

RESEARCH

Open Access



Comparison of tumor and two types of paratumoral tissues highlighted epigenetic regulation of transcription during field cancerization in non-small cell lung cancer

Qiushi Wang¹, Libo Wu¹, Jiaying Yu¹, Guanghua Li¹, Pengfei Zhang¹, Haozhe Wang², Lin Shao², Jinying Liu² and Weixi Shen^{3*}

Abstract

Background: Field cancerization is the process in which a population of normal or pre-malignant cells is affected by oncogenic alterations leading to progressive molecular changes that drive malignant transformation. Aberrant DNA methylation has been implicated in early cancer development in non-small cell lung cancer (NSCLC); however, studies on its role in field cancerization (FC) are limited. This study aims to identify FC-specific methylation patterns that could distinguish between pre-malignant lesions and tumor tissues in NSCLC.

Methods: We enrolled 52 patients with resectable NSCLC and collected resected tumor (TUM), tumor-adjacent (ADJ) and tumor-distant normal (DIS) tissue samples, among whom 36 qualified for subsequent analyses. Methylation levels were profiled by bisulfite sequencing using a custom lung-cancer methylation panel.

Results: ADJ and DIS samples demonstrated similar methylation profiles, which were distinct from that of TUM. Comparison of TUM and DIS profiles led to identification of 1740 tumor-specific differential methylated regions (DMRs), including 1675 hypermethylated and 65 hypomethylated (adjusted $P < 0.05$). Six of the top 10 tumor-specific hypermethylated regions were associated with cancer development. We then compared the TUM, ADJ, and DIS to further identify the progressively aggravating aberrant methylations during cancer initiation and early development. A total of 332 DMRs were identified, including a predominant proportion of 312 regions showing stepwise increase in methylation levels as the sample drew nearer to the tumor (i.e. $DIS < ADJ < TUM$) and 20 regions showing a stepwise decrease pattern. Gene set enrichment analysis (GSEA) for KEGG and GO terms consistently suggested enrichment of DMRs located in transcription factor genes, suggesting a central role of epigenetic regulation of transcription factors in FC and tumorigenesis.

Conclusion: We revealed distinct methylation patterns between pre-malignant lesions and malignant tumors, suggesting the essential role of DNA methylation as an early step in pre-malignant field defects. Moreover, our study also identified differentially methylated genes, especially transcription factors, that could potentially be used as markers for lung cancer screening and for mechanistic studies of FC and early cancer development.

Keywords: NSCLC, Field cancerization, DNA methylation, Bisulfite sequencing, Epigenetics

*Correspondence: Shenweixi1982@163.com

³ Department of Medical Oncology, The Second Affiliated Hospital of Harbin Medical University, Harbin 150001, China

Full list of author information is available at the end of the article



Background

Lung cancer is the leading cause of cancer deaths worldwide. Its high mortality is partially attributable to the scarce knowledge of molecular mechanisms mediating lung cancer pathogenesis and the late diagnosis of the majority of lung cancers [1]. Non-small cell lung cancer (NSCLC), representing the majority of diagnosed lung cancers, is a complex malignancy that develops through progressive pathologic changes driven by an interplay of a variety of molecular pathways including both genetic and epigenetic mechanisms [2]. Therefore, mapping genetic and epigenetic changes in normal tissue at high risk of malignant transformation is critically important for understanding the mechanism of carcinogenesis, identifying early causal drivers and predicting cancer risk.

Field cancerization (FC), also referred to as pre-malignant field defect, is the process in which a population of normal or pre-malignant cells is affected by oncogenic alterations leading to progressive molecular changes that drive their malignancy [2, 3]. The acquirement of tumor-primed genetic alterations (such as *EGFR* and *KRAS* mutations, loss of heterogeneity of chromosomal regions 3p and 9p, and genomic instability) has been described in histologically normal bronchial epithelia adjacent to the lung carcinoma [4–6]. On the other hand, DNA methylation is a primary epigenetic modification in the mammalian genome. Aberrant DNA methylation has been implicated in early cancer development, including lung cancer [7]. Belinsky and colleagues reported aberrant promoter methylation of p16, which commonly occurred in lung tumors [8], in bronchial epithelial sites from 44% of lung cancer patients and cancer-free smokers [9]. The aberrant methylation of various frequently methylated genes in lung cancer, including retinoic acid receptor 2 β (*RAR- β 2*), H-cadherin, adenomatous polyposis coli (*APC*), and Ras association domain family member 1 (*RASSF1A*), has also been described in bronchial epithelial cells of heavy smokers [10]. Despite a number of studies [11, 12] have suggested the phenomenon of epigenetic FC in lung cancer, most of them interrogated the methylation profile in pre-malignant lesions (such as basal cell hyperplasia, squamous metaplasia, dysplasia), lacking appropriated subject-matched controls for both normal and malignant tissues. Furthermore, the vast majority of these studies focused on a limited number of candidate genes, the methylation of which had been often observed in lung cancer; the methylation profile was often assessed qualitatively not quantitatively. These limitations would attenuate the measured magnitude of epigenetic differences and inhibit the ability to identify the earliest methylation alterations that occur in carcinogenesis. Given the ubiquitous inter-patient heterogeneity, the extent to

which DNA methylation profiles modify the FC effect of individuals may be largely obscured by such study design.

In the present study, we aimed to identify tumor-specific methylation patterns that could distinguish between pre-malignant normal lesions and tumor tissues and could potentially be developed as biomarkers, using subject-matched surgically-resected tumor, tumor-adjacent normal (ADJ) and tumor-distant normal (DIS) tissue samples.

Methods

Patients' information and study design

A total of 52 patients with early-stage resectable NSCLCs from 2018 and 2019 were enrolled in this study. Matched surgically-resected tumor, tumor-adjacent normal (ADJ) and tumor-distant normal (DIS) tissue samples were collected from each patient during surgery. ADJ tissues were biopsied 2 cm distant from resection margins, which allowed for both proximity to the tumor and a low chance of tumor cell contamination in case of a R2 resection margin; DIS samples were biopsied 5 cm distant from resection margins to make sure sample collection was feasible for centrally located stage T4 tumors treatment with surgery of curative intent, which usually sets the margin ~5 cm outside of the tumor zone. Samples underwent histopathological assessment. Tumor tissues with a tumor cell fraction < 10% and normal tissues with any visible tumor cell were excluded. Eventually, thirty-six patients had all three types of samples subjected to bisulfite DNA sequencing for methylation profiling that was used for subsequent analyses to identify differential methylation signatures. The histopathological and clinical characteristics of patients were collected. The study was approved by the institutional review board of The Second Affiliated Hospital of Harbin Medical University. All patients provided written informed consent, in accordance with the Declaration of Helsinki.

DNA isolation

Genomic DNA were extracted tissue samples using a QIAamp DNA FFPE tissue kit, according to the manufacturer's standard protocol (Qiagen, Hilden, Germany). DNA was quantified using the Qubit dsDNA assay (Life Technologies, Carlsbad, CA, USA).

Bisulfite targeted sequencing

DNA was sequenced using a brELSATM method as described previously [13]. Briefly, purified DNA was converted to single-strand DNA by sodium bisulfite treatment. The converted single-strand DNA was subsequently ligated to a splinted adapter, and amplified by a uracil-tolerating DNA polymerase to generate whole-genome bisulfite sequencing (BS-seq) libraries. Target

enrichment was performed using custom-designed lung-cancer methylation profiling RNA baits covering 80,672 CpG sites, spanning 1.05 megabases of the human genome (Burning Rock Biotech, Guangzhou, China). The target libraries were finally quantified by real-time PCR (Kapa Biosciences, Wilmington, MA, USA) and sequenced on a NovaSeq 6000 (Illumina, San Diego, CA, USA) using 2×150 bp cycles. The targeted methylation panel was designed as previously described [14, 15]. Briefly, many differentially methylated loci (DMLs) were selected from the 450 K microarray data of NSCLC/ adjacent tissue and normal plasma samples downloaded from TCGA dataset.

Data analysis

Bisulfite sequencing data analysis was performed using an optimized pipeline. Trimmomatic (v.0.32) was used to remove custom adaptor sequences and low-quality bases. Paired-end reads were aligned to C to T- and G to A-transformed hg19 genome using BWA-meth (v.0.2.2) [16]. After alignment, duplicate reads were marked by sambalster (v.0.1.20) [17], and low mapping quality (MAPQ < 20) or improper pairing reads were removed by sambamba (v.0.4.7) [18] from downstream analyses. Paired reads were merged by clipping overlapping reads to avoid double-counting of methylation calls.

Identification of differential methylation regions (DMRs)

The 80,672 CpG sites included in the panel were grouped into 8312 methylation blocks using an algorithm as described previously [15]. Specifically, we applied a region-defined algorithm with co-methylation effect

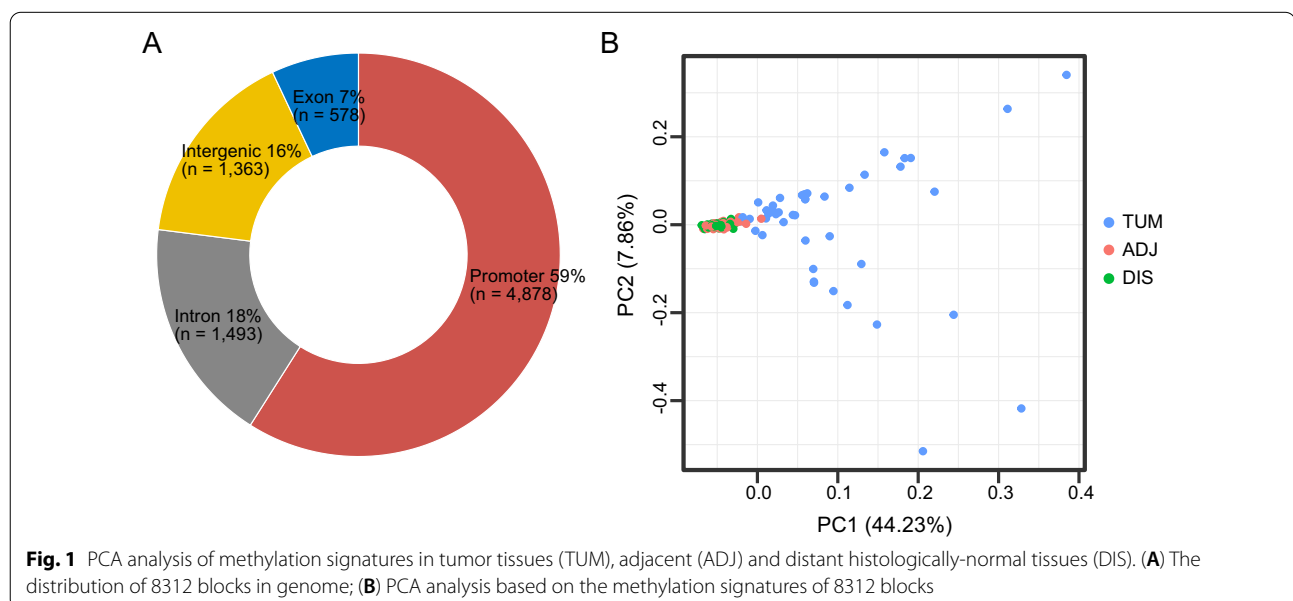
between adjacent CpG sites in consideration [14]. To estimate the predefined coefficients of the algorithm, we used a series of methylation data of different tissues with the same panel mentioned in this study. Methylation blocks were defined as the genomic region consisting of the neighboring CpG which were not only close on distance but also correlated on methylation level. Briefly, the difference among the methylation frequencies of each pair of CpG sites was calculated by Pearson's correlation analysis, and normalized by the difference in genomic distance and methylation level. Within 8312 blocks, 84% were annotated in genes with 59% in promoter regions, 7% in exons and 18% in introns (Fig. 1A).

A block-wise statistic methylMean was generated for downstream analyses. Besides the CpG sites, the information of CHH (H denotes A, T or C) sites were also included to estimate the background error in sequencing which would help to correct the methylMean value of CpG sites. We denote M_j/U_j as the number of methylated/unmethylated read counts for the j th CpG sites within a methylation block. And Me_k/Ue_k denotes the number of methylated/unmethylated read counts for the k th CHH sites within a methylation block. The corrected methylMean was defined as

$$\text{methylMean} = \frac{\sum_j M_j}{\sum_j (M_j + U_j)}$$

$$\text{error} = \frac{\sum_j Me_j}{\sum_j (Me_j + Ue_j)}$$

$$\text{corrected methylMean} = \frac{\text{methylMean} - \text{error}}{1 - \text{error}}$$



The differential methylated regions (DMRs) were identified by comparing corrected methylMean values of blocks between different groups using “limma” package in R software. Blocks with significant difference (threshold $\text{abs}(\log_2\text{FC}) > 0.1$, adjust $p\text{-value} < 0.05$) were chosen. Bonferroni correction was applied to for multiple comparisons. The volcano plot and heatmap were drawn using R software.

Functional enrichment analyses

Gene Set Enrichment Analysis (GSEA) [19, 20] was performed for the functional annotation of DMRs using the Molecular Signatures Database (version 7.4) [21]. For KEGG terms, *c2.cp.kegg.v7.4.entrez.gmt* and *c2.cp.v7.4.entrez.gmt* were used separately [22]. Gene Ontology (GO) Enrichment Analysis was also performed for DMRs [23]. The cut-off of two-sided adjusted p value (i.e. false discovery rate) was set to 0.05.

Statistical analysis

Statistical analysis was performed using R version 3.3.3 software. Principal component analysis (PCA) [24] and hierarchical clustering analysis were performed for clustering samples according to their methylation profiles using all 8312 blocks or tumor specific block. Differential methylation analysis was performed with the “limma” package. Differences were evaluated with Fisher’s exact test for proportions of categorical variables across groups, with Pearson’s correlation analysis for 2 continuous variables, and with paired Student’s t -test for DNA methylation levels between 2 groups and multiple paired t -test for 3 groups. For other continuous variables between 2 groups, the Wilcoxon rank sum test was used for comparison, and ANOVA was performed for continuous variables across 3 groups. Statistical significance was defined as two-sided P values < 0.05 .

Results

Demographic and clinicopathological characteristics of patients

Three types of tissue samples, including surgically-resected tumor (TUM), tumor-adjacent normal (ADJ) and tumor-distant normal tissue (DIS) were collected from 52 enrolled NSCLC patients. Among them, 36 generated sequencing data with sufficient quality for all 3 sample types samples and therefore underwent further analyses. The demographic and clinicopathological characteristics of the 36 patients were summarized in Table 1. The median age of the cohort was 58.6 years, ranging from 30 to 73. Male and female patients comprised 52.8% and 47.2% of the cohort, respectively. Of the 36 patients, 16 (44.4%) had no smoking history, and 6 (16.7%) and 14 (38.9%) patients were former and

Table 1 Characteristics of the 36 patients with qualified bisulfite sequencing data for matched TUM, ADJ and DIS samples

Characteristic	No. of patients (%)
Age, years (Median [Range])	58.6 [30–73]
Sex	
Male	19 (52.8)
Female	17 (47.2)
Smoking status	
Never	16 (44.4)
Former	6 (16.7)
Current	14 (38.9)
Tumor size (cm ²)	
< 5	10 (27.8)
5–9	12 (33.3)
10–20	11 (30.6)
> 20	3 (8.3)
Histology	
ADC	18 (50.0)
SCC	7 (19.4)
Others	11 (30.6)
T stage	
T1	12 (33.3)
T2	11 (30.6)
T3	6 (16.7)
T4	6 (16.7)
Unknown	1 (2.8)
N stage	
N0	22 (61.1)
N1	10 (27.8)
N2	3 (8.3)
Unknown	1 (2.8)

The 52 enrolled patients underwent exclusion by tumor cell fraction within their TUM, ADJ, and DIS samples, and 36 were eligible for subsequent analyses. ADJ—tumor-adjacent normal tissue. DIS—tumor-distant normal tissue. TUM—surgically-resected tumor

current smokers, respectively. Ten patients (27.8%) had their tumors measuring < 5 cm²; 12 (33.3%) had tumors measuring 5–9 cm²; and the tumor size of 11 patients (30.6%) ranged from 10 to 20 cm². Only 3 patients had tumors > 20 cm². The majority of the patients (50%) had adenocarcinomas, 19.4% were diagnosed with squamous cell carcinomas, and 30.6% with tumors of other histology. The T stage and N stage were also summarized in Table 1.

Distinct methylation profile of tumor tissues

PCA analysis was first performed based on 8312 blocks and demonstrated the distinct methylation profile of TUM as compared to both ADJ and DIS tissues

(Fig. 1B). Meanwhile, ADJ and DIS tissues had a similar methylation profile. Furthermore, heterogeneous methylation profiles were observed within tumor tissues.

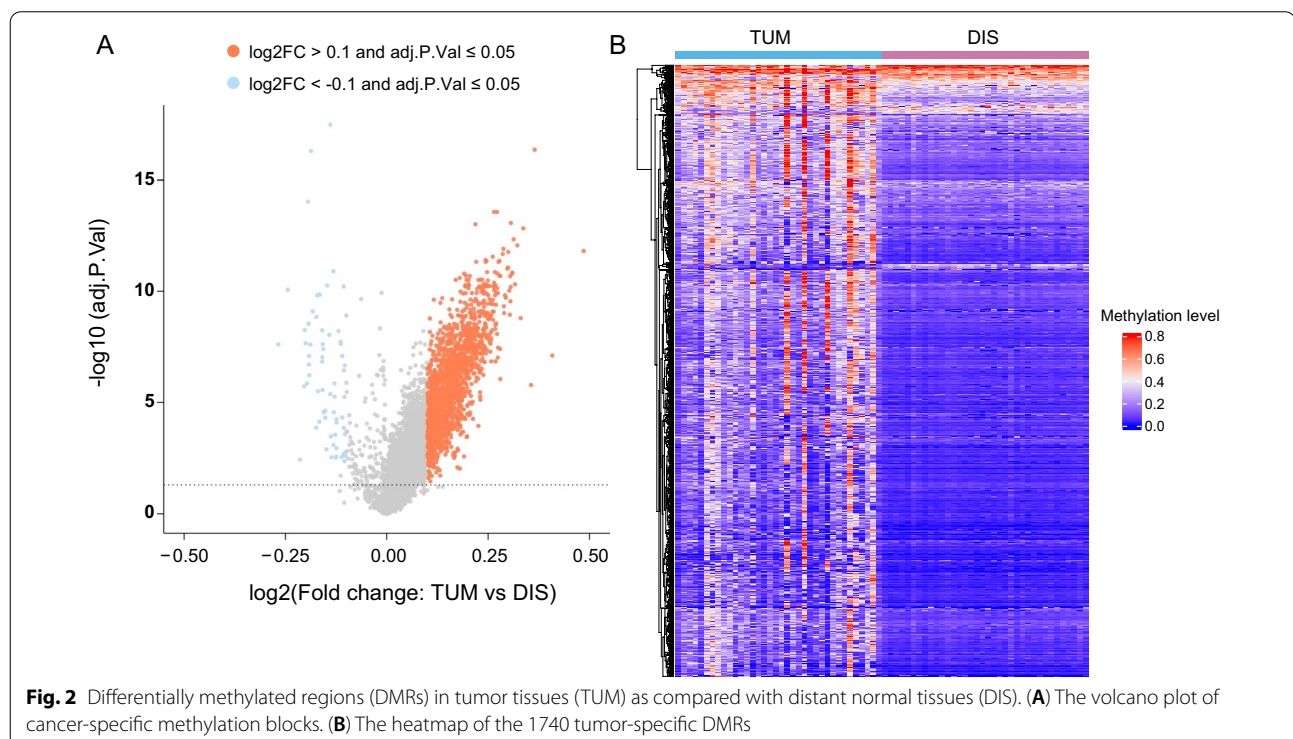
A total of 1740 tumor-specific DMRs, including 1675 hypermethylated and 65 hypomethylated DMRs, spanning 626 genes were found to be differentially methylated in TUM as compared to DIS (abs (log₂FC) > 0.1, adjusted p-value < 0.05, Fig. 2, Additional file 1: Table S1). Six of the top 10 differentially hypermethylated genes have been associated with lung cancer (Table 2), including *BARHL2*, *DMRTA2*, *OTX1*, *OTX2*, *MIR124* and *HOXA9*.

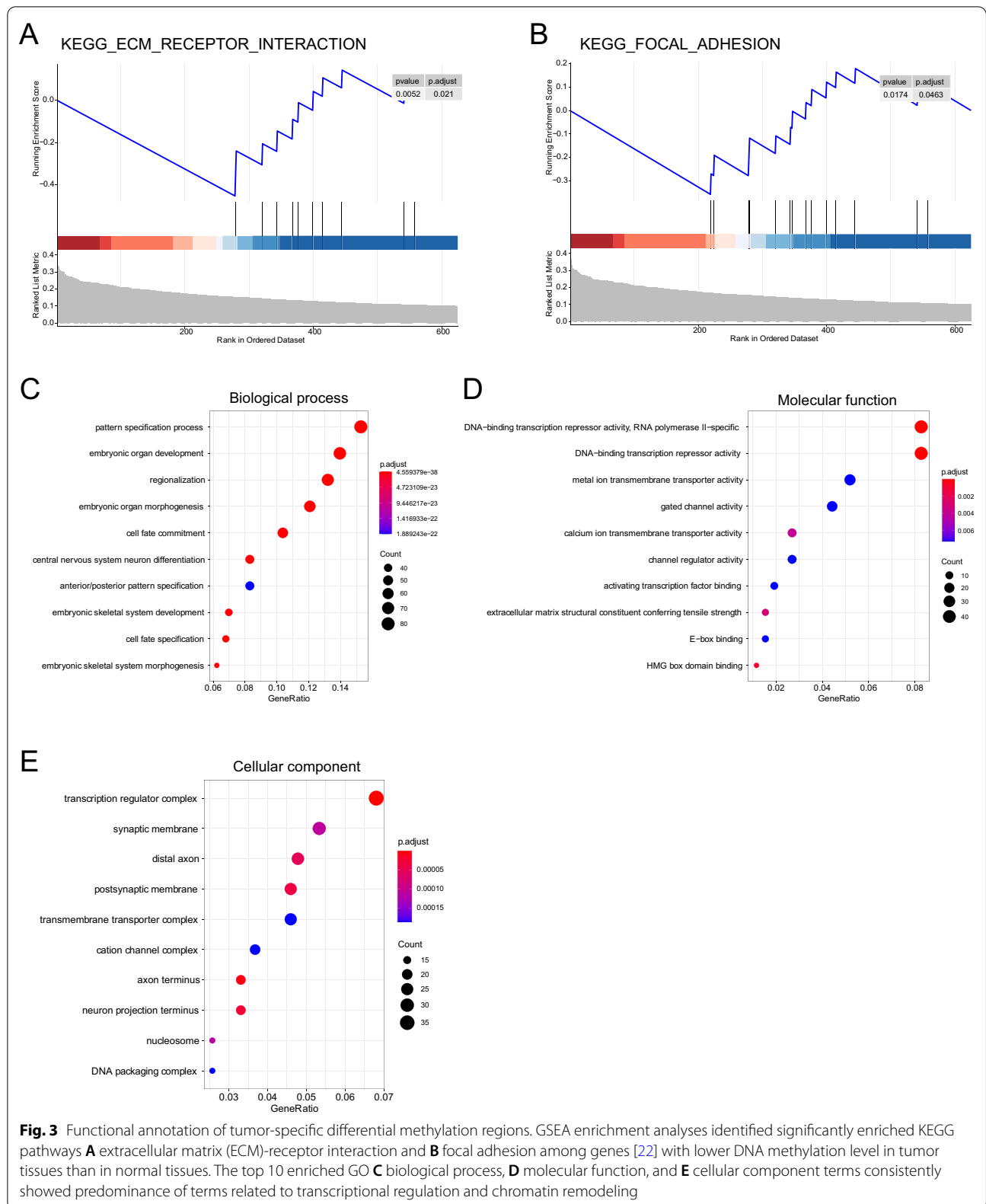
Next, we performed both GSEA and GO enrichment analyses for the functional annotation of the DMRs. GSEA analysis demonstrated that genes in the extracellular matrix (ECM)-receptor interaction (NES = -1.89, P = 0.005, adjusted P = 0.021) and focal adhesion (NES = -1.67, P = 0.017, adjusted P = 0.046) pathways were less commonly methylated in tumor tissues than in normal tissues (Fig. 3A, B; Additional file 1: Fig. S1 and Table S2). Besides, GO analysis identified a total of 521 biological processes (BP), 30 cellular components (CC), and 19 molecular functions (MF) enriched among the DMRs. The most significantly enriched BP terms appeared to be related to cell differentiation, including pattern specification process, regionalization, and cell fate commitment (Fig. 3C). The transcriptional machinery was strongly implicated in enriched MF terms (Fig. 3D), which was in line with CC terms enriched in

Table 2 The top 10 hypermethylated genes in tumor tissues

Gene	log ₂ (Fold change)	p value	Adjusted p value
<i>BARHL2</i>	0.486	2.42E-15	1.55E-12
<i>MIR124-3</i>	0.408	4.31E-09	7.68E-08
<i>DMRTA2</i>	0.337	1.60E-16	1.48E-13
<i>ESPN</i>	0.330	3.32E-11	1.63E-09
<i>OTX2</i>	0.322	1.14E-15	8.60E-13
<i>YAE1D1</i>	0.316	2.83E-13	4.90E-11
<i>SKOR1</i>	0.313	5.53E-16	4.60E-13
<i>ZNF876P</i>	0.310	4.89E-14	1.45E-11
<i>HOXA9</i>	0.305	3.57E-12	2.85E-10
<i>OTX1</i>	0.305	9.17E-14	2.12E-11

transcriptional regulation complex, transmembrane transportation, and chromatin remodeling (Fig. 3E). Together, these terms depicted a rough picture in which the DMRs participated in an orchestrated program of transcriptional regulation, thereby highlighting the significance of transcription factors and chromatin remodelers in tumor initiation and development. We also performed gene set enrichment separately in genes harboring hyper- and hypomethylated DMRs. Similar enriched terms were observed for the former set (n = 626; Additional file 1: Fig. S2), while no term was significantly enriched among the latter, perhaps due to the small set size (n = 52).



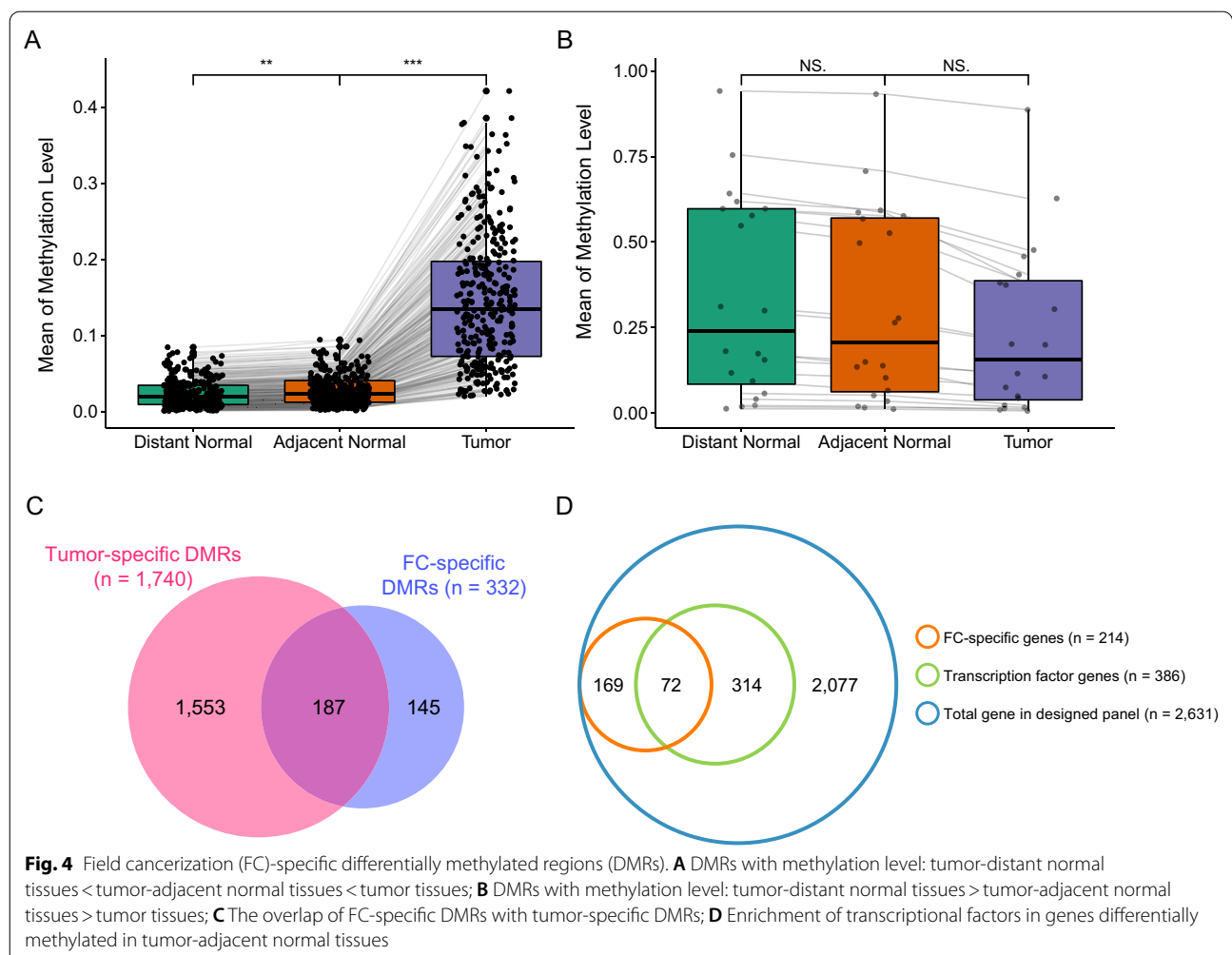


The identification of field cancerization (FC)-specific DMRs

In order to further identify aberrant methylation during different steps of cancer development, we compared the methylation level of each block among TUM, ADJ, and DIS by multiple paired-t test. A total of 332 DMRs were found to be differentially methylated among the three tissue types, indicating pre-malignant field-related methylation patterns. The methylation levels from 312 DMRs were significantly lower in DIS as compared to ADJ and also lower in ADJ than in TUM (Fig. 4A). Meanwhile, methylation levels from 20 DMRs were higher in DIS than ADJ, and also higher in ADJ as compared with TUM (Fig. 4B). Among the 332 FC-specific DMRs, 187 (56.3%) were overlapped with tumor-specific DMRs (Fig. 4C). Among the top 15 FC-specific hypermethylated genes, the methylation of *ZSCAN31* [25], *KCNA3* [26] and *CDO1* [27, 28] were reported to be associated with lung cancer development (Table 3). Besides, methylation of *DRD4* [29, 30], *ZNF132* [31] and *ZNF43* [32] have been reported to play roles in other cancer types. Due to the

Table 3 The top 15 hypermethylated genes in field cancerization

Gene	log2 (Fold change)	P value	adjusted p value
<i>ZSCAN31</i>	5.079	<0.001	<0.001
<i>ZNF345</i>	4.431	0.040	0.022
<i>DRD4</i>	4.312	<0.001	<0.001
<i>RAI1</i>	4.097	<0.001	<0.001
<i>ZNF132</i>	4.083	<0.001	<0.001
<i>ZNF175</i>	3.870	0.040	0.023
<i>ZNF43</i>	3.861	0.016	0.002
<i>SNX32</i>	3.847	<0.001	<0.001
<i>FAM19A2</i>	3.742	<0.001	<0.001
<i>HIST1H2BE</i>	3.651	0.040	0.007
<i>FABP5</i>	3.541	0.040	0.010
<i>NTMT1</i>	3.501	0.040	0.040
<i>ENPP2</i>	3.440	0.040	0.032
<i>KCNA3</i>	3.401	0.035	0.005
<i>CDO1</i>	3.395	<0.001	<0.001



small number of the FC-specific DMRs identified, functional enrichment analyses failed to identify any enriched pathways.

The total of 8312 blocks included in the panel span 2631 genes, among which 385 (14.6%) are transcription factor genes. On the other hand, the 332 FC-specific DMRs were annotated in 241 genes. Among the 241 genes, 72 (29.9%) were transcription factor genes (Fig. 4D; Additional file 1: Table S3). Hypergeometric analysis revealed that these differentially methylated genes were enriched with transcription factor genes ($P = 4.729e - 11$), which were consistent with GO enrichment results. Transcription factors also account for a similar proportion in differentially methylated tumor-specific genes (30%). These remarkable percentages suggested the role of epigenetic regulation of transcription factors as a key step in driving malignancy and FC.

Finally, we evaluated the associations between methylation levels of FC-specific DMRs in tumor samples and clinical characteristics by PCA analysis. We found that

Table 4 PCA of the methylation levels of field cancerization-specific DMRs in tumor samples revealed significant associations (in bold) with some clinical features

P value	PC1	PC2	PC3
Age	0.007	1.081e - 13	3.091e - 05
Sex	0.374	0.1621	0.0026
T-stage	0.1714	0.9696	0.054
N-stage	0.2857	0.2769	0.9209
Histology	0.0728	0.0003	0.0792
Smoking status	0.2904	0.7241	0.3701
Tumor size	0	0.223	0

ADJ—tumor-adjacent normal tissue, DIS—tumor-distant normal tissue, DMR—differential methylation region, TUM—surgically-resected tumor

age, histology, and tumor size were significantly associated with DMR methylation level (Table 4). These associations were confirmed with further analyses (Fig. 5). Among the genes most intensely methylated in tumor samples, patients with squamous cell carcinoma had significantly higher methylation level than those with adenocarcinoma ($P = 0.024$; Fig. 5A), and methylation was also significantly correlated with tumor size ($R = 0.38$, $P = 0.023$) and age ($R = 0.59$, $P < 0.001$; Fig. 5B, C). Among the hypomethylated genes, men showed lower methylation levels than women ($P = 0.042$; Fig. 5D).

Discussion

DNA methylation, occurring very early in the process of carcinogenesis, has been widely recognized as an important cancer-related biomarker. In the present study, we identified 1675 hypermethylated and 65 hypomethylated tumor-specific DMRs, which were annotated by 626 genes. Among these differentially methylated genes, some have been confirmed to be regulated by methylation in lung cancer development. *BARHL2* [33, 34], *DMRTA2* [33, 35, 36], *OTX1* [33, 34] and *OTX2* [33, 37] were identified as DNA methylation markers for lung cancer. Increased methylation of *MIR124* has also been found in NSCLC [38]. Methylation of *HoxA9* has been demonstrated as a reliable prognostic marker for NSCLC [28, 39–41]. We further performed functional enrichment analysis to clarify the role of methylation in NSCLC. We found a trend of methylation down-regulation for genes in ECM-receptor interaction pathway (adjusted $p = 0.021$). ECM constitutes the main part of the extracellular microenvironment. Its synthesis, distribution, and degradation are closely linked to the differentiation, proliferation, invasion, and metastasis of malignant tumors. Overexpression of the ECM-receptor

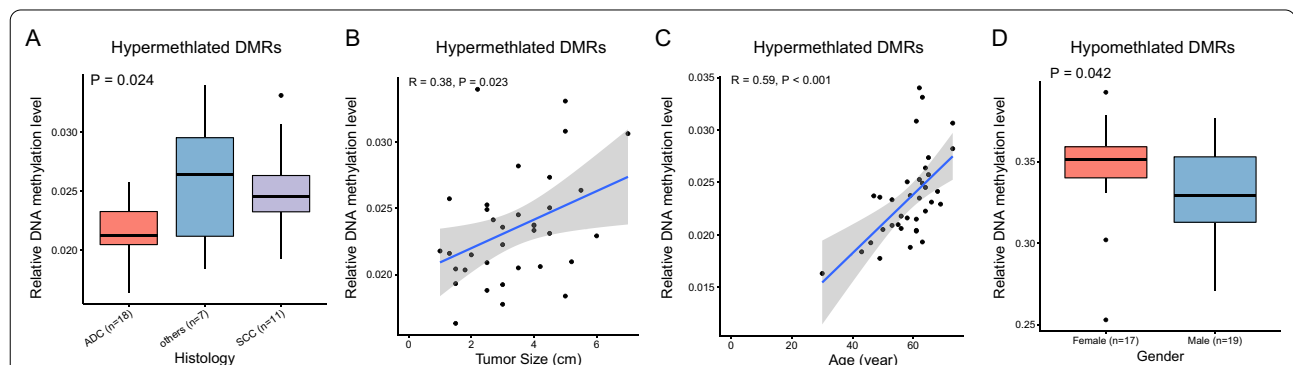


Fig. 5 Association between DNA methylation levels of field cancerization-specific differentially methylated regions (DMRs) and clinicohistologic characteristics. (A) Relative DNA methylation levels per histology among hypermethylated DMRs. (B) Correlation between relative DNA methylation levels and tumor size or (C) patient age among hypermethylated DMRs. (D) Relative DNA methylation levels per sex among hypomethylated DMRs. Hypermethylated DMRs refer to those showing a methylation level pattern of tumor-distant normal tissues < tumor-adjacent normal tissues < tumor tissues, and hypomethylated DMRs refer to those with a complete reversed pattern. ADC adenocarcinoma, SCC squamous cell carcinoma

(*hyaluronan receptor HMMR*) has been found primarily in LUAD and was connected with an inflammatory molecular signature and poor prognosis [42]. Lim et al. developed a signature based on the expression of 29 ECM-associated genes to predict the prognosis of the patients at the early stage of NSCLC [43].

More interestingly, we also identified 332 field-cancerization specific DMRs spanning 241 genes, which were differentially methylated in ADJ as compared with TUM and DIS. Compared with tumor-specific DMRs, these FC-cancerization specific DMRs may represent earlier methylation alterations that occur in carcinogenesis and serve as more sensitive biomarkers for early-detection and risk assessment for lung cancers. Of the 15 genes where the top differentially methylated DMRs residue (Table 3), *ZSCAN31* is thought to function as a transcription factor which is involved in airway structure or remodeling [44]. *ZSCAN31* has also been reported to be significantly hypermethylated in lung cancer [25]. *CDO1* silencing promotes proliferation of NSCLC by limiting the futile metabolism of cysteine. Methylation of *CDO1* has been identified as a specific marker for lung cancer diagnosis [27, 28]. *KCNA3* functions in voltage-gated potassium channels, which play a variety of roles in cancer progression. *KCNA3* inactivation via promoter hypermethylation has been found across multiple cancer types including lung, breast, pancreas, ovarian, kidney, prostate, and colon [26, 45]. Besides, *DRD4*, *ZNF132* and *ZNF43* have been linked with other non-lung cancer types: *DRD4*, encoding dopamine receptor, is involved in early brain development and epigenetically repressed in pediatric CNS tumors [30]; it also has been identified as a potential epigenetic driver in hepatocellular carcinoma [29]. *ZNF132* belongs to C2H2 zinc finger protein family and plays an important role in ESCC development as a tumor suppressor gene. It has been identified as a novel hypermethylation biomarker in ESCC [31]. Hypermethylated *ZNF43* has been reported as a biomarker for colorectal cancer [32]. We reported the first clinical evidence that methylation of *DRD4*, *ZNF132* and *ZNF43* may be also involved in lung cancer development. Furthermore, we also identified several novel methylation biomarkers that have not previously been reported in cancer. Some of them are transcription factors, such as *ZNF345* and *ZNF175* (Table 3). Hypergeometric analysis further revealed an enrichment of transcription factor genes ($p=4.729e-11$) in the 241 differentially methylated genes. Consistently, the most significantly enriched GO BP, CC, and MF terms highlighted a sizable proportion of genes participating in transcription regulation and chromatin remodeling. Collectively, these findings suggested the role of epigenetic regulation of transcription factors as a key step in driving malignancy and FC.

Different cell type composition could be a powerful source of DNA methylation level changes between surgical samples, and immune infiltration cell is a prominent cause of cell type composition perturbations. To evaluate the extent to which immune infiltration could have affected methylation level changes, we compiled a list of 782 immune cell marker genes from literature. Twelve of these markers harbored regions among the 1675 hypermethylated regions in this study, and 2 markers for the 65 hypomethylated.

Overall, DM immune cell markers accounted for 0.8% of all DM genes. Additionally, no GO or KEGG terms related to immune-related biological processes or functions were significantly enriched differentially methylated genes (Fig. 3). These findings suggested that immune cell infiltration was present but did not remarkably interfere with the identification of differentially methylated genes. Apart from cell type makeup, functional enrichment of genes identified with a targeted approach may be largely affected by the targeted panel. We compared the enriched GO BP terms among the genes that harbored the targeted methylation blocks and those among randomly sampled subsets (Additional file 1: Fig. S3). The different enrichment results suggested no inherent concentration of specific GO terms among the 2613 targeted genes.

Compared with previous works that identified genes differentially methylated in TUM and ADJ [15, 46, 47], strength of this study partly stems from a novel design that used two normal samples at defined distances away from resection margins and therefore allowed identification of genes that showed FC-specific dynamics of DNA methylation levels. These genes showed progressively enhanced or attenuated methylation as the location drew nearer to the tumor, thereby providing candidate markers for tracking tumorigenesis and early development. On the other hand, the major limitation of this study is the lack of validation of the prognostic values of DMRs. This is largely due to the short follow-up time after the surgery so that the relapse-free survival data of patients remains immature. This study was also limited by the lack of a well-defined consensus on the area undergoing cancerization, as there is clinical evidence suggesting that lung tissues deemed non-tumorous patients NSCLC patients may already be under FC due to carcinogen exposure [46, 47]. Therefore, although the DM genes with progressive increase/decrease as location sample drew nearer to the tumor remain FC-specific, it is unclear how their methylation levels change during early cancer development without a validated non-cancerous control sample. Besides, functional studies should be performed to confirm the roles of those newly identified biomarkers in field cancerization of lung cancer.

Conclusions

In conclusion, our data revealed distinct methylation patterns between pre-malignant lesions and malignant tumors, suggesting the essential role of DNA methylation as an early step in pre-malignant field defects. Moreover, our study also identified cancer-specific methylation blocks that could potentially be used as markers for lung cancer screening.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-022-01192-1>.

Additional file 1. Fig. S1: Significantly enriched KEGG pathways terms [1] among tumor-specific differentially methylated genes using a different gene set annotation file showed similar results as in Figures 3A and 3B. All analysis parameters were the same except for the gene set file (c2.cp.v7.4.entrez.gm in this analysis).

Additional file 2. Fig. S2: Significantly enriched GO terms among genes harboring hypermethylated field cancerization-specific differentially methylated regions.

Additional file 3. Fig. S3: Enriched GO biological process terms among (A) all 2613 genes that harbored the targeted methylation blocks and those among randomly sampled subsets of (B) 100, (C) 200, and (D) 500 genes.

Additional file 4. Table S1: Details of tumor-specific differentially methylated regions.

Additional file 5. Table S2: Results of GSEA analysis using different MsigDB gene sets. Analysis 1 used "KEGG gene sets as NCBI (Entrez) Gene IDs" (c2.cp.kegg.v7.4.entrez.gmt), and the corresponding significantly enriched KEG pathways are shown in Figures 3A and 3B. Analysis 2 used "canonical pathways as NCBI (Entrez) Gene IDs" (c2.cp.v7.4.entrez.gmt), and the corresponding significantly enriched KEGG pathways are shown in Figure S1).

Additional file 6. Table S3: List of 72 transcription factor genes the field cancerization-specific differentially methylated regions spanned and significantly enriched GO terms.

Acknowledgements

We are grateful to Xiao Zou, Wenjie Sun, Bing Li, Wei Xue, and Jinying Liu from Burning Rock Biotech for helpful advice on manuscript revision.

Authors' contributions

QW designed the study, collected the data and revised the manuscript; LW, JY, GL and PZ acquired the materials and collected the data; HW analyzed and interpreted the data; LS drafted the manuscript; JL collected the data; WS made substantial contributions to the conception. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the NODE repository (<https://www.biosino.org/node/>) (ID: OEP002659).

Declarations

Ethics approval and consent to participate

The study was approved by the institutional review board of the Second Affiliated Hospital of Harbin Medical University. All patients provided written informed consent, in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable as no human identity revealing data is used in the study.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Thoracic Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China. ²Burning Rock Biotech, Guangzhou, China. ³Department of Medical Oncology, The Second Affiliated Hospital of Harbin Medical University, Harbin 150001, China.

Received: 8 July 2021 Accepted: 18 February 2022

Published online: 21 March 2022

References

- Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med*. 2008;359:1367–80. <https://doi.org/10.1056/NEJMra0802714>.
- Kadara H, Wistuba II. Field cancerization in non-small cell lung cancer: implications in disease pathogenesis. *Proc Am Thorac Soc*. 2012;9:38–42. <https://doi.org/10.1513/pats.201201-004MS>.
- Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization. *Nat Rev Cancer*. 2018;18:19–32. <https://doi.org/10.1038/nrc.2017.102>.
- Tang X, Varella-Garcia M, Xavier AC, Massarelli E, Ozburn N, Moran C, et al. Epidermal growth factor receptor abnormalities in the pathogenesis and progression of lung adenocarcinomas. *Cancer Prev Res*. 2008;1:192–200. <https://doi.org/10.1158/1940-6207.CAPR-08-0032>.
- Nelson MA, Wymmer J, Clements N Jr. Detection of K-ras gene mutations in non-neoplastic lung tissue and lung cancers. *Cancer Lett*. 1996;103:115–21. [https://doi.org/10.1016/0304-3835\(96\)04202-4](https://doi.org/10.1016/0304-3835(96)04202-4).
- Wistuba II, Gazdar AF. Lung cancer preneoplasia. *Annu Rev Pathol*. 2006;1:331–48. <https://doi.org/10.1146/annurev.pathol.1.110304.100103>.
- Kerr KM, Galler JS, Hagen JA, Laird PW, Laird-Offringa IA. The role of DNA methylation in the development and progression of lung adenocarcinoma. *Dis Markers*. 2007;23:5–30. <https://doi.org/10.1155/2007/985474>.
- Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, et al. Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc Natl Acad Sci USA*. 1998;95:11891–6. <https://doi.org/10.1073/pnas.95.20.11891>.
- Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Res*. 2002;62:2370–7.
- Zochbauer-Muller S, Lam S, Toyooka S, Virmani AK, Toyooka KO, Seidl S, et al. Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. *Int J Cancer*. 2003;107:612–6. <https://doi.org/10.1002/ijc.11458>.
- Kanai Y, Hirohashi S. Alterations of DNA methylation associated with abnormalities of DNA methyltransferases in human cancers during transition from a precancerous to a malignant state. *Carcinogenesis*. 2007;28:2434–42. <https://doi.org/10.1093/carcin/bgm206>.
- Denisov EV, Schegoleva AA, Gervas PA, Ponomaryova AA, Tashireva LA, Boyarko VV, et al. Premalignant lesions of squamous cell carcinoma of the lung: The molecular make-up and factors affecting their progression. *Lung Cancer*. 2019;135:21–8. <https://doi.org/10.1016/j.lungcan.2019.07.001>.
- Kang G, Chen K, Yang F, Chuai S, Zhao H, Zhang K, et al. Monitoring of circulating tumor DNA and its aberrant methylation in the surveillance of surgical lung cancer patients: protocol for a prospective observational study. *BMC Cancer*. 2019;19:579. <https://doi.org/10.1186/s12885-019-5751-9>.
- Liang N, Li B, Jia Z, Wang C, Wu P, Zheng T, et al. Ultrasensitive detection of circulating tumour DNA via deep methylation sequencing aided by machine learning. *Nat Biomed Eng*. 2021;5:586–99. <https://doi.org/10.1038/s41551-021-00746-5>.
- Yang L, Zhang J, Yang G, Xu H, Lin J, Shao L, et al. The prognostic value of a methylome-based malignancy density scoring system to predict recurrence risk in early-stage lung adenocarcinoma. *Theranostics*. 2020;10:7635–44. <https://doi.org/10.7150/thno.44229>.

16. Pedersen BS, Eyring K, De S, Yang IV, Schwartz DA. Fast and accurate alignment of long bisulfite-seq reads. arXiv:14011129v2 [q-bio.GN]. 2014;
17. Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics*. 2014;30:2503–5. <https://doi.org/10.1093/bioinformatics/btu314>.
18. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics*. 2015;31:2032–4. <https://doi.org/10.1093/bioinformatics/btv098>.
19. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545–50. <https://doi.org/10.1073/pnas.0506580102>.
20. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34:267–73. <https://doi.org/10.1038/ng1180>.
21. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1:417–25. <https://doi.org/10.1016/j.cels.2015.12.004>.
22. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.
23. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res*. 2019;47:D419–26. <https://doi.org/10.1093/nar/gky1038>.
24. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016;374:20150202. <https://doi.org/10.1098/rsta.2015.0202>.
25. Kettunen E, Hernandez-Vargas H, Cros MP, Durand G, Le Calvez-Kelm F, Stuoopelyte K, et al. Asbestos-associated genome-wide DNA methylation changes in lung cancer. *Int J Cancer*. 2017;141:2014–29. <https://doi.org/10.1002/ijc.30897>.
26. Kim JH, Karnovsky A, Mahavisno V, Weymouth T, Pande M, Dolinoy DC, et al. LRpath analysis reveals common pathways dysregulated via DNA methylation across cancer types. *BMC Genomics*. 2012;13:526. <https://doi.org/10.1186/1471-2164-13-526>.
27. Wrangle J, Machida EO, Danilova L, Hulbert A, Franco N, Zhang W, et al. Functional identification of cancer-specific methylation of CDO1, HOXA9, and TAC1 for the diagnosis of lung cancer. *Clin Cancer Res*. 2014;20:1856–64. <https://doi.org/10.1158/1078-0432.CCR-13-2109>.
28. Ooki A, Maleki Z, Tsay JJ, Goparaju C, Brait M, Turaga N, et al. A panel of novel detection and prognostic methylated DNA markers in primary non-small cell lung cancer and serum DNA. *Clin Cancer Res*. 2017;23:7141–52. <https://doi.org/10.1158/1078-0432.CCR-17-1222>.
29. Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Mendez-Gonzalez J, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. *Hepatology*. 2015;61:1945–56. <https://doi.org/10.1002/hep.27732>.
30. Unland R, Kerl K, Schlosser S, Farwick N, Plagemann T, Lechtape B, et al. Epigenetic repression of the dopamine receptor D4 in pediatric tumors of the central nervous system. *J Neurooncol*. 2014;116:237–49. <https://doi.org/10.1007/s11060-013-1313-1>.
31. Jiang D, He Z, Wang C, Zhou Y, Li F, Pu W, et al. Epigenetic silencing of ZNF132 mediated by methylation-sensitive Sp1 binding promotes cancer progression in esophageal squamous cell carcinoma. *Cell Death Dis*. 2018;10:1. <https://doi.org/10.1038/s41419-018-1236-z>.
32. Kel A, Boyarskikh U, Stegmaier P, Leskov LS, Sokolov AV, Yevshin I, et al. Walking pathways with positive feedback loops reveal DNA methylation biomarkers of colorectal cancer. *BMC Bioinformatics*. 2019;20:119. <https://doi.org/10.1186/s12859-019-2687-7>.
33. Rauch TA, Wang Z, Wu X, Kernstine KH, Riggs AD, Pfeifer GP. DNA methylation biomarkers for lung cancer. *Tumour Biol*. 2012;33:287–96. <https://doi.org/10.1007/s13277-011-0282-2>.
34. Rauch TA, Zhong X, Wu X, Wang M, Kernstine KH, Wang Z, et al. High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc Natl Acad Sci USA*. 2008;105:252–7. <https://doi.org/10.1073/pnas.0710735105>.
35. Carvalho RH, Hou J, Haberle V, Aerts J, Grosveld F, Lenhard B, et al. Genomewide DNA methylation analysis identifies novel methylated genes in non-small-cell lung carcinomas. *J Thorac Oncol*. 2013;8:562–73. <https://doi.org/10.1097/JTO.0b013e3182863ed2>.
36. Heller G, Babinsky VN, Ziegler B, Weinzierl M, Noll C, Altenberger C, et al. Genome-wide CpG island methylation analyses in non-small cell lung cancer patients. *Carcinogenesis*. 2013;34:513–21. <https://doi.org/10.1093/carcin/bgs363>.
37. Daugaard I, Dominguez D, Kjeldsen TE, Kristensen LS, Hager H, Wojdacz TK, et al. Identification and validation of candidate epigenetic biomarkers in lung adenocarcinoma. *Sci Rep*. 2016;6:35807. <https://doi.org/10.1038/srep35807>.
38. Heller G, Altenberger C, Steiner I, Topkian T, Ziegler B, Tomasich E, et al. DNA methylation of microRNA-coding genes in non-small-cell lung cancer patients. *J Pathol*. 2018;245:387–98. <https://doi.org/10.1002/path.5079>.
39. Robles AI, Arai E, Mathe EA, Okayama H, Schetter AJ, Brown D, et al. An integrated prognostic classifier for stage I lung adenocarcinoma based on mRNA, microRNA, and DNA methylation biomarkers. *J Thorac Oncol*. 2015;10:1037–48. <https://doi.org/10.1097/JTO.0000000000000560>.
40. Sandoval J, Mendez-Gonzalez J, Nadal E, Chen G, Carmona FJ, Sayols S, et al. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. *J Clin Oncol*. 2013;31:4140–7. <https://doi.org/10.1200/JCO.2012.48.5516>.
41. Lissa D, Ishigame T, Noro R, Tucker MJ, Bliskovsky V, Shema S, et al. HOXA9 methylation and blood vessel invasion in FFPE tissues for prognostic stratification of stage I lung adenocarcinoma patients. *Lung Cancer*. 2018;122:151–9. <https://doi.org/10.1016/j.lungcan.2018.05.021>.
42. Stevens LE, Cheung WKC, Adua SJ, Arnal-Estape A, Zhao M, Liu Z, et al. Extracellular matrix receptor expression in subtypes of lung adenocarcinoma potentiates outgrowth of micrometastases. *Cancer Res*. 2017;77:1905–17. <https://doi.org/10.1158/0008-5472.CAN-16-1978>.
43. Lim SB, Tan SJ, Lim WT, Lim CT. An extracellular matrix-related prognostic and predictive indicator for early-stage non-small cell lung cancer. *Nat Commun*. 2017;8:1734. <https://doi.org/10.1038/s41467-017-01430-6>.
44. Li X, Hawkins GA, Ampleford EJ, Moore WC, Li H, Hastie AT, et al. Genome-wide association study identifies TH1 pathway genes associated with lung function in asthmatic patients. *J Allergy Clin Immunol*. 2013;132(313–20): e15. <https://doi.org/10.1016/j.jaci.2013.01.051>.
45. Brevet M, Fucks D, Chatelain D, Regimbeau JM, Delcenserie R, Sevestre H, et al. Deregulation of 2 potassium channels in pancreas adenocarcinomas: implication of KV1.3 gene promoter methylation. *Pancreas*. 2009;38:649–54. <https://doi.org/10.1097/MPA.0b013e3181a56ebf>.
46. Sato T, Arai E, Kohno T, Tsuta K, Watanabe S, Soejima K, et al. DNA methylation profiles at precancerous stages associated with recurrence of lung adenocarcinoma. *PLoS ONE*. 2013;8:e59444. <https://doi.org/10.1371/journal.pone.0059444>.
47. 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;135:319–34.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

