31×10^{-1}	−0.036 ± 0.069	0.008 ± 0.021	$\underline{6.34\times10^{-3}}$	0.002 ± 0.051	−0.047 ± 0.068	$\underline{1.14\times10^{-2}}$
cg05109049	EVI2B	−0.057 ± 0.106	−0.041 ± 0.090	7.56×10^{-1}	−0.023 ± 0.093	−0.062 ± 0.084	1.32×10^{-1}	−0.044 ± 0.090	−0.009 ± 0.102	5.07×10^{-1}	−0.002 ± 0.054	−0.058 ± 0.099	$\underline{1.72\times10^{-2}}$
cg07973967	CD79B	−0.034 ± 0.105	−0.028 ± 0.075	9.16×10^{-1}	−0.017 ± 0.060	−0.042 ± 0.094	2.83×10^{-1}	−0.031 ± 0.080	0.007 ± 0.021	$\underline{2.06\times10^{-2}}$	−0.007 ± 0.067	−0.039 ± 0.081	1.69×10^{-1}

(B) Hallmark genes for Cluster II and III

		DNA methylation level in non-cancerous lung tissue (N) samples[5] (mean ± SD)										
		Lymphatic invasion			Nodal status				Pathological Tumor-Node-Metastasis stage			
Target ID [4]	Gene symbol	Negative	Positive	*p*-Value[6]	N0	N1	N2-3	*p*-Value[6]	IA-IB	IIA-IIB	IIIA-IV	*p*-Value[6]
cg26606064	*EI24*	0.018 ± 0.098	0.098 ± 0.113	$\underline{3.01\times10^{-2}}$	0.004 ± 0.098	0.010 ± 0.094	0.118 ± 0.091	$\underline{1.27\times10^{-3}}$	0.009 ± 0.102	−0.008 ± 0.082	0.118 ± 0.091	$\underline{1.14\times10^{-3}}$
cg17872476	*VTI1A*	−0.035 ± 0.134	−0.126 ± 0.137	$\underline{4.32\times10^{-2}}$	−0.023 ± 0.132	0.022 ± 0.127	−0.149 ± 0.121	$\underline{5.69\times10^{-3}}$	−0.020 ± 0.116	−0.016 ± 0.166	−0.149 ± 0.121	$\underline{6.70\times10^{-3}}$
cg21063899	*SCEL*	−0.043 ± 0.093	−0.109 ± 0.110	6.08×10^{-2}	−0.024 ± 0.082	−0.024 ± 0.090	−0.141 ± 0.096	$\underline{2.81\times10^{-4}}$	−0.021 ± 0.089	−0.033 ± 0.062	−0.141 ± 0.096	$\underline{2.60\times10^{-4}}$
cg14074641	*ABCC12*	−0.017 ± 0.118	−0.083 ± 0.117	8.60×10^{-2}	0.005 ± 0.104	−0.006 ± 0.112	−0.118 ± 0.114	$\underline{2.09\times10^{-3}}$	0.000 ± 0.111	0.014 ± 0.084	−0.118 ± 0.114	$\underline{2.00\times10^{-3}}$

[1]Probe ID for the Infinium HumanMethylation27 Bead Array.
[2]$\Delta\beta_{N\text{-}averageC}$.
[3]*p* values (Welch's *t*-test) of <0.05 are underlined.
[4]Probe ID for the Infinium HumanMethylation27 Bead Array.
[5]$\Delta\beta_{N\text{-}averageC}$.
[6]*p* values (Welch's *t*-test) of <0.05 are underlined.