Published in final edited form as:

Cancer Genet. 2015 May ; 208(5): 253–260. doi:10.1016/j.cancergen.2014.12.002.

DNA Methylation and RNA Expression Profiles in Lung Adenocarcinomas of Never-Smokers

Aaron S. Mansfield, MD¹, Liang Wang, MD, PhD⁷, Julie M. Cunningham, PhD^{2,3}, Jin Jen, MD, PhD^{2,3,4}, Christopher P. Kolbert³, Zhifu Sun, MD⁵, and Ping Yang, MD, PhD^{4,6}

Aaron S. Mansfield: mansfield.aaron@mayo.edu; Liang Wang: liwang@mcw.edu; Julie M. Cunningham: cunningham.julie@mayo.edu; Jin Jen: jen.jin@mayo.edu; Christopher P. Kolbert: kolbert.christopher@mayo.edu; Zhifu Sun: sun.zhifu@mayo.edu; Ping Yang: yang.ping@mayo.edu

¹Division of Medical Oncology, Mayo Clinic, Rochester, MN

²Division of Experimental Pathology and Laboratory Medicine, Mayo Clinic, Rochester, MN

³Medical Genome Facility, Mayo Clinic, Rochester, MN

⁴Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, MN

⁵Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN

⁶Division of Epidemiology and Department of Medical Genetics, Mayo Clinic, Rochester, MN

⁷Department of Pathology, Medical College of Wisconsin, Milwaukee, WI

Abstract

Lung cancer occurs in never-smokers. Epigenetic changes in lung cancer potentially represent important diagnostic, prognostic, and therapeutic targets. We compared DNA methylation profiles of 28 adenocarcinomas of the lungs of never-smokers with paired adjacent nonmalignant lung tissue. We correlated differential methylation changes with gene expression changes from the same 28 samples. We observed a distinct separation in methylation profiles between tumor and adjacent nonmalignant lung tissue using principal component analysis. Tumors were generally hypomethylated compared with adjacent nonmalignant tissue. Of 1,906 differentially methylated CpG sites between tumor and nonmalignant tissue, 1,198 were within classically defined CpG islands where tumors were hypermethylated compared with nonmalignant tissue. A total of 708 sites were outside CpG islands where tumors were hypomethylated compared with nonmalignant tissue. There were significant differences in expression of 351 genes (23%) of the 1,522 genes matched to the differentially methylated CpG sites. Genes that were not significantly differentially

© 2014 Elsevier Inc. All rights reserved.

Corresponding Author: Ping Yang, MD, PhD, Mayo Clinic 200 First St SW, Rochester, MN 55905, yang.ping@mayo.edu.

Reprints: Ping Yang, MD, PhD, Mayo Clinic, 200 First St SW, Rochester, MN 55905, yang.ping@mayo.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A portion of this work was presented as a poster at the International Association for the Study of Lung Cancer, Sydney, Australia, October 29, 2013.

Conflict of Interest

The authors have no conflicts of interest to disclose with this manuscript.

expressed and were hypermethylated within CpG sites were enriched for homeobox genes. These results suggest that the methylation profiles of lung adenocarcinomas of never-smokers and adjacent nonmalignant lung tissue are significantly different. Despite the differential methylation of homeobox genes, no significant changes in expression of these genes were detected.

Keywords

cell cycle; DNA methylation; epigenetics; gene expression; homeobox; lung adenocarcinoma

Introduction

Worldwide, lung cancer is the most common cancer and the most common cause of cancer-related death [1]. Although lung cancer is most frequently related to tobacco exposure, 10% to 15% of cases occur in never-smokers [2]. Lung cancers in never-smokers are enriched for mutations for which targeted therapy exists such as with epidermal growth factor receptor (EGFR) inhibitors [3] and anaplastic lymphoma kinase inhibitors [4]; however, most of these patients still receive nontargeted cytotoxic therapy because of the lack of an identifiable mutation for which a therapeutic agent exists. In contrast to the static nature of somatic mutations, DNA methylation is potentially reversible. A better understanding of the methylation profiles in nonsmoking patients with lung cancer may provide a rationale for exploring epigenetic therapeutics in this patient population.

DNA methylation at CpG dinucleotides is an important epigenetic mechanism regulating gene expression. Aberrant DNA methylation is known to be associated with cancer and other human diseases [5], and it is well known that DNA methylation and RNA expression profiles are profoundly altered in human cancers [6]. Human cancers frequently demonstrate genome-wide hypomethylation and site-specific hypermethylation [6]. Furthermore, methylation influences DNA strand separation and can result in errors in chromosome segregation [7]. The methylation of CpG shores, which are regions of less CpG density near CpG islands, controls expression of genes relevant to tissue differentiation and carcinogenesis [8]. Accordingly, methylation inside or outside CpG islands can influence the physical properties of DNA and control many genes relevant to carcinogenesis.

To understand the relationship of methylation and gene expression in tumor cells, we performed an unbiased analysis on a genome-wide scale to help determine potential diagnostic and therapeutic targets. In this study, we performed profiling analysis on both DNA methylation and RNA expression of adenocarcinomas of the lungs of never-smokers and their adjacent nonmalignant lung tissue.

Materials and Methods

Patients

Patients with lung cancer were identified and recruited between January 1997 and September 2008 from patients who were seen at Mayo Clinic, Rochester, MN. A detailed explanation of the recruitment process has been reported previously [9–11]. Briefly, patients with newly diagnosed lung cancer were identified by a daily electronic pathology reporting

system. Once identified, patients were invited to participate, and they were enrolled after they provided written informed consent. After enrollment, interviews were conducted and patients' medical records were abstracted. Patients must have been never-smokers, defined as having smoked fewer than 100 cigarettes in their lifetime. Patient lung tumor and matched nonmalignant lung tissue samples were obtained at surgery. Tumor specimens with 70% or more viable tumor, and non-malignant lung specimens with 0% tumor from the same patient were utilized. The Mayo Clinic Institutional Review Board approved this project.

DNA Extraction and Modification

We extracted DNA from tissue samples with use of an automated genomic DNA isolation system (AutoGen Flex Star; AutoGen, Inc, Holliston, Massachusetts) and modified the genomic DNA specimens using a kit that combined bisulfite conversion and DNA cleaning (EZ DNA Methylation Kit; Zymo Research Corporation, Orange, California). We modified 1 mcg of genomic DNA from lung tissue (both tumor and matched normal tissue) according to the manufacturer's recommendation. Treated DNA specimens were stored at -20°C and were assayed within 2 weeks.

DNA Methylation Profiling Analysis

Methylation assay (Infinium HumanMethylation27 BeadChip; Illumina, Inc) was used following the manufacturer's procedures. The arrays were imaged (BeadArray Reader scanner; Illumina, Inc), and the proportion methylated (β value) at each CpG site was calculated (BeadStudio Software; Illumina, Inc) after subtracting background intensity, computed from negative controls at each analytical data point. The β value represented each methylation data point as a ratio of fluorescent signals between the methylated and sum of methylated and unmethylated alleles, which ranged continuously from 0 (unmethylated) to 1 (fully methylated).

RNA Extraction and Microarray Analysis

After RNA extraction (miRNeasy Mini Kit; Qiagen, Valencia, California), gene expression profiling was performed (Illumina Human WG DASL BeadChip; Illumina, Inc). Three batches of raw intensity data were generated, and each batch was loaded into a data analysis software module (BeadStudio 3.1 gene expression module 3.4; Illumina, Inc) for quality control and calculation of gene or probe level intensity for each sample. Average signal intensity, number of detected genes, and signal-to-noise ratio (ratio of 95th vs 5th percentile signal intensity) were assessed for each sample. Samples with a signal-to-noise ratio less than 10 were excluded or repeated. Correlation between the replicates was assessed to ensure the reproducibility of the experiment. Samples that passed quality control were normalized together using the R package in the faster cyclic loess function [12]. After normalization, the results were assessed for batch effect and adequacy of normalization with principal component analysis, unsupervised clustering, and sample replicates. We did not identify a noticeable batch effect, and correlation among the replicates across batches was high ($r^2 > 0.95$).

Twenty-eight adenocarcinoma tumor samples and matched adjacent nonmalignant lung tissue samples were ultimately used for gene expression analysis. The expression data consisted of transcript levels for 24,526 microarray probes, representing 18,626 unique genes [13]. These data are available using the accession number GSE62950 through the Gene Expression Omnibus for the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/geo/>).

Gene Ontology and Pathway Analysis

To explore whether a group of selected genes shared a common biological function, we searched for overrepresentation in gene ontology (GO) categories using the Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.7 [14,15].

The Cancer Genome Atlas

To validate our findings, we downloaded the methylation data of 24 patients with adenocarcinoma of the lung and their paired normal tissue. Specifically, we analyzed the following subjects: The Cancer Genome Atlas (TCGA)-38-4628, TCGA-38-4629, TCGA-38-4630, TCGA-44-3918, TCGA-44-3919, TCGA-44-4112, TCGA-49-4486, TCGA-49-4494, TCGA-49-4501, TCGA-49-4505, TCGA-49-4506, TCGA-49-4507, TCGA-49-4510, TCGA-55-1592, TCGA-55-1594, TCGA-55-1595, TCGA-55-1596, TCGA-73-4659, TCGA-73-4662, TCGA-73-4666, TCGA-73-4668, TCGA-73-4670, TCGA-73-4675, and TCGA-73-4677. Methylation analysis was done (Infinium HumanMethylation27 BeadChip; Illumina, Inc), and the raw signal intensities and calculated β values were available for download from <http://cancergenome.nih.gov>. These subjects were not all never-smokers based on the available clinical information. These subjects from the TCGA were chosen based on the availability of DNA methylation and RNA expression data at the time of analysis.

Statistics

Principal component analysis and unsupervised hierarchical clustering were performed (Partek, St. Louis, Missouri). The Mann-Whitney U test was used to compare methylation profiles between groups (JMP software; SAS Institute Inc, Cary, North Carolina). Bonferroni corrected P values were used to determine CpG sites and genes that were significantly differentially methylated or expressed. We focused on CpG sites that were significantly differentially methylated in tumor compared with adjacent nonmalignant lung tissue for our analyses. Concordance of methylation and gene expression between groups was compared using the Fisher exact test with JMP. Curve-fitting software (Prism; GraphPad Software, Inc, La Jolla, California) was used to create Figure 1 and 2.

Results

Patient Characteristics and DNA Methylation Quality

We analyzed 28 tumors and their adjacent nonmalignant lung tissue. Our cohort was primarily female (22 women [79%]), and the majority of patients had stage I disease (18 patients [64%]) (Table 1). Half of the malignant samples had EGFR mutations. We exported

β values from each of 27,578 CpG sites and removed 420 CpG sites with a detection P value $>.05$. The remaining 27,158 CpG sites were used for data analysis.

Global Methylation Difference Between Adenocarcinoma and Adjacent Tissue

To examine the global methylation patterns, we applied an unsupervised principal component analysis and hierarchical cluster analysis using the 27,158 CpG sites. There was a distinct separation of methylation patterns between tumor and adjacent normal tissue (Figure 3A), but there was no separation of patterns between tumors with or without an EGFR mutation (Figure 3B). The global methylation levels (summarized β values in all CpG sites tested) were lower in tumors ($\beta=0.15$, 0.06-0.52 [median, interquartile range]) than in adjacent nonmalignant lung tissue ($\beta=0.17$, 0.07-0.51; $P=.02$; Figure 1A). To test whether different methylation status exists for sites within and outside classically defined CpG islands [16], we analyzed the 2 groups separately: sites within CpG islands (In-CpGs) and outside CpG islands (Out-CpGs). The In-CpGs generally showed lower methylation levels when compared with highly methylated Out-CpGs. Tumors were hypermethylated at In-CpGs ($n=19,911$; $\beta=0.11$, 0.06-0.25) compared with In-CpGs of adjacent lung tissue ($\beta=0.09$, 0.05-0.22; $P=7.7E-8$; Figure 1B). Tumors were hypomethylated at Out-CpGs ($n=7,206$; $\beta=0.59$, 0.41-0.73) compared with Out-CpGs of adjacent lung tissue ($\beta=0.60$, 0.43-0.74; $P=5.96E-4$; Figure 1B).

CpG Site Methylation Differences Between Adenocarcinoma and Adjacent Normal Tissues

To identify differentially methylated CpG sites, we compared methylation levels at each of the 27,158 CpG sites between tumor and adjacent nonmalignant lung tissue. We observed 1,906 CpG sites with significant differences (Bonferroni corrected P value $<.05$) in methylation. Among these sites, 1,198 were located at In-CpGs, and 708 were located at Out-CpGs. For the 1,198 In-CpGs, the overall methylation status was significantly higher in tumors ($\beta=0.38$, 0.27-0.52) than in adjacent lung tissue ($\beta=0.22$, 0.12-0.42; $P<.001$; Figure 1C). The overall methylation status was significantly lower for the Out-CpGs in tumor tissue ($\beta=0.49$, 0.37-0.63) than in adjacent lung tissue ($\beta=0.56$, 0.43-0.67; $P<.001$; Figure 1C).

Validation of Differentially Methylated CpGs Using The Cancer Genome Atlas Data Set

To ensure reproducibility of the differentially methylated CpGs, we analyzed the methylation profiles from 24 paired tissue samples in patients with adenocarcinoma from the TCGA data set. Of the 1,906 significantly differentially methylated CpGs in our data set, 1,180 also showed significant differences between tumor and adjacent normal tissue in the TCGA data set when determined with Bonferroni corrected P value $<.05$. All 1,180 differentially methylated CpGs from the TCGA data set showed the same direction of methylation changes as in our data set. There were 1,841 (96.6%) significantly differentially methylated CpGs in the TCGA data set when the statistical stringency was relaxed to an uncorrected P value $<.05$. Again, all 1,841 significantly changed CpGs in the TCGA data set changed in the same direction as in our patient data set (Figure 2).

Gene Ontology Analysis in Genes With Differentially Methylated CpGs

To elucidate potential biological pathways, we used the DAVID with medium classification stringency to examine GO term enrichment in the genes with differentially methylated CpGs. The 1,906 differentially methylated CpG sites represented 1,483 unique genes, which were enriched with the terms glycoprotein, signal, plasma membrane part, homeobox, pattern specification process, and cell adhesion (Table 2). We also performed a separate GO term enrichment analysis for genes at In-CpGs and Out-CpGs. For 901 unique genes associated with the 1,198 differentially methylated In-CpG sites, we found significant enrichment in homeobox (false discovery rate [FDR]= $2.5E-31$), regionalization (FDR= $5.2E-12$), and plasma membrane part (FDR= $1.5E-10$) in addition to others (Table 3). Hypermethylated genes exclusively drove these enrichments. The hypomethylated genes were not significantly enriched in any GO terms. For 607 unique genes associated with the 708 differentially methylated Out-CpGs, we observed significant enrichments in signal (FDR= $3.6E-27$), defense response (FDR= $9.8E-11$), and cell adhesion (FDR= $9.1E-5$). The direction of methylation did not affect the enrichments for Out-CpG sites.

Correlation Between CpG Methylation and Gene Expression

CpG methylation status usually influences gene expression levels such that hypermethylated CpG sites tend to inactivate a gene, whereas hypomethylated CpG sites tend to promote expression of a gene. Accordingly, we sought to determine if methylation levels were inversely correlated with gene expression. We matched 1,522 genes from the gene expression array (Illumina Human WG DASL BeadChip; Illumina, Inc, San Diego, California) with the genes associated with our differentially methylated CpG sites. We then extracted expression values in these matched genes from the same 28 paired tissues. We found significant differences in expression (based on Bonferroni corrected P value $<.05$) in 351 (23%) of these genes. Of these genes, 282 (80%) were inversely correlated with methylation. In cases where the tumor was hypomethylated relative to adjacent lung tissue, 72% of the genes were inversely correlated with methylation. In cases where tumor was hypermethylated relative to adjacent lung tissue, 87% of the sites were inversely correlated with methylation. The difference in the degree of correlation between hypermethylated sites and hypomethylated genes and their methylation status was significant ($P<.001$ by the Fisher exact test).

Dissociation Between Methylation and Expression in Genes With and Without Differential Methylation

To further define the relationship between CpG site methylation and gene expression, we performed GO enrichment analysis for the genes in 2 groups: the genes with differential methylation (Bonferroni corrected $P<.05$) and genes without differential methylation (uncorrected $P>.05$) at In-CpGs and Out-CpGs. We further divided our analysis by genes that were hypermethylated and hypomethylated. This analysis identified distinct GO enrichment categories. Genes that were significantly differentially expressed and hypermethylated at Out-CpGs were enriched for signal-related genes (FDR= $3.2E-9$). Genes that were not significantly differentially expressed but were hypermethylated at In-CpGs

were enriched for homeobox genes (FDR=1.2E-27). The other categories were not enriched with enrichment scores greater than our cutoff value of 5.

Discussion

In this study, we demonstrate that there are distinct methylation profiles between adenocarcinomas of the lungs of never-smokers and their adjacent nonmalignant lung tissue. Although there were similar frequencies of β values between tumor and adjacent lung tissue, principal component analysis and analysis of the β values by location within CpG islands demonstrated distinct differences in the distribution of methyl groups between these tissues. For both tumor and adjacent nonmalignant lung tissue, the degree of global methylation was lower in classically defined CpG islands than outside these islands. Further analysis showed that tumor tissue was more likely to be hypermethylated at In-CpG sites compared with adjacent lung tissue, which is consistent with previous reports that tumor cells exhibit regional (promoter region) hypermethylation and global (nonpromoter region) hypomethylation [5]. When comparing methylation differences between nonmalignant lung tissue and tumor tissue at individual CpG sites, we identified differential methylation in 1,906 CpG sites. Subsequent analysis using 24 normal-tumor pairs from the TCGA methylation data set confirmed the methylation difference in at least 1,841 of the 1,906 CpG sites, validating our data.

For these differentially methylated genes, we found significant enrichment in several GO categories, including glycoprotein, plasma membrane, secreted protein, homeobox, and embryonic morphogenesis. In fact, both glycoprotein and plasma membrane are closely related to signal proteins while homeobox and embryonic morphogenesis are associated with developmental proteins. Interestingly, the enrichment of signal proteins was seen only when methylation changes occurred outside classically defined CpG islands. We were not able to correlate the differential methylation of homeobox genes with a change in expression of the genes. Although there is an association with aberrant DNA methylation and alteration of gene expression, exceptions to this general rule clearly exist. An example is the positive correlation between methylation of the telomerase reverse transcriptase promoter sequence and gene expression and telomerase activity [17]. Chromatin modifications may also affect gene expression [18]. Thus, other epigenetic phenomena not accounted for with our methodology could explain the discrepancies we observed between methylation and gene expression.

Homeobox genes are a group of transcription factors that regulate embryonic morphogenesis [19]. Some homeobox genes exhibit varying degrees of altered expression in various human cancers, including kidney [20], colon [21], bladder [22], melanoma [23], and prostate cancers [24]. Homeobox genes have been studied in human lung cancer. The homeobox A9 (HOXA9) promoter has been reported to exhibit high levels of DNA methylation in stage I squamous cell lung carcinomas [25]. HOXA9 was found to be hypermethylated in patients with lung cancer compared with healthy controls, and this hypermethylation was correlated with reduced gene expression [26]. More recently, homeobox A7 (HOXA7), HOXA9, and homeobox B4 (HOXB4) were found to be hypermethylated in adenocarcinoma of the lung compared with nontumor lung [27]. Although we also detected hypermethylation of

homeobox-related genes, including HOXA7, HOXA9, and HOXB4, we did not observe significant differences in expression. Similarly, we found that differential methylation of PAX genes (e.g. PAX3, PAX4, PAX5, PAX6, PAX7, PAX8, and PAX9) did not correlate with changes in expression whereas, no change was observed in the methylation or expression of SIX1. The homeobox gene NANOG was not covered by the platform we used.

One recent study on DNA methylation profiling of lung adenocarcinoma reported that approximately one-third of differentially methylated genes are associated with functional changes in gene expression [27]. Similarly, 23% of the differentially methylated CpG sites in our study were associated with differential expression of their corresponding genes. The small difference between these 2 studies might be attributable to our use of the Bonferroni correction rather than the Benjamini-Hochberg correction.

CpG island methylation has been studied in cancer with findings that hypermethylation of tumor suppressor genes and hypomethylation of oncogenes contribute to the process of carcinogenesis. In our study, there was an inverse correlation between methylation and gene expression in 80% of the differentially methylated and differentially expressed samples. There was a stronger inverse correlation at CpG sites that were hypermethylated in tumors relative to adjacent lung tissue. Accordingly, about 20% of the differentially methylated and expressed genes in our study did not follow the classic paradigm of how methylation affects gene expression. Consequently, this finding supports the notion that there are more complexities to epigenetic regulation of gene expression, and our correlations with CpG site methylation and gene expression may not capture the full complexity of the effects of methylation on gene expression.

Recent studies have added to our knowledge of abnormal methylation patterns in lung cancer. In one report, *CPEB1* and *EMILIN2* were methylated more frequently in current smokers compared with never smokers; whereas *CST6*, *LAYN*, and *MARVELD3* were more frequently methylated in never smokers than current or former smokers [28]. A smaller study focused on the methylation patterns of tumor suppressor genes reported that *p16* was more commonly methylated in never smokers than current or former smokers [29]. A different study showed that heavy smokers, light smokers and never smokers separated into different groups based on an unsupervised hierarchical clustering analysis of DNA methylation levels [30]. Additionally, work from The Cancer Genome Atlas identified three sets of CpG island methylator phenotype (CIMP) samples: high, intermediate and low. Tumors in the CIMP-high subset often had methylation of following genes *CDKN2A*, *GATA2*, *GATA4*, *GATA5*, *HIC1*, *HOXA9*, *HOXD13*, *RASSF1*, *SFRP1*, *SOX17* and *WIF1* [31], some are overlapping with our findings.

Epigenetic patterns can be cell-type specific. Furthermore, tumors can have heterogeneous gene expression signatures [32]. We did not test regions of tumor separately and cannot account for tumor heterogeneity in our study. Additionally, we used adjacent nonmalignant lung tissue as a comparison. We are not certain what differences would have been detected if lung tissue farther away from the malignant cells had been used, although this would have been difficult to obtain. As there is more than one cell type in adjacent lung tissue, our results likely reflect an averaging of these cell types. This same issue arises in gene

expression data, and a method to decompose whole tissue expression into cell-specific components [33] may prove useful in identifying cell-specific epigenetic marks in complex tissues.

In summary, we identified distinct methylation profiles between adenocarcinoma of the lung in never-smokers and adjacent nonmalignant lung tissue. These epigenetic changes were associated with differential gene expression in about a quarter of the differentially methylated CpG sites; however, there was no significant change in expression of homeobox-related genes. These profiles may represent novel biomarkers or targets in lung adenocarcinomas of never-smokers.

Acknowledgments

This work was supported by NIH R01 84354, Mayo Foundation Research Fund (P.Y.), and Advancing a Healthier Wisconsin fund (L.W., Project 5520227). The study sponsors had no role in the conduct of this study, data interpretation, or writing of the manuscript. We thank the Mayo Clinic Genotyping and Gene Expression Cores for performing the genomic analyses and providing technical support, specifically Debra A. Schultz, Jeffrey Meyer and Jason Vollenweider. These cores are supported by NCI CA15093 Mayo Clinic Cancer Center CCSG. We would also like to thank Bobbi Jebens for her administrative support.

Abbreviations

DAVID	Database for Annotation, Visualization and Integrated Discovery
EGFR	epidermal growth factor receptor
FDR	false discovery rate
GO	gene ontology
HOXA7	homeobox A7
HOXA9	homeobox A9
HOXB4	homeobox B4
In-CpG	inside CpG islands
Out-CpG	outside CpG islands
TCGA	The Cancer Genome Atlas

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*. 2010 Dec 15; 127(12):2893–2917. [PubMed: 21351269]
2. Thun MJ, Hannan LM, Adams-Campbell LL, Boffetta P, Buring JE, Feskanich D, et al. Lung cancer occurrence in never-smokers: an analysis of 13 cohorts and 22 cancer registry studies. *PLoS Med*. 2008 Sep 30.5(9):e185. Epub 2008 Sep 9. [PubMed: 18788891]
3. Laurie SA, Goss GD. Role of epidermal growth factor receptor inhibitors in epidermal growth factor receptor wild-type non-small-cell lung cancer. *J Clin Oncol*. 2013 Mar 10.31(8):1061. Epub 2013 Feb 11. [PubMed: 23401452]
4. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med*. 2010 Oct 28.363(18):1693. Erratum in: *N Engl J Med*. 2011 Feb 10;364(6):588. [PubMed: 20979469]

5. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*. 2004 May 27; 429(6990):457–463. [PubMed: 15164071]
6. Azad N, Zahnow CA, Rudin CM, Baylin SB. The future of epigenetic therapy in solid tumours: lessons from the past. *Nat Rev Clin Oncol*. 2013 May; 10(5):256–266. Epub 2013 Apr 2. [PubMed: 23546521]
7. Severin PM, Zou X, Gaub HE, Schulten K. Cytosine methylation alters DNA mechanical properties. *Nucleic Acids Res*. 2011 Nov 1.39(20):8740. Epub 2011 Jul 20. [PubMed: 21775342]
8. Yang IV, Schwartz DA. Epigenetic control of gene expression in the lung. *Am J Respir Crit Care Med*. 2011 May 15; 183(10):1295–1301. [PubMed: 21596832]
9. Yang P, Wentzlaff KA, Katzmman JA, Marks RS, Allen MS, Lesnick TG, et al. Alpha1-antitrypsin deficiency allele carriers among lung cancer patients. *Cancer Epidemiol Biomarkers Prev*. 1999 May; 8(5):461–465. [PubMed: 10350443]
10. Yang P, Bamlet WR, Sun Z, Ebbert JO, Aubry MC, Krowka MJ, et al. Alpha1-antitrypsin and neutrophil elastase imbalance and lung cancer risk. *Chest*. 2005 Jul.128(1):445. Erratum in: *Chest*. 2006 Jan;129(1):216. [PubMed: 16002971]
11. Yang P, Sun Z, Krowka MJ, Aubry MC, Bamlet WR, Wampfler JA, et al. Alpha1-antitrypsin deficiency carriers, tobacco smoke, chronic obstructive pulmonary disease, and lung cancer risk. *Arch Intern Med*. 2008 May 26; 168(10):1097–1103. [PubMed: 18504338]
12. Ballman KV, Grill DE, Oberg AL, Therneau TM. Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics*. 2004 Nov 1.20(16):2778. Epub 2004 May 27. [PubMed: 15166021]
13. April C, Klotzle B, Royce T, Wickham-Garcia E, Boyaniwsky T, Izzo J, et al. Whole-genome gene expression profiling of formalin-fixed, paraffin-embedded tissue samples. *PLoS One*. 2009 Dec 3.4(12):e8162. [PubMed: 19997620]
14. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4(1):44–57. [PubMed: 19131956]
15. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009 Jan; 37(1):1–13. Epub 2008 Nov 25. [PubMed: 19033363]
16. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. *J Mol Biol*. 1987 Jul 20; 196(2):261–282. [PubMed: 3656447]
17. Kumari A, Srinivasan R, Vasishta RK, Wig JD. Positive regulation of human telomerase reverse transcriptase gene expression and telomerase activity by DNA methylation in pancreatic cancer. *Ann Surg Oncol*. 2009 Apr; 16(4):1051–1059. Epub 2009 Feb 5. [PubMed: 19194757]
18. Kouzarides T. Chromatin modifications and their function. *Cell*. 2007 Feb 23; 128(4):693–705. [PubMed: 17320507]
19. Garcia-Fernandez J. The genesis and evolution of homeobox gene clusters. *Nat Rev Genet*. 2005 Dec; 6(12):881–892. [PubMed: 16341069]
20. Okuda H, Toyota M, Ishida W, Furihata M, Tsuchiya M, Kamada M, et al. Epigenetic inactivation of the candidate tumor suppressor gene HOXB13 in human renal cell carcinoma. *Oncogene*. 2006 Mar 16; 25(12):1733–1742. [PubMed: 16278676]
21. Freund JN, Domon-Dell C, Kedinger M, Duluc I. The Cdx-1 and Cdx-2 homeobox genes in the intestine. *Biochem Cell Biol*. 1998; 76(6):957–969. [PubMed: 10392709]
22. Marra L, Cantile M, Scognamiglio G, Perdona S, La Mantia E, Cerrone M, et al. Deregulation of HOX B13 expression in urinary bladder cancer progression. *Curr Med Chem*. 2013 Feb 1; 20(6): 833–839. [PubMed: 23276138]
23. Maeda K, Hamada J, Takahashi Y, Tada M, Yamamoto Y, Sugihara T, et al. Altered expressions of HOX genes in human cutaneous malignant melanoma. *Int J Cancer*. 2005 Apr 10; 114(3):436–441. [PubMed: 15551325]
24. Waltregny D, Alami Y, Clausse N, de Leval J, Castronovo V. Overexpression of the homeobox gene HOXC8 in human prostate cancer correlates with loss of tumor differentiation. *Prostate*. 2002 Feb 15; 50(3):162–169. [PubMed: 11813208]
25. Rauch T, Wang Z, Zhang X, Zhong X, Wu X, Lau SK, et al. Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery

- assay. *Proc Natl Acad Sci U S A*. 2007 Mar 27;104(13):5527. Epub 2007 Mar 16. [PubMed: 17369352]
26. Son JW, Jeong KJ, Jean WS, Park SY, Jheon S, Cho HM, et al. Genome-wide combination profiling of DNA copy number and methylation for deciphering biomarkers in non-small cell lung cancer patients. *Cancer Lett*. 2011 Dec 1;311(1):29. Epub 2011 Jun 24. [PubMed: 21757291]
27. Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, et al. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res*. 2012 Jul; 22(7):1197–1211. Epub 2012 May 21. [PubMed: 22613842]
28. Tessema M, Yingling CM, Liu Y, Tellez CS, Van Neste L, Baylin SS, et al. Genome-wide unmasking of epigenetically silenced genes in lung adenocarcinoma from smokers and never smokers. *Carcinogenesis*. 2014 Jun; 35(6):1248–1257. [PubMed: 24398667]
29. Scesnaite A, Jarmalaite S, Mutanen P, Anttila S, Nyberg F, Benhamou S, et al. Similar DNA methylation pattern in lung tumours from smokers and never-smokers with second-hand smoke exposure. *Mutagenesis*. 2012 Jul; 27(4):423–429. [PubMed: 22217548]
30. Sato T, Arai E, Kohno T, Takahashi Y, Miyata S, Tsuta K, et al. Epigenetic clustering of lung adenocarcinomas based on DNA methylation profiles in adjacent lung tissue: Its correlation with smoking history and chronic obstructive pulmonary disease. *Int J Cancer*. 2014 Jul 15; 135(2): 319–334. [PubMed: 24921089]
31. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinomas. *Nature*. 2014 Jul 31; 511(7511):543–550. [PubMed: 25079552]
32. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012 Mar 8;366(10):883. Erratum in: *N Engl J Med*. 2012 Sep 6;367(10):976. [PubMed: 22397650]
33. Shen-Orr SS, Tibshirani R, Khatri P, Bodian DL, Staedtler F, Perry NM, et al. Cell type-specific gene expression differences in complex tissues. *Nat Methods*. 2010 Apr; 7(4):287–289. Epub 2010 Mar 7. [PubMed: 20208531]

Figure 1.

Summary of methylation profiles. The distributions of β values are shown above for A, all CpG sites; B, all CpG sites by whether they are inside or outside CpG islands (In-CpG or Out-CpG); and C, differentially methylated CpG sites by In-CpG and Out-CpG. In panels B and C, the In-CpG site frequencies are plotted against the left y-axis, and the Out-CpG site frequencies are plotted against the right y-axis.

β value correlations

Figure 2.

Correlation of differentially methylated CpG sites. The differences of the β values (N-T) of the 1,841 differentially methylated CpG sites from The Cancer Genome Atlas (TCGA) (y-axis) are plotted against the differences of the β values (N-T) of the Mayo Clinic cohort (x-axis). All CpG sites changed in the same direction.

Figure 3. Principal component analysis (PCA) of methylation profiles. A, The PCA of the methylation profiles of tumors (blue) and adjacent, nonmalignant lung tissue (red) demonstrates separation between these 2 groups. B, No separation of methylation profiles is observed between tumors (lower group of dots connected by lines) or adjacent lung tissue (upper group of dots connected by lines), based on the presence (magenta spheres) or absence (teal spheres) of epidermal growth factor receptor mutations (EGFR mut).

Table 1

Patient Characteristics

Characteristic	Value ^a
Age, median (interquartile range), y	65 (57–72)
Sex	
Male	6 (21)
Female	22 (79)
Histology	
Adenocarcinoma	23 (82)
Adenosquamous	4 (14)
Bronchioloalveolar	1 (4)
Stage ^b	
IA	6 (21)
IB	12 (43)
IIA	1 (4)
IIB	4 (14)
IIIA	3 (11)
IIIB	1 (4)
IV	1 (4)
EGFR	
Mutant	14 (50)

Abbreviation: EGFR, epidermal growth factor receptor.

^a Values are number (percentage) unless indicated otherwise.

^b Percentages do not total 100% due to rounding.

Table 2

Functional Annotation of 1,483 Differentially Methylated Genes^a

Cluster	Enrichment Score	Terms	Fold Change	P Value	FDR
1	20.32	Glycoprotein	1.5	9.6E-26	1.5E-22
		Signal	1.6	1.2E-23	1.8E-20
3	13.2	Plasma membrane part	1.5	4.5E-16	6.4E-13
4	10.96	DNA: homeobox	4.4	3.6E-25	6.6E-22
		DNA: homeobox	3.8	1.9E-22	2.8E-19
6	7.47	Pattern specification process	2.7	1.3E-12	2.4E-9
		Embryonic morphogenesis	2.5	2.6E-12	4.8E-9
7	6.56	Cell adhesion	1.8	1.3E-8	2.4E-5
		Biological adhesion	1.8	1.4E-8	2.6E-5

Abbreviation: FDR, false discovery rate.

^aThe most relevant nonredundant annotation clusters with enrichment scores higher than 6 are listed with their top 2 nonredundant terms.

Table 3

Functional Annotation of Differentially Methylated Genes at In-CpGs^a

Cluster	Enrichment Score	Terms	Fold Change	P Value	FDR
1	18.05	DNA-binding region: homeobox	6.9	1.4E-34	2.5E-31
		DNA-binding region: homeobox	5.8	7.4E-31	1.1E-27
2	9.46	Regionalization	4.2	2.8E-15	5.2E-12
		Pattern specification process	3.5	3.0E-14	5.5E-11
3	6.6	Plasma membrane part	1.7	1.1E-13	1.5E-10
		Integral to plasma membrane	1.9	4.7E-12	6.6E-9
4	6.34	Lipid moiety-binding region: S-palmitoyl cysteine	3.4	2.6E-8	4.5E-5
		Lipid moiety-binding region: palmitate	2.9	7.4E-7	1.1E-3
5	6.34	Cell-cell signaling	2.2	2.3E-9	4.1E-6
		Transmission of nerve impulse	2.4	1.4E-7	2.5E-4
6	6.1	Regulation of transcription from RNA polymerase II promoter	2.1	6.6E-10	1.2E-6
		Positive regulation of transcription from RNA polymerase II promoter	2.3	2.7E-7	4.9E-4
7	5.21	Embryonic morphogenesis	4.3	3.3E-11	6.0E-8
		Sensory organ development	2.7	1.0E-6	1.8E-3

Abbreviations: FDR, false discovery rate; In-CpG, inside CpG islands.

^aThe most relevant nonredundant annotation clusters with enrichment scores higher than 5 are listed with their top 2 nonredundant terms.