Research Article

Improving CT scan for lung cancer diagnosis with an integromic signature

Jipei Liao, Pushpawallie Dhilipkannah, Feng Jiang*

Department of Pathology, University of Maryland School of Medicine, 10 S. Pine St. Baltimore, MD 21201, United States of America

Abstract

Lung cancer is the leading cause of cancer-related mortality globally, making early detection crucial for reducing death rates. Low-dose computed tomography (LDCT) screening helps detect lung cancer early but often identifies indeterminate pulmonary nodules (PNs), leading to potential overtreatment. This study aimed to develop a diagnostic test that accurately differentiates malignant from benign PNs detected on LDCT scans by analyzing non-coding RNAs, DNA methylation, and bacterial DNA in patient samples. Using droplet digital polymerase chain reaction, we analyzed samples from a training set of 150 patients with malignant PNs and 250 smokers with benign PNs. Individual biomarkers in plasma and sputum showed moderate effectiveness, with sensitivities ranging from 62% to 77% and specificities from 54% to 87%. We developed an integromic signature by combining two plasma biomarkers and one sputum biomarker, along with additional clinical data, which demonstrated a sensitivity of 90% and specificity of 95%. The signature's diagnostic performance was further validated in a cohort consisting of 30 patients with malignant PNs and 50 smokers with benign PNs. The integromic signature showed high sensitivity and specificity in distinguishing malignant from benign PNs identified through LDCT. This tool has the potential to significantly lower both mortality and health-care costs associated with the overtreatment of benign nodules, offering a promising approach to improving lung cancer screening protocols.

Keywords: Diagnosis, Early stage, Lung cancer, Plasma, Sputum

1. INTRODUCTION

Non-small cell lung cancer (NSCLC) represents the most common type of lung cancer and the leading cause of cancer deaths. NSCLC is a heterogeneous disease that mainly includes adenocarcinoma (AC) and squamous cell carcinoma (SCC) [1,2]. AC typically originates in cells that secrete mucus and other substances. It usually begins in the peripheral lung tissue, which refers to the outer areas of the lungs. SCC, on the other hand, typically starts in the squamous cells that line the inside of the airways in the lungs. It tends to be located centrally in the lung, often in the bronchi, which are the two main airways that branch off from the trachea (windpipe) to the lungs [3]. Timely and precise identification of lung cancer can make a significant difference in reducing mortality rates [4-6]. Low-dose computed tomography (LDCT) is now being used for screening lung cancer, which has been shown to reduce mortality [3]. However, over a quarter of patients screened with LDCT have indeterminate pulmonary nodules (PNs), with 4% of these cases diagnosed as lung tumors and over 95% being benign [3]. Patients with indeterminate PNs are often subjected to expensive and invasive biopsies and treatments, leading to significant morbidity and mortality. Clinical efforts have been made to distinguish malignant from benign PNs. The lung imaging reporting and data system has successfully reduced the false-positive rate by 50%, significantly cutting down on unnecessary treatments and associated expenses [7]. Despite this improvement, 44% of patients still undergo invasive procedures with only a 5% chance of having malignancy. In addition, 35% of surgical resections are found to be benign. This highlights the unmet clinical need for developing biomarkers that can accurately distinguish between benign and malignant PNs.

We have previously developed separate sputum and plasma biomarker panels for lung cancer using non-coding RNAs (miRNAs, lncRNAs, and snoRNAs), DNA methylation, and bacterial DNA methylation [8,9].

***Corresponding author:**
Feng Jiang (fjiang@som.umaryland.edu)

Received: 12 July 2024; Revision received: 28 July 2024;
Accepted: 14 August 2024; Published: 06 September 2024

How to cite this article: Liao J, Dhilipkannah P, Jiang F. Improving CT scan for lung cancer diagnosis with an integromic signature. *J Biol Methods*. 2024;11(3):e99010023. DOI: 10.14440/jbm.2024.0028

Although showing promising, diagnostic performance of the individual types of molecular biomarkers is not sufficient to be used in clinical practice for distinguishing between benign and malignant PNs. Due to the heterogeneity of NSCLC, which develops from various molecular types, a single class of biomarkers tested in one sample type may not provide sufficient diagnostic significance for early lung cancer detection. In addition, biomarkers in sputum are derived from cells shed from the bronchial epithelium in the main bronchi or large airways, where SCC frequently occurs. Conversely, biomarkers in plasma are from flowing molecules discharged by lung tumors and free-floating cancer cells, rendering sputum biomarkers less sensitive for ACs, and plasma biomarkers less sensitive for SCCs. In this study, we determined if the integration of multiple molecular biomarkers can improve the separation of benign from malignant growths.

2. MATERIALS AND METHODS

2.1. Patient cohorts and research design

We conducted this study under the approval of the University of Maryland's Institutional Review Boards (ethical approval code: IRB HP-00040666). Eligible participants were current and former smokers aged 50–80 with PNs detected through CT scan. PNs were defined according to standard clinical guidelines as nodules <3 cm in diameter identified on LDCT scans. We collected demographic and clinical data from medical records, including age, gender, race, and detailed smoking history. Malignant diagnoses were confirmed through pathological examination of tissues obtained through surgery or biopsy. Benign diagnoses were established either through specific pathological confirmation or by clinical and radiographic stability of the PNs over a 2-year follow-up period with multiple assessments.

2.2. Collection and preparation of specimens

In the collection of sputum, to decrease the proportion of oral epithelial cells, participants were instructed to clear their nasal passages, rinse their mouths thoroughly, and drink water. Sputum was harvested into sterile containers and managed in 1 h. Paque or solid parts were isolated with forceps. The samples were treated with PBS and dithiothreitol and then filtered. Blood was taken, and plasma was immediately prepared within an hour of collection using a standard clinical protocol, as described previously [9].

2.3. DNA and RNA isolation

DNA and RNA were purified using standardized protocols, with all samples promptly stored at −80°C for future use [8,9].

2.4. Droplet digital polymerase chain reaction (ddPCR) analysis of DNA methylation and bacterial DNA

To analyze DNA methylation, we performed bisulfite conversion on DNA using the Zymo EZ DNA methylation kit according to the manufacturer's protocol. A 22 μL volume of PCR mix, including primers and bisulfite-treated DNA, was loaded onto a plate. We utilized the primers for RASSF1A as described in a cited study. The automated droplet generator (Bio-Rad) was used to prepare droplets of the PCR reaction, which were then transferred to a 96-well PCR plate. The plate was then placed in a thermal cycler for amplification, including the following steps: activation at 95°C for 10 min, denaturation at 94°C for 30 s, and annealing/extension at 60°C for 1 min, repeated for 40 cycles, followed by a final extension at 98°C for 10 min. Following amplification, we put the plate in a Bio-Rad droplet reader and calculated the copy number of methylated DNA per microliter, by employing Poisson distribution analysis based on the fraction of positive reactions. To detect bacterial DNA, we used ddPCR following the method detailed in our previous study, utilizing specific PCR primers to amplify bacterial genera DNA. We combined 20 μL of PCR reaction with 70 μL of droplet generation oil for probes (Bio-Rad) and generated droplets using the droplet generator. The amplification was carried out as described above. The ddPCR system's software was employed for data acquisition, determining the concentration of DNA in copies per microliter using Poisson distribution analysis based on the fraction of positive reactions.

2.5. ddPCR analysis of miRNA, lncRNAs, and snoRNAs

To prepare cDNA from RNA, one μL of RNA was used for reverse transcription (RT) with the TaqMan miRNA RT Kit (Applied Biosystems, Foster City, CA) and gene-specific primers. For the ddPCR setup, a 20 μL reaction mixture was prepared, consisting of 5 μL cDNA, 10 μL Supermix, and 1 μL TaqMan primer/probe mix. This mixture was combined with droplet generation oil in a cartridge and processed using the QX100 droplet generator. The resulting droplets were then transferred to a 96-well PCR plate. PCR amplification was conducted on a T100 thermal cycler with the following protocol: initial enzyme activation at 95°C for 10 min, 40 cycles of denaturation at 94°C for 30 s, annealing/extension at 60°C for 1 min, and a final enzyme deactivation at 98°C for 10 min. The concentration of the original target was accurately determined by analyzing the number of positive reactions using Poisson distribution. The Bio-Rad software facilitated data acquisition and calculated the target RNA concentration in copies per microliter.

2.6. Clinical models

We conducted an analysis to assess the accuracy of two clinical models, namely the Mayo Clinic model and the VA model, in estimating the probability of NSCLC using a specific equation [10,11]. The Mayo Clinic model is characterized by the following equations: pre-test probability of a malignant PN = exp(x)/(1 + exp(x)) x = 26.8272 + (0.0391 * age) + (0.7917 * smoke) + (1.3388 * cancer) + (0.1274 * diameter) + (1.0407 * spiculation) + (0.7838 * upper). The VA model is defined by the following equation: pre-test probability of a malignant PN = exp(x)/(1 + exp(x)) x = 28.404 + (2.061 * smoke) + (0.779 * age10) + (0.112 * diameter^2) + (0.567 * yearsquit10).

2.7. Statistical analysis

To develop a panel of biomarkers for lung cancer, our cohort of cases and controls was randomly divided into two sets: a training set and a validation set, following the recommended guidelines of the National Cancer Institute [12]. The training set comprised 80% of the cohort to ensure a sufficient sample size for robust model development and parameter estimation. The remaining 20% formed the validation set, serving as an independent dataset for model evaluation and assessment of generalizability. Within the training set, feature selection was performed using the least absolute shrinkage and selection operator (LASSO) method on candidate biomarkers to develop a logistic regression model. In the logistic regression analysis, a constant term was used to represent the estimated log odds of the event (presence of neoplastic mass) when all predictor variables are zero. Coefficients were used to quantify the relationship between each predictor variable and cancer status. Positive coefficients indicated a positive association, while negative coefficients suggested a negative association with the event of interest. Lower p-values (<0.05) indicated stronger evidence against the null hypothesis, supporting the presence of a significant association. To evaluate the performance of the signature model, 10-fold cross-validation was employed during training, and validation was conducted using a separate validation set. Bootstrapping was utilized to randomly select multiple subgroups of the training set, reducing the impact of outlier data and training an ensemble model. Variable importance was assessed using the mean decrease in the Gini impurity, and the false discovery rate control was applied to adjust for multiple testing and identify differentially abundant biomarkers between the groups. For the specific selection of the lung cancer model, an empirical Bayes linear model was employed, considering clinical covariates. Discriminatory performance was evaluated using receiver-operator characteristic (ROC) analysis, with the area under the curve (AUC) reported along with 1000 bootstrap bias-corrected 95% confidence intervals. Confidence intervals for performance variables were calculated using various statistical methods. For AUC, methods such as the DeLong method or bootstrapping were used. Sensitivity and specificity were estimated using Agresti–Coull interval based on the binomial distribution. To assess the sensitivity of the signature to changes in the training cohort, cross-validation or bootstrapping techniques were employed. These rigorous evaluation methods involved repeatedly resampling the training data to assess the performance. The model acquired using the training set was further confirmed using the separate validation set for final analysis. In addition, a likelihood ratio test was conducted to compare our new signature model with our previous biomarker panels and the Mayo Clinic and VA Models in terms of performance and predictive accuracy.

3. RESULTS

3.1. Patients and controls

We recruited a total of 400 smokers with PNs for our study. Among the participants, 150 were diagnosed with neoplastic masses, while the remaining 250 had benign PNs. The benign PNs were further categorized based on their diagnoses: of 125 patients with granulomatous infection, 73 demonstrated common infection, and 52 had other lung diseases. To facilitate the development and validation of our diagnostic tools, the entire cohort was randomly divided into two distinct sets: 80% of the patients were assigned to the training set and the remaining 20% were allocated to the validation set. Detailed demographic and clinical characteristics of these sets are presented in Tables 1 and 2, respectively.

3.2. The diagnostic performance of the individual plasma and sputum bacterial biomarkers for distinguishing between malignant and benign PNs

A total of 14 potential molecular biomarkers that were identified in our previous studies were analyzed in specimens of the cases and controls using ddPCR [8,9]. The 14 potential biomarkers included sputum miRNA-31-5p, sputum miRNA-210-3p, sputum lncRNA-SNHG1, sputum lncRNA-H19, sputum lncRNA-HOTAIR, sputum snoRD66, sputum snoRD78, plasma miRNA-205-5p, plasma miRNA-126-5p, plasma lncRNA-SNHG1, plasma lncRNA-RMRP, sputum RASSF1A methylation, sputum Acidovorax biomarker, and sputum Veillonella biomarker (Table 3). ddPCR analysis of the ncRNAs, DNA methylation, and bacterial DNA in sputum and/or plasma created more than 10,000 droplets per well of plates. The sputum biomarkers, including two miRNAs, three lncRNAs, two snoRNAs, and two bacteria, displayed a different level in the training set of 120 patients and 200 controls (all $P \leq 0.05$) (Table 3). The specific sputum biomarkers showed an AUC of 0.68–0.82, creating 62.1–77.5% sensitivities and 58.8–87.0% specificities (Table 3). The sputum biomarkers demonstrated a strong association with SCC (all $P \leq 0.05$) and showed higher diagnostic efficacy for SCC, with sensitivities ranging from 70.8% to 87.5% and specificities between 64.0% and 89.5%. In contrast, the

Table 1. A training cohort of subjects with neoplastic masses (NSCLC) and controls

Patient information	NSCLC cases (*n*=120)	Controls (*n*=200)	*P*-value
Age	66.29 (SD 11.25)	65.38 (SD 10.23)	0.30
Sex			0.29
Female	43	72	
Male	77	128	
Race			0.35
African Americans	38	64	
White Americans	82	136	
Smoking pack-years (median)	32.9	31.7	0.12
PN size (mm)	23.5±7.32	12.6±5.32	<0.01
The number of spiculated PNs	72	40	<0.01
The location of PNs in the upper lobe	62	108	0.32
The number of PNs with <20 mm	48	76	<0.01
The number of PNs with ≥20 mm	72	124	<0.01
Stage			
Stage I	66		
Stage II	23		
Stage III-IV	31		
Histological type			
Adenocarcinoma	72		
Squamous cell carcinoma	48		

NSCLC: Non-small cell lung cancer; SD: Standard deviation; PNs: Pulmonary nodules.

Table 2. A validation cohort of subjects with neoplastic masses (NSCLC) and controls

Patient information	NSCLC cases (*n*=30)	Controls (*n*=50)	*P*-value
Age	66.82 (SD 11.35)	65.82 (SD 10.56)	0.29
Sex			0.30
Female	11	18	
Male	19	32	
Race			0.32
African Americans	10	16	
White Americans	20	34	
Smoking pack-years (median)	34.7	32.8	0.27
Nodule size (mm)	24.3±7.61	12.7±5.26	<0.01
The number of spiculated PNs	18	10	<0.01
The location of PNs in the upper lobe	16	25	0.43
The number of PNs with <20 mm	12	31	<0.01
The number of PNs with ≥20 mm	18	19	<0.01
Stage			
Stage I	16		
Stage II	9		
Stage III-IV	5		
Histological type			
Adenocarcinoma	18		
Squamous cell carcinoma	12		

NSCLC: Non-small cell lung cancer; SD: Standard deviation; PNs: Pulmonary nodules.

diagnostic performance for AC was lower, with sensitivities of 59.7–70.8% and specificities of 56.0–84.5% ($P < 0.05$) (Table 4).

Plasma biomarkers, which included two miRNAs (miR-205-5p and miR-126-5p) and two lncRNAs (SNHG1 and RMRP), exhibited significantly altered levels in patients with neoplastic masses compared to healthy controls (all $P \leq 0.001$) (Table 3). Each of these biomarkers demonstrated an AUC between 0.71 and 0.80. Their diagnostic performance for detecting NSCLC showed sensitivities ranging from 68.3% to 75.0% and specificities between 54.0% and 82.9% (Table 3). Furthermore, the plasma biomarkers were significantly associated with AC (All $P \leq 0.05$). Consequently, the plasma biomarkers had a higher diagnostic value for AC with 76.0–77.8% sensitivities and 56.0–86.0% specificities compared with SCC (58.3–70.0% sensitivities and 52.5–78.0% specificities, all $P < 0.05$) (Table 4).

The analysis using logistic regression showed that plasma-miRN-205-5p, sputum miRNA-210-3p, sputum lncRNA-SNHG1, sputum lncRNA-H19, sputum snoRD66, sputum lncRNA-HOTAIR, sputum RASSF1A mutation, and sputum Veillonella were associated with smoking status ($P < 0.05$) (Supplementary Table S1). Plasma lncRNA-SNHG1, sputum miRNA-31-5p, and sputum miR-126-5p were related to patient gender ($P < 0.05$) (Supplementary Table S2). Sputum snoRD78 and sputum snoRD66 were associated with patient race ($P < 0.05$) (Supplementary Table S3). Sputum lncRNA-SNHG1, sputum lncRNA-H19, sputum lncRNA-HOTAIR, sputum snoRD66, plasma lncRNA-RMRP, sputum RASSF1A methylation, and sputum Acidovorax were associated with chronic obstructive pulmonary disease (COPD) in patients ($P < 0.05$) (Supplementary Table S4).

3.3. An integromic signature for identifying malignant PNs

In addition to the molecular biomarkers, factors such as smoking history, the diameter and spiculation of the PNs, and their location in the upper lobes were also linked to malignancy (Supplementary Table S5). To develop a comprehensive diagnostic signature, we employed logistic regression models with constrained parameters. This model incorporated biomarkers, as well as clinical and radiological features, and was constructed using the LASSO method based on the ROC criterion. We identified plasma miR-205-5p, plasma miR-210-3p, sputum methylation of RASSF1A, smoking pack-year, and diameter of PN as the most informative predictors of cancer and used them to develop a prediction signature. The signature had a 0.97 AUC in distinguishing malignant from benign PNs (Figure 1A). Adding other biomarkers and imaging/clinical variables did not improve the performance of the signature for predicting malignant PNs. Furthermore,

Table 3. The individual biomarkers are relative/related to lung cancer

Biomarkers	Coefficients	*P*-value	AUC	Sensitivity, %	Specificity, %
Sputum miRNA-31-5p	0.208	0.0100	0.819	65.8	87.0
Sputum miRNA-210-3p	0.167	0.0357	0.805	77.5	78.0
Sputum lncRNA-SNHG1	0.478	<0.0001	0.775	71.7	73.0
Sputum lncRNA-H19	0.486	<0.0001	0.786	70.0	80.0
Sputum lncRNA-HOTAIR	0.177	0.0423	0.715	71.7	71.0
Sputum snoRD66	0.349	<0.0001	0.798	71.7	86.0
Sputum snoRD78	0.410	<0.0001	0.762	74.1	70.0
Plasma miRNA-205-5p	0.339	<0.0001	0.766	75.0	71.0
Plasma miRNA-126-5p	0.337	<0.0001	0.708	68.3	66.5
Plasma lncRNA-SNHG1	0.366	<0.0001	0.725	70.0	54.0
Plasma lncRNA-RMRP	0.288	0.0001	0.799	73.3	82.0
Sputum DNA methylation, RASSF1A	0.350	<0.0001	0.717	70.8	71.5
Sputum bacterial biomarker, Acidovorax	0.321	<0.0001	0.716	62.1	79.0
Sputum bacterial biomarker, Veillonella	0.329	<0.0001	0.679	69.2	58.8

AUC: Area Under the Curve

Table 4. Performance of individual biomarkers

Biomarkers	Diagnostic performance for AC		Diagnostic performance for SCC		Diagnostic performance for NSCLC	
	Sensitivity, %	Specificity, %	Sensitivity, %	Specificity, %	Sensitivity, %	Specificity, %
Sputum miRNA-31-5p	59.7	84.5	75.0	89.5	65.8	87.0
Sputum miRNA-210-3p	70.8	75.0	87.5	82.0	77.5	78.0
Sputum lncRNA-SNHG1	66.7	69.5	79.2	76.0	71.7	73.0
Sputum lncRNA-H19	65.3	78.0	77.1	83.0	70.0	80.0
Sputum lncRNA-HOTAIR	68.1	69.0	77.1	73.0	71.7	71.0
Sputum snoRD66	69.4	83.5	75.0	89.5	71.7	86.0
Sputum snoRD78	69.4	67.0	81.3	74.0	74.1	70.0
Plasma miRNA-205-5p	75.0	71.0	75.0	70.0	75.0	71.0
Plasma miRNA-126-5p	75.0	69.0	58.3	64.5	68.3	66.5
Plasma lncRNA-SNHG1	75.0	56.0	60.4	52.5	70.0	54.0
Plasma lncRNA-RMRP	77.8	86.0	66.7	78.0	73.3	82.0
Sputum RASSF1A methylation	69.4	69.5	72.9	73.0	70.8	71.5
Sputum Acidovorax	59.7	79.0	70.8	82.0	62.1	79.0
Sputum Veillonella	62.5	56.0	79.2	64.0	69.2	58.8

AC: Adenocarcinoma; SCC: Squamous Cell Carcinoma; NSCLC: Non-small cell lung cancer.

several models that are based on only radiological and clinical characteristics of smokers have been developed for predicting the probability of malignant PNs, [10,11,13,14]. The diagnostic signature achieved an AUC of 0.97 in differentiating malignant from benign PNs (Figure 1A). Incorporating additional biomarkers, imaging, or clinical variables did not enhance the signature's predictive accuracy for malignancy. In addition, multiple models focusing solely on the radiological and clinical features of smokers have been created to estimate the likelihood of malignant PNs, [10,11,13,14] of which, the Mayo Clinic model and VA model are commonly used ones. [10,11] We applied the models [10,11] in the training set. The Mayo Clinic model and VA model create AUCs of 0.82 and 81 (Figure 1A). The direct comparison of the two methods demonstrated that the integromic signature outperformed the Mayo Clinic and VA models, with a significantly higher AUC (0.97), sensitivity (89.2%), and specificity (94.7%) compared to the models (AUC 0.81–82, sensitivity 78.1–81.2%, specificity 73.2–78.5%, $P < 0.05$) (Figure 1A). Moreover, there were no significant differences in the diagnostic performance of the biomarkers across different lung cancer types, but this was not the case for tumor stages. Consequently, the integromic signature demonstrated significantly lower sensitivity (83%) for stage I NSCLC compared to Stage II and Stage III-IV (88% and 92%, respectively) ($P < 0.05$) while maintaining a specificity of 95% (Figure 1B). In addition, the integromic signature showed a higher AUC for the diagnosis of lung cancer in larger nodules (≥20 mm) compared to smaller nodules (<20 mm) (0.98, 0.93, $P = 0.043$), resulting in significantly higher sensitivity (90.3%) and specificity (96.0%) for lung cancer detection in the larger nodules as

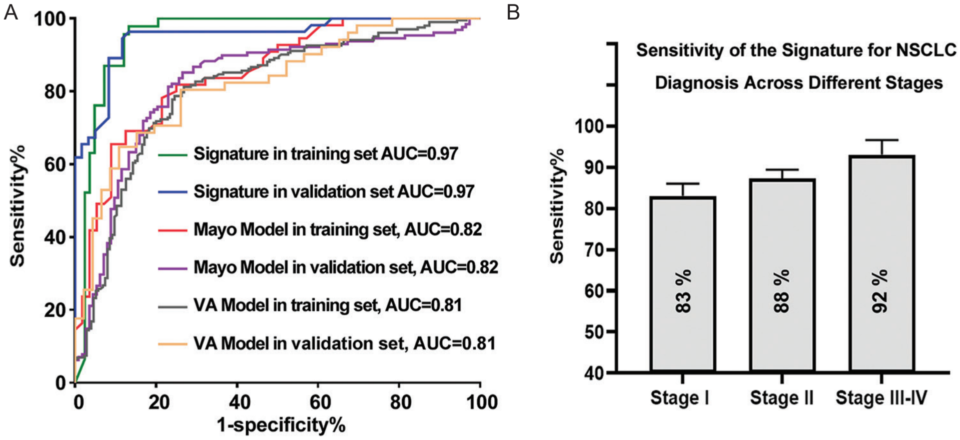

Figure 1. The signature in distinguishing malignant from benign PNs. (A) ROC curves showing the accuracy of the signature and Mayo Clinic and VA models for diagnosing malignant PNs. The signature produced an AUC of 0.97 in the training and validation sets, with a higher AUC value than the Mayo Clinic model and VA model. (B) The signature exhibited lower sensitivity for Stage I NSCLC (83%) compared with Stage II and Stage III-IV (88% and 92%) while having 95%.

PNs: Pulmonary nodules; ROC: Receiver-operator characteristic; AUC: Area under the curve; NSCLC: Non-small cell lung cancer

compared to the smaller ones (85.4%, 91.2%, all $P < 0.05$) (Supplementary Table S6).

3.4. Biomarker validation

The integromic signature was evaluated in a validation cohort comprising 30 cases and 50 controls. The results showed a comparable AUC (0.97) for diagnosing lung cancer, similar to that obtained in the training set (Figure 1A). The integromic signature demonstrated a 90.0% sensitivity and 94.0% specificity, which were not significantly different from the performance observed in the training set. In agreement with the results in the training set, the signature was unrelated with cancer types (all $P > 0.05$) but dependent on tumor stages ($P < 0.05$). The signature demonstrated a reduced sensitivity for Stage I NSCLC compared to advanced stages, as indicated by a statistical significance of $P < 0.05$. Consistent with the results in the first cohort, the signature revealed a higher AUC for diagnosing lung cancer in larger nodules (≥20 mm) compared to smaller nodules (<20 mm). This difference resulted in a higher diagnostic sensitivity (89.0%) and specificity (95.0%) in larger nodules compared to smaller ones (sensitivity of 83.3% and specificity of 90.3%, $P < 0.05$) (Supplementary Table S6).

4. DISCUSSION

This study revealed that integrating various molecular biomarkers with patients' smoking history and nodule size enhanced diagnostic accuracy compared to assessing a single biomarker type in isolation. Unlike plasma biomarkers, which are more specific to AC, and sputum biomarkers, which are more specific to SCC, this combined analysis transcends NSCLC histological differences. Furthermore, the signature outperformed established clinical models, offering greater diagnostic value.

The integrated signature includes miR-205-5p, miR-210-3p, RASSF1A, and *Veillonella.* Alterations in miR-205-5p facilitate tumorigenesis by regulating TP53INP1, RB1, and P21 [15-19]. miR-210-3p can promote lung cancer development and metastasis by disrupting USF1's activation of PCGF3, providing insights into the underlying mechanisms of lung cancer progression [20]. Abnormal miR-210-3p expression in body fluids can be used to diagnose various types of malignancies [20-26]. RASSF1A, a tumor suppressor gene, is frequently inactivated by promoter hypermethylation in lung cancer, leading to uncontrolled cell proliferation and reduced apoptosis [27-32]. Bacterial infections may contribute to the development and progression of lung cancer [33-44]. Our earlier research revealed that bacterial infections could promote tumorigenesis by activating NF-kB pathways through the binding of PspC to PAFR [45]. Recently, we have demonstrated that analyzing the presence of Acidovorax and Veillonella in sputum samples could enhance lung cancer detection [46]. In this study, we found that the combined use of *Veillonella* with ncRNA and DNA methylation had complementary function to improve the diagnosis of NSCLC.

This study offered important insights while also pointing out the need for further investigation. One limitation is the relatively small sample size, especially in the validation cohort. To address this, we intend to initiate a new study aimed at prospectively validating these biomarkers for early lung cancer detection in a larger population. Although our current analysis considered multiple variables, additional stratification could yield more in-depth understanding. We will analyze the performance of the signature specifically in subgroups based on smoking status, COPD presence, and nodule size to refine its diagnostic accuracy and clinical application. The sensitivity of the signature was significantly better for advanced lung cancer stages and larger tumors, suggesting current findings are more applicable to advanced and larger cancers. We are validating these biomarkers in the context of LDCT screening for smaller nodules and focusing on identifying additional biomarkers specific to early-stage disease. The critical role of smoking cessation in lung cancer prevention and early detection, as well as the potential benefits of biomarkers, will also need to be investigated. The patients with malignant nodules were selected from a clinical series, resulting in larger nodule sizes compared to those typically detected through screening. We will further validate the biomarkers in screening contexts to ensure applicability to smaller, screening-detected nodules.

5. CONCLUSION

We created a signature by combining biomarkers with clinical and radiological features to distinguish lung cancer in patients with PNs. This signature has the potential to reduce unnecessary invasive procedures for individuals with benign nodules while ensuring timely and appropriate therapy for those with NSCLC. To standardize this diagnostic assay, we need further validation in larger multi-center cohorts, regulatory approval, and a streamlined workflow for sample processing and analysis, in collaboration with clinical laboratories and regulatory bodies.

ACKNOWLEDGMENTS

We thank all the subjects who participated in this study.

FUNDING

This work was supported by Food and Drug administration (grant no.: 5U01FD005946-06).

CONFLICT OF INTEREST

Authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Conceptualization: Feng Jiang
Data curation: Pushpawallie Dhilipkannah
Formal analysis: Jipei Liao, Feng Jiang
Investigation: Jipei Liao, Feng Jiang
Methodology: Jipei Liao, Feng Jiang
Writing – original draft: Jipei Liao, Feng Jiang
Writing – review & editing: All authors

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the University of Maryland's Institutional Review Boards (approval no.: IRB HP-00040666). All participants provided informed consent before their inclusion in the study.

CONSENT FOR PUBLICATION

All the participants gave consent to publish their date in this study.

AVAILABILITY OF DATA

The datasets utilized and analyzed in this study are available from the corresponding author on reasonable request. However, individual subject data are not publicly accessible to protect participant consent and confidentiality.

REFERENCES

1. Rashidi A, Kao R, Echeverria R, Sadigh G. Lung cancer screening updates: Impact of 2023 American Cancer Society's guidelines for lung cancer screening. *Clin Imaging*. 2024;13:110229. doi: 10.1016/j.clinimag.2024.110229
2. Wolf AMD, Oeffinger KC, Shih TYC, *et al*. Screening for lung cancer: 2023 guideline update from the American Cancer Society. *CA Cancer J Clin*. 2024;74:50-81. doi: 10.3322/caac.21811
3. Aberle DR, van der Aalst CM, de Jong PA, *et al*. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395-409. doi: 10.1056/NEJMoa1911793
4. Bonney A, Malouf R, Marchal C, *et al.* Impact of low-dose computed tomography (LDCT) screening on lung cancer-related mortality. *Cochrane Database Syst Rev*. 2022;8:CD013829. doi: 10.1002/14651858.CD013829.pub2
5. Oudkerk M, Liu S, Heuvelmans MA, Walter JE, Field JK. Lung cancer LDCT screening and mortality reduction - evidence, pitfalls and future perspectives. *Nat Rev Clin Oncol*. 2021;18:135-151. doi: 10.1038/s41571-020-00432-6
6. Reich JM. Estimated impact of LDCT-identified stage IA non-small-cell lung cancer on screening efficacy. *Lung Cancer*. 2006;52:265-271. doi: 10.1016/j.lungcan.2006.02.007
7. McKee BJ, Regis SM, McKee AB, Flacke S, Wald C. Performance of ACR lung-RADS in a clinical CT lung screening program. *J Am Coll Radiol*. 2016;13:R25-R29.
8. Su Y, Fang H, Jiang F. Integrating DNA methylation and microRNA biomarkers in sputum for lung cancer detection. *Clin Epigenetics*. 2016;8:109. doi: 10.1186/s13148-016-0275-5
9. Zhou H, Liao J, Leng Q, Chinthalapally M, Dhilipkannah P, Jiang F. Circulating bacterial DNA as plasma biomarkers for lung cancer early detection. *Microorganisms.* 2023;11:582. doi: 10.3390/microorganisms11030582
10. Gould MK, Ananth L, Barnett PG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary

pulmonary nodules. *Chest*. 2007;131:383-388. doi: 10.1378/chest.06-1261

11. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med*. 1997;157:849-855.
12. Feng Z, Pepe MS. Adding rigor to biomarker evaluations-EDRN experience. *Cancer Epidemiol Biomarkers Prev*. 2020;29:2575-2582. doi: 10.1158/1055-9965.EPI-20-0240
13. McWilliams A, Tammemagi MC, Mayo RC, *et al*. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med*. 2013;369:910-919. doi: 10.1056/NEJMoa1214726
14. Schultz EM, Sanders GD, Trotter PR, *et al*. Validation of two models to estimate the probability of malignancy in patients with solitary pulmonary nodules. *Thorax*. 2008;63:335-341. doi: 10.1136/thx.2007.084731
15. Liang G, Li G, Wang Y, Lei W, Xiao Z. Aberrant miRNA expression response to UV irradiation in human liver cancer cells. *Mol Med Rep*. 2014;9:904-910. doi: 10.3892/mmr.2014.1901
16. Lu LG, Zhang GM. Serum miR-205-5p level for non-small-cell lung cancer diagnosis. *Thorac Cancer*. 2022;13:1102-1103. doi: 10.1111/1759-7714.14356
17. Zhao YL, Zhang JX, Yang JJ, *et al*. MiR-205-5p promotes lung cancer progression and is valuable for the diagnosis of lung cancer. *Thorac Cancer*. 2022;13:832-843. doi: 10.1111/1759-7714.14331
18. Xu LB, Xiong J, Zhang YH, *et al*. miR2053p promotes lung cancer progression by targeting APBB2. *Mol Med Rep*. 2021;24:588. doi: 10.3892/mmr.2021.12227
19. Zhu H, Xu Y, Li M, Chen Z. Inhibition sequence of miR-205 hinders the cell proliferation and migration of lung cancer cells by regulating PETN-mediated PI3K/AKT signal pathway. *Mol Biotechnol*. 2021;63:587-594. doi: 10.1007/s12033-021-00321-y
20. Chen Q, Zhang H, Zhang J, *et al*. miR-210-3p promotes lung cancer development and progression by modulating USF1 and PCGF3. *Onco Targets Ther*. 2021;14:3687-3700. doi: 10.2147/OTT.S288788
21. Eilertsen M, Andersen S, Al-Saad S, *et al*. Positive prognostic impact of miR-210 in non-small cell lung cancer. *Lung Cancer*. 2014;83:272-278. doi: 10.1016/j.lungcan.2013.11.005
22. Puissegur MP, Mazure NM, Bertero T, *et al*. miR-210 is overexpressed in late stages of lung cancer and mediates mitochondrial alterations associated with modulation of HIF-1 activity. *Cell Death Differ*. 2011;18:465-478. doi: 10.1038/cdd.2010.119
23. Hisakane K, Seike M, Sugano T, *et al*. Exosome-derived miR-210 involved in resistance to osimertinib and epithelial-mesenchymal transition in EGFR mutant non-small cell lung cancer cells. *Thorac Cancer*. 2021;12:1690-1698. doi: 10.1111/1759-7714.13943
24. Wang L, He J, Hu H, *et al*. Lung CSC-derived exosomal miR-210-3p contributes to a pro-metastatic phenotype in lung cancer by targeting FGFRL1. *J Cell Mol Med*. 2020;24: 6324-6339. doi: 10.1111/jcmm.15274
25. Grosso S, Doyen J, Parks SK, *et al*. MiR-210 promotes a hypoxic phenotype and increases radioresistance in human lung cancer cell lines. *Cell Death Dis*. 2013;4:e544. doi: 10.1038/cddis.2013.71
26. Wang H, Bian S, Yang CS. Green tea polyphenol EGCG suppresses lung cancer cell growth through upregulating miR-210 expression caused by stabilizing HIF-1alpha. *Carcinogenesis*. 2011;32:1881-1889. doi: 10.1093/carcin/bgr218
27. Xu W, Ye J, Cao Z, Zhao Y, Zhu Y, Li L. Glucocorticoids in lung cancer: Navigating the balance between immunosuppression and therapeutic efficacy. *Heliyon*. 2024;10:e32357. doi: 10.1016/j.heliyon.2024.e32357
28. Bai Y, Wang Y, Qin J, *et al*. Systematic pan-cancer analysis identified RASSF1 as an immunological and prognostic biomarker and validated in lung cancer. *Heliyon*. 2024;10:e33304. doi: 10.1016/j.heliyon.2024.e33304
29. Mashayekhi M, Asadi M, Hashemzadeh S, *et al*. Promoter methylation levels of RASSF1 and ATIC genes are associated with lung cancer in Iranian patients. *Horm Mol Biol Clin Investig*. 2023;44:145-152. doi: 10.1515/hmbci-2022-0007
30. Walter RFH, Rozynek P, Casjens S, *et al*. Methylation of L1RE1, RARB, and RASSF1 function as possible biomarkers for the differential diagnosis of lung cancer. *PLoS One*. 2018;13:e0195716. doi: 10.1371/journal.pone.0195716
31. Xiao G, Zhang T, Yao J, Ren J, Cao W, Wu G. The association between RASSF1 gene polymorphisms and lung cancer susceptibility among people in Hubei Province of China. *J Huazhong Univ Sci Technolog Med Sci*. 2009;29:646-649. doi: 10.1007/s11596-009-0522-5
32. Buckingham L, Faber LP, Kim A, *et al*. PTEN, RASSF1 and DAPK site-specific hypermethylation and outcome in surgically treated stage I and II nonsmall cell lung cancer patients. *Int J Cancer*. 2010;126:1630-1639. doi: 10.1002/ijc.24896
33. Hridoy HM, Hossain MP, Ali MH, *et al. Alocasia macrorrhiza* rhizome lectin inhibits growth of pathogenic bacteria and human lung cancer cell *in vitro* and Ehrlich ascites carcinoma Zcell *in vivo* in mice. *Protein Expr Purif*. 2024;219:106484. doi: 10.1016/j.pep.2024.106484
34. Wang W, Liang X, Kong H, *et al*. Correlation analysis of lung mucosa-colonizing bacteria with clinical features reveals metastasis-associated bacterial community structure in non-small cell lung cancer patients. *Respir Res*. 2023;24:129. doi: 10.1186/s12931-023-02420-7
35. Shi H, Chen L, Liu Y, *et al*. Bacteria-driven tumor microenvironment-sensitive nanoparticles targeting hypoxic regions enhances the chemotherapy outcome of lung cancer. *Int J Nanomedicine*. 2023;18:1299-1315. doi: 10.2147/IJN.S396863

36. Wong-Rolle A, Dong Q, Zhu Y, *et al*. Spatial meta-transcriptomics reveal associations of intratumor bacteria burden with lung cancer cells showing a distinct oncogenic signature. *J Immunother Cancer*. 2022;10:e004698. doi: 10.1136/jitc-2022-004698
37. Qian X, Zhang HY, Li QL, *et al*. Integrated microbiome, metabolome, and proteome analysis identifies a novel interplay among commensal bacteria, metabolites and candidate targets in non-small cell lung cancer. *Clin Transl Med*. 2022;12:e947. doi: 10.1002/ctm2.947
38. Chen Y, Wen F, Chen H, *et al*. Analysis of the pathogenic bacteria, drug resistance, and risk factors of postoperative infection in patients with non-small cell lung cancer. *Ann Palliat Med*. 2021;10:10005-10012. doi: 10.21037/apm-21-2364
39. Liu H, Liu B, Zheng F, Chen X, Ye L, He Y. Distribution of pathogenic bacteria in lower respiratory tract infection in lung cancer patients after chemotherapy and analysis of integron resistance genes in respiratory tract isolates of uninfected patients. *J Thorac Dis*. 2020;12:4216-4223. doi: 10.21037/jtd-20-928
40. Gui Q, Li H, Wang A, *et al*. The association between gut butyrate-producing bacteria and non-small-cell lung cancer. *J Clin Lab Anal*. 2020;34:e23318. doi: 10.1002/jcla.23318
41. Sun M, Bai Y, Zhao S, *et al*. Gram-negative bacteria facilitate tumor progression through TLR4/IL-33 pathway in patients with non-small-cell lung cancer. *Oncotarget*. 2018;9:13462-13473. doi: 10.18632/oncotarget.24008
42. Liu J, Wu Q, Li L, *et al.* Discovery of phylloseptins that defense against gram-positive bacteria and inhibit the proliferation of the non-small cell lung cancer cell line, from the skin secretions of *Phyllomedusa* frogs. *Molecules*. 2017;22:1428. doi: 10.3390/molecules22091428
43. Ye M, Gu X, Han Y, Jin M, Ren T. Gram-negative bacteria facilitate tumor outgrowth and metastasis by promoting lipid synthesis in lung cancer patients. *J Thorac Dis*. 2016;8: 1943-1955. doi: 10.21037/jtd.2016.06.47
44. Chow SC, Gowing SD, Cools-Lartigue JJ, *et al*. Gram negative bacteria increase non-small cell lung cancer metastasis via Toll-like receptor 4 activation and mitogen-activated protein kinase phosphorylation. *Int J Cancer*. 2015;136:1341-1350. doi: 10.1002/ijc.29111
45. Li N, Zhou H, Holden VK, *et al*. *Streptococcus pneumoniae* promotes lung cancer development and progression. *iScience*. 2023;26:105923. doi: 10.1016/j.isci.2022.105923
46. Leng Q, Holden VK, Deepak J, Todd NW, Jiang F. Microbiota biomarkers for lung cancer. *Diagnostics* (*Basel*). 2021;11:407. doi: 10.3390/diagnostics11030407