Cellular and Molecular Biology

Journal homepage: www.cellmolbiol.org

Significance of screening sensitive methylation sites using whole-genome sequencing in early diagnosis of non-small cell lung cancer

Yu Zhang*, Ning Lu, Shunchang Pu, Kui Mu

Department of Biology and Food Engineering, Bozhou University, Bozhou, 236800, China

ARTICLE INFO

Original paper

Article history:
Received: August 20, 2021
Accepted: November 17, 2021
Published: December 15, 2021

Keywords:
whole-genome sequencing; non-small cell lung cancer; methylation; diagnostic value

ABSTRACT

The current study aimed to screen the sensitive methylation sites of non-small cell lung cancer by whole-genome sequencing and construct an early warning system for lung cancer. For this purpose, from June 2017 to December 2020, fresh NSCLC tissues and paired adjacent NSCLC tissues from 45 patients were collected. DNA and total RNA were extracted from non-small cell lung cancer (NSCLC) and paired non-cancerous lung tissues. The DNA library combined with a biotinylated probe was collected by Dynabeads m270 streptavidin beads. The concentration of the final library was determined by qubit dsDNA HS assay. Quantitative analysis of DMR methylation in 45 paired tumor and normal lung tissues was performed. RT qPCR and Western blot were used to verify the mRNA expression of candidate genes. Results showed that the methylation rate of CpG 7 in stxbp6 in stage III NSCLC was higher than that in stage I and early-stage II NSCLC; The methylation rates of cpg1 and 38-39 units in fzd10 were higher in stage I NSCLC than in stage II and III NSCLC; The methylation rates of CpG 6 in stxbp6 and CpG 4 and 20-21 in bcl6b in patients with tumor diameter > 3cm were higher than those in patients with tumor diameter < 3cm; Methylation of CpG unit 3 in stxbp6 is associated with age. Stxbp6, bcl6b, fzd10 and hspb6 mRNA expression were down-regulated in patients under 45 years old. The methylation rates of CpG 7 in stxbp6, CPG 6 in stxbp6 and CpG 4 and 20-21 in bcl6b were negatively correlated with the survival time of patients; The methylation rates of CpG 1 and 38-39 units in fzd10 were positively correlated with survival time (P<0.05). It was concluded that the methylation rates of CpG 7 in stxbp6, CPG 6 in stxbp6 and CpG 4 and 20-21 in bcl6b are valuable for early diagnosis.

DOI: http://dx.doi.org/10.14715/cmb/2021.67.5.30

Introduction

At present, non-small-cell lung cancer (NSCLC) has become the leading cause of malignant tumor death and a public health-threatening problem globally (1). Although using a combination of radiotherapy, chemotherapeutics, surgery and targeted treatment resulted in a higher survival rate in NSCLC patients, the prognosis is still very low (2). One of the important causes is that it is difficult to detect early NSCLC, and most cases have reached an advanced stage after diagnosis (3).

DNA methylation is a process by which methyl - S-adenosine methionine is used as a donor under the catalysis of DNA methyl to transfer methyl to a specific base. At present, the methylation of lung cancer-related DNA gene as a molecular marker for early diagnosis of lung cancer has become a research hotspot, and its clinical diagnostic effect has attracted much attention. DNA methylation is one of the most in-depth genetic modifications. It fills the carbon position of cytosine with a methyl group, and the methylation mainly occurs in CpG dinucleotide (4). It is a dynamic process that is strictly regulated under typical physiological situations and is equilibrated by transmethylase and demethylase of DNA, contributing to gene expression regulation and maintaining heterochromatin configuration during development (5). Several studies have found a direct association between abnormal DNA methylation and tumorigenesis promotion through direct mutagenesis of methylated cytosine, focal hypermethylation of multi-genomic regions (mainly CpG islands) and overall hypomethylation (6). In addition, it has been found that DNA methylation changes have occurred during the development of NSCLC, even prior to the development of atypical adenomatous hyperplasia (7).

*Corresponding author. E-mail: cauzhangyu@163.com

According to all these evidences, DNA methylation can be considered as a biomarker with a high potential in the early detection of lung cancer.

During the occurrence and development of lung cancer, tumor cells have to face different cellular pressures, such as abnormal accumulation of various toxic and harmful substances in cells and oxygen stress. Under these conditions, the methylation of lung cancer-related genes is involved in the whole development and progression process of lung cancer. At present, the incidence of lung cancer and other malignant tumors has increased sharply. Although the traditional pathological examination is still the gold standard for its diagnosis, the traditional pathological examination can no longer meet the needs of early, accurate and rapid clinical diagnosis in some cases. Gene molecular diagnosis is the application of molecular biological methods to detect changes in the structure or expression level of genetic material in patients and make a diagnosis of the technology. This method can not only diagnose individual genetic diseases but also assist in pathomorphological diagnosis (4-7). Therefore, in this study, whole-genome sequencing was used to screen the sensitive methylation sites in NSCLC and explore its significance in the early diagnosis of lung cancer. Detection of lung cancer-related gene methylation is helpful for the diagnosis of lung cancer. Combined detection of multi-gene methylation can improve the detection rate and predictive efficiency of early lung cancer.

Materials and methods

Tissue samples

From June 2017 to December 2019, matched adjacent non-cancerous lung tissues and fresh NSCLC tissues were collected from a total of 45 patients (Table 1). Non-cancerous lung tissue is defined as tissue with the least distance of 5 cm from cancerous tissue. To confirm the existence of cancer cells, the samples were evaluated by practiced pathologists. The collected tissues were stored at -70℃ following stabilization in RNA later dilution (Qiagen, Hilden, Germany). Clinicopathologic features of patients were collected from their medical records. No one of the patients had immune diseases, chronic wasting diseases or other types of malignancies, and none of them had ever experienced preoperative therapy, e.g., chemotherapeutics or radiotherapy. Informed consent were obtained from all participants.

Table 1. Clinicopathologic features of 45 patients

Male		Number of patients
Female		25/45 (55.56%)
		20/45 (44.44%)
Age (years)	≤45	16/45 (35.56%)
	≥46	29/45 (64.44%)
Smoking history	Yes	11/45 (24.44%)
	No	34/45 (75.56%)
Pathologic stage	I	20/45 (44.44%)
	II	16/45 (35.56%)
	III	9/45 (20.00%)
Lymphatic metastasis	Yes	25/45 (55.56%)
	No	20/45 (44.44%)
Tumor size (cm)	0-3	23/45 (51.11%)
	3.1-5	18/45 (40.00%)
	≥5.1	4/45 (8.89%)
Survival time (months)	1-12	6/45 (13.33%)
	13-24	15/45 (33.33%)
	25+	24/45 (53.33%)

DNA bisulfite modification, cDNA synthesis, DNA, RNA and protein extraction

The extraction and purification processes of DNA (gDNA) and RNA from NSCLC and matched non-cancerous lung tissues were carried out using QIAamp DNA and RNeasy Mini kits (Qiagen, Hilden, Germany) in accordance with the manufacturer's guidance. Radioimmunoassay (RIPA) was used to separate total protein from tissues according to the instructions provided by the manufacturer (Beyotime, Shanghai City, China). The separated samples were then protected with phenylmethylsulfonyl fluoride (PMSF). The integrity of RNA was examined using standard denaturing agarose gel electrophoresis, and NanoDrop ND-1000 spectrophotometer (CapitalBio Nano Q, Beijing, China) was applied for gDNA and total RNA quantification.

Transcriptor first-strand cDNA synthesis kit (Roche, Mannheim, Germany) was applied to synthesize cDNA by RNA reverse transcription. Following the instructions provided by the manufacturer, the Epitect Fast DNA Bisulfite kit (Qiagen, Hilden, Germany), was used to modify and purify gDNA in10 ng/L bisulfite solution.

Whole-genome methylation sequencing

Zymo Lightning Conversion Reagent was used for bisulfite conversion. 130μL of Lightning Conversion Reagentwas added to 20μL of DNA sample. Mixture incubation was done inside a thermal cycler according to the following procedure: 98℃ for 8 minutes, 54 ℃ for 60 minutes and 4℃ for 20 hours. A mixture of bisulfite-conversed DNA and M-Binding buffer was prepared and passed through Zymo-Spin (TM) IC column for sulfonation. Then, it was washed and eluted in 17μL of M-Elution buffer. The index PCR kit of AnchorDx EpiVisioTM (AnchorDx, Cat# A2DX00025) and methylation library preparation kit of AnchorDx EpiVisioTM (AnchorDx, Cat# A0UX00019) were used to pre-build the AnchorIRISTM library. Terminal repair of bisulfite-conversed DNA was carried out at 37℃ for 30 minutes using the MEE1 enzyme. DNA denaturing was conducted at 95℃ for 5 min and quickly cooled on ice. MLE1 and MLE5 enzymes were used to ligate the 3'-terminal adaptor at 37 ℃ for 30 minutes. MAE3 enzyme was quickly used for the first amplification to produce reverse complementary DNA (cDNA) through the following PCR procedures: one 3 minutes cycle at 95℃, four cycles at 95℃ for 30 seconds, 60 ℃ for 30 seconds, 68 ℃ for 1 minute and one 5 minutes cycle at 68℃. Purified by AMB1 magnetic beads, the amplified DNA was eluted in a volume of 20 μ L. Next, 3'-terminal adaptor ligation of reverse complementary DNA was carried out at 37℃ for 30 minutes using MSE1 and MSE5 enzymes. The MIB1 PCR premix was immediately used for index PCR (i5 and i7) according to the following PCR procedures: one cycle at 98℃ for 45s, 14 cycles at 98℃ for 15 s, 60℃ for 30 s, 72℃ for 30 s, and one 5 minutes cycle at 72℃. IPB1 magnetic beads were then applied to purify the amplified library and Qubit™dsDNA HS analysis kit to determine its concentration. This prebuilt library with over than 400ng DNA was regarded as eligible for target enrichment.

Target enrichment was done using the AnchorDx EpiVisioTM target enrichment kit (AnchorDx, catalog number A0UX00031). The use of our customized 10K methylation panel to target enrichment resulted in several 1000 ng DNA involving up to 4 pre-libraries being collected. The blocking agents of HBA, HBB and HE were then introduced to the 1000 ng combined prebuilt library. Then, the combined prebuilt library was completely dried using a heated vacuum and rebuilt in 7.5μL of MHB1 hybridization buffer and 3μL of MHE1 hybridization enhancer. After denaturing at 95 ℃ for 10 minutes, the reconstructed prebuilt library immediately transferred to a hybridization box at 47 ℃. Probes were introduced to prebuilt libraries and immediately transferred to the thermal cycler for hybridization incubation according to the procedures of the manufacturer.

Dynabeads M270 streptavidin beads (Thermo Fisher Scientific, Cat#65306) was used to pull down the DNA pre-library attached to biotinylated probes after hybridization. In brief, each pre-library washed in 30 μ L of streptavidin beads. After washing two times with a Binding Wash Buffer of 1X, they were suspended in 60 μ L of binding wash buffer for the second time. The pre-pool was introduced and completely mixed with the beads using repeated pipetting. Mixture incubation was performed on a spinner at 47℃ for 45 minutes. After the beads were combined, 100μL of preheated 1X Transfer Buffer was introduced to the resulting mixture. Immediately after clarification, the supernatant was eliminated, and the beads were eluted twice in preheated 1X stringent washing buffer. Afterward, the beads were re-suspended at room temperature in 200μL of 1X Wash Buffer I and mixed completely. After removing the supernatant, the beads were eluted by 1X Wash Buffer II and 1X Wash Buffer III and finally with 23μL of H2O following the same procedure. KAPA HiFi HotStart Ready Mix (KAPA Biosystems, catalog number KK2602) was applied to further amplify the libraries by P5 and P7 primers as the PCR procedure below: one cycle at 98℃ for 45s, twelve cycles at 98℃ for 15s+60℃ for 30s+72℃ for the 30s, and one cycle at 72℃ for 1 minute. After purification with Agencourt AMPure XP Magnetic Beads (Beckman Coulter, Cat#A63882), PCR products were washed in 40 μ L of EB buffer. Qubit dsDNA HS assay was applied to measure the final concentration of the library.

Preparation analysis of methylation detection library

The analysis of AnchorIRISTM performance was performed by introducing cfDNA inputs from 1 ng to 10 ng. CfDNA was separated from the plasma of three patients suffering from ovarian cancer. Qubit HS DNA assay was used to measure DNA concentration and combine cfDNA. Based on the total volume and amount of DNA, the combined cfDNA concentration was computed. After that, bisulfite conversion with a final concentration of 0.667 ng/μL was prepared and used to elute DNA following the instructions of the manufacturer. The bisulfite-conversed cfDNA was equally divided and used for various titrations introduced to the duplicate (i.e., 1ng, 3 ng, 5 ng, and 10 ng). Based on that, measurement differences between samples introduced in bisulfite conversion were avoided.

IRIS library was built according to the above method, while SWIFT library was built by denaturing bisulfite-converted DNA at 95 ℃ for 2 minutes and quickly cooling it on ice. Adaptor-1 was immediately ligated at 3'-terminal using the adaptase reaction mixture according to the following culture procedures: 37 ℃for one minute, 62℃for two minutes, and 65℃for five minutes. To synthesize the reverse complementary sequence of ssDNAs, an extension program and enzyme Y2 were used. The incubation procedures were: 1 minute at 98℃, 2 2 minutes at 62 ℃, 5 minutes at 65℃. Molecules of dsDNA were obtained. After being After by SPRI select beads, dsDNA molecules were eluted in 15μL of low EDTATE. Adaptor 2 ligation was performed at 25℃ for 15 minutes using Enzyme B3. The ligation product was purified by SPRI select beads and then amplified by Indexing PCR according to PCR procedures: one cycle at 98℃ for 30 s and one cycle at 98℃ for 10s+60℃ for 30s+68℃ for 1 minute. The number of PCR cycles was determined according to the input DNA value of SWIFT and IRIS analysis. Purified with beads, PCR products were eluted in 40μL of EB buffer and quantified by Qubit. According to the above method, the same 10K panel was used for the target enrichment of two assays.

Quantitative DNA methylation analysis

As mentioned above, the quantitative analysis of DMR methylation in candidate genes was carried out on 45 matched tumors and normal tissues of the lung by the use of the Sequenom MassARRAY (CapitalBio) platform. The platform adopts matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) and base-special RNA cleavage. The design of MassARRAY PCR primers was performed using http://www.epidesigner.com (S3 Table). Epityper software V1.0 (Sequenom, San Diego, California, U.S.A) was used to obtain the methylation rate of spectra.

Fluorescence-based real-time quantitative PCR (RT-qPCR)

To verify the mRNA expressions of candidate genes, RT-qPCR was carried out on 39 normal lung tissues and paired tumors. The remained six pairs of samples were not sufficient for the present analysis. Utilizing Quantinova SYBR Green PCR Master Mix (Qiagen), the analysis was conducted on the Applied Biosystems 7300 Real-Time PCR System (Applied Biosystems, Inc., Foster City, California, U.S.A). Ct value comparison approach was applied to calculate the relative expression value of tumor and normal tissue of lung. S4 Table summarizes primers of RT-qPCR.β-actin expression was considered as a measure to normalize other genes' expressions. Experiments were conducted in repetitions.

Western bolt

After separation with 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), the same amounts of protein in samples were submitted to polyvinylidene fluoride membrane (Merck Millipore, BilleriCA, MA, U.S.A) Rad, Hercules, ca, U.S.A) by Bio-Protocol and sealed with 5% skim milk. The membrane probe was performed with a monoclonal antibody against GAPDH and candidate gene at 4℃ overnight. Table S5 shows information on the antibody. After being washed with TBST, the membrane was incubated for 2 hours by IgG conjugated with horseradish peroxidase at room temperature. Electrochemiluminescence detection reagent (Millipore, Bedford, Massachusetts, U.S.A) was applied to visualize protein bands on ImageQuant LAS 500(GE Healthcare, Pittsburgh, Pennsylvania, U.S.A). Image J software was used to quantify the intensity of bands. For each candidate gene, the band

intensity was subtracted from its background intensity. The result was standardized according to the intensity of related GAPDH and compared with the band intensity of paired samples. Experiments were conducted in repetitions.

Statistical analysis

The data in this study were processed by SPSS20.0 statistical analysis software (IBM). Measurement data were expressed as "mean ± standard deviation" ($\bar{x}$ ±s). One-way ANOVA or repeated measurement ANOVA was used for comparison between groups, and LSD t-test was used for pair comparison between groups; Enumeration data were expressed as a percentage (%), and χ^2 analysis was used for comparison between groups; The difference was considered statistically significant when $P<0.05$.

Results and discussion

DMR identification from paired samples of patients

According to whole-genome DNA methylation analysis of paired NSCLC and normal lung tissues, there were 6899 DMR in tumor tissues compared with normal tissues, including 5,788 and 1,111 hyper- and hypo-methylation regions, respectively (microarray data was fed to Gene Expression Omnibus (GEO) with an accession number of GSE113432). Among hypermethylation regions, 560 regions were in the upstream region of the gene, and 561 were in the downstream region of the gene. 1,315 regions were in the gene and 3,352 regions were in the promoter region. Among hypomethylation regions, 22 regions were in the upstream region, and 20 regions were in the downstream region. 11 regions were in the gene and 1,058 were in the promoter region. Therefore, most DMR (63.92%) were in the promoter region, mainly on chromosomes 19, 1, 17 and 11 (Table 1). 581 DMR (495 hypermethylation regions and 86 hypomethylation regions) which completely overlapped with the promoter region were selected as candidate DMRs. The candidate DMRs were searched in the USCS genome browser to determine 388 annotated genes with the potential ability to change tumor-specific methylation. The GO analysis showed the contributing role of these genes contribute to several physiological procedures, like cellular processes, biological process regulation, metabolism, catalytic activity, binding, biological regulation, development and transport. The of pathway showed that these genes also contribute in several pathways, such as cancer pathway, metabolic pathway, cAMP signal pathway, GABA synapse, PI3K-Akt signaling, glutamate synapse, neuroactive ligand-receptor interaction, estrogen signaling, longevity, melanogenesis, calcium signaling, alcoholism, and HTLV-1 infection. The heat map showed 1000 methylated regions randomly selected to represent representative NSCLC and benign tissue samples. The level of methylation calculated for each region was considered as the reading score of co-methylation. Signals were displayed in a linear proportion of colors, with red indicating hypermethylation signals and green indicating hypomethylation signals (Fig 1 and Table 2).

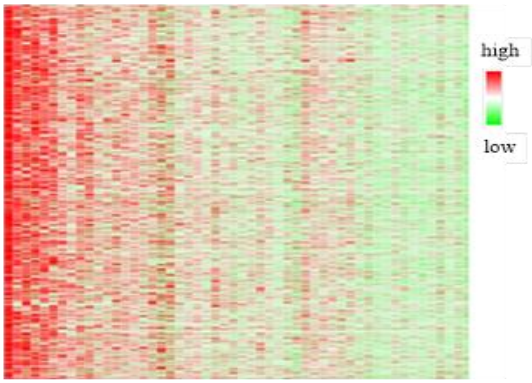

Figure 1. Characteristics of hypermethylation signal in non-small cell lung cancer tissues (NSCLC)

DEG identification from paired samples of patients

The microarray analysis of gene expression showed that paired tumors had different gene expressions compared to normal lung tissues. Although 11861 genes were expressed in the same way, 5129 genes were recognized as DEG. Among these DEGs, 1881 were down-regulated while the remaining 3248 were up-regulated. According to the cluster analysis, normal and tumorous lung tissue had different gene expressions. GO analysis showed that these types of DEGs are associated with a variety of cancer-related procedures, such as DNA replication, cell proliferation, cell adhesion, and cell division. Moreover, pathway analysis also indicated the role of these DEGs in several pathways, such as hedgehog signaling, Jak-STAT signaling, p53 signaling, MAPK, and NSCLC related pathways (Table 3).

Table 2. Distribution of 4,410 DMRs in the promoter region

Chromosome	Quantity of hypermethylation DMR	Quantity of hypomethylation DMR
1	321	109
2	146	59
3	169	49
4	75	21
5	133	28
6	108	47
7	113	51
8	117	18
9	173	20
10	120	74
11	246	77
12	108	75
13	16	19
14	75	23
15	90	46
16	236	41
17	282	74
18	28	8
19	377	94
20	122	45
21	38	18
22	110	40
X	138	21
Y	11	1
Total	3352	1058

Table 3. Microarray analysis of gene expression

Gene	Number of up-regulated genes	Number of down-regulated genes
Differential gene	3248	1881
Non-differential gene	18861	

Verification of methylation of candidate genes

MS-HRM analysis was performed on 45 paired NSCLC and non-cancerous lung tissue samples to verify the promoter methylation of four candidate genes. The whole-genome sequencing data was supported. All four candidate genes showed tumor-specific methylation. MS-HRM results for the test regions showed that NSCLC tumors had higher rates of hypermethylation than that of paired normal samples in all the four candidate genes (57.78% vs.17.78% and P < 0.001 for STXBP6; 71.11% vs. 37.78% and P= 0.001 for BCL6B; 93.33% vs. 68.89% and P = 0.006 for FZD10; and 91.11% vs. 62.22% and P = 0.001 for HSPB6) (Table 4).

Table 4. Verification of promoter methylation of candidate genes

Group	STXBP6		BCL6B		FZD10		HSPB6	
	High	Low	High	Low	High	Low	High	Low
NSCLC group	26	19	32	13	42	3	41	4
Control group	8	37	17	28	31	14	28	17
χ^2 value	15.315		10.08		8.775		10.497	
P value	<0.001		0.001		0.006		0.001	

RT-qPCR verification of candidate gene expression

Hypermethylation of promoters may affect gene expression. RT-qPCR was used to further verify the expressions of four candidate genes on 39 paired samples. In most tumor tissues, the level of mRNA expression of four candidate genes was down-regulated compared to paired normal tissues (Table 5).

Table 5. MiRNA expressions of four candidate genes in 39 paired samples

Group	STXBP6	BCL6B	FZD10	HSPB6
NSCLC group	3.63±0.45	3.56±0.31	2.78±0.23	3.44±0.18
Control group	8.54±0.31	7.22±0.87	7.83±0.64	6.54±0.31
χ^2 value	56.114	24.748	46.373	54.006
P value	<0.001	<0.001	<0.001	<0.001

Western blotting verification of candidate gene expression

The expressions of four candidate genes were further verified by western blotting on 39-paired samples. In most tumor tissues, the level of protein expression was down-regulated in four candidate genes. RT-qPCR and western blotting indicated that there is a negative correlation between gene expression and the overall methylation level of the promoter region (Table 6).

Table 6. Western blotting expressions of 4 candidate genes in 39 paired samples

Group	STXBP6	BCL6B	FZD10	HSPB6
NSCLC group	2.27±0.42	4.84±0.16	2.87±0.37	2.16±0.12
Control group	7.89±0.35	8.53±0.17	6.88±0.17	7.15±0.13
χ^2 value	64.196	98.710	61.501	176.141
P value	<0.001	<0.001	<0.001	<0.001

Relationship between methylation and gene expression with the age of NSCLC patients

According to correlation analysis between gene methylation and age, the expressions of BCL6B (P = 0.040), STXBP6 (P = 0.029), HSPB6 (P = 0.034) and FZD10 (P = 0.043) were down-regulated in patients under 45 years old (Fig 2).

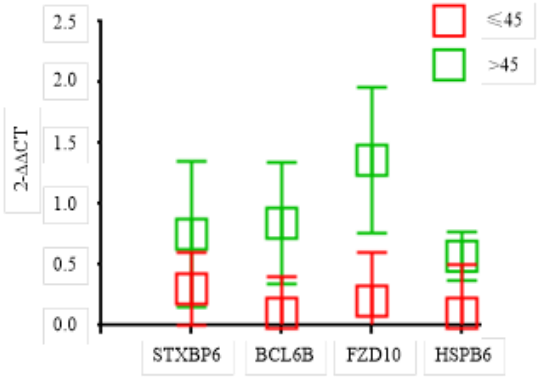

Figure 2. Correlation analysis between four candidate genes and patients' age

Relationship between gene methylation and gene expression and survival in patients with non-small cell lung cancer

The methylation rates of CpG 7 in STXBP6, CpG 6 in STXBP6, and CpG 4 and 20-21 in BCL6B were negatively correlated with the survival time of patients; The methylation rates of CpG 1 and 38-39 in FZD10 were positively correlated with the survival time of patients, (P<0.05).

Up to now, early detection of cancer is the most effective and economical means to decrease mortality from cancer (8). NSCLC is the most common type of cancer worldwide. Due to the difficulty in performing a diagnostic biopsy and the existence of a high rate of false-positive rate in low-dose CT screening, early screening of NSCLC has been a challenging process for a long time (9-11). As a result, it would be favorable to develop a sensitive diagnostic method that is non-invasive and able to differentiate between benign diseases and malignant pulmonary nodules in patients with positive low-dose lung CT screening (12-14). Sensitive methylation site has become one of the most attractive methods in clinical application.

Compared with somatic mutation spectrum analysis of early patients, methylation spectrum analysis can achieve higher sensitivity and specificity because: 1) a larger number of markers can be used at the same time to improve sensitivity (15). 2) There are multiple CpG loci in each selected site for the joint query of the region to obtain a "cancer-specific methylation pattern" to improve specificity (16). In addition, methylation profile analysis can be used to distinguish origin tissues from cancer subtypes. Some recent studies have shown that it is feasible to identify patients with malignant diseases by bisulfite sequencing of plasma DNA, but all of them focus on the advanced stage (17). According to an early NSCLC study, serum DNA treated with bisulfite was subjected to methylation-specific PCR, and six genes were selected according to the TCGA data set. The sensitivities of adenocarcinoma and squamous cell carcinoma at the Ia stage were reported to be 72.1% and 60% respectively (18). PCR was used for the other 6 genes identified in the TCGA data set, showing the specificity of 74-84% and sensitivity of 65-76% (19). In this study, markers were picked up from the TCGA database where the control group was normal tissue beside the tumor, and microarray was used to generate data. Candidate markers were generated by distinguishing tumor from adjacent non-tumor normal. This study mainly focused on early NSCLC, combined with a new targeted methylation spectrum analysis method. It has excellent library conversion efficiency and detection sensitivity. Two main technical obstacles of bisulfite sequencing include 1) Having limited means to achieve targeted enrichment, and 2) low efficiency of library conversion, leading to some limitations in the application of low input specimens (e.g., plasma DNA). In combined with simplified targeted enrichment and high-efficiency library conversion procedures, IRIS analysis can perform deep sequencing on preselected highly informative regions in the clinical situation. In our clinical cohort, the mean coverage rate of unique target is more than 180X. Compared with previous studies that used a shallow sequencing method, it greatly enhances the ability to detect low-frequency DNA.

Compared with normal lung tissues, several 6899 tumor-specific DMRs were identified in tumors by MeDIP chip analysis, including 5788 hypermethylation regions and 1,111 hypomethylation regions. This indicated that tumor-specific DMRs were prevalent in NSCLC, and hypermethylation

regions were more common in these tumors compared to hypomethylation regions (5.2 vs. 1). Both of chromosome distribution and genome location of DMR were heterogeneous.

We selected four hypermethylation genes having not been reported in NSCLC, namely HSPB6, FZD10, BCL6B, and STXBP6, for further study. STXBP6, located at 14q12, encodes the binding component of the SNARE complex, which plays a key role in fusion and vesicle targeting in eukaryotic cells. BCl6B is placed on chromosome 17p13.1 and is a homologue of BCL6. Its potential as a tumor suppressor is well recognized, and its methylation is considered as an indication of poor prognosis in many human-related cancers, such as hepatocellular carcinoma, colorectal and gastric cancers (20-23). As a member of the curly gene family, FZD10 is placed on chromosome 12q24.33. The cell surface receptor of Wnt pathway molecules is encoded by FZD10. FZD10 is highly up-regulated in many cancerous cells and cancers, e.g., cervical cancer cells, glioblastoma cells, primary colon cancer, and synovial sarcoma. It may have a leading role in the incidence and progression of tumors. Also known as HSP20, HSPB6 is placed on chromosome 19q13.12 and is a member of heat shock protein family B. It is responsible for encoding a small group of heat shock proteins (HSPs) that are widely expressed in different tissues with multiple functions. Up-regulated expression of HSP20 can cause myocardial angiogenesis by activating VEGFR signal cascade reaction, thus protecting the heart from a variety of stress motives. HSP20 may be involved in protection against the occurrence and development of many human cancers and has the potential to become a new objective for predicting and treating these types of cancers. In most tumor samples, the expressions of these four hypermethylated genes were down-regulated, indicating the effect of methylation in the regulation of these gene expressions, and the ability of methylation to be used as a marker in NSCLC. In a few cases, the up-regulation of hypermethylation genes and down-regulation of hypomethylation genes observed might be interpreted by other types of regulatory mechanisms of gene expression, like histone modification and non-coding RNA regulation. The methylation rates of CpG 7 in STXBP6, CpG 6 in STXBP6, and CpG 4 and 20-21 in BCL6B were negatively correlated with the survival time of patients; the methylation rates of CpG 1 and 38-39 in FZD10 were positively correlated with the survival time of patients, which are valuable in early diagnosis.

Conclusions

In summary, we identified novel DNA methylation markers associated with non-small cell lung cancer by a comprehensive genome-wide analysis of DNA methylation and gene expression in tumors. The methylation rates of CpG 7 in STXBP6, CPG 6 in STXBP6 and CpG 4 and 20-21 in BCL6B are valuable in early diagnosis.

Acknowledgements

The research is supported by: Support program for outstanding young talents in colleges and universities in Anhui Province - Determination of polysaccharides and triterpenes in Ganoderma lucidum spore powder and evaluation of its antitumor activity in vitro (No. gxyp2020071).

Interest conflict

The authors declare no conflict of interest.

References

1. Hoffman RM and Sanchez R. Lung Cancer Screening. Med Clin North Am. 2017;101:769-785.
2. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ, Wu YL and Paz-Ares L.Lung cancer: current therapies and new targeted treatments.Lancet, 2016;389:299-311
3. Reck M and Rabe KF. Precision Diagnosis and Treatment for Advanced Non-Small-Cell Lung Cancer. N Engl J Med. 2017;377:849-861.
4. Lu Y, Li S, Zhu S andGong YB. Methylated DNA/RNA in Body Fluids as Biomarkers for Lung Cancer. Biol Proced Onlin. 2017;19:2.
5. Agha G, Mendelson MM, Ward-Caviness CK, Joehanes R, Huan TX, Gondalia R, Salfati E, Brody JA, Fiorito G, Bressler J, Chen BH, Ligthart S and Guarrera S. Blood leukocyte DNA methylation predicts risk of future myocardial infarction and coronary heart disease. Circulation, 2019;140: 645-657.
6. Bacolod MD. Barany F and Fisher PB. Can CpG methylation serve as surrogate markers for

immune infiltration in cancer. Adv Cancer Res, 2019; 143: 351-384.
7. Farghaly TA, Masaret GS, Muhammad ZA and Harras MF. Discovery of thiazole-based-chalcones and 4-hetarylthiazoles as potent anticancer agents: Synthesis, docking study and anticancer activity. Bioorg Chem, 2020; 98:103761.
8. Smith RA, Andrews KS, Brooks D, Fedewa SA,Manassaram-Baptiste D and Saslow D. Cancer screening in the United States, 2017: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin. 2017 ;67;100-121
9. Lu J, Tan T, Zhu L, Dong H and Xian R. Hypomethylation causes MIR21 overexpression in tumors. Mol Ther Oncolytics, 2020;18:47-57.
10. Ning Y, Liu W, Guan X, Xie X and Zhang Y. CPSF3 is a promising prognostic biomarker and predicts recurrence of NSCLC. Oncol Lett, 2019:18;2835-2844.
11. Anh DT, Thuan NT, Hai PT and Thi Thu HL. Design, synthesis and evaluation of novel 3/4-((Substituted benzamidophenoxy) methyl)-N-hydroxybenzamides/propenamides as Histone Deacetylase Inhibitors and Antitumor Agents. Anticancer Agents Med Chem, 2019;19:546-556.
12. Öner D, Ghosh M, Coorens R,Bové H, Moisse M, Lambrechts D, Ameloot M and Godderis L. Induction and recovery of CpG site specific methylation changes in human bronchial cells after long-term exposure to carbon nanotubes and asbestos. Environ Int, 2020:137;105530.
13. Jurmeister P, Schöler A, Arnold A, Klauschen F, Lenze D, Hummel M, Schweizer L, Bläker H, Pfitzner BM, Mamlouk S, Sers CH and Denkert C. DNA methylation profiling reliably distinguishes pulmonary enteric adenocarcinoma from metastatic colorectal cancer. Mod Pathol, 2019;32:855-865.
14. Deutschmeyer V, Breuer J, Walesch SK, Sokol, AM, Graumann J, Bartkuhn M and Boettger T. Epigenetic therapy of novel tumour suppressor ZAR1 and its cancer biomarker function, Clin Epigenetics, 2019:11;182.
15. Alizadeh N, Asadi M and Shanehbandi D. Evaluation of the Methylation of MIR129-2 Gene in Gastric Cancer, J Gastrointest Cancer, 2020;51:267-270.
16. Liu L, Zeng P, Yang S and Yuan Z. Leveraging methylation to identify the potential caU.S.Al genes associated with survival in lung adenocarcinoma and lung squamous cell carcinoma, Oncol Lett, 2020;20:193-200.
17. Zhang J, Luo L, Dong J, Liu M, Zhai D, Huang D, Ling L and Jia X. A prognostic 11-DNA methylation signature for lung squamous cell carcinoma. J Thorac Dis, 2020;12:2569-2582.
18. Böttcher J, Dilworth D, Reiser U, Neumuller R, Schleicher M, Petronczki M, Zeeb M, Mischerikow N, Allali hassani A and Szewczyk M. Fragment-based discovery of a chemical probe for the PWWP1 domain of NSD3. Nat Chem Biol, 2019;15:822-829.
19. Alnabulsi SM and Al-Shar'i NA. Hit identification of SMYD3 enzyme inhibitors using structure-based pharmacophore modeling. Future Med Chem, 2019;11:1107-1117.
20. Kazemi E, Zargooshi J, Kaboudi M, Heidari P, Kahrizi D, Mahaki B, Mohammadian Y, Khazaei H, Ahmed K. A genome-wide association study to identify candidate genes for erectile dysfunction. Brief Bioinforma 2021;22(4):bbaa338. https://doi.org/10.1093/bib/bbaa338.
21. Tourang M, Fang L, Zhong Y, Suthar R. Association between Human Endogenous Retrovirus K gene expression and breast cancer. Cell Mol Biomed Rep 2021; 1(1): 7-13.
22. Tourang M, Fang L, Zhong Y, Suthar R. Association between Human Endogenous Retrovirus K gene expression and breast cancer. Cell Mol Biomed Rep 2021; 1(1): 7-13.
23. Tang X, Jiang J, Zhu J, He N and Tan J. HOXA4-regulated miR-138 suppresses proliferation and gefitinib resistance in NSCLC.Mol Genet Genomics, 2019; 294:85-93.