

A Prediction Model Based on DNA Methylation Biomarkers and Radiological Characteristics for Identifying Malignant From Benign Pulmonary Nodules

Wenqun Xing

Henan Cancer Hospital Department of Thoracic Surgery

Haibo Sun

Henan Cancer Hospital Department of Thoracic Surgery

Chi Yan

Henan Cancer Hospital Department of Molecular Pathology

Chengzhi Zhao

Henan Cancer Hospital Department of Molecular Pathology

Dongqing Wang

Henan Cancer Hospital Department of Molecular Pathology

Mingming Li

Excellen Medical Technology Co., Ltd.

Jie Ma (✉ majie_fzbl@163.com)

Department of Molecular Pathology, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, No 127, Dongming Road, Zhengzhou, Henan 450008, China <https://orcid.org/0000-0001-5346-8457>

Research article

Keywords: CT, DNA methylation, biomarkers, lung cancer, pulmonary nodules

Posted Date: October 29th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-97310/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on March 10th, 2021. See the published version at <https://doi.org/10.1186/s12885-021-08002-4>.

Abstract

Background

Lung cancer remains the leading cause of cancer deaths across the world. Early detection of lung cancer by low-dose computed tomography (LDCT) can reduce the mortality rate. However, making a definitive preoperative diagnosis of malignant pulmonary nodules (PNs) found by LDCT is a clinical challenge. This study aimed to develop a prediction model based on DNA methylation biomarkers and radiological characteristics for identifying malignant pulmonary nodules from benign PNs.

Methods

We assessed three DNA methylation biomarkers (*PTGER4*, *RASSF1A*, and *SHOX2*) in a training cohort of 110 individuals with PNs. Using univariate and multivariate logistic regression analysis, we developed a prediction model based on the three DNA methylation biomarkers and one radiological characteristic for identifying malignant from benign PNs. The performance of the prediction model with that of the methylation biomarkers and the Mayo Clinic model were compared using the non-parametric approach of DeLong et al. with the area under a receiver operator characteristic curve (AUC) analysis.

Results

The developed prediction model achieved a sensitivity of 87.3% and a specificity of 95.7% with an AUC value of 0.951 in malignant PNs diagnosis, being significantly higher than the three DNA methylation biomarkers (84.1% sensitivity and 89.4% specificity, $p=0.013$) or clinical/radiological characteristics (76.2% sensitivity and 87.2% specificity, $p=0.001$) alone. Validation of the prediction model in the testing cohort of 100 subjects with PNs confirmed the diagnostic value.

Conclusion

We have shown that integrating DNA methylation biomarkers and radiological characteristics could more accurately identify lung cancer in subjects with CT-found PNs. The prediction model developed in our study may provide clinical utility in combination with LDCT to improve the over-all diagnosis of lung cancer.

1. Background

Lung cancer is the second most common cancer globally and the leading cause of cancer mortality worldwide [1]. In 1987, it surpassed breast cancer as the leading cause of cancer-related deaths of women. By 2020, Lung cancer is expected to account for 22% of all female cancer deaths and 23% of all male cancer deaths [1].

The exceptional high mortality of lung cancer can be attributed to a high degree by late diagnosis. The 5-year survival rate of lung cancer is only 15–19% at all stages, outcomes can be significantly better at an

early-stage diagnosis, especially for stage I, the 5-year survival rate can increase up to 81%-85% [2, 3]. Thus, it seems reasonable to improve lung cancer screening at earlier stages. Low-dose computed tomography (LDCT) is widely accepted as a reliable screening tool for early detection of lung cancer. The National Lung Screening Trial (NLST) reported that LDCT decreases the mortality rate of 20% in high-risk people [4, 5]. However, due to the widespread use of LDCT, pulmonary nodules (PNs) are encountered with increased frequency in asymptomatic individuals, the diagnostic accuracy of LDCT screening is limited by its high false-positive rate and overdiagnosis. The National Lung Screening Trial showed that in heavy smokers, the positive rate of indeterminate PNs detected by LDCT was 24.2%; however, 96.4% of these PNs were ultimately confirmed to be false positives over the three rounds of screening [5]. Currently, to predict the malignancy probability of PNs found by LDCT, a series of examination techniques have been proposed, including noninvasive and invasive approaches [6]. Each approach has advantages and disadvantages. Noninvasive approaches include follow-up with positron emission tomography, LDCT, or magnetic resonance imaging for up to 2 years to determine whether it is a benign lesion. These noninvasive approaches often result in unnecessary radiation exposure, anxiety, procedures, and additional cost for subjects with benign lesions. A CT-guided transthoracic needle biopsy can establish a specific benign or malignant diagnosis but is invasive, potentially risky, and sometimes non-diagnostic [7]. Thus, it is clinically significant to develop new approaches to accurately identify patients with malignant from benign PNs safely and cost-effectively.

Analysis of lung tumor-associated molecular changes in body fluids may provide a safe and cost-effective approach for detecting lung cancer. DNA methylation is a relatively stable biochemical modification; it can be detected not only from tissue but also in serum and plasma [8]. Assessment of DNA methylation in plasma offers a potentially cost-effective method in discriminating malignant from benign PNs. Prostaglandin E receptor 4 gene (*PTGER4*), ras association domain family 1A (*RASSF1A*), and short stature homeobox gene two (*SHOX2*) methylation have been separately identified as valuable biomarkers for lung cancer diagnosis in several research studies [9–12]. However, investigating whether the three methylation biomarkers are useful in distinguishing lung cancer among individuals with LDCT-detected PNs has hardly been reported.

Previous studies also showed that, based on subjects' demographic characteristics and radiological features of PNs on CT images, the constructed predictive models could identify malignant from benign PNs [13–16]. For example, Swensen et al. developed a Mayo Clinic model based on six independent predictors (patients' age, smoking history, cancer history, nodule diameter, upper lobe position, and spiculation), which had an AUC of 0.83 for the diagnosis of malignant PNs [13]. Gould et al. established another prediction model, which yielded 0.78 AUC based on age, smoking history, nodule diameter, and smoking cessation [14, 15]. Recently, McWilliams et al. also developed two similar prediction models, with AUCs of 0.89–0.91 [16]. Although these clinical/radiological characteristics-based models are promising in identifying malignant PNs, the sensitivity and specificity still need improvement.

Considering the complex tumor microenvironment and clonal selection in lung cancer development, using circulating biomarkers alone or clinical/radiological factors alone might not have sufficient diagnostic

accuracy for lung cancer. We aimed to investigate if combining DNA methylation biomarkers with clinical/radiological characteristics could more efficiently distinguish malignant from benign lung nodules detected by LDCT.

2. Materials And Methods

2.1. Ethics

We enrolled participants in Henan Cancer Hospital, the Affiliated Cancer Hospital of Zhengzhou University. All participants signed the informed consent before blood collection, and they were informed of the usage of plasma and the test results. The current study has been received approval from the Medical Ethics Committee of Henan Cancer Hospital (2018157).

2.2. Patient cohorts and study design

The participants were enrolled with lung nodules newly detected in Henan Cancer Hospital from January 2019 to December 2019. We obtained blood samples from all subjects who met the selection criteria. The inclusion criteria were: (I) subjects detected pulmonary nodules on CT scans in 2019. (II) LDCT-derived nodule diameter between 4 and 35 mm; (III) the participants' clinical information should be complete. The exclusion criteria were: (I) pregnancy or lactation; (II) current pulmonary infection; (III) surgery within six months; (IV) radiotherapy within one year; and (V) life expectancy of < 1 year. We collected the general characteristics, nodules, and clinical characteristics information of participants from the hospital information system. General characteristics included age, gender, smoking behavior (smoking status, pack-years, and the number of years since quitting), and cancer history. Nodule radiographic characteristics comprised the maximum transverse size; the location in the lungs; and nodule type (nonsolid or ground-glass opacity, perifissural, part-solid, solid, and spiculation). Malignant diagnosis of PNs was verified based on the pathologic examination of tissues obtained via surgery or biopsy. Furthermore, it was determined as benign if a specific benign etiology was confirmed pathologically, or after a 2-year follow-up with multiple examinations, the PNs were clinically and radiographically stable based on the Fleischner Society guidelines [17]. The surgical pathologic staging was determined based on the TNM guidelines classification criteria [18]. According to the World Health Organization classification to determine the histopathologic classification [19].

2.3. Sample collection and storage

Plasma samples were collected from outpatients and inpatients of Henan Cancer Hospital, and the sample information was recorded in sample collection forms. 5 ml of peripheral blood from the subject was drawn in a 5-ml K2EDTA anticoagulant tube (BD biosciences, Franklin Lakes, NJ, USA). The plasma sample's storage and transportation followed the instructions of the Nucleic Acid Extraction Reagent (Excellen Medical Technology Co., Ltd.).

2.4. DNA Isolation and Bisulfite Conversion

Blood samples were collected before surgery, anesthesia, and adjuvant therapy. The collected specimens were processed within 4 hours by centrifuging at 3,000 g for 10 minutes at 4 °C. Then, we transferred the collected plasma to a new tube and stored at -80 °C until use. DNA was extracted from plasma using the Nucleic Acid Extraction Reagent (Excellen Medical Technology Co., Ltd.) according to the instructions. In Brief, circulating DNA was extracted from 2 mL of plasma utilizing magnetic beads, then converted the unmethylated cytosine residue to uracil residue in DNA by a bisulfite reaction. After further purification, bisulfite-converted DNA (bisDNA) was eluted in 35 µL and ready for real-time PCR use.

2.5. DNA Methylation Analysis

DNA methylation analysis was performed according to the diagnostic kit's instructions (Excellen Medical Technology Co., Ltd.). The eluted DNA was used as a template for fluorescent real-time PCR. Each PCR reaction mixture has a total reaction volume of 25 µL, including 12.5 µL reaction buffer, 2.5 µL primer mix, and 10 µL eluted DNA. Fluorescence PCR amplifications were performed on 96-well plates of Applied Biosystems 7500 Fast Real-Time PCR Systems. Each sample was carried out in triplicate. In addition to subject DNA samples, each plate also included positive controls (in vitro methylated leukocyte DNA), negative controls (normal leukocyte DNA or DNA from a known unmethylated cell line), and water blanks. The thermal profile for amplification reactions was 98 °C for 5 minutes, followed by 45 cycles at 95 °C for 10 seconds and 63 °C for 5 seconds to 58 °C for 30 seconds. In the PCR reaction, the primers and probes were designed to amplify the methylated sequences preferentially. During the PCR process, the methylated target sequence can be exclusively identified from unmethylated DNA. Increased inflorescent emission of the reporter dye can be detected on fluorescence channels of FAM, HEX, Texas Red, and CY5. The resulting data were analyzed by Applied Biosystems 7500 Fast Real-Time PCR System Sequence Detection Software v1.4.1.

2.6. Statistical Analysis

We used univariate analysis to test the differences in radiological variables between subjects with malignant and benign nodules. For differed significantly variables, we then analyzed by using multivariate logistic regression models with stepwise regression based on the receiver operator characteristic (ROC) curve to construct a prediction model for identifying malignant PNs. We reported the area under the ROC curve (AUC) and 95% confidence intervals (CI) and obtained the optimal cutoff value using the Youden index [20]. The ROC curve was constructed to assess each model's diagnostic accuracy by calculating the area under the ROC curve (AUC) and 95% confidence intervals (CI). The non-parametric approach of DeLong et al. was used to compare the performance of the prediction model with that of the plasma biomarkers and the Mayo Clinic model [21]. The prediction model developed in a cohort's training set and blindly validated in an additional set of subjects by comparing the calculated results with the final clinical diagnosis and the AUCs. All statistical analyses were performed using the IBM SPSS Statistics 24 and MedCale Statistics. P values < 0.05 were considered to indicate statistical significance.

3. Results

3.1. Clinical characteristics of subjects

Altogether, we recruited 110 participants with PNs as a training cohort and 100 participants with PNs as a validated cohort. Of the training cohort, 63 had malignant PNs and were diagnosed with LC, and 47 had benign PNs (Table 1). Subjects with lung cancer were generally older than subjects with benign nodules (58 vs 55 years). Of the subjects, 63 (57.3%) were male, and 69 (62.7%) were non-smokers. The 63 subjects with malignant PNs were diagnosed with adenocarcinomas (n = 37), squamous cell carcinomas (n = 14), small cell lung cancer (n = 8) and, unclassified lung cancer patients (n = 4). The LC patients consisted of 17 stage I, 21 stage II, and 25 stage III to IV cases. One hundred subjects with PNs were used as a validated cohort to confirm the prediction model for the differentiation of malignant from benign PNs. The cohort consisted of 57 subjects with malignant PNs (LC) and 43 subjects with benign PNs (Table 2). Of the patients with malignant PNs, 32 were diagnosed with adenocarcinomas, 14 were diagnosed with squamous cell carcinomas, 2 were diagnosed with small cell lung cancer, and 9 were unclassified lung cancer patients. The demographic and clinical parameters, including detailed information about the two cohorts' nodule characteristics, are shown in Tables 1 and 2, respectively.

Table 1—Subjects' Characteristics of Training Study		
Characteristics	Subjects with Malignant PNs (n=63)	Subjects with Benign PNs (n=47)
Clinical		
Age (Years)		
Median Age	58	55
Age Range	36-77	26-70
Sex		
Male	36	27
Female	27	20
Smoking history		
Non-smoker	30	39
Ex-smoker	12	3
Current smoker	21	5
Smoking pack-years		
Mean Pack-years (Smokers only)	36.52	20
Years quit (Smoker sonly)	8.9	3
Histology subtype		
Adenocarcinoma	37	
Squamous cell carcinoma	14	
Small cell lung cancer	8	
Other	4	
Stage		
I	17	
II	21	
III-IV	25	
Radiological		
Nodule size (mm)	21.46 (SD 10.52)	11.89 (SD 6.81)
Nodule location		

Left lower lobe	9	6
Left upper lobe	17	12
Right lower lobe	15	15
Right middle lobe	3	4
Right upper lobe	19	10
Nodule type (number)		
Nonsolid or ground-glass opacity	17	13
Perifissural	6	5
Part-solid	9	8
Solid	13	11
Spiculation	18	10
Abbreviations: PN, pulmonary nodule; SD, standard deviation		

Table 2—Subjects' Characteristics of Validation Study		
Characteristics	Subjects with Malignant PNs (n=57)	Subjects with Benign PNs (n=43)
Clinical		
Age (Years)		
Median Age	62	54
Age Range	38-78	27-72
Sex		
Male	39	29
Female	18	14
Smoking history		
Non-smoker	25	30
Ex-smoker	12	4
Current smoker	20	9
Smoking pack years		
Mean Pack-years (Smokers only)	40.77	21.64
Years quit (Smoker sonly)	7.35	3.75
Histology subtype		
Adenocarcinoma	32	
Squamous cell carcinoma	14	
Small cell lung cancer	2	
Other	9	
Stage		
I	18	
II	20	
III-IV	19	
Radiological		
Nodule size (mm)	21.83 (SD 10.88)	11.22 (SD 7.56)
Nodule location		

Left lower lobe	12	10
Left upper lobe	12	10
Right lower lobe	11	11
Right middle lobe	5	4
Right upper lobe	17	8
Nodule type (number)		
Nonsolid or ground-glass opacity	17	13
Perifissural	8	6
Part-solid	9	7
Solid	13	10
Spiculation	10	7
Abbreviations: PN, pulmonary nodule; SD, standard deviation		

3.2. Diagnostic accuracy of the three methylation biomarkers for identifying malignant PNs

To determine the diagnostic values of the three methylation biomarkers, we quantitative analysis of promoter methylation in the plasma DNA samples from the training cohort of 110 subjects using the diagnostic kit for the methylated gene of lung cancer (Excellen Medical Technology Co., Ltd.). We performed receiver operating characteristic (ROC) curve analysis to analyze the diagnostic efficacy of using the three the methylation (*PTGER4*, *RASSF1A*, and *SHOX2*) in blood plasma. As shown in Fig. 1, the three DNA methylation used in combination yielded 0.912 AUC in identifying malignant from benign PNs and generated a sensitivity of 84.1% and a specificity of 89.4% (Fig. 1a). No statistically significant association was observed between the logistic model with subjects' age, gender, and smoking history (all $p > 0.05$).

3.3. Developing a prediction model based on the methylation biomarkers and radiographic features of PNs for distinguishing malignant from benign PNs

Although the combined use of the three DNA methylation showing promise with a sensitivity of 84.1% and 89.4% specificity, it is not sufficient for identifying malignant PNs in the clinic. Using the univariate analysis, we analyzed 110 subjects of the training cohort to determine which clinical and radiological variables were associated with malignant PNs. As shown in Table 3, smoking history and the diameter of the PNs were related to malignant PNs. Then, we used multivariate logistic regression models with stepwise regression based on the ROC curve to construct a prediction model, including *PTGER4*,

RASSF1A, and *SHOX2* diameter of PNs. The prediction model produced 0.951 AUC in identifying malignant PNs from benign (Fig. 1b), and adding other clinical/ radiological variables to the prediction model would not improve the performance. It has been reported that several prediction models based on PNs parameters on CT images and clinical characteristics of subjects developing to predict the probability of malignant PNs [13–16], of which the Mayo Clinic model is a commonly used one. We also applied the equation of the Mayo Clinic model [13] in the training cohort of 110 subjects to predict malignant PNs, as shown in Fig. 1c. The AUC value obtained by the Mayo Clinic model was 0.823, and the value was similar to the previous reports [13–15]. The AUC value of the prediction model (0.951, 95% CI: 0.892–0.983) was significantly higher than the panel of the three methylation biomarkers (0.912, 95% CI: 0.843–0.958) used alone and the Mayo Clinic model (0.823, 95% CI: 0.739–0.890) (Figs. 1ac). Our prediction model achieved 87.3% sensitivity and 95.7% specificity for identifying malignant PNs, which were significantly higher compared with the biomarkers (84.1% sensitivity and 89.4% specificity, $p = 0.013$) and the Mayo Clinic model (76.2% sensitivity and 87.2% specificity, $p = 0.001$) (Figs. 1ac).

Table 3
Association of clinical and radiological variables with malignant PNs

Variables	OR	95% CI	P value
Age (year)	1.04	0.10–1.08	0.055
Sex	1.01	0.47–2.17	0.975
Smoking history	5.36	2.16–13.29	0.0001
Nodule diameter on CT	1.13	1.07–1.2	< 0.0001
Nodule spiculation on CT	1.48	0.61–3.60	0.386
Upper lobe locations of PNs	1.52	0.71–3.24	0.283
Abbreviations: PN, pulmonary nodule.			

3.4. Validating the prediction model for identifying malignant PNs in an independent cohort

The expression levels of the three gene methylation (*PTGER4*, *RASSF1A* and, *SHOX2*) panel were further confirmed in an independent cohort. The integration of the three gene methylation showed a similar change tendency in the validated cohort as in the training cohort, which indicated that the gene methylation could be reproducibly measured. As shown in Table 4, the AUC value of the prediction model in the validated cohort (0.948) was similar as in the training cohort (0.951). We used the optimal cut-offs obtained in the training set to determine the three models' diagnostic performance in the validated cohort. The prediction model produced a sensitivity of 89.5% and specificity of 95.4%, significantly higher than the panel of the biomarkers (82.5% sensitivity and 86.0% specificity) and the Mayo Clinic model (75.4% sensitivity and 86.0% specificity) in identifying malignant from benign PNs. Taken together, these results

confirmed that the prediction model had the potential for estimating of malignant PNs among individuals with CT-detected PNs.

Table 4
Comparison of the prediction model, panel of the three DNA methylation biomarkers, and Mayo Clinic model for distinguishing malignant from benign PNs in validation cohorts of subjects*

Approaches	Sensitivity (95% CI)	Specificity (95% CI)	AUC
The prediction model	89.5% (78.5%-96.0%)	95.4% (84.2%-99.4%)	0.948
The biomarker panel	82.5% (70.1%-91.3%)	86.0% (72.1%-94.7%)	0.912
The Mayo Clinic model	75.4% (62.2%-85.9%)	86.0% (72.1%-94.7%)	0.829
Abbreviations: CI, confidence interval; AUC, area under the ROC curve			

4. Discussion

Low-dose spiral computed tomography (LDCT), a reliable screening tool for early detection of lung cancer, was severely limited by its low specificity [4, 5]. LDCT dramatically increases the number of indeterminate pulmonary nodules (PNs), whereas most PNs are ultimately false positives [22]. It is clinically significant to develop new methods that can precisely identify malignant from benign PNs safely and cost-effectively.

Some clinical/radiological characteristics-based models have shown the potential to identify malignant PNs [13–15]. The finding from our present study confirmed the previous observations. However, the moderate sensitivity and specificity of these models limit the application in clinical. DNA methylation plays a vital role in tumorigenesis at an early stage [23–25], which makes DNA methylation alterations among the most promising candidates in biomarker research. In the present study, we first combined detection of *PTGER4*, *RASSF1A*, and *SHOX2* methylation biomarkers for estimating malignant from benign PNs in a training cohort. The three methylation biomarkers used in combination produced an AUC value of 0.912. Despite showed promising, the diagnostic accuracy also needed to be further improved. We developed a novel lung nodule risk prediction model by integrating the three DNA methylation biomarkers with one radiological variable of PNs to estimate the probability of malignancy in PNs. The prediction model has higher sensitivity and specificity than the Mayo Clinic model or the panel of biomarkers used alone. Furthermore, in an independent cohort, the prediction model's performance validated, further confirming the tremendous potential for detecting malignant PNs. In addition, the predictive model is a convenient analytic method with a simple equation and a single cutoff value. Our current findings suggested that the prediction model with three DNA methylation biomarkers and the diameter of PNs may potentially guide the management of CT screening results.

Based on the Food and Drug Administration criteria, a disease with a 5% prevalence, the screening test should have a sensitivity exceeding 95% when the specificity \leq 95%, and vice versa [26]. The prevalence

of lung cancer in high-risk populations is 1–3%, while LDCT has about 90% sensitivity and only 61% specificity, which is prone to produce a high false-positive rate. The ideal prediction model should have > 95% specificity and appropriate sensitivity for identifying malignant PNs, thus could augment the performance of LDCT for lung cancer screening [27]. Our result appears promising; the developed prediction model achieved a sensitivity of 87.3% and a specificity of 95.7% with an AUC value of 0.951 in malignant PNs diagnosis, which suggested that the prediction model does possess the required diagnostic performance for routine clinical application.

However, our study also has some limitations. The sample size is small. The exact number of subjects in some histological subtype groups, such as small cell lung cancer, may be insufficient. Large sample size is needed in further studies to confirm the results. Furthermore, subjects in this study were recruited from hospital-based patients with PNs. The subjects might not be representative of a population-based LDCT screening setting for lung cancer. We will conduct a large trial of population-based LDCT screening to confirm the prediction model's performance in identifying malignant PNs.

5. Conclusions

In summary, we developed a simple prediction model based on DNA methylation biomarkers with radiological characteristics that could identify malignant from benign nodules detected by LDCT. Future use of the prediction model could reduce costs and avoid invasive diagnostic procedures for patients with benign PNs while allowing immediate treatment for lung cancer patients. This prediction model could be used in combination with LDCT to improve the over-all diagnosis of lung cancer. Nevertheless, undertaking a prospective study of the prediction model for malignant PNs in an extensive population-based LDCT screening is required.

Abbreviations

LDCT

Low-dose spiral computed tomography

LC

Lung cancer

PNs

Pulmonary nodules

PTGER4

prostaglandin E receptor 4

RASSF1A

Ras association domain family 1A

SHOX2

Short stature homeobox gene two

ROC

Receiver operating characteristic

AUC

Area under a receiver operator characteristic curve

CI

Confidence interval

Declarations

Ethics approval and consent to participate

We collected participation samples with written, informed consent with human ethics approval from the Medical Ethics Committee of Henan Cancer Hospital (2018157). All participants provided written informed consent.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request. Subjects data are not publicly available for containing information that could compromise participants' consent and confidentiality.

Competing interests

ML is an employee of Excellen Medical Technology Co., Ltd. All other authors have no conflicts of interest to declare.

Funding

This work was supported by Major Science and Technology Projects of Henan, 161100311500, and supported by Project Construction of Precision Medical Data Platform on Lung Cancer and other Common Malignancy.

Author contributions

WX and JM conceived and designed the paper. HS collected the data and processed subject samples. CY, CZ, and DW analyzed and interpreted the data. ML provided technical support. All authors contributed to the preparation, editing, review of the manuscript and approved the final manuscript.

Acknowledgment

We thank all the participants in this study.

References

- [1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(1):7-30. doi:10.3322/caac.21590
- [2] Begum S, Brait M, Dasgupta S, et al. An epigenetic marker panel for detection of lung cancer using cell-free serum DNA. *Clin Cancer Res.* 2011;17(13):4494-4503. doi:10.1158/1078-0432.CCR-10-3436
- [3] Blandin Knight S, Crosbie PA, Balata H, Chudziak J, Hussell T, Dive C. Progress and prospects of early detection in lung cancer. *Open Biol.* 2017;7(9):170070. doi:10.1098/rsob.170070.
- [4] Patz EF Jr, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med.* 2014;174: 269–74.
- [5] National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med.* 2011;365(5):395-409. doi:10.1056/NEJMoa1102873
- [6] Diederich S, Das M. Solitary pulmonary nodule: detection and management. *Cancer Imaging.* 2006;6(Spec No A): S42-S46. Published 2006 Oct 31. doi:10.1102/1470-7330.2006.9004.
- [7] National Lung Screening Trial Research Team, Aberle DR, Berg CD, et al. The National Lung Screening Trial: overview and study design. *Radiology.* 2011;258(1):243-253. doi:10.1148/radiol.10091808.
- [8] Hulbert A, Jusue-Torres I, Stark A, et al. Early Detection of Lung Cancer Using DNA Promoter Hypermethylation in Plasma and Sputum. *Clin Cancer Res.* 2017;23(8):1998-2005. doi:10.1158/1078-0432.CCR-16-1371.
- [9] Weiss G, Schlegel A, Kottwitz D, König T, Tetzner R. Validation of the SHOX2/PTGER4 DNA Methylation Marker Panel for Plasma-Based Discrimination between Patients with Malignant and Nonmalignant Lung Disease. *J Thorac Oncol.* 2017;12(1):77-84. doi:10.1016/j.jtho.2016.08.123.
- [10] Zhang C, Yu W, Wang L, et al. DNA Methylation Analysis of the SHOX2 and RASSF1A Panel in Bronchoalveolar Lavage Fluid for Lung Cancer Diagnosis. *J Cancer.* 2017;8(17):3585-3591. Published 2017 Sep 30. doi:10.7150/jca.21368.
- [11] Kneip C, Schmidt B, Seegebarth A, et al. SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. *J Thorac Oncol.* 2011;6(10):1632-1638. doi:10.1097/JTO.0b013e318220ef9a.
- [12] Hu H, Zhou Y, Zhang M, Ding R. Prognostic value of RASSF1A methylation status in non-small cell lung cancer (NSCLC) patients: A meta-analysis of prospective studies. *Biomarkers.* 2019;24(3):207-216. doi:10.1080/1354750X.2019.1583771.

- [13] 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(8):849-855.
- [14] Gould MK, Ananth L, Barnett PG; Veterans Affairs SNAP Cooperative Study Group. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest*. 2007;131(2):383-388. doi:10.1378/chest.06-1261.
- [15] 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(4):335-341. doi:10.1136/thx.2007.084731.
- [16] McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med*. 2013;369(10):910-919. doi:10.1056/NEJMoa1214726.
- [17] MacMahon H, Austin JH, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology*. 2005;237(2):395-400. doi:10.1148/radiol.2372041887.
- [18] Ettinger DS, Wood DE, Akerley W, et al. Non-small cell lung cancer, version 1.2015. *J Natl Compr Canc Netw*. 2014;12(12):1738-1761. doi:10.6004/jnccn.2014.0176.
- [19] Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol*. 2015;10(9):1243-1260. doi:10.1097/JTO.0000000000000630.
- [20] Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*. 2005;16(1):73-81. doi:10.1097/01.ede.0000147512.81966.ba.
- [21] Cui X, Heuvelmans MA, Han D, et al. Comparison of Veterans Affairs, Mayo, Brock classification models and radiologist diagnosis for classifying the malignancy of pulmonary nodules in Chinese clinical population. *Transl Lung Cancer Res*. 2019;8(5):605-613. doi:10.21037/tlcr.2019.09.17.
- [22] Blanchon T, Bréchet JM, Grenier PA, et al. Baseline results of the Depiscan study: a French randomized pilot trial of lung cancer screening comparing low dose CT scan (LDCT) and chest X-ray (CXR). *Lung Cancer*. 2007;58(1):50-58. doi:10.1016/j.lungcan.2007.05.009
- [23] Locke WJ, Guanzone D, Ma C, et al. DNA Methylation Cancer Biomarkers: Translation to the Clinic. *Front Genet*. 2019;10:1150. Published 2019 Nov 14. doi:10.3389/fgene.2019.01150
- [24] Fukushige S, Horii A. DNA methylation in cancer: a gene silencing mechanism and the clinical potential of its biomarkers. *Tohoku J Exp Med*. 2013;229(3):173-185. doi:10.1620/tjem.229.173.

[25] Klutstein M, Nejman D, Greenfield R, Cedar H. DNA Methylation in Cancer and Aging. *Cancer Res.* 2016;76(12):3446-3450. doi:10.1158/0008-5472.CAN-15-3278.

[26] Ma J, Guarnera MA, Zhou W, Fang H, Jiang F. A Prediction Model Based on Biomarkers and Clinical Characteristics for Detection of Lung Cancer in Pulmonary Nodules. *Transl Oncol.* 2017;10(1):40-45. doi:10.1016/j.tranon.2016. 11.001

[27] Lin Y, Leng Q, Jiang Z, et al. A classifier integrating plasma biomarkers and radiological characteristics for distinguishing malignant from benign pulmonary nodules. *Int J Cancer.* 2017;141(6):1240-1248. doi:10.1002/ijc.30822

Figures

Figure 1

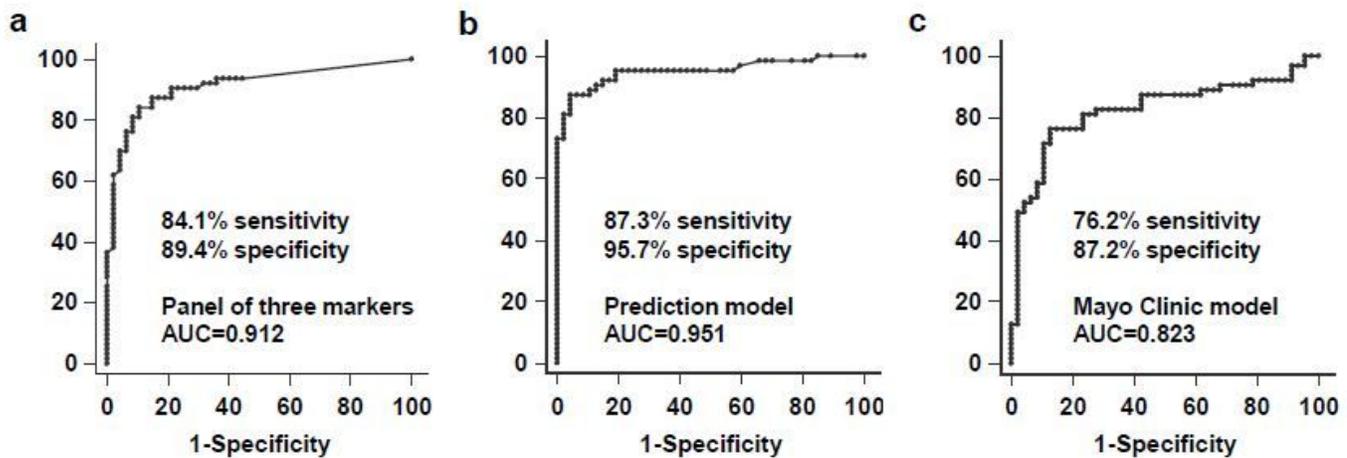


Figure 1

Receiver-operator characteristic (ROC) curve analysis of the three models in a training cohort. The area under the ROC curve (AUC) for each model conveys its accuracy for diagnosing malignant PNs. The prediction model produced a higher AUC value (b) for identifying malignant PNs, compared with the panel of the three DNA methylation biomarkers (a) and the Mayo Clinic model (c).