Genome-scale meta-analysis of DNA methylation during progression of lung adenocarcinoma

Q.H. Zhang[1,2], X.H. Dai[1,2], Z.M. Dai[1,2] and Y.N. Cai[1,2]

[1]School of Information Science and Technology, Sun Yat-Sen University, Higher Education Mega Center, Guangzhou, China
[2]SYSU-CMU Shunde International Joint Research Institute, Shunde, Guangdong, China

Corresponding author: Q.H. Zhang
E-mail: sunnydoll78@163.com

ABSTRACT. Identification of epigenetic alterations in tumors has become a common method for identifying genes critical to cancer development and progression. Thus, we identified DNA methylation alterations on the genome scale during lung adenocarcinoma (LADC) progression to understand the carcinogenic process and identify clinically relevant biomarkers. We found that epigenetic alterations in LADC mainly occur during the early stage of LADC progression, and there are no significant methylation differences between early-stage and late-stage LADCs. This suggests that DNA methylation alterations characterize a turning point of early events in LADC progression. By comparing DNA methylation between early-stage LADCs and normal lung tissues, we further identified 940 genes with significant alterations in DNA methylation. Sixty-seven genes were found to exhibit strong correlation between methylation alterations and expression changes, based on associated gene expression data. According to gene ontology analysis, these genes are involved in lung development, respiratory

system development, cell cycle, histidine metabolism, the Wnt signaling pathway, and the p53 signaling pathway. We also found that genes on chromosome 18 most frequently showed promoter hypermethylation. Moreover, we found that LADC-associated DNA hypomethylation occurred preferentially at neither histone H3 lysine 4 nor histone H3 lysine 27 mark domains in human embryonic stem cells (NMDs) and that hypomethylation of NMDs was associated with a poor prognostic signature in LADC. Our findings have important implications for LADC progression because of the identification of novel epigenetic biomarkers potentially involved in early-stage LADC and for establishing the importance of NMD DNA hypomethylation for predicting prognosis in LADC.

Key words: Carcinogenic process; DNA methylation; Genome; Epigenetic biomarkers; Lung adenocarcinomas

INTRODUCTION

DNA methylation is a chemical modification of DNA involved in gene expression programming (Razin and Riggs, 1980). One of the hallmarks of cancer is aberrant DNA methylation; however, its precise role in carcinogenesis and clinical prognosis remains unclear (Baylin and Jones, 2011). Although multiple studies have attempted to assess the methylated gene profiles in lung cancer (Shiraishi et al., 2002; Divine et al., 2005; Feng et al., 2008; Tsou et al., 2008; Anglim et al., 2008; Buckingham et al., 2010), they have either included a limited number of samples or genes assayed or have focused on a mix of lung cancer histologies and different stage tumors, limiting the ability to identify dynamic changes in DNA methylation during carcinogenesis.

Moreover, both hyper- and hypomethylation are commonly observed in cancer (Jones and Baylin, 2007). Hypermethylation appears to occur locally, preferentially targeting gene promoters. Several studies have reported that a statistically high fraction of these promoters map to stem cell polycomb group target genes (PCGTs) (Ohm et al., 2007; Widschwendter et al., 2007), many of which encode transcription factors needed for differentiation. These genes are normally suppressed in embryonic stem cells through a reversible mechanism mediated by the polycomb repressive complex (Lee et al., 2006). This preferential hypermethylation of PCGTs in cancer supports that the reversible gene repression of PCGTs in stem cells can be replaced by permanent silencing in cancer, potentially impairing the differentiation capacity of cells. In contrast to hypermethylation, studies of preferential hypomethylation are limited.

To address these issues, we established a genomic model of disease development from a pre-cancerous genome (adjacent non-involved tissue), from the early-stage tumor genome to the late-stage tumor genome. This model enabled examination of trends in genome-wide DNA methylation alterations and their effects on disease development. We verified our findings using 2 independent data sets. To identify "passenger" genes and "driver" genes among genes that were aberrantly methylated, we integrated the expression profiles of normal lung tissues (NTLs) and lung adenocarcinomas (LADCs) to identify DNA methylation events with potential functional significance.

MATERIAL AND METHODS

Patient population

Patients were eligible to enter the study as part of either discovery or validation cohorts if they had undergone surgical resection of non-small cell lung carcinoma in any of the international participating institutions. Patients treated with neoadjuvant therapy and/or patients with large cell-carcinoma were not included in the study.

DNA methylation and mRNA expression profiles

We collected human LADC DNA methylation data sets from a public microarray database (GEO, http://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2009). DNA methylation data and expression data for 57 pairs of LADC and matched NTL methylation data were downloaded from GSE32867 (Selamat et al., 2012). Methylation of DNA from primary non-small cell lung carcinoma fibroblast cell lines and normal lung fibroblasts were downloaded from GSE27284 (Navab et al., 2011). Three independent verification sets were downloaded from GSE32867 (Selamat et al., 2012), GSE39279 (Sandoval et al., 2013), and the cancer genome atlas (TCGA, http://tcga-data.nci.nih.gov/tcga).

Selection of cancer-specific DNA methylation markers

For Illumina Infinium DNA methylation data analysis, we masked data points as "NA" for P values >0.05 and replaced them with "NA" values. We also excluded probes that were designed for sequences on either the X or Y chromosome. Next, we analyzed the DNA methylation data set that did not contain any ''NA''-masked data points. The Wilcoxon rank-sum test was used to evaluate the difference in DNA methylation beta-values for each probe between groups tested. False discovery rate adjusted P values for multiple comparisons were calculated using the Benjamini and Hochberg approach. Probes were considered statistically significant difference between the tested groups, if they met the following criteria: groups with a false discovery rate-adjusted P value <0.05 and absolute mean beta-value difference >0.2. This threshold of absolute beta-value difference >0.2 was determined previously as a stringent estimate of beta-value difference detection sensitivity across a range of beta-values (Bibikova et al., 2009).

Integrated analyses of Illumina Infinium DNA methylation and gene expression data

The raw expression values were log2-transformed, and robust spline normalization was performed using the lumi package in R (Du et al., 2008). For genes with multiple probes, we averaged the probes for genes and analyzed differential expression using t-tests and a Benjamini-Hochberg multiple comparison correction. Statistical significance was considered at Benjamini-Hochberg-adjusted P value <0.05 and a mean difference of at least 2-fold change. For genes with multiple probes measuring DNA methylation, we averaged the DNA methylation value of the probes for the genes. The correlation between

gene expression and DNA methylation for each gene was measured using the Spearman correlation coefficient.

RESULTS

DNA methylation changes mainly occurred in early-stage LADC

We specifically examined the dynamics of DNA methylation changes in LADC, and thus analyzed DNA methylation changes during LADC progression. We performed a locus-by-locus DNA methylation analysis of early-stage LADCs *vs* NTLs and early-stage LADCs *vs* late-stage LADCs. The analysis was restricted to 26,420 autosomal CpG probes with a detection $P < 0.05$. By using a criterion of false-discovery rate adjusted P value <0.05 and a minimum mean beta-value difference of 20%, we identified 1211 CpGs corresponding to 940 unique genes that showed significant differential methylation between early-stage LADCs and NTLs, but we found no significant differential methylation pattern between early- and late-stage LADCs. To ensure that our findings were not dependent on the specific population analyzed, we performed additional DNA methylation analyses on 4 independent sample sets of 671 LADCs and 100 NTLs (Table 1). Using the same criteria, we also found no significant difference in methylation between early- and late-stage LADCs. This suggests that the DNA methylation changes mainly occurred in early-stage LADCs, while the methylation of early- and late-stage LADCs were genetically highly similar.

Among the 1211 CpGs showing significantly different methylation between early-stage LADCs and NTLs, 860 CpGs, corresponding to 634 unique genes, were significantly hypermethylated in early-stage LADC, and 351 CpGs, corresponding to 306 unique genes, were significantly hypomethylated. Hierarchical clustering of the 1211 differentially methylated CpG sites showed clear demarcation between early-stage LADCs and NTLs (Figure 1). Gene ontology analysis of the hypermethylated or hypomethylated CpG sites located in the promoter regions and CpG islands (CGIs) in LADCs was performed.

In order to validate differentially methylated genes identified in early-stage LADCs, we analyzed the methylation patterns of 2 independent data sets using the same Infinium platform (Table 1, validation set 1 and validation set 2). Using the same criteria, we found that 85.8 and 95% of hypomethylation probes and 88.6 and 90% of hypermethylation probes overlapped in the 2 verification sets and the discovery set, respectively, supporting our initial findings.

Table 1. Data sets used in this study.

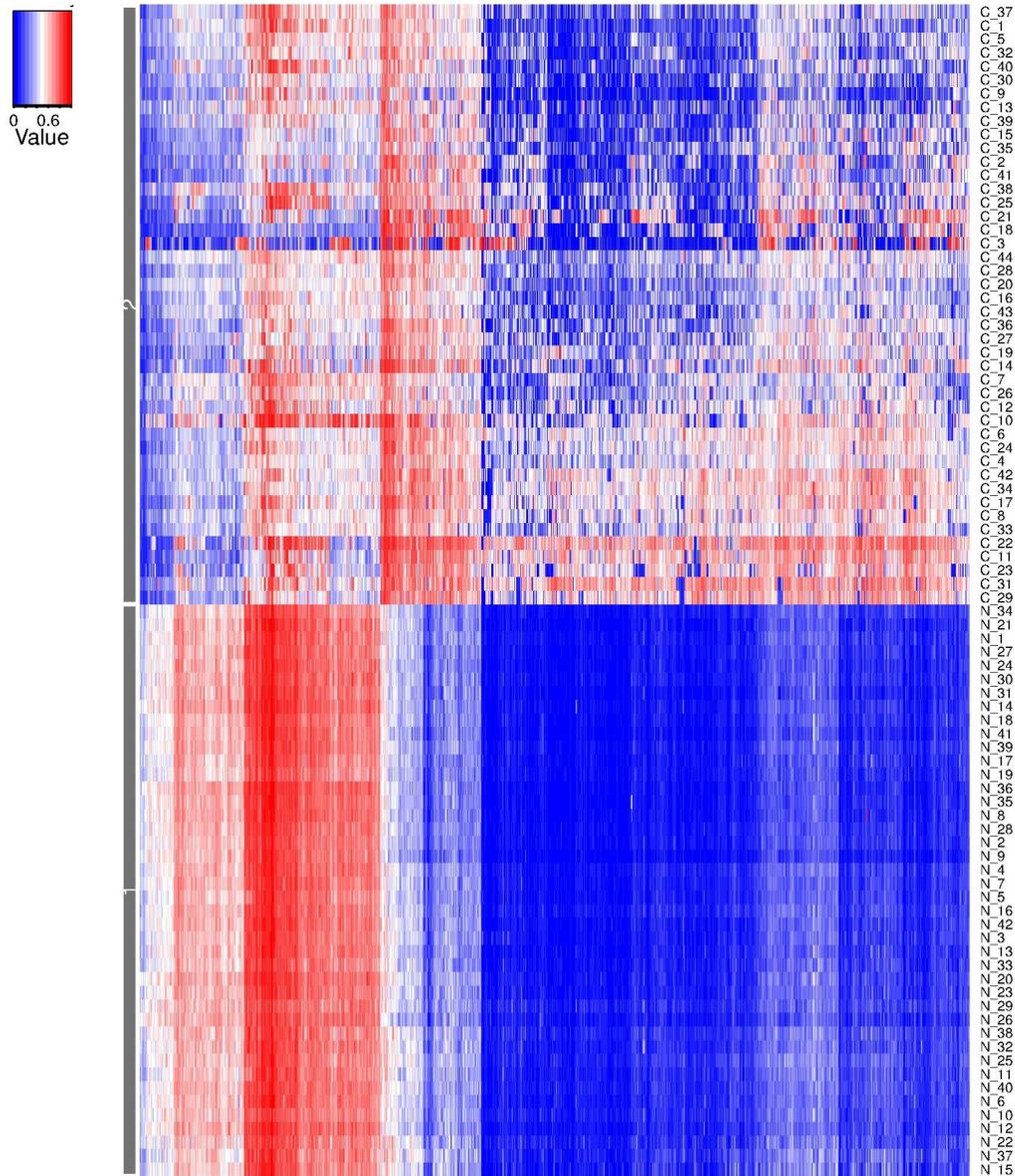| Tissue source	Platform	Sample number	Data ID
LADC (main set)	Illum27K	114	GSE32863
NSCLC fibroblasts	Illum27K	10	GSE27284
LADC (validation set 1)	Illum27K	28	GSE32866
LADC (validation set 2)	Illum27K	149	TCGA
LADC (validation set 3)	Illum450K	145	GSE39279
LADC (validation set 4)	Illum450K	449	TCGA

Figure 1. Hierarchical clustering with the differentially methylated CpG sites separates tumor from normal lung.

Effects of DNA methylation alterations on gene expression

Promoter CGI DNA methylated alteration can lead to transcriptional silencing or activation of the associated gene. However, most cancer-specific CGI hypermethylation or hypomethylation events may occur during gene expression (Figure 2). We found no statistically significant difference in the expression level between hypermethylated promoter and

hypomethylated promoter groups (Table 2). We found 47 epigenetically suppressed genes whose promoters showed increased methylation levels and whose expression levels were reduced in tumors (Table 3). Additionally, we identified 20 epigenetically induced genes whose promoters showed reduced methylation levels and whose expression levels were elevated in tumors (Table 4). Functional analysis revealed that the epigenetically suppressed genes were enriched in normal lung development, gland development, respiratory system development, and vasculature development. Most of these functions are essential for normal lung function and development. In contrast, the epigenetically induced genes were enriched in biological processes critical for tumor progression, survival and motility, differentiation. This suggests that these genes are important for LADC development and progression; however, further molecular studies are needed to confirm this.
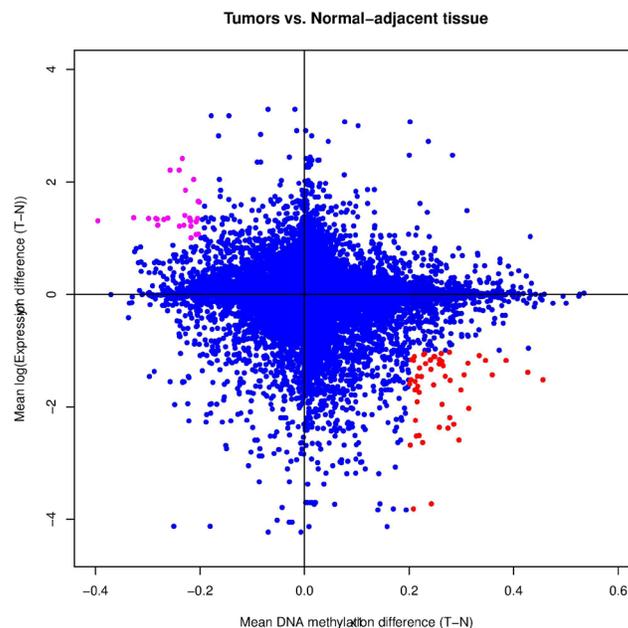

Figure 2. Integrated analysis of gene expression and promoter DNA methylation changes between lung tumors and matched normal adjacent tissues. Mean DNA methylation beta-value differences between lung tumors and matched normal lung tissues are plotted on the x-axis, while mean log2-transformed gene expression value differences are plotted on the y-axis for each gene. Red data points indicate genes that were hypermethylated with a beta-value difference >0.20 and showed more than 2-fold decrease in their gene expression levels in lung tumors. Magenta data points highlight genes that were hypomethylated with a beta-value difference >0.20 and showed more than 2-fold increase in their gene expression levels in lung tumors.

Table 2. Comparison of promoter methylation status with expression level.

mRNA expression (fold change)	Hypermethylation (N = 634)	Hypomethylation (N = 306)
<-0.5	85	24
-0.5-0.5	517	233
>0.5	30	44
Not checked	2	5

P value = 0.3325.

Table 3. Top 20 hypermethylated and down-regulated genes in early-stage LADCs.

Gene	Adjusted P value	Δ_{Exp_NC}	Function
FAM107A	1.4E-05	-3.8	Cell growth suppressor
HBA1	3.96E-09	-3.7	Oxygen transport from the lung to peripheral tissues
MAMDC2	2.2E-09	-2.7	Glycosaminoglycan binding
TEK	4.7E-12	-2.6	Endothelial cell proliferation, differentiation
SPARCL1	3.0E-12	-2.6	Calcium ion binding
CLDN5	3.5E-12	-2.5	Integral membrane protein, tight junction component
TCF21	1.6E-12	-2.5	Tumor suppressor gene
C7	3.5E-11	-2.4	Component of complement system
CLEC14A	2.2E-11	-2.4	Carbohydrate binding
COX7A1	1.6E-12	-2.3	Terminal oxidase in mitochondrial electron transport
HOXA5	3.7E-08	-2.2	Transcription factor; development, up-regulates p53
SOSTDC1	1.5E-11	-2.2	BMP antagonist; Wnt and TGF-beta signaling
CYYR1	1.5E-11	-2.0	Undetermined
JAM2	2.0E-10	-2.0	Tight junctions
SCARA5	1.3E-07	-1.9	Ferritin receptor
HSPA12B	3.4E-09	-1.7	ATP binding
SOCS2	7.0E-05	-1.7	Cytokine transduction; negative regulator in GH/IFG1 signaling
ACVRL1	4.0E-12	-1.6	TGF-beta receptor, serine/threonine kinase
FOXF1	5.4E-10	-1.6	Undetermined
CLEC1A	2.3E-11	-1.6	Regulate dendritic cell function

Table 4. Top 20 hypomethylated and up-regulated genes in early-stage LADCs.

Gene	Adjusted P value	Δ_{Exp_NC}	Function
PROM2	1.6E-12	2.4	Organization of plasma membrane
SFN	1.6E-12	2.2	p53-regulated inhibitor of G2/M progression
SPDEF	1.2E-10	2.1	Androgen-independent transactivator of PSA promoter
KRT8	1.8E-12	1.9	Cellular structure; signal transduction
CFB	4.0E-09	1.7	Component of complement system
PCP4	4.4E-09	1.6	Regulator of calmodulin signaling
SLC22A18AS	3.8E-10	1.4	Undetermined
FAM83A	9.8E-12	1.4	Decreased Wnt reporter activity
HABP2	2.0E-12	1.4	Activates coagulation factor VII
TCN1	1.6E-12	1.4	Vitamin B12-binding protein
TM4SF4	1.3E-11	1.4	Cell proliferation, growth, motility
ABCA4	2.9E-11	1.3	Function as an inward-directed retinoid flippase
MX2	3.2E-11	1.3	Against HIV-1 and simian immunodeficiency virus (SIV-mnd); regulation of nucleocytoplasmic transport and cell-cycle progression
MGC9712	3.6E-12	1.3	Undetermined
ABP1	2.9E-11	1.3	Receptor activity; cation channel activity; sodium channel activity; copper ion binding; calcium ion binding
AGR2	1.9E-12	1.2	Proto-oncogene; cell migration, differentiation and growth
FLJ13841	1.0E-09	1.2	Epidermis differentiation
FLJ10916	3.6E-09	1.1	Decreased TP53 protein express
IGSF9	2.7E-11	1.1	Functions in dendrite outgrowth and synapse maturation
RAB25	1.6E-11	1.0	Cell survival, migration, and proliferation

Sequence and distribution characteristics of cancer-specific DNA methylation gene promoters

We next examined DNA methylation patterns across different genomic regions. We found that hypermethylated CpGs were located mostly in CGI and high-CpG density promoters, while hypomethylated CpGs were typically located outside of the CGI and in intermediate- and low-CpG-density promoters (Figure 3A). Notably, the frequency of methylation loss was highest in intermediate-density promoters, suggesting that weak CpG-density promoters

are more prone to demethylation in carcinogenesis, while high CpG-density promoters are prone to increased methylation in carcinogenesis. Genes with hypermethylated or hypomethylated promoters in LADC were generally found on all chromosomes (Figure 3B). However, there were differences between chromosomes (Figure 3C). For example, chromosome 18 showed the highest rate of hypermethylation, while hypomethylated genes were mostly concentrated on chromosome 21.
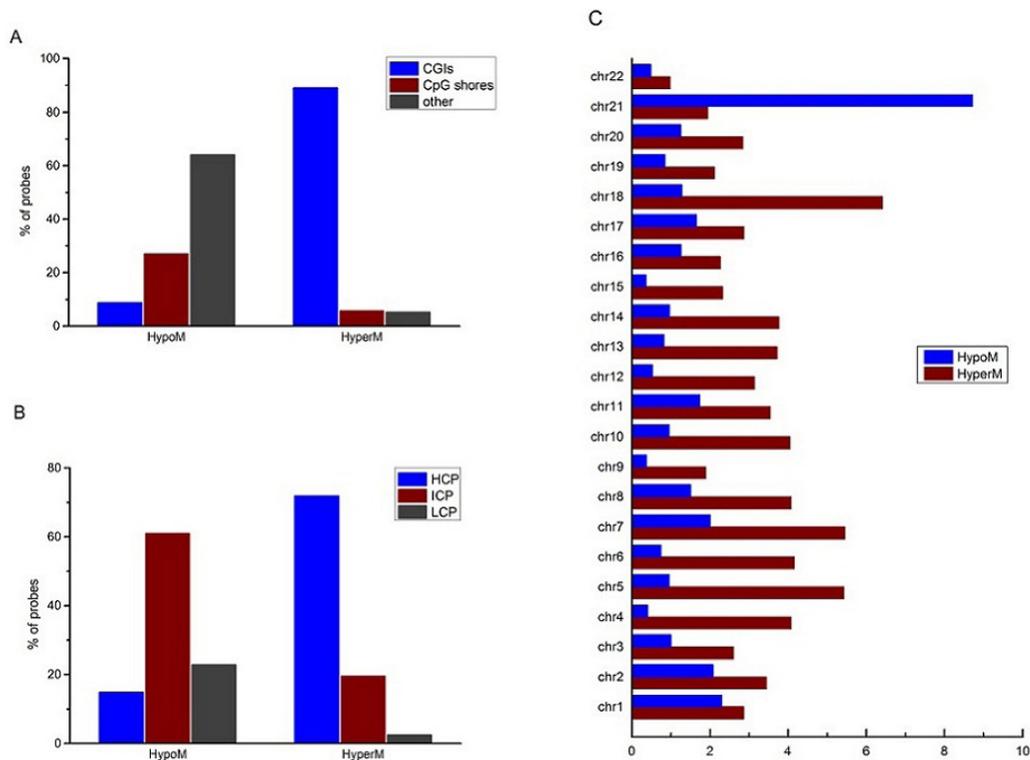

Figure 3. DNA methylation patterns across different genomic regions. **A.** and **B.** CpG location *vs* CGI and CpG location *vs* promoter classes, respectively; **C.** Chromosomal distribution of hypermethylated or hypomethylated promoters in CGI. Percentage indicates number of hypermethylated or hypomethylated CpG sites x 100 / number of total CpG sites located in promoter regions and CpG islands on individual chromosome.

We next explored whether specific sequence motifs or repeat sequences surrounding CGIs play a role in differential DNA hypermethylation, specifically in early-stage LADCs. We carried out *de novo* DNA motif analysis (Segal et al., 2003) on differentially methylated CGIs, as most *de novo* DNA methylation occurred in these regions (Figure 3A). By using non-differentially methylated CGIs as controls, we identified 3 DNA motifs preferentially enriched in hypermethylated CGIs (Figure 4A), including one that resembled the motif bound by the transcription factor Kruppel-like factor 4 (KLF4). Analysis of mRNA expression data showed that KLF4 was down-regulated in early-stage LADCs (Figure 4B). We hypothesize that KLF4 binding to CGIs prevents DNA methylation in LADC, while a loss of KLF4 expression leads to increased DNA methylation.
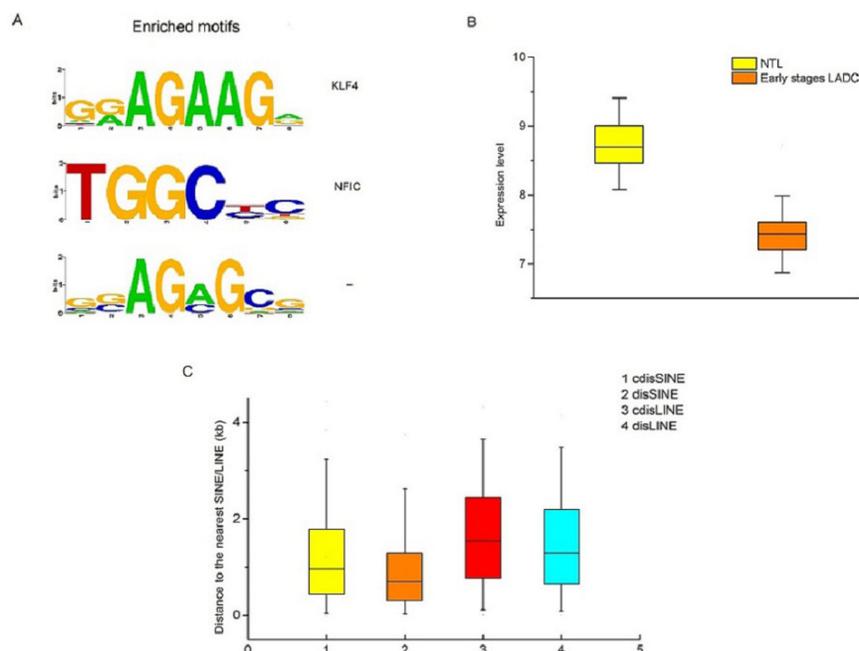
Figure 4. DNA structural and sequence characteristics of DNA methylation profiles in LADC. **A.** Motif analysis of hypermethylated CGIs; **B.** KLF4 mRNA expression between NTLs and early-stage LADCs; **C.** distribution of distances of DNA methylation probes to the nearest ALU and repetitive long interspersed nuclear elements. In each box plot, the top and bottom edges are the 25 and 75th quartiles, respectively. The horizontal line within each box identifies the median.

We found that the distances between the DNA methylation probes and the nearest repetitive short interspersed nuclear elements were significantly different between cancer-specific methylated promoter sequences (median distance: 1014 bp) and those that did not exhibit cancer-specific DNA methylation changes (median distance: 746 bp; P = 9.6E-10, Wilcoxon rank-sum test, Figure 4C), which was consistent with the results of a previous study (Estécio et al., 2010). Similarly, we found that cancer-specific methylated promoter sequences showed a greater median distance to repetitive long interspersed nuclear elements compared with those that did not show cancer-specific DNA methylation changes (1559 *vs* 1330 bp; P = 1.79E-5, Wilcoxon rank-sum test, Figure 4C).

LADC-associated DNA hypomethylation occurs preferentially at domains that are characterized by neither histone H3 lysine 4 (H3K4me3) nor histone H3 lysine 27 (H3K27me3) in human embryonic stem cells (hESCs)

Several studies have examined the relationship between cancer-specific DNA methylation and some stem cell loci, such as PCGTs (Ohm et al., 2007; Widschwendter et al., 2007). Thus, we explored the dynamics of epigenetic changes at stem cell loci throughout all stages of LADC and investigated their potential roles in predicting poor prognosis. First, we examined the LADC-specific DNA methylation distribution with respect to trimethylation status of the H3K4me3 and H3K27me3 chromatin domains in hESCs using a previously published

data set (Ku et al., 2008). We found that the genes hypermethylated in early-stage LADC were highly enriched in embryonic stem cell bivalent domains, while those hypomethylated in early-stage LADC were enriched for domains whose chromatin states with NMDs (Figure 5A). When we explored the dynamics of epigenetic changes at NMDs in hESCs throughout all stages of LADC, we found that CpGs in NMDs were increasingly hypomethylated with cancer progression (P = 7.9E-5 and 2.5E-4 for NTLs *vs* early-stage LADCs and early-stage LADCs *vs* late-stage LADCs, respectively; Figure 5B). Although no gene showed a significant difference in methylation between early-stage and late-stage LADCs, the mean methylation level of NMDs decreased with tumor progression. This suggests that LADC-associated DNA hypomethylation occurs preferentially at NMDs. Moreover, when we compared the DNA methylation of lung cancer-associated fibroblasts to normal lung fibroblasts (Navab et al., 2011), we found that 37 of 77 hypomethylated CpGs in lung cancer fibroblasts were located in NMDs. This further supports that hypomethylation of CpGs in NMDs is an important characteristic of LADC progression, and that it may be a molecular determinant of clinical outcome.
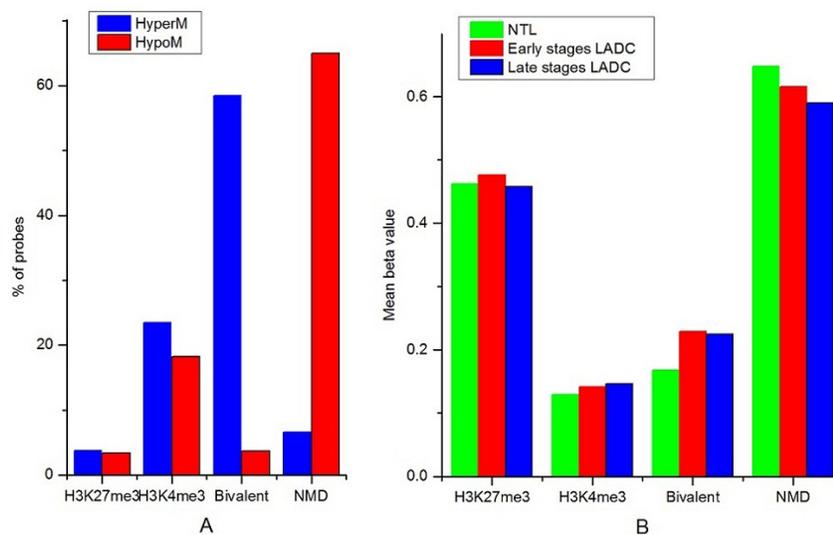

Figure 5. ESC histone marks associated with cancer-specific DNA methylation genes in LADC. **A.** Bar charts show the proportion of gene promoters with occupancy of H3K4me3, H3K27me3, bivalent, and NMDs in hESCs in early-stage LADCs; **B.** dynamics of hESC histone mark domain methylation level in LADCs.

To test the prognostic potential of NMD hypomethylation, we analyzed 2 independent data sets (Table 1, validation set 3 and validation set 4) by using the demethylation instability index (DeMI), which has been described previously (Zhuang et al., 2012). Briefly, we first defined the CpGs as unmethylated (0) (beta-value <0.25), hemimethylated (1) ($0.25 \leq$ beta-value ≤ 0.7), or methylated (2) (beta-value >0.7). Next, we selected stable CpGs with the same methylation state (beta-value >0.7) in all NTLs. For each tumor sample, we computed the DeMI, which reflected the fraction of stable CpGs in NTLs hypomethylated in tumor tissues (beta-value ≤ 0.7). As shown in Figure 6A and B, survival analysis revealed a significant difference between the clinical outcomes of the upper and lower tertiles of the DeMI at NMDs, with a high DeMI at NMDs corresponding to a worse prognosis.
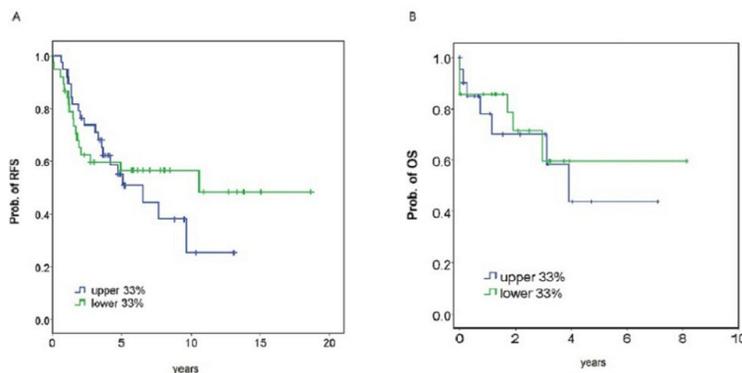
Figure 6. Survival analysis of NMD demethylation instability index in validation of the LADC cohort. Kaplan-Meier survival curves between the upper (blue) and lower (green) tertiles of the DeMI at NMDs in validation set 3 (**A**) and validation set 4 (**B**), respectively. Clinical endpoint used is indicated on the y-axis (OS = overall survival, RFS = relapse-free survival).

To further demonstrate that MESC hypomethylation is an important determinant of a poor outcome in LADC, we tested whether these epigenetic changes progress further in metastatic LADC. Next, we compared the DeMI of primary LADC to that of metastatic LADC and found that the demethylation frequency (from methylated state in NTL to hemimethylated or unmethylated state in LADC and from hemimethylated state in NTL to unmethylated state in LADC) of NMDs was higher in metastatic LADCs than in primary LADCs (Figure 7). These results demonstrate that MND hypomethylation is a progressive process and a key characteristic of advanced LADC.
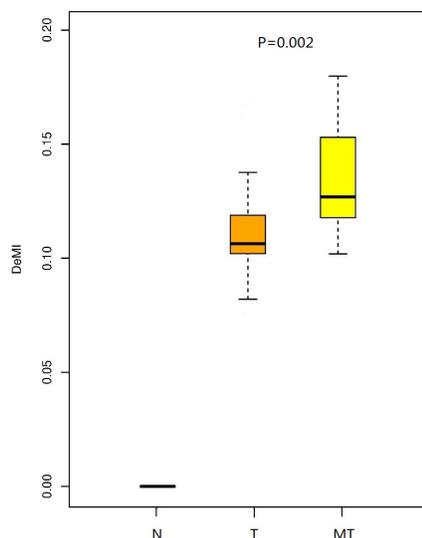

Figure 7. Methylation changes between normal, primary, and metastatic LADCs. Boxplots comparing the frequency of NMD demethylation between normal lung tissues (N) and primary LADCs (T) and between normal lung tissues (N) and metastatic LADCs (MT). Wilcoxon rank sum test P values for the DeMI between primary LADCs and metastatic LADCs is indicated.

DISCUSSION

One of the hallmarks of cancer is aberrant DNA methylation; studying the role of DNA methylation in cancer is considered to be a new and promising approach for cancer therapy. In this study, we analyzed the different DNA methylation states of early-stage LADCs *vs* NTLs and early-stage LADCs *vs* late-stage LADCs. We found that DNA methylation changes mainly occur during early-stage in LADC, while DNA methylation level in early-stage LADC and late-stage LADC were genetically highly similar. This suggests that essential methylation changes required for LADC progression occur during the early-stage before initiation of tumor severity, in contrast to histone modification (Wen et al., 2011; Song et al., 2012).

By integrating the DNA methylation data with gene expression profiles for the same tumors, we found 47 epigenetically suppressed genes whose promoters showed increased methylation levels and whose expression levels were reduced in cancer as well as 20 epigenetically induced genes whose promoters showed reduced methylation levels and whose expression levels were elevated in cancer. The concordance between methylation alterations and expression changes was well under 10% (67/940). There may be 2 explanations for this. First, hypomethylation or hypermethylation may not be sufficient for these genes to be expressed or repressed in response to specific triggers. This likely reflects the fact that many epigenetic and genetic alterations in cancers are passenger events that are not important in the pathogenesis of cancer (Maley et al., 2004). Second, it may be because of the molecular pathogenesis of lung adenocarcinomas, and additional studies are needed to examine this. Among these 67 epigenetically induced or suppressed genes, some genes were previously reported to be related with lung cancer. TCF21, which regulates the mesenchymal cell transition into epithelial cells, has been demonstrated previously to be silenced in most head and neck squamous cell carcinomas and non-small-cell lung cancer tumor cells (Smith et al., 2006). SOSTDC1 has been reported previously to regulate the Wnt receptor signaling pathway, while Wnt has been implicated in skin cancer, lung cancer, bladder cancer, and leukemia, and may be oncogenic in other cancers (Reya and Clevers, 2005; Barker and Clevers, 2006; Polakis, 2007). SFN, which functions as an adapter protein implicated in the regulation of a large spectrum of both general and specialized signaling pathways, may also regulate MDM2 auto-ubiquitination and degradation and thereby activate p53/TP53. SFN has been reported to be overexpressed in early invasive lung adenocarcinoma (Shiba-Ishii et al., 2011). FAM83A, also known as tumor antigen BJ-TSA-9, is a potential biomarker in lung cancer (Li et al., 2005).

Moreover, we also identified several novel epigenetically suppressed genes. FAM107A, which was down-regulated in renal cell carcinoma, FOXF1, which was down-regulated in hepatocellular carcinoma, and SLIT3, whose down-regulation may promote invasion and progression of thyroid cancer, were the first reported epigenetic deregulation genes in early-stage LADC. These genes may be useful for developing biomarkers for diagnostic or prognostic purposes in early-stage LADC.

Hypermethylation in LADC was mostly located in high-CpG density promoters and CGIs, while hypomethylation promoters were enriched in intermediate-density promoters and outside CGIs (Figure 3A), which agrees with the results of previous studies (Dedeurwaerder et al., 2011). Our motif analysis indicated that KLF4 DNA motifs were overrepresented at hypermethylated CGIs in early-stage LADC. KLF4 is a member of the KLF family of transcription factors and regulates the proliferation, differentiation, apoptosis, and somatic cell reprogramming. Evidence also suggests that KLF4 is a tumor suppressor in certain cancers, such as

colorectal cancer (El-Karim et al., 2013). Our analyses also showed that the KLF4 expression level was decreased in LADC. These data suggest that downregulation of KLF4 enables DNA methylation to accumulate at CGIs, perhaps by recruiting CGI-binding protein complexes that include DNA methyltransferase.

Interestingly, we found that promoters on chromosome 18 were most frequently hypermethylated. A previous study showed that the genes on chromosome 18 were most frequently down-regulated in rectal cancer (Lips et al., 2008). In addition, loss of chromosome 18q was found to be associated with the progression of prostate cancer (Padalecki et al., 2003) and the loss of chromosome 18 occurs during early-stage colorectal carcinogenesis (Cardoso et al., 2007). Although more extensive studies are required to investigate the role of chromosome 18 in LADC, our results suggest that aberrant methylation of genes cooperate with genetic alterations to drive the initiation and progression of LADC.

Another important finding in our study is that NMD hypomethylation occurs progressively in LADC, increasing significantly in advanced LADCs. This suggests that the level of NMD hypomethylation in primary LADCs is an important determinant of clinical outcome (Figure 7). This is the first report of a prognostic signature in lung cancer that is based on NMD DNA methylation. When we compared the methylation level of NMDs in fetal lung and adult lung, we found no statistically significant difference between these groups (data not shown), but in tumorigenesis, NMD was hypomethylated. This methylation level decreased with tumor progression, and we hypothesized that hypomethylation of NMDs act as activators of tumor-associated gene activation.

There were some limitations to our study, including the insufficiency of some clinical characteristics and the lack of some experimental conditions. We found no significant associations between methylation and tumor stage, gender, and age, except for a trend in the differences associated with the histological grade of tumors. However, hypomethylation of NMDs in LADC may serve both as a factor responsible for the substantial variability in treatment response among patients and as a tool for assessing tumor aggressiveness.

In summary, we analyzed the DNA methylation profiles of different-stage LADCs and matched adjacent NTLs and reached 3 main conclusions: 1) we identified numerous new LADC-specific DNA methylation changes that may be used as potential LADC biomarkers; 2) we identified characteristic sequence signatures in LADC-associated CpG islands; and 3) from a clinical perspective, the observation that DNA methylation decreases in NMDs with LADC progression provides a new opportunity for detecting and diagnosing advanced LADC and determining clinical outcome.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Anglim PP, Alonzo TA and Laird-Offringa IA (2008). DNA methylation-based biomarkers for early detection of non-small cell lung cancer: an update. *Mol. Cancer* 7: 81.

Barker N and Clevers H (2006). Mining the Wnt pathway for cancer therapeutics. *Nat. Rev. Drug Discov*. 5: 997-1014.

Barrett T, Troup DB, Wilhite SE, Ledoux P, et al. (2009). NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res*. 37: D885-D890.

Baylin SB and Jones PA (2011). A decade of exploring the cancer epigenome - biological and translational implications. *Nat. Rev. Cancer* 11: 726-734.

Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, et al. (2009). Genome-wide DNA methylation profiling using Infinium assay. *Epigenomics* 1: 177-200.

Buckingham L, Penfield Faber L, Kim A, Liptay M, et al. (2010). PTEN, RASSF1 and DAPK site-specific hypermethylation and outcome in surgically treated stage I and II nonsmall cell lung cancer patients. *Int. J. Cancer* 126: 1630-1639.

Cardoso J, Boer J, Morreau H and Fodde R (2007). Expression and genomic profiling of colorectal cancer. *Biochim. Biophys. Acta* 1775: 103-137.

Dedeurwaerder S, Desmedt C, Calonne E, Singhal SK, et al. (2011). DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Mol. Med*. 3: 726-741.

Divine KK, Pulling LC, Marron-Terada PG, Liechty KC, et al. (2005). Multiplicity of abnormal promoter methylation in lung adenocarcinomas from smokers and never smokers. *Int. J. Cancer* 114: 400-405.

Du P, Kibbe WA and Lin SM (2008). lumi: A pipeline for processing Illumina microarray. *Bioinformatics* 24: 1547-1548.

El-Karim EA, Hagos EG, Ghaleb AM, Yu B, et al. (2013). Krüppel-like factor 4 regulates genetic stability in mouse embryonic fibroblasts. *Mol. Cancer* 12: 89.

Estécio MR, Gallegos J, Vallot C, Castoro RJ, et al. (2010). Genome architecture marked by retrotransposons modulates predisposition to DNA methylation in cancer. *Genome Res*. 20: 1369-1382.

Feng Q, Hawe SE, Stern JE, Wiens L, et al. (2008). DNA methylation in tumor and matched normal tissues from non-small cell lung cancer patients. *Cancer Epidemiol. Biomarkers Prev*. 17: 645-654.

Jones PA and Baylin SB (2007). The epigenomics of cancer. *Cell* 128: 683-692.

Ku M, Koche RP, Rheinbay E, Mendenhall EM, et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet*. 4: e1000242.

Lee TI, Jenner RG, Boyer LA, Guenther MG, et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125: 301-313.

Li Y, Dong X, Yin Y, Su Y, et al. (2005). BJ-TSA-9, a novel human cancer-specific gene, has potential as a biomarker of lung cancer. *Neoplasia* 7: 1073-1080.

Lips EH, van Eijk R, de Graaf EJ, Oosting J, et al. (2008). Integrating chromosomal aberrations and gene expression profiles to dissect rectal tumorigenesis. *BMC Cancer* 8: 314.

Maley CC, Galipeau PC, Li X, Sanchez CA, et al. (2004). Selectively advantageous mutations and hitchhikers in neoplasms: p16 lesions are selected in Barrett's esophagus. *Cancer Res*. 64: 3414-3427.

Navab R, Strumpf D, Bandarchi B, Zhu CQ, et al. (2011). Prognostic gene-expression signature of carcinoma-associated fibroblasts in non-small cell lung cancer. *Proc. Natl. Acad. Sci. U. S. A*. 108: 7160-7165.

Ohm JE, McGarvey KM, Yu X, Cheng L, et al. (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat. Genet*. 39: 237-242.

Padalecki SS, Weldon KS, Reveles XT, Buller CL, et al. (2003). Chromosome 18 suppresses prostate cancer metastases. *Urol. Oncol*. 21: 366-373.

Polakis P (2007). The many ways of Wnt in cancer. *Curr. Opin. Genet. Dev*. 17: 45-51.

Razin A and Riggs AD (1980). DNA methylation and gene function. *Science* 210: 604-610.

Reya T and Clevers H (2005). Wnt signalling in stem cells and cancer. *Nature* 434: 843-850.

Sandoval J, Mendez-Gonzalez J, Nadal E, Chen G, et al. (2013). A prognostic DNA methylation signature for stage I non-small-cell lung cancer. *J. Clin. Oncol*. 31: 4140-4147.

Segal E, Yelensky R and Koller D (2003). Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. *Bioinformatics* 19 (Suppl 1): i273-i282.

Selamat SA, Chung BS, Girard L, Zhang W, et al. (2012). Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res*. 22: 1197-1211.

Shiba-Ishii A, Kano J, Morishita Y, Sato Y, et al. (2011). High expression of stratifin is a universal abnormality during the course of malignant progression of early-stage lung adenocarcinoma. *Int. J. Cancer* 129: 2445-2453.

Shiraishi M, Sekiguchi A, Oates AJ, Terry MJ, et al. (2002). HOX gene clusters are hotspots of de novo methylation in CpG islands of human lung adenocarcinomas. *Oncogene* 21: 3659-3662.

Smith LT, Lin M, Brena RM, Lang JC, et al. (2006). Epigenetic regulation of the tumor suppressor gene TCF21 on 6q23-q24 in lung and head and neck cancer. *Proc. Natl. Acad. Sci. U. S. A*. 103: 982-987.

Song JS, Kim YS, Kim DK, Park SI, et al. (2012). Global histone modification pattern associated with recurrence and disease-free survival in non-small cell lung cancer patients. *Pathol. Int*. 62: 182-190.

Tsou JA, Galler JS, Siegmund KD, Laird PW, et al. (2007). Identification of a panel of sensitive and specific DNA methylation markers for lung adenocarcinoma. *Mol. Cancer* 6: 70.

Wen J, Fu J, Zhang W and Guo M (2011). Genetic and epigenetic changes in lung carcinoma and their clinical implications. *Mod. Pathol*. 24: 932-943.

Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, et al. (2007). Epigenetic stem cell signature in cancer. *Nat. Genet*. 39: 157-158.

Zhuang J, Jones A, Lee SH, Ng E, et al. (2012). The dynamics and prognostic potential of DNA methylation changes at stem cell gene loci in women's cancer. *PLoS Genet*. 8: e1002517.