Am J Transl Res 2025;17(11):8939-8952
www.ajtr.org /ISSN:1943-8141/AJTR0168122

Original Article

Application of circulating tumor DNA methylation characteristics in early diagnosis and prognosis monitoring of lung cancer

Panpan Zhao, Ran Hu

Department of Medical Laboratory, People's Hospital of Chongqing Liangjiang New Area, Chongqing 401121, China

Received August 15, 2025; Accepted November 15, 2025; Epub November 15, 2025; Published November 30, 2025

Abstract: Objective: To investigate the value of circulating tumor DNA (ctDNA) methylation for early detection and prognostic monitoring of lung cancer. Methods: A retrospective analysis was conducted on the DNA methylation test results of 150 lung cancer patients and 100 patients with benign lung lesions enrolled from January 2021 to December 2023. Quantitative methylation fluorescence analysis and bisulfite sequencing were used to measure ctDNA gene methylation, with ROC curves assessing early-stage diagnostic value. Lung cancer patients were followed for 1 year, then grouped by survival status. Cox regression identified poor prognosis factors, and ROC curves evaluated ctDNA methylation's prognostic value. An external cohort of 80 lung cancer patients from May 2022 to December 2023 validated the model using ROC and calibration curves. Results: The positivity rates of SHOX2, CDO1, and SOX17 were significantly higher in lung cancer patients than those in controls (all $P<0.05$). The combined diagnostic model of these genes had a higher AUC than single-marker tests ($Z = 2.363$, 2.157, 2.061, all $P<0.05$). Compared with the favorable prognosis group, the unfavorable prognosis group had a higher proportion of stage III-IV tumors and higher positivity rates of SHOX2, CDO1, SOX17, and HOXA7 (all $P<0.05$). SHOX2, CDO1 and SOX17 were identified as independent poor-prognosis risk factors (all $P<0.05$), and their combined prognostic assessment outperformed single-marker tests ($Z = 3.316$, 2.394, 2.696, all $P<0.05$). Kaplan-Meier analysis showed that patients with negative methylation of SHOX2, CDO1, and SOX17 had longer survival (Log-rank $\chi^2 = 6.273$, 4.524, 4.364, $P<0.001$). The model showed good predictive performance (AUC = 0.773), and external validation confirmed its accuracy (AUC = 0.682). Conclusions: Abnormal methylation of SHOX2, CDO1, and SOX17 is prevalent in lung cancer, potentially serving as biomarkers for early diagnosis and prognostic monitoring.

Keywords: Circulating tumor DNA, methylation characteristics, lung cancer, diagnosis, prognosis

Introduction

Lung cancer is one of the most common malignant tumors, with persistently high incidence and mortality rates, imposing a heavy burden on patients and their families. It is reported that there are approximately 2.2 million new cases of lung cancer and 1.8 million deaths annually [1]. Patients with early-stage lung cancer often present with no typical clinical symptoms, and most cases are not diagnosed until an advanced stage, which significantly increases the difficulty of treatment and results in a poor prognosis [2]. Early diagnosis of lung cancer is crucial to enhance patient outcomes. In clinical practice, conventional diagnostic such as chest X-ray and CT scans have limitations in detecting early-stage lung cancer, particularly in identifying tiny lesions [3]. The gold standard for diagnosing lung cancer is histopathological examination, but it has drawbacks including invasiveness and potential for false negatives due to sampling limitations [4]. DNA methylation is an epigenetic modification that primarily occurs on cytosine residues within CpG dinucleotides, whose complex regulation involves various enzymes and proteins [5]. Previous studies have demonstrated that abnormal methylation of specific genes can be detected in the early stages of lung cancer [6, 7]. These abnormally methylated genes may be involved in key biological processes such as proliferation, inva-

https://doi.org/10.62347/ELJ06418

sion, and metastasis of lung cancer cells. However, currently, there is a lack of sufficient evidence to support the application of circulating tumor DNA methylation profiles in the early diagnosis and prognostic monitoring of lung cancer. If these research results can be applied in clinical practice, it is expected to advance non-invasive early screening technology, address the missed diagnosis of small lesions in traditional imaging, establish a dynamic prognostic monitoring system based on methylation characteristics, predict the risk of recurrence through changes in circulating tumor DNA methylation levels after surgery, combine methylation markers with clinical staging to achieve accurate risk stratification of patients, and provide a new path to improve the overall diagnosis and treatment of lung cancer. The aim of this study is to explore the application of circulating tumor DNA methylation characteristics in the early diagnosis and prognosis monitoring of lung cancer, providing new molecular markers and theoretical basis for the early diagnosis and prognosis evaluation of lung cancer with the goal of improving the diagnosis and treatment of lung cancer patients.

Materials and methods

General information

Clinical data of 150 lung cancer patients admitted to People's Hospital of Chongqing Liangjiang New Area from January 2021 to December 2023 were retrospectively analyzed (study group), and another 100 cases with benign lung lesions were enrolled as the control group. This study has been approved by the Ethics Committee of People's Hospital of Chongqing Liangjiang New Area. The study was conducted in strict adherence to the principles of the Declaration of Helsinki.

Inclusion criteria: (1) Conforming to the diagnostic criteria of Chinese Society of Oncology Clinical Practice Guidelines for Lung Cancer [8]: (i) Chest CT showed space-occupying lesions in the lungs with typical imaging features of lung cancer such as lobulation, spiculation, and pleural traction; (ii) Pathological diagnosis was confirmed by pathological section and staining after obtaining lesion tissues via biopsy or other methods; (2) No anti-tumor therapy such as radiotherapy, chemotherapy, or targeted therapy had been administered before enrollment; (3) First diagnosis and treatment in the above hospital; (4) Complete clinical data; (5) Good compliance and willingness to actively cooperate with the study.

Exclusion criteria: (1) Concurrent with other malignant tumors; (2) Concurrent with infectious diseases such as hepatitis B; (3) Concurrent with autoimmune diseases such as systemic lupus erythematosus; (4) With severe dysfunction of liver, kidney and other organs; (5) Concurrent with coagulation dysfunction; (6) Previous history of pulmonary surgery; (7) Inability to complete follow-up.

Methods

General clinical data were collected from the hospital medical record system, including gender, age, body mass index, education level, place of residence, comorbidity history, history of smoking and alcohol consumption, blood pressure, heart rate, tumor stage, differentiation degree, and tumor diameter.

Collection of laboratory indicators: 5 ml of fasting venous blood was collected in the early morning on the day after admission, which was centrifuged at 3000 r/min for 15 minutes. Total cholesterol and triglycerides were measured using an automatic biochemical analyzer (AU5800, Beckman Coulter, USA). Hemoglobin levels were determined using an automatic hematology analyzer (BC-20s, Mindray Medical; China). Serum levels of sodium, potassium, calcium, and phosphorus were detected using an automatic chemiluminescence immunoassay analyzer (AUTOAE 2100, Chongqing Cosmai Biotechnology; China).

Collection of DNA methylation indicators: Cell-free tumor DNA (ctDNA) in plasma was extracted using the QIAamp Circulating Nucleic Acid Kit from QIAGEN, strictly following the kit instructions. The extracted ctDNA was dissolved in 30 μl elution buffer and stored at -20°C for later use. The concentration and purity of ctDNA were determined using a NanoDrop 2000 ultra-micro spectrophotometer, and an A260/A280 ratio between 1.8 and 2.0 was considered qualified. The extracted ctDNA was subjected to bisulfite modification using the EZ DNA Methylation-Gold Kit according to the instructions. The modified ctDNA was used as a template for amplification. PCR reaction condi-

tions: initial denaturation at 95°C for 5 min, denaturation at 95°C for 30 s, annealing at 60°C for 30 s, extension at 72°C for 30 s, with a total of 40 cycles; final extension at 72°C for 5 min. The amplified products were separated by 2% agarose gel electrophoresis, and the results were observed under a gel imaging system. If a specific band was amplified by the methylated primer and no band was amplified by the unmethylated primer, the gene was judged to be methylation-positive; if a specific band was amplified by the unmethylated primer and no band was amplified by the methylated primer, the gene was judged to be methylation-negative; if bands were amplified by both, the gene was judged to be partially methylated. In this study, PAX5, SHOX2, CDO1, SOX17, HOXA7, GATA4, and GATA5 were finally selected as lung cancer-related DNA methylation indicators.

Methylation-specific fluorescence (MethyLight) quantitative analysis was performed: Methylation-specific primers and probes were designed, with COL2A1 as the internal reference gene (its primers and probes were designed from regions lacking CpG dinucleotides in the promoter sequence). PCR reaction conditions: the reaction volume was 10 μl, with initial denaturation at 95°C for 4 min, followed by 35 cycles of denaturation at 95°C for 15 s and annealing/extension at 60°C. Clonal bisulfite sequencing: To analyze the methylation status of CpGs in specific gene promoters (SHOX2, CDO1, SOX17), 5 methylation-positive tumor tissue samples (positive methylation ratio [PMR] >4) and 3 normal samples were randomly selected for clonal bisulfite sequencing. The regions covered by each gene promoter were determined using methyl fluorescence analysis. The amplified products were subjected to gel cutting and purification using a rapid gel extraction kit. The purified PCR products were ligated into the T vector using the InsTAclone™ PCR Cloning Kit, followed by transformation and cloning into Escherichia coli DH5α. Positive transformed clones were obtained through blue-white screening and purified using the NucleoSpin® Plasmid DNA Purification Kit. 10-12 cloned colonies were picked and sequenced on a 3730XL Genetic Analyzer.

Positive determination criteria: Methylation positivity in the gene promoter region was defined as specific fluorescence signals amplified by methylation-specific primers (Ct value ≤ 35) with no signals from unmethylated primers (Ct value >35 or undetectable Ct value). Only non-methylated primers amplified signal was considered negative. Both produce effective amplification (Ct values ≤ 35) was judged as partial methylation. In clone bisulfite sequencing, methylation rate of CpG sites in the promoter region of the target gene exceeding 40% (i.e. methylation CpG number/total CpG number × 100% ≥ 40%) was regarded as methylation positive. These criteria were determined based on the threshold difference in methylation levels between tumor and normal tissues.

Follow-up: Treatment regimens for all patients in the study group were formulated in accordance with guidelines [9], including surgical treatment, induction chemotherapy, radiotherapy, targeted therapy, and immunotherapy. All patients were followed up for 1 year after discharge via outpatient rechecks and telephone interviews. During the follow-up period, the patient survival status was recorded, and the 1-year survival rate was calculated. Based on the follow-up results, patients were divided into the poor prognosis group (deceased) and the good prognosis group (survived).

Research flowchart

Sample size calculation was based on the formula: $n = Z^2_{\alpha/2}\, p(1\text{-}p)/d^2$, where n, d, and p were estimated values of sample size, allowable error, and population rate π, respectively. Assuming P = 0.5, taking 0.05 on both sides of α and z, and 0.1 on d, the sample size was at least 97 cases. However, the actual sample size was adjusted according to clinical conditions, and ultimately 150 cases were included, with an additional 80 patients included as an external validation cohort (**Figure 1**).

Statistical methods

SPSS 26.0 (IBM Corporation, Amonk, New York, USA) was employed for data analysis in this study. Categorical data were expressed as count and percentage, and analyzed with Chi-square test. Quantitative data were tested using Kolmogorov-Smirnov test. Those normally distributed were presented as mean ± standard deviation (SD) and independent sample *t*-test was used for inter-group comparison.

Multivariate Cox regression analysis was performed to identify the factors influencing poor prognosis in lung cancer patients, and forest

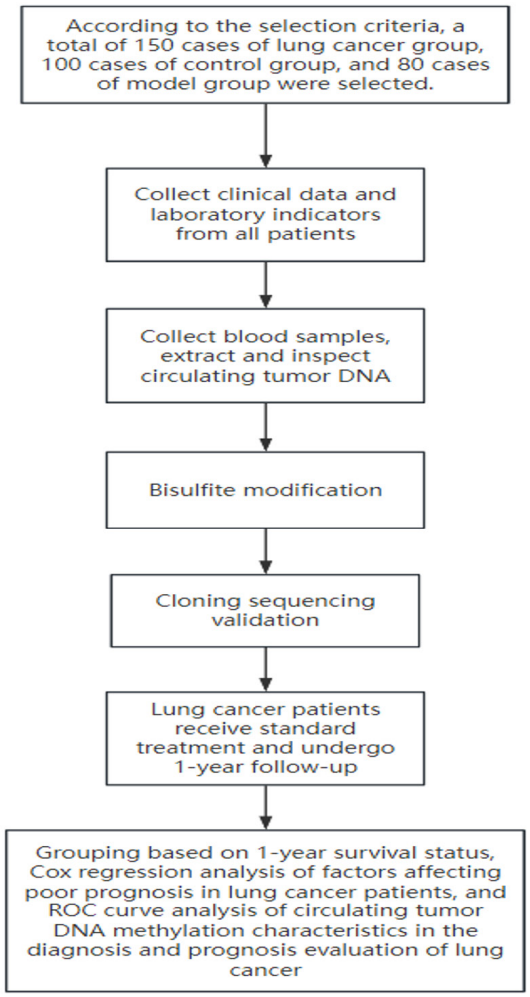

Figure 1. Research flowchart.

plots were generated using GraphPad Prism 8.0 software. Receiver operating characteristic (ROC) curve analysis was conducted to evaluate the utility of circulating tumor DNA methylation signatures in the diagnosis and prognostic assessment of lung cancer. Survival curves were plotted using the Kaplan-Meier method, and the Log-rank test was applied to compare survival differences among patients in different groups. A *P* value <0.05 was considered statistically significant.

Results

Comparison of baseline data between the study and control groups

There was no significant difference in clinical data between the study group and the control group (all *P*>0.05), see **Table 1**.

Comparison of the DNA methylation positive rate between the study and control groups

The positive rate of SHOX2, CDO1 and SOX17 in the study group was higher than that in the control group (all *P*<0.05), as shown in **Table 2**.

Value of circulating tumor DNA methylation characteristics in lung cancer diagnosis

The ROC curve analysis (**Figure 2**) showed that the AUC value of SHOX2, CDO1 and SOX17 combined diagnosis of lung cancer was higher than that of single detection (*Z* = 2.363, 2.157, 2.061, all *P*<0.05, **Table 3**).

Ker announced the results of sulfite sequencing

According to kegliosulfite sequencing, the SHOX2 gene promoter contains 1 CpG site, the CDO1 gene promoter contains 5 CpG sites, and the SOX17 gene promoter contains 3 CpG sites, see **Table 4** for the sequence.

Follow-up results

According to the follow-up data, there were 63 deaths in 150 patients, with a mortality of 42.00%, and 87 patients survived with a 1-year survival rate of 58.00%. Thus, 87 patients were divided into a good prognosis group and 63 patients were divided into a poor prognosis group.

Comparison of general data between the good prognosis and poor prognosis groups

The proportion of stage III-IV tumors, SHOX2, CDO1, SOX17 and HOXA7 positivity rate in the poor prognosis group was higher than those in the good prognosis group (all *P*<0.05), as shown in **Table 5**.

Factors affecting the poor prognosis of lung cancer patients

Cox regression analysis was performed using these significant factors, with patient mortality status (yes = 1, no = 0) as the outcome variable, tumor staging, SHOX2, CDO1, SOX17, and HOXA7 as covariates, and survival time as the time variable. The results demonstrated that SHOX2, CDO1, and SOX17 were independent prognostic risk factors for poor prognosis in

Table 1. Comparison of baseline data between the study and control groups

Variable	The study group (n = 150)	The control group (n = 100)	χ^2/t	P
Sex (m/f)	91/59	64/36	0.171	0.679
Age (years)	62.16±9.76	61.62±9.20	0.325	0.726
Body mass index (kg/m^2)	21.86±1.94	21.73±1.85	0.242	0.810
Educational attainment			0.822	0.365
High school and below	97 (64.67)	59 (59.00)		
College and above	53 (35.33)	41 (41.00)		
Current residence			0.433	0.511
Cities and towns	115 (76.67)	73 (73.00)		
Countryside	35 (23.33)	27 (27.00)		
Hypertension history	79 (52.67)	56 (56.00)	0.268	0.604
Diabetes history	78 (52.00)	55 (55.00)	0.217	0.641
Hyperlipidemia history	46 (30.67)	34 (34.00)	0.306	0.580
Smoking history	75 (50.00)	49 (49.00)	0.024	0.878
Drinking history	63 (42.00)	45 (45.00)	0.220	0.639

Table 2. Comparison of the DNA methylation positive rate between the study group and the control group

Variable	The study group (n = 150)	The control group (n = 100)	χ^2	P
PAX5 (%)	75 (50.00)	53 (53.00)	0.954	0.329
SHOX2 (%)	92 (61.33)	39 (39.00)	54.199	<0.001
CDO1 (%)	94 (62.67)	42 (42.00)	48.447	<0.001
SOX17 (%)	89 (59.33)	42 (42.00)	32.648	<0.001
HOXA7 (%)	104 (69.33)	65 (65.00)	2.998	0.083
GATA4 (%)	87 (58.00)	59 (59.00)	0.124	0.724
GATA5 (%)	65 (43.33)	42 (42.00)	0.161	0.688

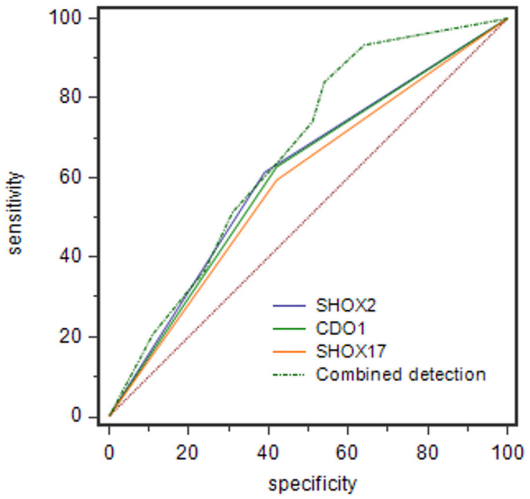

Figure 2. ROC curve of the value of circulating tumor DNA methylation characteristics in lung cancer diagnosis.

lung cancer patients (all P<0.05, **Table 6**). The significant variables identified through Cox regression were visualized in a forest plot based on hazard ratio values (HR), as illustrated in **Figure 3**. A HR <1 indicates a negative correlation between the factor and adverse prognosis, while a HR>1 suggests a positive correlation.

Value of DNA methylation characteristics in lung cancer prognosis assessment

The Cox regression model was directly applied to integrate data for modeling, with risk scores predicted by the Cox model as the test variable. The analysis evaluated the combined assessment of SHOX2, CDO1, and SOX17 for predicting 1-year mortality in lung cancer patients. Results showed that the AUC values (**Figure 4**) for predicting poor prognosis from the combined evaluation of SHOX2, CDO1, and SOX17 were significantly higher than those from single-item testing (Z = 3.316, 2.394, 2.696; *all* P<0.05), as shown in **Table 7**.

Table 3. The value of circulating tumor DNA methylation characteristics in lung cancer diagnosis

Variable	AUC	Standard error	95% CI	Youden index	Sensitivity (%)	Specificity (%)
SHOX2	0.612*	0.032	0.548-0.672	0.223	61.33	61.00
CDO1	0.603*	0.032	0.540-0.664	0.207	62.67	58.00
SOX17	0.587*	0.032	0.523-0.648	0.173	59.33	58.00
Combined detection	0.665	0.036	0.603-0.723	0.300	84.00	46.00

Note: *P<0.05 compared with combined detection.

Table 4. Gene sequence

Gene name	Number of CpG loci	Quantitative analysis of methylation fluorescence
SHOX2	1	Chromosome position: chrl3:158080000-158105000sequence GGCGCGACATTGGTGCTGGCGTTGGCGTCACAGACCCAGGGCTGCGGCGT-GCTTTTTGGCGGCGCGATATTGGTGTTGGCGTTGGCGTTATAGATTTAGGGTTGCGGC-GTGTTTTTTGGT
CDO1	5	Chromosome position: chrl5:115805000-115807500 sequence CCTGGCCCA-CACGGGAGCATATTCTGTCTTGATTCTTCTGCACTGGCCCTCCCAAAGTCATTTGGTT-TATACGGGAGTATATTTTGTTTTGATTTTTTTGTATTGGTTTTTTTAAAGTTA
SXO17	3	Chromosome position: chrl8:54457935-54460892 sequence GTTTA-AAATTAGGGGTGTGTAGCAATTCAAAACCAAAAATACTCTCGTTTAAAATTAGGGGTGTG-TAGCCAATTCAAAACCAAAAATACTCTCTTTAAAATTAGGGGTGTGTAGCAATTCAAAAC-CAAAAATACTCTC
Gene name	Fluorescence quantitative analysis of gene methylation and bisulfite sequencing region sequence	
SHOX2	GCTTTTTGGCTTTCAGTCTGAGATCGGCGATGCTGGAGTTCTTGCTGGTGGTCTTGGCGGCTGCTGCGGC-CGCCACTACCGAGGCGGCGGAAGCCGAATCCGCGGCCAGCGTGGCGAGCGGCAGTCCGAAGGGCG-GTGCTGGGAACATCATGTAGGGCGCGTGCGCGGCCAGGTGCGGATGCAGGTGGTGGTGCGCGTGCGCCA-CAGCGCTGTCCAGCTGCAGCTGCGCCTGAACCTGAAAGGACAAGGGCGTCACGTTGCAATGACTATCCTAGGGT-GACAACAGAATAGAAACAGAGCATTA	
CDO1	CTGGTGCAGGATGCGGATCAGATCAGCCAGGGTCCGTGGCTTCAGCACTTCGGTCTGTTCCATCTC-GTGGGGAGCTGGCTGCGCGCGCGTCTCACTGCTGGGCTGCGGTGGAGGAGCTGAGCGAGC-CAAGGAGCTGGGGGCGAGGGAGCCTAACAGCCCGCTAGACCGCTAAGCAGACACACACGCACAAACCCAG-CATTAGAGTGCCGAAACGTAAGGATGTCGTCGCAGAGACAGCAAGAGACCCACCCCCAGGCCCCTGGCAGCG-CAGTGGA	
SXO17	ATGACTCCGGTGTGAATCTCCCCGACAGCCACGGGGCCATTTCCTCGGTGGTGTCCGACGCCAGCTCCGCGG-TATATTACTGCAACTATCCTGACGTGTGACAGGTC	

Kaplan-meier survival curve analysis

The Kaplan-Meier curve showed that the survival time of patients with negative SHOX2, CDO1 and SOX17 was significantly longer than that of patients with positive SHOX2, CDO1 and SOX17 (Log-rank χ^2 = 6.273, 4.524, 4.364, all P<0.001), as shown in **Figure 5**.

Comparison of clinical data between the model group and external validation group

There were no statistically significant difference in general data between the model group and the external validation group (all P>0.05), indicating good consistency between the model group and the external validation group, as shown in **Table 8**.

Line chart for predicting the risk of poor prognosis in lung cancer patients based on Cox regression analysis

Based on Cox regression analysis, the risk prediction column chart for poor prognosis in lung cancer patients includes three variables, namely SHOX2, CDO1, and SOX17. SHOX2 has a higher weight in the column chart, indicating its important role in risk prediction, as shown in **Figure 6**.

ROC curve and calibration curve were used to evaluate the predicted values and accuracy of the model group. The ROC curve of the model group shows an AUC value of 0.773, indicating that the model had high value in predicting the risk of poor prognosis in lung cancer patients

Table 5. Comparison of general data between the groups with good prognosis and poor prognosis

Variable	poor prognosis (*n* = 63)	good prognosis (*n* = 87)	χ^2/t	*P*
Sex (m/f)	37/26	54/33	0.171	0.679
Age (years)	62.49±10.22	64.92±9.48	0.325	0.726
Body mass index (kg/m^2)	21.83±1.86	21.91±2.10	0.242	0.810
Educational attainment			0.363	0.547
High school and below	39 (61.90)	58 (66.67)		
College and above	24 (38.10)	29 (33.33)		
Current residence			0.259	0.611
Cities and towns	47 (74.60)	68 (78.16)		
Countryside	16 (25.40)	19 (21.84)		
Hypertension history	36 (57.14)	43 (49.43)	0.873	0.350
Diabetes history	34 (53.97)	44 (50.57)	0.168	0.681
Hyperlipidemia history	19 (30.16)	27 (31.03)	0.013	0.909
Smoking history	33 (52.38)	42 (48.28)	0.246	0.620
Drinking history	28 (44.44)	35 (40.23)	0.266	0.606
Systolic pressure (mmHg)	126.1±19.64	125.29±20.68	0.242	0.809
Diastolic pressure (mmHg)	82.32±9.51	80.80±9.21	0.984	0.327
Heart rate (beats/min)	94.70±7.64	93.92±8.74	0.568	0.571
Stage of tumor			5.432	0.020
Stage I-II	39 (61.90)	44 (50.57)		
Stage III-IV	24 (38.10)	43 (49.43)		
The degree of differentiation			0.365	0.833
poorly differentiated	21 (33.33)	27 (31.03)		
moderately differentiated	23 (36.51)	36 (41.38)		
well-differentiated	19 (30.16)	24 (27.59)		
Hemoglobin (g/L)	122.63±9.33	122.68±9.31	0.681	0.497
total cholesterol (mmol/L)	4.18±0.83	4.19±0.74	0.078	0.938
Triglyceride (mmol/L)	1.09±0.37	1.17±0.39	1.267	0.207
Blood potassium (mmol/L)	4.52±0.54	4.58±0.53	0.679	0.498
Blood sodium (mmol/L)	140.03±2.85	140.54±2.96	1.058	0.292
Blood calcium (mmol/L)	2.41±0.95	2.44±0.84	0.204	0.838
Blood phosphorus (mmol/L)	1.26±0.18	1.31±0.17	1.734	0.085
PAX5 (%)	33 (52.38)	42 (48.28)	0.246	0.620
SHOX2 (%)	46 (73.02)	46 (52.87)	6.251	0.012
CDO1 (%)	46 (73.02)	48 (55.17)	4.973	0.026
SOX17 (%)	44 (69.84)	45 (51.72)	4.971	0.026
HOXA7 (%)	47 (74.60)	51 (58.62)	4.121	0.042
GATA4 (%)	35 (55.55)	52 (59.77)	0.266	0.605
GATA5 (%)	29 (46.03)	36 (41.38)	0.322	0.570

(**Figure 7A**). Bootstrap sampling was conducted 1000 times, with gray lines representing the model's predicted values and black lines representing actual observed values. The calibration curve showed that the model exhibited good calibration ability, and the predicted values were consistent with the actual results (**Figure 7B**). The goodness of fit test shows a *P* value of 0.993, indicating good fitting, and the calibration analysis revealed an average absolute error of 0.035, indicating good accuracy.

For the external validation group, the ROC curve had an AUC of 0.682, reflecting good evaluation performance consistent with the model group (**Figure 7C**). The calibration curve also indicated good calibration with consistent predicted and actual results (**Figure 7D**). With a

Table 6. Factors affecting the poor prognosis of lung cancer patients

Variable	*B*	*SE*	*Wald*	*P*	*HR*	*95% CI*
Stage of tumor	-0.352	0.261	1.818	0.178	0.703	0.422-1.173
SHOX2	0.625	0.287	4.040	0.029	1.868	1.064-3.279
CDO1	0.572	0.285	3.876	0.044	1.772	1.014-3.097
SOX17	0.545	0.277	0.410	0.049	1.725	1.002-2.967
HOXA7	0.178	0.292	0.410	0.522	1.205	0.680-2.136

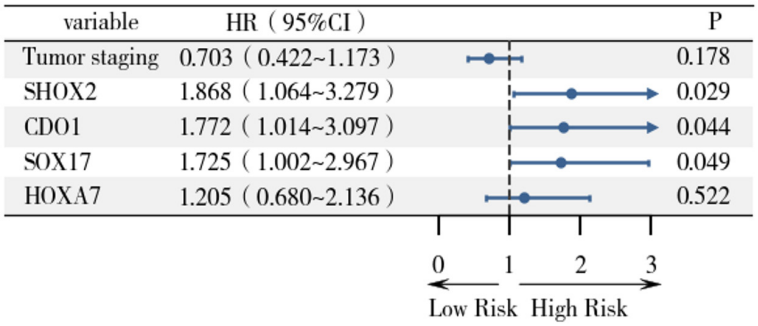

Figure 3. Cox regression significant variable forest.

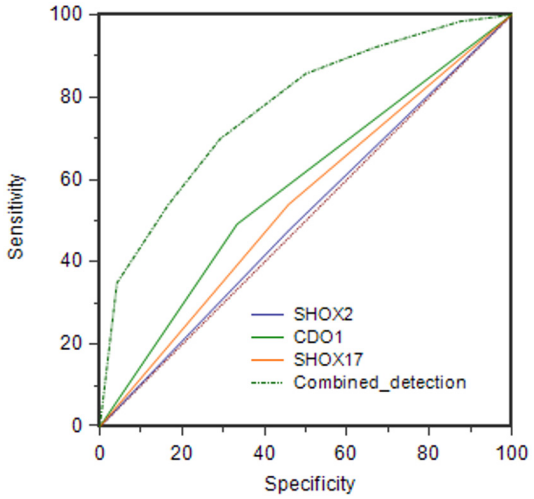

Figure 4. ROC curve of the value of DNA methylation characteristics in the prognostic evaluation of lung cancer.

goodness-of-fit *P*-value of 0.464 and MAE of 0.039, the model's accuracy was further verified.

Discussion

ctDNA refers to DNA fragments released into the bloodstream during tumor cell apoptosis, necrosis, or active secretion. During tumor development, these fragments undergo characteristic alterations, including gene mutations, copy number variations, and abnormal methylation modifications [10]. As a crucial epigenetic modification mechanism, DNA methylation plays a pivotal role in gene expression regulation, cellular differentiation, embryonic development, and tumorigenesis [11]. In tumor progression, DNA methylation patterns change with reduced overall genomic methylation and hypermethylation of specific CpG islands in gene promoters. These alterations dysregulate tumor-related gene expression, thereby promoting cancer cell proliferation, invasion, and metastasis [12].

This study innovatively explores the application value of DNA methylation characteristics in early lung cancer diagnosis and prognostic monitoring. By screening and validating novel ctDNA methylation biomarkers for lung cancer, we aim to provide new strategies for clinical management, ultimately improving early detection rates and patient survival.

The study findings indicate that the positive rates of SHOX2, CDO1, and SOX17 in the study group were higher than those in the control group. The AUC for combined diagnosis of lung cancer was greater than that of single-gene detection, suggesting that these three genes can serve as biomarkers for early lung cancer diagnosis. This may be attributed to SHOX2 (a key developmental regulatory gene) undergoing abnormal methylation leading to gene silencing, which disrupts the Hedgehog signaling pathway, impairs cell cycle progression and differentiation balance, and promotes malignant transformation [13]. Cystine dioxygenase (encoded by CDO1) maintains intracellular redox balance; hypermethylation in its promoter region may suppress transcription, reducing

Table 7. Value of DNA methylation characteristics in lung cancer prognosis assessment

Variable	AUC	Standard error	95% CI	Youden index	Sensitivity (%)	Specificity (%)
SHOX2	0.509*	0.061	0.399-0.618	0.018	47.62	54.17
CDO1	0.579*	0.059	0.469-0.684	0.159	49.21	66.67
SOX17	0.541*	0.061	0.430-0.648	0.081	53.97	54.17
Combined detection	0.773	0.054	0.670-0.856	0.30/	69.84	60.92

Note: *P<0.05 compared with combined detection.

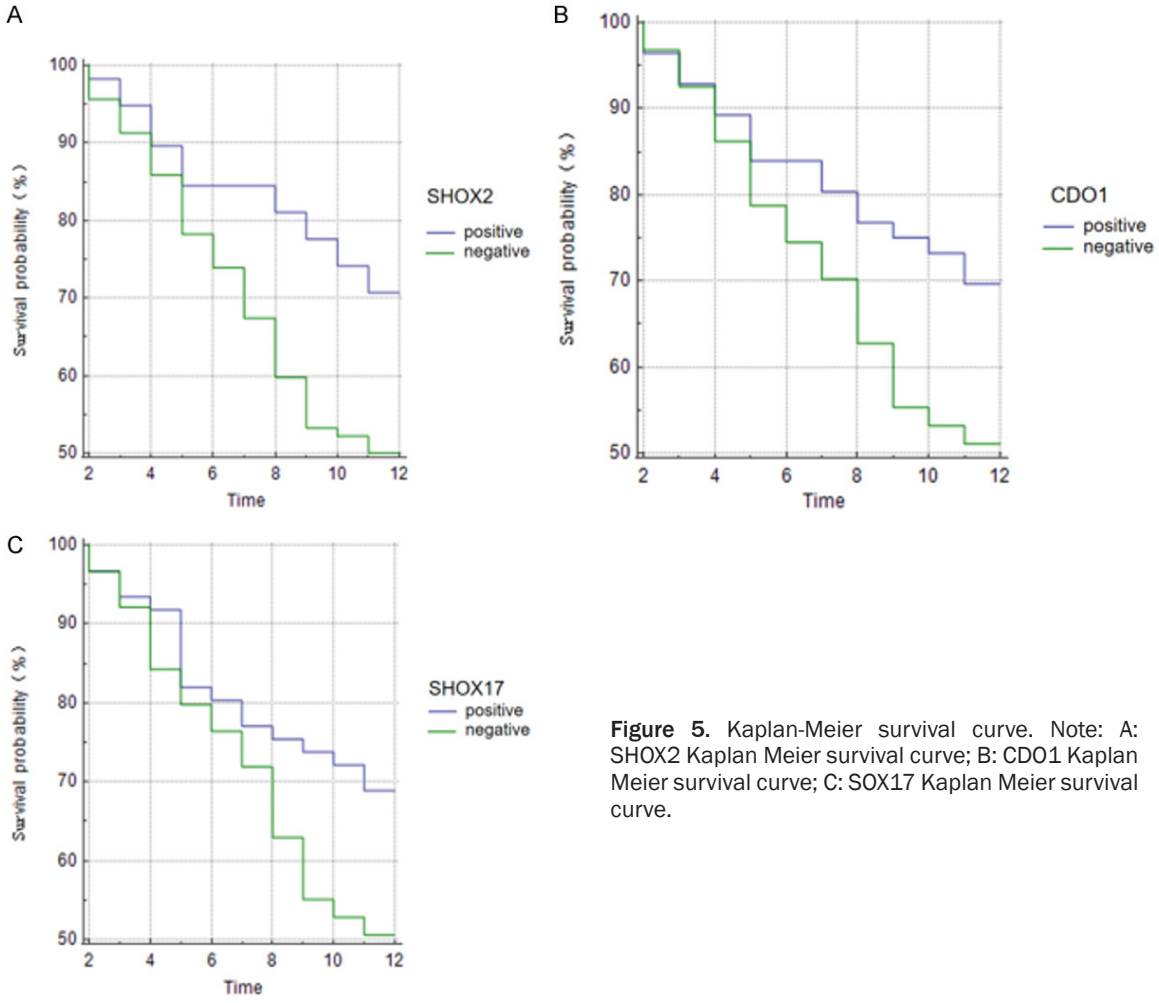

Figure 5. Kaplan-Meier survival curve. Note: A: SHOX2 Kaplan Meier survival curve; B: CDO1 Kaplan Meier survival curve; C: SOX17 Kaplan Meier survival curve.

cells' ability to eliminate reactive oxygen species, accumulating DNA oxidative damage, and accelerating tumorigenesis [14]. SOX17 (involved in embryonic development and tumor suppression) may have altered methylation, disrupting normal signaling pathways and promoting carcinogenesis [15]. In early lung cancer, abnormally methylated three genes are released into the bloodstream via ctDNA, increasing detection positivity. Single-gene methylation testing may be confounded by tumor heterogeneity, leading to false negatives or positives. Combined detection integrates methylation information from multiple targets, reflects the molecular characteristics of lung cancer, reducing individual variations and detection errors, thereby improving diagnostic accuracy.

Xie et al. [16] demonstrated that SHOX2 holds diagnostic value in early-stage lung cancer, as its regulation can influence tumor development and metastasis. Li et al. [17] proposed that DNA methylation profiles may assist in predict-

Table 8. Comparison of clinical data between model group and external validation group [($\bar{x}$±sd), *n* (%)]

Variable	Model Group (*n* = 150)	validation group (*n* = 80)	χ^2/t	*P*
Sex (m/f)	91/59	46/34	0.171	0.679
Age (years)	64.59+9.28	64.33±9.15	0.203	0.839
Body mass index (kg/m^2)	21.86±1.93	21.75±1.84	0.418	0.676
Educational attainment			0.106	0.745
High school and below	97 (64.67)	50 (62.50)		
College and above	59 (39.33)	30 (37.50)		
Current residence			0.486	0.486
Cities and towns	115 (76.67)	58 (72.50)		
Countryside	35 (23.33)	22 (27.50)		
History of hypertension	79 (52.67)	45 (56.25)	0.270	0.604
History of diabetes	78 (52.00)	39 (48.75)	0.221	0.639
History of hyperlipidemia	46 (30.67)	28 (35.00)	0.449	0.503
Smoking history	75 (50.00)	37 (46.25)	0.284	0.589
Drinking history	63 (42.00)	38 (47.50)	0.641	0.424
Systolic pressure (mmHg)	125.73±19.92	125.44±19.73	0.106	0.916
Diastolic pressure (mmHg)	81.62±9.32	81.77±9.46	0.116	0.908
Heart rate (beats/min)	94.15±7.93	94.55±8.06	0.362	0.717
Stage of tumor			0.464	0.496
Stage I-II	83 (55.33)	48 (60.00)		
Stage III-IV	67 (44.67)	32 (40.00)		
The degree of differentiation			0.248	0.883
poorly differentiated	48 (32.00)	25 (31.25)		
moderately differentiated	59 (39.33)	34 (42.50)		
well-differentiated	43 (28.67)	21 (26.25)		
Hemoglobin (g/L)	122.65±9.30	122.26±9.13	0.305	0.761
total cholesterol (mmol/L)	4.19±0.78	4.20±0.76	0.009	0.926
Triglyceride (mmol/L)	1.12±0.38	1.15±0.37	0.575	0.566
Blood potassium (mmol/L)	4.55±0.52	4.59±0.55	0.679	0.498
Blood sodium (mmol/L)	140.37±2.89	140.13±2.75	1.058	0.292
Blood calcium (mmol/L)	2.42±0.91	2.37±0.86	0.204	0.838
Blood phosphorus (mmol/L)	1.29±0.18	1.30±0.17	1.734	0.085
PAX5 (%)	75 (50.00)	43 (53.75)	0.294	0.588
SHOX2 (%)	92 (61.33)	53 (66.25)	0.541	0.462
CDO1 (%)	94 (62.67)	55 (68.75)	0.846	0.359
SOX17 (%)	99 (66.00)	49 (61.25)	0.513	0.474
HOXA7 (%)	98 (65.33)	58 (72.50)	1.228	0.267
GATA4 (%)	87 (58.00)	43 (53.75)	0.383	0.536
GATA5 (%)	65 (43.33)	39 (48.75)	0.618	0.432

ing cancer prognosis. Our findings reveal that the proportion of stage III-IV tumors and positive rates of SHOX2, CDO1, SOX17, and HOXA7 were significantly higher in the poor prognosis group compared with the good prognosis group. Notably, SHOX2, CDO1, and SOX17 are independent prognostic risk factors for lung cancer, indicating that their methylation status is a critical molecular marker for prognosis assessment. The underlying mechanism may be that SHOX2 (a cell proliferation regulator) activates downstream oncogenic signaling pathways via abnormal methylation, accelerating malignant proliferation and tissue infiltration, and incre-

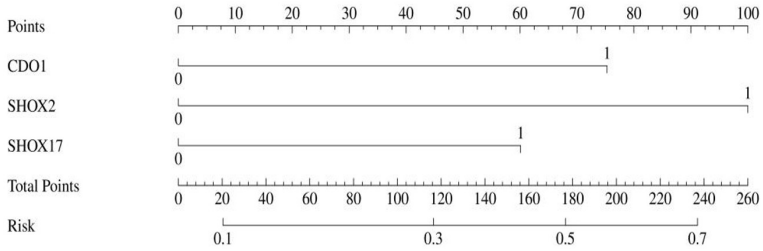

Figure 6. Line chart for predicting the risk of poor prognosis in lung cancer patients based on Cox regression analysis.

asing the risk of advanced staging (III-IV) and poor prognosis [18]. CDO1, a classic tumor suppressor gene, is silenced by methylation, leading to protein deficiency, impairing its inhibitory effect on abnormal cell proliferation and apoptosis induction, enabling tumor cells to evade immune surveillance and therapeutic interventions, and enhancing drug resistance and recurrence [19]. Abnormal methylation of SOX17, critical for cell differentiation and signaling regulation, disrupts intercellular adhesion, facilitating metastasis through the hematogenous or lymphatic systems [20]. Hypermethylation of HOXA7 may regulate tumor angiogenesis-related genes, providing nutritional support for tumor growth and accelerating progression [21].

The hypoxic and inflammatory microenvironment of advanced tumors may further induce or exacerbate gene methylation disorders, forming a vicious cycle: abnormal methylation → tumor progression → worsening microenvironment → further abnormal methylation [22]. As independent prognostic factors, the methylation status of SHOX2, CDO1, and SOX17 is not affected by clinical factors such as age and comorbidities, reflecting the malignant biological behavior of tumor cells. However, tumor staging showed no significant association with prognosis in Cox regression analysis, which may be due to insufficient sample size diluting the prognostic impact of staging [23]. Jung et al. [23] demonstrated that SHOX2 methylation can predict postoperative mortality risk, with higher methylation levels indicating poorer outcomes. Harada et al. [24] reported that high CDO1 methylation is associated with poor prognosis, consistent with our findings.

Using risk scores predicted by the Cox model as the test variable, we found that the AUC of combined evaluation of SHOX2, CDO1, and SOX17 for predicting poor prognosis in lung cancer patients was higher than that of single-gene detection. Kaplan-Meier curves showed that patients with negative results for SHOX2, CDO1, and SOX17 had significantly longer survival times than those with positive plasma levels of these three genes. This indicates that combined testing of SHOX2, CDO1, and SOX17 achieves higher accuracy in prognostic assessment of lung cancer patients. The negative status of all three genes correlates with longer survival, making them reliable prognostic indicators.

SHOX2, CDO1, and SOX17 are involved in different stages of lung cancer development, including cell proliferation, inhibition of apoptosis, and regulation of metastasis. Single-gene testing only reflects specific biological characteristics of tumors, while combined testing integrates multi-dimensional molecular information to capture the malignant potential of tumors. Consistent with this, the combined evaluation in the Cox model yielded a higher AUC, reflecting greater accuracy. In patients with positive results, abnormal methylation of SHOX2, CDO1, and SOX17 causes dysfunction, promoting uncontrolled tumor cell proliferation, enhanced anti-apoptotic capacity, and increased metastatic potential, thereby shortening the survival. The combined negative or positive status of these genes more stably reflects the overall molecular phenotype of tumors, reducing potential false positives or false negatives associated with single-gene testing.

This study constructed a prognostic risk model for poor outcomes in lung cancer based on SHOX2, CDO1, and SOX17, and its predictive performance was evaluated using ROC curves. With an AUC of 0.773, the model exhibited good predictive ability. Findings further validated by the external validation group confirmed the model's universality and reliability. The consistency of predictive factors between the external validation and model groups supports its clinical application. In clinical practice, combined testing of SHOX2, CDO1, and SOX17 pro-

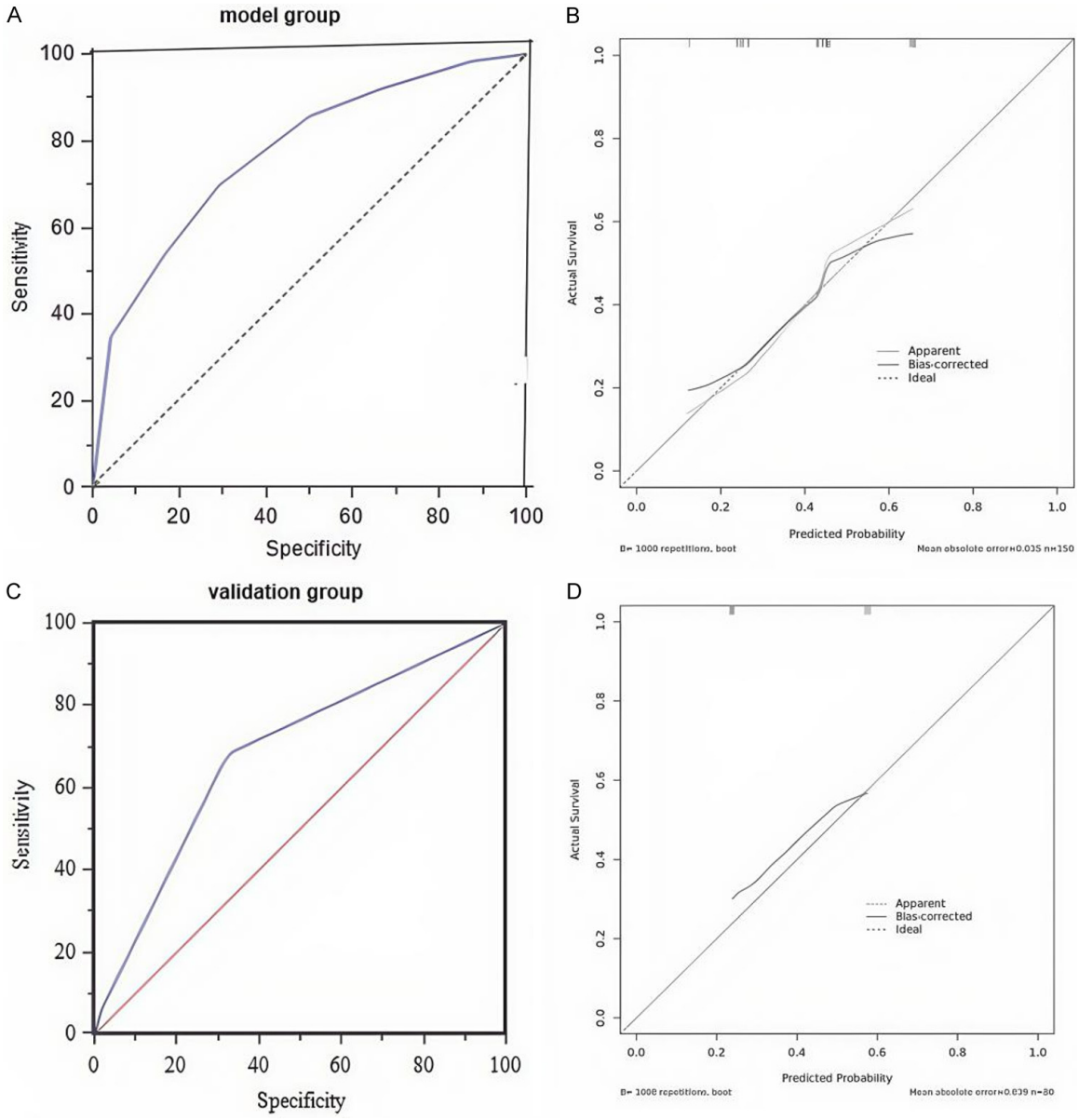

Figure 7. ROC curves and calibration curves for the model group and validation group. Note: A: ROC curve of the model group; B: Model group calibration curve; C: ROC curve of validation group; D: Validation group calibration curve.

vides a reliable molecular subtyping tool for lung cancer prognosis assessment. Patients with negative results for all three genes have a favorable prognosis and can be suitable for conservative treatment, while patients with positive results indicate a high risk of recurrence and require close follow-up with consideration of adjuvant therapy.

In summary, abnormal methylation of SHOX2, CDO1, and SOX17 is relatively common in lung cancer patients and can serve as potential biomarkers for early diagnosis and prognosis. However, this study has certain limitations: the sample size may be relatively limited, and the focus has been concentrated on a single or a few medical centers, leading to regional and selection biases. Additionally, the study only conducted a one-year follow-up, which is relatively short. Future research should extend the observation period to explore the application value of DNA methylation characteristics in the long-term prognosis of lung cancer patients.

Disclosure of conflict of interest

None.

Address correspondence to: Ran Hu, Department of Medical Laboratory, People's Hospital of Chongqing Liangjiang New Area, Chongqing 401121, China. E-mail: m13983094983_1@163.com

References

[1] Li C, Lei S, Ding L, Xu Y, Wu X, Wang H, Zhang Z, Gao T, Zhang Y and Li L. Global burden and trends of lung cancer incidence and mortality. Chin Med J (Engl) 2023; 136: 1583-1590.

[2] Casal-Mouriño A, Ruano-Ravina A, Lorenzo-González M, Rodríguez-Martínez Á, Giraldo-Osorio A, Varela-Lema L, Pereiro-Brea T, Barros-Dios JM, Valdés-Cuadrado L and Pérez-Ríos M. Epidemiology of stage III lung cancer: frequency, diagnostic characteristics, and survival. Transl Lung Cancer Res 2021; 10: 506-518.

[3] Bradley SH, Abraham S, Callister ME, Grice A, Hamilton WT, Lopez RR, Shinkins B and Neal RD. Sensitivity of chest X-ray for detecting lung cancer in people presenting with symptoms: a systematic review. Br J Gen Pract 2019; 69: e827-e835.

[4] Nooreldeen R and Bach H. Current and future development in lung cancer diagnosis. Int J Mol Sci 2021; 22: 8661.

[5] Ramasamy D, Deva Magendhra Rao AK, Rajkumar T and Mani S. Non-CpG methylation-a key epigenetic modification in cancer. Brief Funct Genomics 2021; 20: 304-311.

[6] Tan T, Shi P, Abbas MN, Wang Y, Xu J, Chen Y and Cui H. Epigenetic modification regulates tumor progression and metastasis through EMT (Review). Int J Oncol 2022; 60: 70.

[7] Hoang PH and Landi MT. DNA methylation in lung cancer: mechanisms and associations with histological subtypes, molecular alterations, and major epidemiological factors. Cancers (Basel) 2022; 14: 961.

[8] Chinese Medical Association; Oncology Society of Chinese Medical Association; Chinese Medical Association Publishing House. Chinese medical association guidelines for clinical diagnosis and treatment of lung cancer (2019 Edition). Zhonghua Zhong Liu Za Zhi 2020; 42: 257-287.

[9] Detterbeck FC, Mazzone PJ, Naidich DP and Bach PB. Screening for lung cancer: diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. Chest 2013; 143 Suppl: e78S-e92S.

[10] Bronkhorst AJ, Ungerer V and Holdenrieder S. Early detection of cancer using circulating tumor DNA: biological, physiological and analytical considerations. Crit Rev Clin Lab Sci 2019; 57: 253-269.

[11] Bai L, Hao X, Keith J and Feng Y. DNA methylation in regulatory T cell differentiation and function: challenges and opportunities. Biomolecules 2022, 12: 1282.

[12] Tan T, Shi P, Abbas MN, Wang Y, Xu J, Chen Y and Cui H. Epigenetic modification regulates tumor progression and metastasis through EMT (Review). Int J Oncol 2022; 60: 70.

[13] Li N, Zeng Y and Huang J. Signaling pathways and clinical application of RASSF1A and SHOX2 in lung cancer. J Cancer Res Clin Oncol 2020; 146: 1379-1393.

[14] Chen M, Zhu JY, Mu WJ and Guo L. Cysteine dioxygenase type 1 (CDO1): its functional role in physiological and pathophysiological processes. Genes Dis 2022; 10: 877-890.

[15] Jasim SA, Farhan SH, Ahmad I, Hjazi A, Kumar A, Jawad MA, Pramanik A, Altalbawy MAF, Alsaadi SB and Abosaoda MK. A cutting-edge investigation of the multifaceted role of SOX family genes in cancer pathogenesis through the modulation of various signaling pathways. Funct Integr Genomics 2025; 25: 6.

[16] Xie B, Dong W, He F, Peng F, Zhang H and Wang W. The combination of SHOX2 and RASSF1A DNA methylation had a diagnostic value in pulmonary nodules and early lung cancer. Oncology 2024; 102: 759-774.

[17] Li P, Liu S, Du L, Mohseni G, Zhang Y and Wang C. Liquid biopsies based on DNA methylation as biomarkers for the detection and prognosis of lung cancer. Clin Epigenetics 2022; 14: 118.

[18] Yang W, Chen H, Ma L, Dong J, Wei M, Xue X, Li Y, Jin Z, Xu W and Ji Z. SHOX2 promotes prostate cancer proliferation and metastasis through disruption of the Hippo-YAP pathway. iScience 2023; 26: 107617.

[19] Chen X and Poetsch A. The role of CDO1 in ferroptosis and apoptosis in cancer. Biomedicines 2024; 12: 918.

[20] Qin S, Liu G, Jin H, Chen X, He J, Xiao J, Qin Y, Mao Y and Zhao L. The dysregulation of SOX family correlates with DNA methylation and immune microenvironment characteristics to predict prognosis in hepatocellular carcinoma. Dis Markers 2022; 2022: 2676114.

[21] Jasim SA, Farhan SH, Ahmad I, Hjazi A, Kumar A, Jawad MA, Pramanik A, Altalbawy FMA, Alsaadi SB and Abosaoda MK. Role of homeobox genes in cancer: immune system interactions, long non-coding RNAs, and tumor progression. Mol Biol Rep 2024; 51: 964.

[22] Que HT, Chen HS and Yan AT. Correlation between c-met and tumor stage, pathological grade, lymph node metastasis and postoperative recurrence in lung adenocarcinoma. J Diagn Pathol 2024; 31: 1145-1149.

[23] Jung M, Ellinger J, Gevensleben H, Syring I, Lüders C, de Vos L, Pützer S, Bootz F, Landsberg J, Kristiansen G and Dietrich D. Cell-free SHOX2 DNA methylation in blood as a molecular staging parameter for risk stratification in renal cell carcinoma patients: a prospective observational cohort study. Clin Chem 2019; 65: 559-568.

[24] Harada H, Hosoda K, Moriya H, Mieno H, Ema A, Ushiku H, Washio M, Nishizawa N, Ishii S, Yokota K, Tanaka Y, Kaida T, Soeno T, Kosaka Y, Watanabe M and Yamashita K. Cancer-specific promoter DNA methylation of Cysteine dioxygenase type 1 (CDO1) gene as an important prognostic biomarker of gastric cancer. PLoS One 2019; 14: e0214872.