Epigenetics Insights

ARTICLE

Open Access

https://doi.org/10.48130/epi-0026-0005

Epigenetics Insights **2026**, 19: e008

Detection of lung cancer through a method of enzymatic methylation sequencing in cell-free DNA

Congyi Xie[1#], Zhenyang Lin[2#], Dan Wang[3#], Huijun Zhang[1,4], Xiong Xiao[1], Hailing Zheng[3], Yulu Liu[3], Xingqiang Zhu[3], Huan Xu[3], Zhi He[3], Hongliang Chen[3,5*] and Hongni Jiang[1,4,6*]

[1] *Department of Pulmonary Medicine, Zhongshan Hospital (Xiamen), Fudan University, Xiamen, Fujian 361015, China*
[2] *Department of Thoracic Surgery, Zhongshan Hospital (Xiamen), Fudan University, Xiamen, Fujian 361015, China*
[3] *Xiamen Vangenes Biotechnology Co., Ltd., Xiamen, Fujian 361015, China*
[4] *Department of Pulmonary Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, China*
[5] *School of Life Sciences, Xiamen University, Xiamen, Fujian 361102, China*
[6] *Fudan Zhangjiang Institute, Shanghai 201203, China*
[#] Authors contributed equally: Congyi Xie, Zhenyang Lin, Dan Wang
* Correspondence: hlchen@xmu.edu.cn (Chen H); jiang.hongni@zsxmhospital.com (Jiang H)

Abstract

Tumor-specific DNA methylation carried by circulating tumor DNA (ctDNA) is a promising marker for early cancer screening. This study endeavors to create, verify, and deploy a liquid biopsy method based on ctDNA methylation for the early identification of lung cancer (LC). Firstly, we established an LC-specific methylation profile by analyzing the differential expression of DNA methylation between 69 pairs of cancerous and paracancerous healthy tissues. A total of 2,680 differentially methylated positions (DMPs) between lung cancer tissues and paracancerous healthy tissues were identified. Enrichment analysis unveiled associations of these DMPs with various biological processes. The heatmap representation of LC-specific DMPs underscores distinct methylation patterns observed between lung cancer and paracancerous tissues. Subsequently, 127 plasma samples from stage I lung cancer patients and unaffected individuals were split into training and test groups to establish a model incorporating 22 of these LC-specific DMPs. This model yielded an 88.57% sensitivity and 78.43% specificity in the training data, with an area under the curve (AUC) of 0.8913. The corresponding performance in the test set was 60% sensitive and 76.92% specific, with an AUC of 0.7025. These outcomes indicate the promise of ctDNA methylation markers in enhancing the potential for early lung cancer detection.

Citation: Xie C, Lin Z, Wang D, Zhang H, Xiao X, et al. 2026. Detection of lung cancer through a method of enzymatic methylation sequencing in cell-free DNA. *Epigenetics Insights* 19: e008 https://doi.org/10.48130/epi-0026-0005

Introduction

Lung cancer stands as the predominant cause of cancer-related fatalities, posing a substantial public health concern[1]. The majority of lung cancer (LC) diagnoses occur at advanced stages, exceeding 80%, which contributes significantly to the disease's high mortality rate[2]. The prognosis of LC is heavily dependent on the stage at diagnosis, ranging from an optimistic 84% five-year survival rate for stage IA to a concerning 13% for stage IV[3,4]. Therefore, early screening and diagnosis hold promise for improving the prognosis of LC and mitigating mortality rates[5,6]. Presently, low-dose spiral computed tomography (LDCT) is recognized as the most effective screening method for LC. However, its propensity for false positives may lead to overdiagnosis, hindering its seamless integration into routine clinical practice[7,8]. For early detection of LC, extensive investigations have been conducted on blood-based protein biomarkers. Tumor-associated antigens (TAAs), such as cytokeratin 19 fragment (CYFRA 21-1)[9], carcinoembryonic antigen (CEA)[10], neuron-specific enolase (NSE)[11], and squamous cell carcinoma antigen (SCC-Ag)[12], have shown promise, but their suboptimal sensitivity hinders their efficacy in early diagnosis. Tumor-associated autoantibodies (TAAbs) represent the immune system's responses to peptides shed by tumors or present on the surface of tumor cells[13]. At present, the combination of TAAbs and low-dose computed tomography (LDCT) has been used for lung cancer screening. Although studies have shown that the specificity of TAAbs in the diagnosis of lung cancer is 91.60%, its low sensitivity limits its wide clinical application (only 56.53%)[14]. Besides, alterations in tumor signaling pathways play a pivotal role from the very early stages of tumorigenesis. For example, dysregulation of the TGF-β signaling co-receptor TGFBR3 has been reported to promote malignant progression by influencing proliferation, apoptosis, and metastatic potential of lung cancer cells, highlighting the importance of tumor-derived molecular changes in disease initiation[15]. Meanwhile, the rapid evolution of immunotherapy demonstrates that the immune system actively participates in shaping the tumor microenvironment throughout lung cancer development, further emphasizing the unmet clinical need for effective biomarkers capable of capturing tumor-specific molecular characteristics[16]. Consequently, there is an urgent requirement for the development of a sensitive, robust, and minimally invasive approach to detecting early-stage LC.

Cell-free DNA (cfDNA) refers to extracellular DNA fragments circulating in plasma, released primarily through programmed cell death (apoptosis) or uncontrolled cell rupture (necrosis). A subset of these fragments, circulating tumor DNA (ctDNA), originates specifically from cancerous cells[17,18], and contains a wealth of tumor-associated molecular alterations, such as genetic mutations[19], methylation modifications[20], end motifs[21], fragment length profiles[22], etc. The detection of these tumor-derived signals in ctDNA has significant implications for developing liquid biopsy-based diagnostic approaches in oncology[23,24].

Among epigenetic modifications, CpG methylation—the addition of a methyl group to cytosine residues in CpG dinucleotides—is

a fundamental regulatory mechanism in eukaryotic cells[25]. This modification is crucial for normal cellular differentiation, tissue development, and organ function. Due to its widespread dysregulation across cancer types and its early appearance during tumorigenesis, aberrant DNA methylation in ctDNA represents a highly informative biomarker for cancer detection[26].

For methylation analysis, bisulfite conversion remains the gold standard, as it differentiates methylated and unmethylated cytosines at single-base resolution by chemically converting unmethylated cytosines to uracil while leaving methylated cytosines intact[27]. However, the harsh chemical treatment can cause significant DNA degradation, posing challenges for low-input cfDNA studies. In contrast, Enzymatic Methyl-seq (EM-seq) employs a gentler enzymatic approach to profile genome-wide methylation. Our previous work[28] directly compared both methods on cfDNA and demonstrated that EM-seq minimizes DNA damage and, for the same input amount, achieves significantly higher sequencing depth and more uniform coverage compared to bisulfite conversion. This ability to generate robust methylation data from limited and fragmented cfDNA makes EM-seq particularly suited for liquid biopsy applications and motivated its selection for this study. Moreover, compared to other gentle conversion methods like TET-assisted pyridine borane sequencing (TAPS), EM-seq utilizes a single-tube enzymatic reaction that minimizes sample loss and handling errors, which is particularly beneficial when working with the limited quantities and fragile nature of cfDNA. This robustness in low-input conditions directly supports the high conversion efficiency and sequencing uniformity observed in our plasma cohort.

In this research, our objective was to discern lung cancer-specific differentially methylated positions (DMPs) and assess the diagnostic ability of these cfDNA methylation biomarkers to distinguish early-stage LC from benign nodules. First, we commenced by contrasting the methylation signatures of LC tissues and paraneoplastic tissues from The Cancer Genome Atlas (TCGA) to pinpoint LC-specific DMPs. Then we designed a lung cancer detection panel of LC-specific DMPs. To develop a high-performance LC screening model, methylation analysis data were obtained from plasma samples of 127 individuals with early-stage lung cancer or non-lung cancer through targeted EM-seq analysis.

Materials and methods

Study design and participants

Lung cancer patients and non-lung cancer (NL) controls were recruited at Zhongshan Hospital (Xiamen) of Fudan University. Tumor staging was assigned according to the 8th edition of the IASLC Lung Cancer Staging System. The study was conducted in compliance with the Declaration of Helsinki and approved by the institutional review board. Written informed consent was provided by each participant or their legally authorized representative.

Blood sample preparation and cfDNA extraction

For each subject, 10 mL of whole blood was collected into cell-free DNA preservation tubes (Vangenes). Plasma was separated within 2 h through a two-step centrifugation procedure, as described in our previous report[28], and stored at −80 °C. Circulating cell-free DNA (cfDNA) was extracted from 2 mL of thawed plasma using the Concert Plasma Cell-Free DNA Extraction Kit and quantified with the Qubit dsDNA HS Assay Kit (Thermo Fisher). Although no dedicated fragment analysis was performed upfront, post-sequencing insert size metrics displayed the canonical cfDNA profile with a median length of approximately 165 bp and minimal high-molecular-weight contamination, confirming adequate sample quality for methylation analysis (Supplementary Fig. S1).

EM-seq library construction

EM-seq libraries were prepared following an adapted version of the protocol by Guo et al.[28]. In brief, cfDNA inputs underwent end repair and adapter ligation with the VAHTS Universal DNA Library Prep Kit for Illumina V3 (Vazyme). The resulting fragments were treated with the EpiArt DNA Methylation Kit (Vazyme) to convert unmethylated cytosines. Library amplification was carried out with KAPA HiFi HotStart Uracil+ ReadyMix (Roche) and indexed using NEBNext Unique Dual Index Primers (NEB). To monitor conversion efficiency, two spike-in controls were added to every sample: unmethylated λ DNA and hypermethylated pUC-19 DNA. The proportion of non-converted λ DNA (theoretically 0%) and pUC-19 DNA (theoretically 100%) was assessed for each library. Across all 127 analyzed samples, the median λ non-conversion rate was 0.02% (range 0.00%–0.31%), and the median pUC-19 non-conversion rate was 98.55% (range 96.05%–99.58%), indicating robust conversion performance. Detailed per-sample quality metrics are provided in Supplementary Table S1.

Targeted capture and sequencing

For target enrichment, we pooled up to 24 barcoded libraries and hybridized them to a custom NAD oligonucleotide panel (Nanodigmbio) using NADprep Hybrid Capture Reagents. Following stringent washes, captured molecules were amplified on-bead with VAHTS HiFi Amplification Mix (Vazyme) and sequenced on an Illumina NovaSeq 6000 system (2 × 150 bp paired-end). To avoid capture bias related to the original methylation status, a five-probe design was employed for each CpG site[28], the probes targeted: (1) the unconverted forward strand; (2) the methylated forward strand after C-to-T conversion; (3) the methylated reverse strand; (4) the unmethylated forward strand; and (5) the unmethylated reverse strand.

LC-specific methylation markers selection and panel design

Methylation data from 69 paired lung adenocarcinoma and adjacent normal tissues were retrieved from The Cancer Genome Atlas (TCGA) (Supplementary Table S2). We performed differential methylation analysis on all autosomal CpGs using the limma R package, applying a threshold of $|\log_2(\text{fold change})| > 0.35$ and $p < 10^{-15}$. The $|\log_2(\text{fold change})| > 0.35$ threshold corresponds to approximately 1.28-fold change in methylation level, representing a commonly accepted standard for TCGA 450K array differential methylation analysis that balances sensitivity and specificity for biomarker discovery. The stringent cutoff of $p <$ 10e-15 was selected to ensure extremely high statistical confidence. This threshold is well justified given the analysis of 69 paired samples with substantial statistical power, and it effectively minimizes false positives when screening across > 480,000 CpG sites. Together, these thresholds yielded 2,680 DMPs (~0.56% selection rate), consistent with the expected proportion of true differentially methylated sites in cancer vs normal tissue comparisons. To maximize the panel's clinical relevance, we supplemented these DMPs with additional CpG loci that had previously shown value for lung cancer detection[29–31].

Data processing

Raw reads were processed with an updated version of the pipeline described in Guo et al.[28]. Adapter trimming and quality filtering were performed using Fastp (v0.23.2)[32]. Clean reads were aligned to the human reference genome (hg38) with Bismark (v0.23.0)[33] and deduplicated. To exclude incompletely converted fragments, any read containing > 3 methylated cytosines in a non-CpG context (CH) was discarded[28]. Methylation counts at each CpG were extracted with bismark_methylation_extractor using the flags --bedGraph --no_overlap --comprehensive --counts.

Development of LC screening model

We first estimated the individual discriminatory power of each CpG in the training set by calculating its area under the receiver operating characteristic curve (AUC). Markers with an AUC > 0.7 or < 0.3 were carried forward. Feature selection was then refined using Random Forest and least absolute shrinkage and selection operator (Lasso) regression. The resulting marker set was used to train a logistic regression classifier. Performance on the held-out test set was evaluated using the AUC metric, implemented with the roc_auc_score function from scikit-learn[34].

Data and code availability

The beta-value matrix for the 22 diagnostic CpGs across all 127 plasma specimens is available in Supplementary Table S3. The custom R script for feature selection and model training is publicly available at https://github.com/liaohongyue/Methylation-predicts-lung-cancer Additional data and materials can be requested from the corresponding author.

Results

Differential methylation marker for LC detection

To identify methylation sites pertinent to LC, the DNA methylation data of 69 pairs of cancerous and paracancerous healthy tissues were acquired from the TCGA database and analyzed. The specific information is provided in Supplementary Table S2. Differential methylation analysis revealed the presence of 2,680 LC-specific DMPs (Fig. 1a; Supplementary Table S4). They locate on 1,555 genes, among these DMPs, 1,417 (53%) exhibited elevated methylation levels in tumor tissues, while the remaining 1,263 (47%) demonstrated reduced methylation levels in tumor tissues (Fig. 1a). Enrichment analysis unveiled associations of these DMPs with various biological processes, including the processes of embryonic organogenesis and the establishment of anterior/posterior positional patterns, cellular components such as neuronal cell body, transporter complexes, and ion channel complexes, as well as molecular functions such as transcriptional modulation of DNA-binding activator function, characteristic of RNA polymerase II, and monoatomic ion channel activity (Supplementary Fig. S2). A predominance of LC-specific DMPs exhibited heightened methylation ratios within CpG islands (27.5%) and CpG shores (25.2%) (Fig. 1b). Furthermore, the heatmap representation of LC-specific DMPs underscores distinct methylation patterns observed between lung cancer and paracancerous tissues (Fig. 1c).

Cohort characteristics

There were 127 plasma samples collected, including 50 from LC patients and 77 from non-lung-cancer individuals. The workflow is

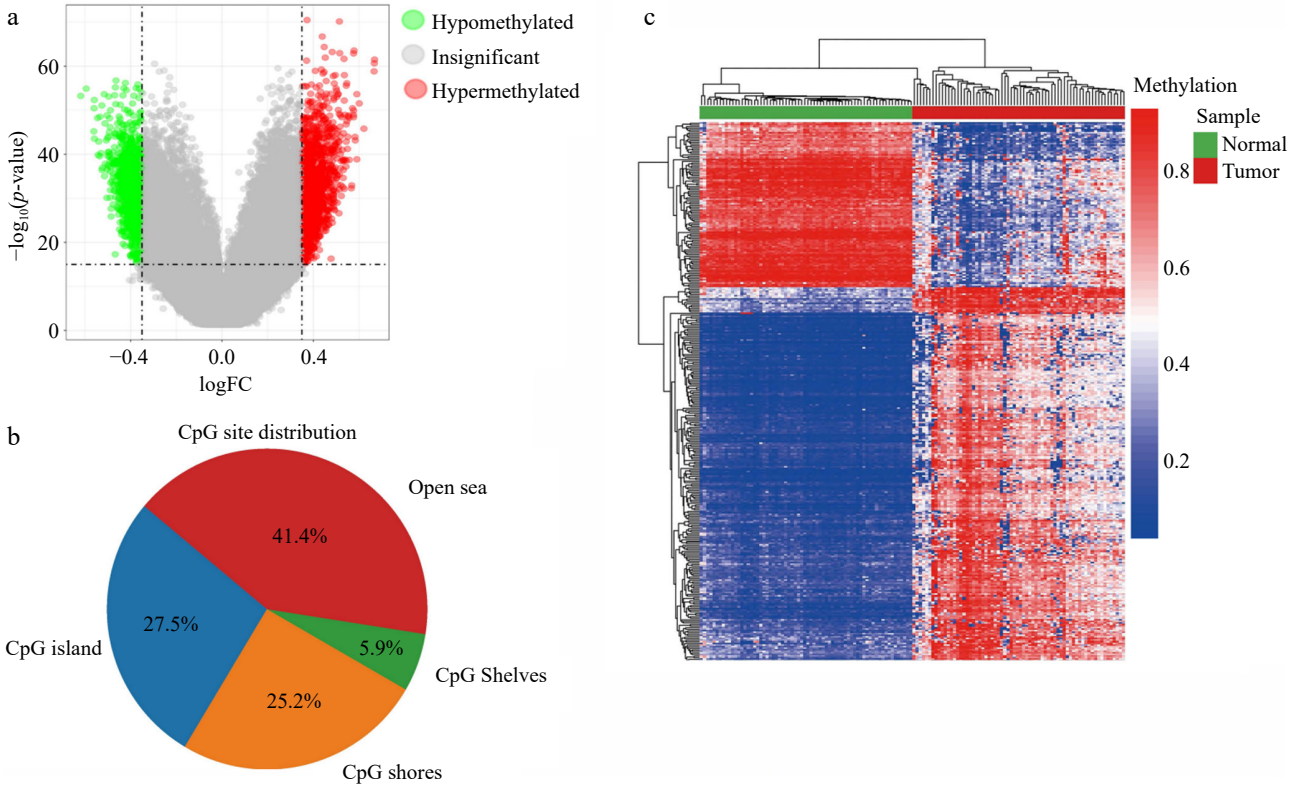

Fig. 1 Selection of methylation markers. (a) Display of the pronounced differences in methylation levels between lung cancer (LC) samples and corresponding healthy paraneoplastic tissues. (b) The distribution of the LC-specific methylation sites. (c) Heatmap visualization depicting the prevalence of hypomethylated and hypermethylated DMPs contrasting lung cancerous with paraneoplastic healthy tissues.

delineated in Fig. 2a and b. The lung cancer (LC) cohort consisted of 50 treatment-naïve patients with stage I disease, predominantly comprising adenocarcinoma (86.7%, n = 43), with a smaller subset of squamous cell carcinoma (13.3%, n = 7). The non-lung cancer (NL) control group consisted of 77 individuals with various benign pulmonary conditions or non-pulmonary diseases. The detailed clinical characteristics of this cohort are summarized in Supplementary Table S5. This cohort included a spectrum of patients with benign pulmonary nodules (e.g., infectious granulomas, inflammatory nodules, and benign tumors), non-nodular pulmonary diseases (e.g., pneumonia, interstitial lung disease), and a few individuals with non-pulmonary conditions. The inclusion of such a heterogeneous control group, particularly those with conditions that mimic lung cancer radiologically, was designed to rigorously test the specificity of our diagnostic model in a clinically relevant setting.

LC detection using plasma samples

All plasma samples underwent targeted enzymatic methyl sequencing, yielding an average of 21 million reads and an average unique read depth of 852× (Supplementary Table S1). Using 35 LC and 51 NL samples for training, we established a diagnostic model based on 22 optimally selected methylation features (Supplementary Table S6). This model demonstrated strong performance in the training set, with 88.57% sensitivity, 78.43% specificity, and an AUC of 0.8913 (Fig. 3a). When tested on the remaining 15 LC and 26 NL samples, it achieved 60% sensitivity, 76.92% specificity, and an AUC of 0.7025 (Fig. 3b).

Functional enrichment analysis of the 16 genes harboring these 22 features revealed several significantly enriched KEGG pathways (p.adjust < 0.01), including cellular senescence, cell cycle, oocyte meiosis, and the PPAR signaling pathway. Additional pathways related to various cancers (e.g., gastric cancer, ductal carcinoma *in situ*) were also enriched (p.adjust < 0.05) (Fig. 3c).

Gene-level annotation of the 22 methylation features revealed that they are distributed across 16 genes, including known tumor suppressor genes (BRCA1, DLC1, MGMT, TFPI2) and oncogenes (CCND2, TLX3), as well as genes involved in transcriptional regulation, signal transduction, and metabolic processes (SDCCAG8, RASGRF1, ZNF536, etc.) (Supplementary Table S6). Notably, five of these genes (MGMT, CCND2, BRCA1, SYT5, DLC1) have been previously reported as methylation markers in lung cancer, supporting

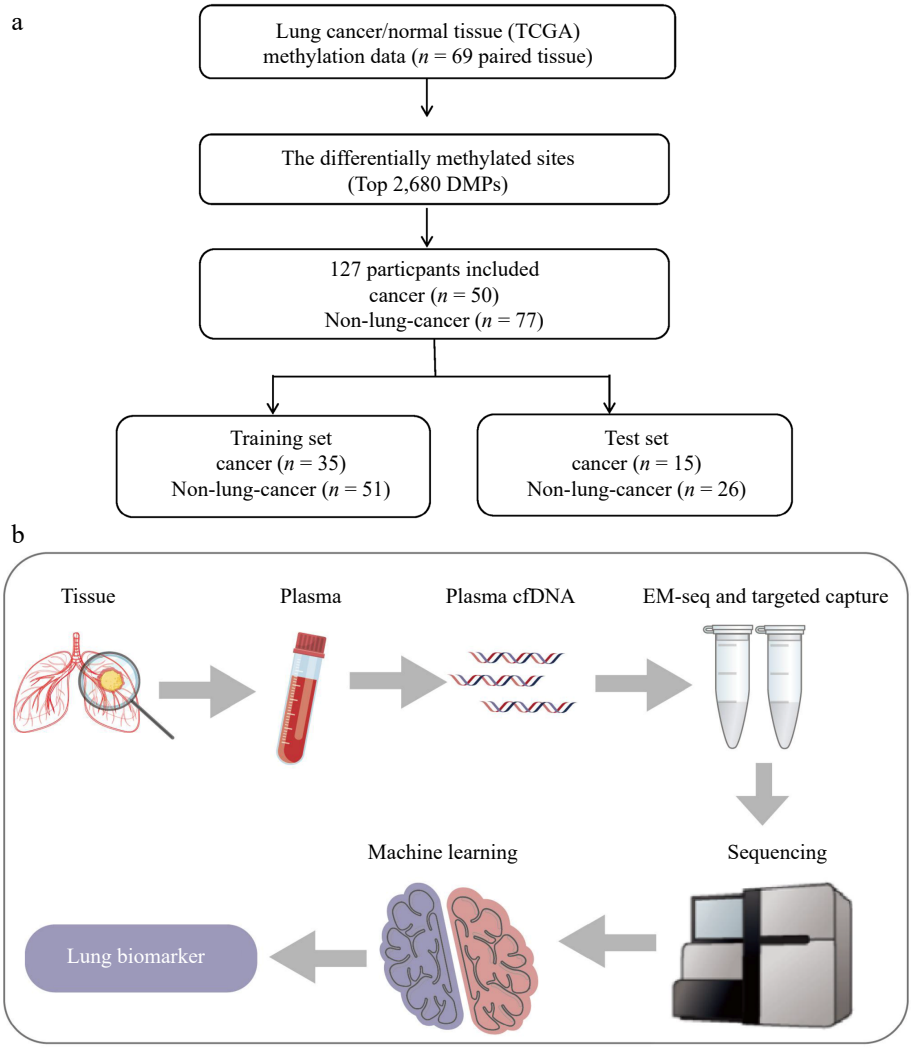

Fig. 2 Schematic illustration of the study. (a) Study design. A total of 127 participants were included in this study. Researchers performed methylation sequencing of cfDNA in plasma. Eighty-six participants (35 LC, 51 NL) were assigned to training to build a machine learning model based on a logistic regression algorithm. Forty-one participants (15 LC, 26 NL) were assigned to testing to confirm the performance of the model. (b) Schematic diagram of the plasma cell free DNA testing process.

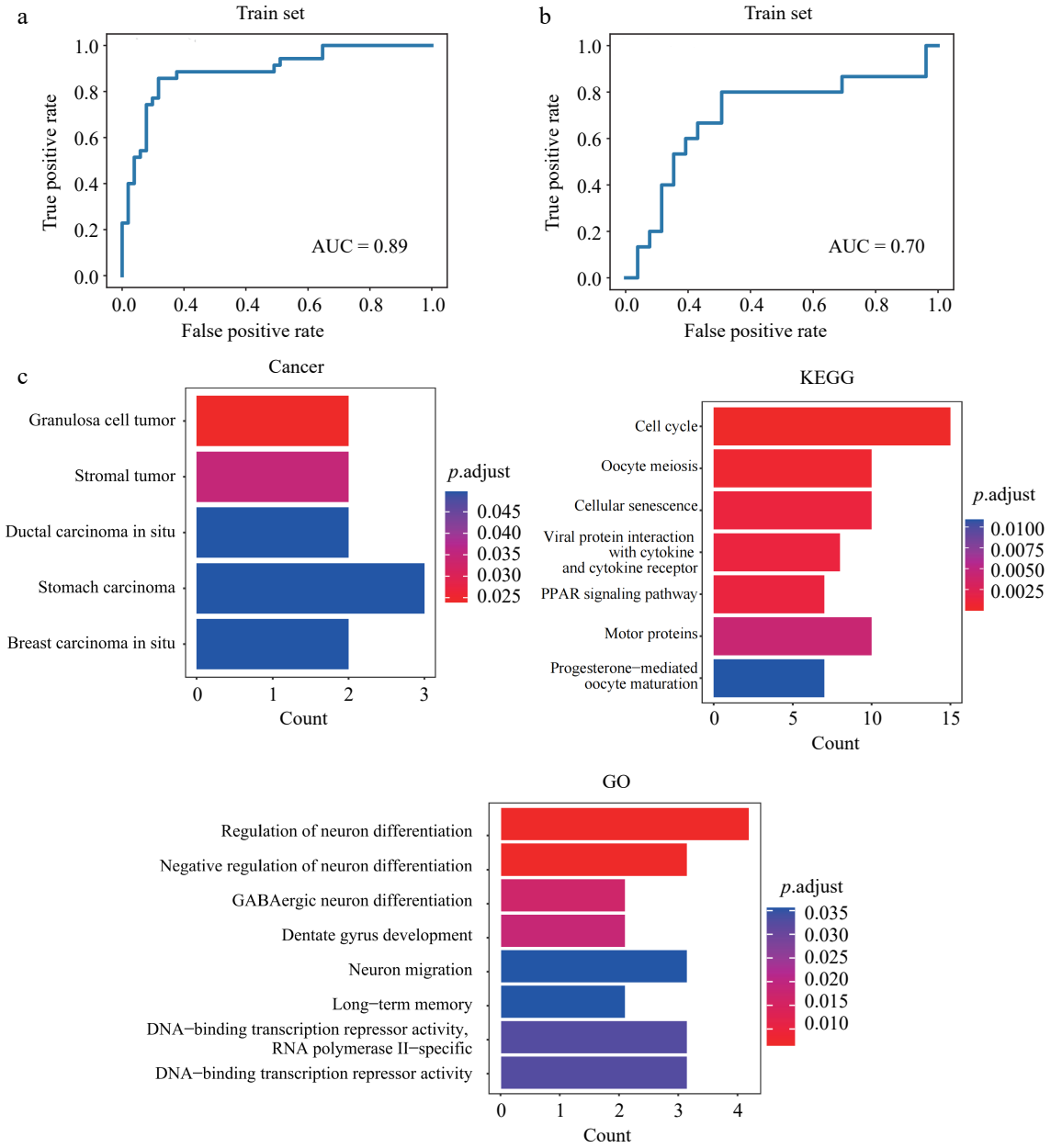

Fig. 3 Verification and construction of the LC screening model. (a) ROC curve for the model in the training dataset. (b) ROC curve for the model in the test dataset. (c) Gene ontology analyses applied to the genes characterized by the model features.

the biological relevance of our panel[31,35–39]. The remaining genes represent novel candidates that may enhance the specificity of early-stage lung cancer detection.

No significant differences in methylation levels (beta values) at 22 features were observed across plasma sample subgroups stratified by patient characteristics (Supplementary Figs. S3–S6).

Discussion

Early diagnosis of LC will significantly improve the 5-year survival rate. Yet, current screening methods, notably LDCT, exhibit limitations including overdiagnosis, cost-effectiveness issues, and radiation exposure[8,40,41]. The invention of cfDNA-based early screening techniques holds promise as a revolutionary alternative[42]. In this study, LC-specific DMPs from plasma cfDNA were identified from the TCGA database and used to construct an early diagnostic model with 60% sensitivity and 76.92% specificity. The biological rationale for LC detection via plasma cfDNA was indirectly corroborated by the enrichment of methylation markers in pathways related to transcriptional regulation, organismal development, and intracellular signaling.

A key methodological strength of our approach is the use of enzymatic methyl-sequencing (EM-seq) for methylation conversion. Compared to conventional bisulfite treatment, EM-seq minimizes DNA damage, reduces conversion errors, and provides higher coverage, particularly for low-input cfDNA samples[28,43]. These features are critical for early-stage cancer detection, where the tumor-derived cfDNA fraction is often extremely low. Recent advances have demonstrated that enzymatic conversion better preserves nucleic acid integrity and enhances detection sensitivity by reducing fragmen-

tation and base-conversion artifacts[28,43], directly supporting the superior performance of our assay.

The biological plausibility of cfDNA methylation as an early diagnostic biomarker is reinforced by the underlying molecular events in tumorigenesis. Oxidative stress-induced apoptosis, characterized by the dysregulation of Bax/Bcl-2 and caspase cascades[44], is a pivotal driver of lung cancer progression and is closely linked to aberrant epigenetic regulation. Such molecular perturbations may give rise to methylation alterations that are shed into the circulation, providing a detectable signal in cfDNA. Thus, the methylation signatures captured by our model may reflect early tumorigenic changes, strengthening the rationale for their use in non-invasive detection.

Compared with previous cfDNA methylation studies, our model demonstrated a relatively higher sensitivity for stage I lung cancer. For instance, Liu et al. reported a sensitivity of only 23% for stage I LC using cfDNA methylation testing[45], whereas our EM-seq-based classifier achieved 60%. It is noteworthy that the effectiveness of cfDNA methylation screening can vary by histological subtype; studies have shown higher detection rates in squamous cell carcinoma than in adenocarcinoma, possibly due to differences in necrosis levels[46,47]. In our cohort, 86.67% of stage I patients had adenocarcinoma, and the observed 60% sensitivity is therefore largely representative of this subtype. Recent work has also reported promising sensitivities for stage I lung adenocarcinoma, underscoring the clinical value of early detection in this population[48]. Additionally, conventional serum protein biomarkers such as CEA and CYFRA 21-1 have insufficient sensitivity for early-stage disease (e.g., CYFRA 21-1 sensitivity as low as 5% at a common cutoff)[9]. In contrast, our methylation-based classifier, which directly detects tumor-derived epigenetic alterations, offers improved sensitivity for early-stage lung adenocarcinoma and holds promise as a next-generation non-invasive biomarker strategy.

Despite these encouraging results, several limitations must be acknowledged. First and foremost, our study lacks an independent external validation cohort, which is the gold standard for confirming the robustness and generalizability of a diagnostic model. This absence directly limits the translational confidence of our classifier, as model performance can decline substantially when applied to populations with different demographic or clinical characteristics. We are actively collecting samples for a large-scale independent validation to address this critical gap. Second, the overall sample size was modest, which increases the risk of overfitting, as indicated by the drop in area under the curve (AUC) from the training set (0.8913) to the test set (0.7025). Third, the lung cancer cohort was predominantly composed of adenocarcinoma cases (86.67%), with only seven squamous cell carcinoma samples. Therefore, the reported sensitivity primarily reflects the detection of lung adenocarcinoma, and broad applicability across all non-small cell lung cancer subtypes remains to be validated in a histologically balanced cohort. Fourth, most participants were symptomatically diagnosed rather than identified through screening, which may overestimate the classifier's performance in an asymptomatic screening setting. Finally, biological and technical variables—such as inter-individual differences in cfDNA tumor fraction (often < 0.1% in early-stage disease)[49], smoking history, sequencing depth, and pre-analytical handling—can influence methylation signals and may have contributed to the observed variability. Future assay iterations could integrate fragmentomic features or methylation haplotype burden as internal calibrators to account for variations in ctDNA shedding[47].

Regarding future directions, prospective evaluation in LDCT-screened populations is essential. LDCT remains the cornerstone of lung cancer screening but suffers from a high false-positive rate (> 90%), leading to unnecessary invasive procedures[50]. Combining cfDNA methylation biomarkers with CT imaging features has been shown to improve pulmonary nodule risk stratification[51]. We therefore plan to investigate the integration of our methylation panel with radiological assessment to enhance diagnostic accuracy for indeterminate nodules. Furthermore, expanding the probe panel to include additional lung cancer-related markers and performing cross-validation in more diverse cohorts will be crucial to improve both sensitivity and generalizability.

Conclusions

In conclusion, this study demonstrates that a plasma cfDNA methylation signature, detected by the EM-seq technique, holds significant promise as a non-invasive tool for the early detection of lung cancer, particularly for adenocarcinoma, which represented the dominant histological type in our stage I cohort. Our diagnostic model, comprising 22 lung cancer-specific DMPs, achieved clinically informative sensitivity in this predominantly adenocarcinoma stage I population, a group for which current diagnostic options remain limited. The biological relevance of the selected markers is reinforced by their enrichment in cancer-associated pathways. Given the limited representation of squamous cell carcinoma and other subtypes, further validation in a more histologically diverse cohort is necessary to assess the generalizability of this cfDNA-based approach for early lung cancer diagnosis.

Ethical statements

The study was conducted in accordance with the Declaration of Helsinki, and all procedures were approved by the Ethics Committee of the Zhongshan Hospital (Xiamen), Fudan University, identification number: B2023-020, approval date: 2023/05/19.

Author contributions

The authors confirm their contributions to the paper as follows: study conception and design, manuscript review and editing, supervision, project administration, funding acquisition: Chen H, Jiang H; methodology: Xie C, Lin Z, Wang D; software: Xie C, Liu Y; validation: Xu H, Zhang H, Xiao X, He Z; formal analysis: Zheng H, Zhu X, Liu Y, Xie C; investigation: Lin Z, Wang D; resources: Lin Z, Zhang H, Xiao X, Xie C; data curation: Xie C, Lin Z, Xu H, Wang D; draft manuscript preparation: Xie C, Lin Z, Wang D; visualization: Wang D. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The underlying raw data that substantiate the study's findings are accessible to the public at the National Genomics Data Center (NGDC) under accession number OMIX016413 (https://ngdc.cncb.ac.cn/omix/release/OMIX016413).

Acknowledgments

The authors would like to acknowledge all the healthy and patient volunteers. This study was supported by the Shanghai Science and Technology Innovation Action Plan (No. 21140902700),

the Major Medical and Health Projects of Xiamen (No. 3502Z20204008) and the Natural Science Foundation of Fujian Province (No. 2023J011690). These findings are partly derived from data provided by the TCGA Research Network: visit www.cancer.gov/tcga for more information.

Conflicts of interest

Although authors XD Wang, H Zheng, Y Liu, X Zhu, H Xu and Z He are employees of Vangenes, Inc., the work presented herein is independent academic research and is not related to the commercial interests of the company. The authors declare that no financial or other contractual agreements between the company and the authors or their institutions influenced the design, outcome, or reporting of this study.

Supplementary information accompanies this paper online at: https://doi.org/10.48130/epi-0026-0005.

Dates

Received 15 August 2025; Revised 21 May 2026; Accepted 4 June 2026; Published online 25 June 2026

References

[1] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, et al. 2024. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 74:229–263

[2] Wang T, Nelson RA, Bogardus A, Grannis FW Jr. 2010. Five-year lung cancer survival: which advanced stage nonsmall cell lung cancer patients attain long-term survival? *Cancer* 116:1518–1525

[3] Woodard GA, Jones KD, Jablons DM. 2016. Lung cancer staging and prognosis. In *Lung Cancer*, ed. Reckamp KL. Cham: Springer. pp. 47–75 doi: 10.1007/978-3-319-40389-2_3

[4] Maeda R, Yoshida J, Ishii G, Hishida T, Nishimura M, et al. 2011. Risk factors for tumor recurrence in patients with early-stage (stage I and II) non-small cell lung cancer: patient selection criteria for adjuvant chemotherapy according to the seventh edition TNM classification. *Chest* 140:1494–1502

[5] Huang J, Deng Y, Tin MS, Lok V, Ngai CH, et al. 2022. Distribution, risk factors, and temporal trends for lung cancer incidence and mortality: a global analysis. *Chest* 161:1101–1111

[6] Li C, Lei S, Ding L, Xu Y, Wu X, et al. 2023. Global burden and trends of lung cancer incidence and mortality. *Chinese Medical Journal* 136:1583–1590

[7] Oudkerk M, Liu S, Heuvelmans MA, Walter JE, Field JK. 2021. Lung cancer LDCT screening and mortality reduction—evidence, pitfalls and future perspectives. *Nature Reviews Clinical Oncology* 18:135–151

[8] Bonney A, Malouf R, Marchal C, Manners D, Fong KM, et al. 2022. Impact of low-dose computed tomography (LDCT) screening on lung cancer-related mortality. *The Cochrane Database of Systematic Reviews* 8:CD013829

[9] Okamura K, Takayama K, Izumi M, Harada T, Furuyama K, et al. 2013. Diagnostic value of CEA and CYFRA 21-1 tumor markers in primary lung cancer. *Lung Cancer* 80:45–49

[10] Wu LX, Li XF, Chen HF, Zhu YC, Wang WX, et al. 2019. Combined detection of CEA and CA125 for the diagnosis for lung cancer: a meta-analysis. *Cellular and Molecular Biology* 64:67–70

[11] Xu CM, Luo YL, Li S, Li ZX, Jiang L, et al. 2019. Multifunctional neuron-specific enolase: its role in lung diseases. *Bioscience Reports* 39:BSR20192732

[12] Wang L, Wang D, Zheng G, Yang Y, Du L, et al. 2016. Clinical evaluation and therapeutic monitoring value of serum tumor markers in lung cancer. *The International Journal of Biological Markers* 31:e80–e87

[13] Tan EM, Zhang J. 2008. Autoantibodies to tumor-associated antigens: reporters from the immune system. *Immunological Reviews* 222:328–340

[14] Du Q, Yu R, Wang H, Yan D, Yuan Q, et al. 2018. Significance of tumor-associated autoantibodies in the early diagnosis of lung cancer. *The Clinical Respiratory Journal* 12:2020–2028

[15] Deng X, Ma N, He J, Xu F, Zou G. 2024. The role of TGFBR3 in the development of lung cancer. *Protein & Peptide Letters* 31:491–503

[16] Li CL, Ma XY, Yi P. 2025. The role of immunotherapy in lung cancer treatment: current strategies, future directions, and insights into metastasis and immune microenvironment. *Current Gene Therapy* 25:453–466

[17] Luo H, Wei W, Ye Z, Zheng J, Xu RH. 2021. Liquid biopsy of methylation biomarkers in cell-free DNA. *Trends in Molecular Medicine* 27:482–500

[18] Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, et al. 2017. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nature Reviews Cancer* 17:223–238

[19] Tébar-Martínez R, Martín-Arana J, Gimeno-Valiente F, Tarazona N, Rentero-Garrido P, et al. 2023. Strategies for improving detection of circulating tumor DNA using next generation sequencing. *Cancer Treatment Reviews* 119:102595

[20] Leal A, Sidransky D, Brait M. 2020. Tissue and cell-free DNA-based epigenomic approaches for cancer detection. *Clinical Chemistry* 66:105–116

[21] Jiang P, Sun K, Tong YK, Cheng SH, Cheng THT, et al. 2018. Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* 115:E10925–E10933

[22] Ding SC, Dennis Lo YM. 2022. Cell-free DNA fragmentomics in liquid biopsy. *Diagnostics* 12:978

[23] Dennis Lo YM, Han DSC, Jiang P, Chiu RWK. 2021. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. *Science* 372:eaaw3616

[24] Song P, Wu LR, Yan YH, Zhang JX, Chu T, et al. 2022. Limitations and opportunities of technologies for the analysis of cell-free DNA in cancer diagnostics. *Nature Biomedical Engineering* 6:232–245

[25] Breiling A, Lyko F. 2015. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics & Chromatin* 8:24

[26] Ibrahim J, Peeters M, Van Camp G, Op de Beeck K. 2023. Methylation biomarkers for early cancer detection and diagnosis: current and future perspectives. *European Journal of Cancer* 178:91–113

[27] Krueger F, Kreck B, Franke A, Andrews SR. 2012. DNA methylome analysis using short bisulfite sequencing data. *Nature Methods* 9:145–151

[28] Guo P, Zheng H, Li Y, Li Y, Xiao Y, et al. 2023. Hepatocellular carcinoma detection via targeted enzymatic methyl sequencing of plasma cell-free DNA. *Clinical Epigenetics* 15:2

[29] Hubers AJ, Heideman DAM, Burgers SA, Herder GJM, Sterk PJ, et al. 2015. DNA hypermethylation analysis in sputum for the diagnosis of lung cancer: training validation set approach. *British Journal of Cancer* 112:1105–1113

[30] Ooki A, Maleki Z, Tsay JJ, Goparaju C, Brait M, et al. 2017. A panel of novel detection and prognostic methylated DNA markers in primary non–small cell lung cancer and serum DNA. *Clinical Cancer Research* 23:7141–7152

[31] Xu W, Lu J, Zhao Q, Wu J, Sun J, et al. 2019. Genome-wide plasma cell-free DNA methylation profiling identifies potential biomarkers for lung cancer. *Disease Markers* 2019:4108474

[32] Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890

[33] Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27:1571–1572

[34] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, et al. 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830

[35] Li X, Yang T, Li CS, Song Y, Wang D, et al. 2020. Polymerase chain reaction - surface-enhanced Raman spectroscopy (PCR-SERS) method for gene methylation level detection in plasma. *Theranostics* 10:898–909

[36] Begum S, Brait M, Dasgupta S, Ostrow KL, Zahurak M, et al. 2011. An epigenetic marker panel for detection of lung cancer using cell-free serum DNA. *Clinical Cancer Research* 17:4494–4503

[37] Mohammed F, Baydaa Abed Hussein A, Ahmed T. 2022. Evaluation of methylation panel in the promoter region of p16INK4a, RASSF1A, and MGMT as a biomarker in sputum for lung cancer. *Archives of Razi Institute* 77:1075–1081

[38] Hung CS, Wang SC, Yen YT, Lee TH, Wen WC, et al. 2018. Hypermethylation of CCND2 in lung and breast cancer is a potential biomarker and drug target. *International Journal of Molecular Sciences* 19:3096

[39] Zhang Y, Wang R, Song H, Huang G, Yi J, et al. 2011. Methylation of multiple genes as a candidate biomarker in non-small cell lung cancer. *Cancer Letters* 303:21–28

[40] Chang HT, Wang PH, Chen WF, Lin CJ. 2022. Risk assessment of early lung cancer with LDCT and health examinations. *International Journal of Environmental Research and Public Health* 19:4633

[41] Maller B, Tanvetyanon T. 2022. Emerging approaches to complement low-dose computerized tomography for lung cancer screening: a narrative review. *Cureus* 14:e27309

[42] Chen M, Zhao H. 2019. Next-generation sequencing in liquid biopsy: cancer screening and early detection. *Human Genomics* 13:34

[43] Vaisvila R, Ponnaluri VKC, Sun Z, Langhorst BW, Saleh L, et al. 2021. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Research* 31:1280–1289

[44] Ullah A, Razzaq A, Alfaifi MY, Elbehairi SEI, Menaa F, et al. 2024. Sanguinarine attenuates lung cancer progression via oxidative stress-induced cell apoptosis. *Current Molecular Pharmacology* 17:e18761429269383

[45] Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, et al. 2020. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Annals of Oncology* 31:745–759

[46] Shen SY, Singhania R, Fehringer G, Chakravarthy A, Roehrl MHA, et al. 2018. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* 563:579–583

[47] Mathios D, Johansen JS, Cristiano S, Medina JE, Phallen J, et al. 2021. Detection and characterization of lung cancer using cell-free DNA fragmentomes. *Nature Communications* 12:5060

[48] Xue R, Li X, Yang L, Yang M, Zhang B, et al. 2024. Evaluation and integration of cell-free DNA signatures for detection of lung cancer. *Cancer Letters* 604:217216

[49] Jamshidi A, Liu MC, Klein EA, Venn O, Hubbell E, et al. 2022. Evaluation of cell-free DNA approaches for multi-cancer early detection. *Cancer Cell* 40:1537–1549.e12

[50] Wang X, Chen Y, Ma C, Bi L, Su Z, et al. 2025. Current advances and future prospects of blood-based techniques for identifying benign and malignant pulmonary nodules. *Critical Reviews in Oncology/Hematology* 207:104608

[51] Liang W, Tao J, Cheng C, Sun H, Ye Z, et al. 2024. A clinically effective model based on cell-free DNA methylation and low-dose CT for risk stratification of pulmonary nodules. *Cell Reports Medicine* 5:101750