

## A Prediction Model Based on Biomarkers and Clinical Characteristics for Detection of Lung Cancer in Pulmonary Nodules<sup>1,2</sup>



Jie Ma<sup>\*,†</sup>, Maria A. Guarnera<sup>†</sup>, Wenxian Zhou<sup>‡</sup>,  
HongBin Fang<sup>‡</sup> and Feng Jiang<sup>†</sup>

<sup>\*</sup>Department of Clinical Biochemistry, Jiangsu University School of Medicine, Xuefu Road 301, Zhenjiang, Jiangsu Province, 212013, China; <sup>†</sup>Department of Pathology, University of Maryland School of Medicine, 10 S. Pine St., Baltimore, MD 21201, USA; <sup>‡</sup>Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University Medical Center, 4000 Reservoir Road, N.W., Washington, D.C. 20057, USA

### Abstract

Lung cancer early detection by low-dose computed tomography (LDCT) can reduce the mortality. However, LDCT increases the number of indeterminate pulmonary nodules (PNs), whereas 95% of the PNs are ultimately false positives. Modalities for specifically distinguishing between malignant and benign PNs are urgently needed. We previously identified a panel of peripheral blood mononucleated cell (PBMC)-miRNA (miRs-19b-3p and -29b-3p) biomarkers for lung cancer. This study aimed to evaluate efficacy of integrating biomarkers and clinical and radiological characteristics of smokers for differentiating malignant from benign PNs. We analyzed expression of 2 miRNAs (miRs-19b-3p and -29b-3p) in PBMCs of a training set of 137 individuals with PNs. We used multivariate logistic regression analysis to develop a prediction model based on the biomarkers, radiographic features of PNs, and clinical characteristics of smokers for identifying malignant PNs. The performance of the prediction model was validated in a testing set of 111 subjects with PNs. A prediction model comprising the two biomarkers, spiculation of PNs and smoking pack-year, was developed that had 0.91 area under the curve of the receiver operating characteristic for distinguishing malignant from benign PNs. The prediction model yielded higher sensitivity (80.3% vs 72.6%) and specificity (89.4% vs 81.9%) compared with the biomarkers used alone (all  $P < .05$ ). The performance of the prediction model for malignant PNs was confirmed in the validation set. We have for the first time demonstrated that the integration of biomarkers and clinical and radiological characteristics could efficiently identify lung cancer among indeterminate PNs.

*Translational Oncology (2017) 10, 40–45*

### Introduction

Lung cancer is the leading cancer killer in both men and women in the United States. Tobacco smoking is the major cause of the disease. Histologically, there are two major types of lung cancer: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC mainly comprises adenocarcinoma (AC), squamous cell carcinoma (SCC), and large cell carcinoma (LCC). A National Cancer Institute National Lung Screening Trial showed that the early detection of lung cancer using low-dose computed tomography (LDCT) significantly reduced the mortality [1]. LDCT is now used for lung cancer screening in smokers [2]. Medicare pays for lung cancer screening with LDCT. However, the CT scan has dramatically increased the number of

Address all correspondence to: Feng Jiang, Department of Pathology, The University of Maryland School of Medicine, 10 South Pine Street, MSTF 7th floor, Baltimore, MD 21201-1192, USA.

E-mail: [fjiang@som.umaryland.edu](mailto:fjiang@som.umaryland.edu)

<sup>1</sup>Funding sources: This work was supported in part by NCI R21CA205746, VA Merit Award 101 CX000512, award from the Geaton and JoAnn DeCesaris Family Foundation (F.J.), LUNGevity/Upstage Foundation Early Detection Award, University of Maryland Cancer Epidemiology Alliance Seed Grant, and UMD-UMB Research and Innovation Seed Grant (F.J.).

<sup>2</sup>The authors report no conflicts of interest.

Received 5 October 2016; Revised 31 October 2016; Accepted 4 November 2016

© 2016 The Authors. Published by Elsevier Inc. on behalf of Neoplasia Press, Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). 1936-5233/17

<http://dx.doi.org/10.1016/j.tranon.2016.11.001>

indeterminate pulmonary nodules (PNs) in asymptomatic individuals. To manage the indeterminate PNs, a workup consisting of noninvasive and invasive techniques is proposed [3]. Noninvasive techniques include follow-up with LDCT, positron emission tomography, or magnetic resonance imaging for up to 2 years to confirm if it is a benign lesion. These noninvasive techniques frequently lead to unnecessary procedures, radiation exposure, anxiety, cost, and low accuracy for subjects with benign lesions. Furthermore, observation with serial chest radiographs may delay appropriate diagnosis and treatment when malignancy is really existent. Invasive techniques comprise CT-guided transthoracic needle and transbronchial biopsies. There is a risk of pneumothorax, hemorrhage, and a false-negative result, which may lead to an unacceptable therapeutic delay in early-stage lung cancer [4]. In the end, a surgical procedure is often required to establish the final diagnosis, which should be avoided in cases of benign growths.

The National Lung Screening Trial results also showed that 24.2% of heavy smokers had indeterminate PNs detected by LDCT, whereas 96.4% of these PNs were ultimately confirmed as false positives [5]. CT screening trial yielded a sensitivity of more than 90% but a specificity of 61% [1]. Therefore, LDCT screening for lung cancer in heavy smokers may cause a high level of false-positive rate or overdiagnosis [5]. Given the high false-positive rate of LDCT, there will be an enormous number of referrals for the invasive and expensive 2-year multiple follow-up examinations that carry their own morbidities and mortalities. Thus, it is clinically imperative to develop new modalities that can accurately distinguish malignant from benign PNs in a safe and cost-effective manner to prevent individuals with benign growths from the biopsies and follow-up examinations while allowing effective treatments to be immediately initiated for lung cancer.

Analysis of biomarkers in body fluids may provide a safe and cost-effective approach for diagnosing lung cancer. It has been suggested that tumors could be recognized by the immune system and that cancer cells can evade the elimination of immunological surveillance [6]. Therefore, immune evasion might occur as an early event in tumorigenesis [7,8]. Peripheral blood mononucleated cells (PBMCs) mainly consist of monocytes, T cells, B cells, granulocytes, and natural killer cells [9] and act as the first line of defense against malignancy in the immune system [9]. Therefore, the determination of molecular changes of PBMCs may open a surrogate window into cancer status [8]. MicroRNAs (miRNAs) have important function in the regulation of gene expression in various biological processes [10]. Alterations of miRNAs play crucial roles in tumorigenesis [10]. Furthermore, miRNAs are involved in the escaping events of cancer cells from immunological surveillance [11]. Differential miRNA expression patterns of PBMCs have been found in patients with malignancies, suggesting that a PBMC-based miRNA profile could be of use as cancer biomarkers [12,13]. We recently identify two PBMC miRNA biomarkers (miRs-19b-3p and -29b-3p) that can diagnose lung cancer with 72.6% sensitivity and 82.6% specificity [12]. However, the sensitivity and specificity of the biomarkers are not sufficient to be used in clinical settings for differentiating between malignant and benign PNs.

Previous studies have shown that predictive models based on patient and PN variables could help discriminate lung cancer from benign growths [14–18]. However, the diagnostic performance of the prediction models also suffered from moderate sensitivity and specificity for predicting malignant PNs [14–19]. Here we aimed to investigate if combined analysis of the biomarkers, radiographic

features of PNs, and clinical characteristics of smokers could efficiently identify lung cancer among the indeterminate PNs.

## Materials and Methods

### Patient Cohorts

This retrospective cohort study was approved by the Institutional Review Boards of the University of Maryland Medical Center and the Baltimore VA Medical Center. We consented the patients who visited the Lung Nodule Clinic in the medical centers. Inclusion criteria were current and former smokers who had LDCT-detected PNs and were between the ages of 55 and 74 years. A PN was defined as a solitary, round, or oval lesion in the lung parenchyma in the absence of adenopathy, atelectasis, or pneumonia. Exclusion criteria included age < 21 years, pregnancy or lactation, current pulmonary infection, thoracic surgery within 6 months, radiotherapy to the chest within 1 year, and life expectancy of <1 year. We reviewed the medical records for their demographic and clinical variables about age, gender, race, ethnicity, history of cancer, and smoking behavior (smoking history, smoking status, pack-years, and number of years since quitting). Furthermore, radiographic characteristics of the PNs including the maximum transverse size, the visually determined type (nonsolid or with ground-glass opacity, part-solid or subsolid, or solid or periferfissural, spiculation), and the location in the lung were obtained. The ground truth of a definitive malignant diagnosis was established and verified based on pathologic examination of tissues obtained via surgery or biopsy. Furthermore, a definitive benign diagnosis was established when a specific benign etiology was confirmed pathologically or if the PNs were clinically and radiographically stable (such as CT imaging) after a 2-year follow-up with multiple examinations based on the Fleischner Society guidelines [20]. The surgical pathologic staging was determined according to the TNM classification of the International Union Against Cancer with the American Joint Committee on Cancer and the International Staging System for Lung Cancer. Histopathologic classification was determined according to the World Health Organization classification.

### PBMC Preparation and Quantitative Reverse Transcriptase Polymerase Chain Reaction (qRT-PCR) Analysis of miRNAs

We collected peripheral blood in BD Vacutainer spray-coated K<sub>2</sub>EDTA Tubes (BD, Franklin Lakes, NJ) as described in our previous report [12]. We isolated PBMCs by using Ficoll gradient centrifugation [8]. We transferred PBMCs into RNeasy Lysis Buffer (Qiagen Technologies, Grand Island, NY). We extracted RNA by using the Qiagen miRNeasy Mini Kit (Qiagen Inc., Valencia, CA) as described in our previous study [8]. We performed qRT-PCR analysis of two miRNAs (miRs-19b-3p and -29b-3p) as previously described [8]. Briefly, qRT-PCR was carried out on a CFX96 thermocycler (Bio-Rad) at 95°C for 10 minutes followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minutes. qPCR data were analyzed by using the Manager software (Bio-Rad) with an automatic cycle threshold (Ct) setting for assigning the baseline and threshold for Ct determination. Relative expression of a targeted miRNA in a given sample was computed using the equation  $2^{-\Delta Ct}$ , where  $\Delta Ct = Ct$  (targeted miRNA) - Ct (miR-423-3p), in which miR-423-3p was used as an internal control for normalization of the target miRNAs [12]. In our previous study [12], to identify an internal control for the normalization of the miRNA qPCR data, we screened the expression of all miRNAs represented in an analysis of array-based panel to find the miRNAs with the minimal variation between lung cancer cases

and cancer-free individuals. miR-423-3p was chosen as an internal control because its expression was very stable in all PBMC samples with no significant difference between the groups determined by normalizer. All assays were performed in triplicates.

### Statistical Analysis

We used univariate and multivariate analyses to determine which of the biomarkers and clinical and radiological variables were associated with malignant PNs. The significantly associated factors were then analyzed by using multivariate logistic regression models with stepwise regression based on receiver-operator characteristic (ROC) curve to select an optimal prediction model for malignant PNs. The optimal cutoff value was generated using the Youden index. The 95% confidence intervals in the ROC plot for proportions were estimated. To compare the performance of the biomarkers only and with added clinical/imaging variables, we used the method of Hanley and McNeil [21]. The performance was evaluated by tabulating the percentage of paired tests indicating a significant difference. We also generated a 95% confidence interval for the difference in the area under the ROCs (AUCs) by using the bootstrap [21]. The predictive model developed in a training set of cohort was validated in a testing set by comparing the calculated results with final clinical diagnosis and the area under the AUC.

## Results

### The Clinical and Radiological Predictors for Malignant PNs

From December 1, 2011, to December 31, 2014, we enrolled 636 heavy smokers who had PNs. Among the subjects, 124 had malignant PNs and 512 had benign PNs. Of the patients with malignant PNs, 108 were diagnosed with NSCLC, whereas 16 were diagnosed with SCLC. The NSCLC patients consisted of 33 stage I, 34 stage II, and 41 stage III to IV cases. Of the NSCLC patients, 51 were diagnosed to have AC, 47 have SCC, and 10 have LC. Of the 16 SCLC patients, 8 were diagnosed with limited stage, whereas 8 were diagnosed with extensive stage. From the 512 cancer-free smokers, we randomly selected 124 subjects with benign PNs as controls. The 248 patients with either malignant or benign PNs were randomly split into a training set and an internal testing set by using a validated random number generator. The training set included 68 cancer cases and 69 cancer-free controls (Table 1) and was used for developing a prediction model for malignant PNs. The testing set included 56 cancer cases and 55 cancer-free controls (Table 2) and was used for the validation of the prediction model. Univariate analysis showed that history of cancer and smoking pack-years were the clinical predictors, whereas the diameter, spiculation, and upper lobe location of the PNs were radiological predictors of malignant PNs (all  $P < .05$ ) (Table 3). Furthermore, multivariate analysis indicated that the smoking pack-years of the subjects and the diameter and spiculation of the PNs were predictors of malignant PNs (all  $P < .05$ ) (Table 3).

### A Prediction Model for Malignant PNs

The two PBMC miRNAs (miRs-19b-3p and -29b-3p) displayed a significantly higher expression level in individuals with malignant PNs compared with subjects with benign PNs (all  $P < .05$ ). Furthermore, the expression level of the PBMC miR-29b-3p was correlated with the size of PNs ( $P = .02$ ). The expression levels of both miR-19b-3p and miR-29b-3p were associated with smoking history of subjects ( $P = .01$ ). In addition, the expression of the two

**Table 1.** Characteristics of Patients with Malignant or Benign PNs in a Training Set

Characteristics	Patients with Malignant PNs ( $n = 68$ )	Patients with Benign PNs ( $n = 69$ )
Clinical		
Age	67.23 (SD 9.99)	66.25 (SD 8.12)
Sex		
Male	45	46
Female	23	23
Race		
African American	20	21
White	48	48
Smoking history		
Current smoker	40	41
Former smoker	28	28
Pack-years	44.76 (SD 13.27)	23.69 (SD 12.46)
Years quit	13.26 (SD 8.98)	11.38 (SD 8.68)
History of cancer	8	2
Stage of non-small cell cancer		
Stage I	18	
Stage II	18	
Stage III-VI	22	
Histological type		
AC	28	
SCC	25	
LC	5	
SCLC	10	
Radiological		
Nodule size (mm)	20.39 (SD 11.27)	12.56 (SD 8.38)
Nodule location		
Left lower lobe	9	14
Left upper lobe	25	18
Right lower lobe	15	19
Right middle lobe	4	6
Right upper lobe	15	9
Nodule type (number)		
Nonsolid or ground-glass opacity	17	19
Perifissural	6	8
Part-solid	8	7
Solid	13	12
Spiculation	22	3

miRNAs was more associated with SCC compared with other subtypes of lung cancer. Therefore, the two PBMC miRNAs could be also predictors of malignant PNs.

### A Prediction Model Based on the Biomarkers, Radiographic Features, and Clinical Characteristics of Smokers for Malignant PNs

Because the two PBMC miRNAs; history of cancer and smoking pack-years of the patients; and the diameter, spiculation, and upper lobe location of the PNs were associated with malignant PNs, we used multivariate logistic regression models with stepwise regression to develop a prediction model for lung cancer. A prediction model was developed: probability of malignant PNs =  $eU/(1 + eU)$ , where  $e$  is the base of the natural logarithm and  $U = -11.26 + 3.56 \times \log(\text{miRs-19b-3p}) - 2.82 \times \log(\text{miR29b-3p}) + 0.18 \times \text{smoking pack-years} + 3.12 \times \text{spiculation of PN}$ . Smoking pack-years = (number of cigarettes smoked per day/20)  $\times$  number of years smoked. Spiculation = 1 if the edge of the PN has spicules (otherwise = 0). The estimated AUC of the prediction model had 0.91 AUC in distinguishing malignant from benign PNs, which was significantly higher than that (0.81) of combined use of the two miRNA biomarkers (Figure 1) ( $P = .02$ ). The prediction model had 80.3% sensitivity and 89.4% specificity for detection of malignant PNs, which were also significantly higher compared with those (72.6% sensitivity and 81.9% specificity) of the panel of the two biomarkers (miRs-19b-3p and -29b-3p) (all  $P < .05$ ) (Supplementary Table 1).

**Table 2.** Characteristics of Patients with Malignant or Benign PNs in a Testing Set

Characteristics	Patients with Malignant PNs (n = 56)	Patients with Benign PNs (n = 55)
Clinical		
Age	66.98 (SD 9.35)	65.29 (SD 8.32)
Sex		
Male	37	36
Female	19	19
Race		
African American	17	16
White	39	39
Smoking history		
Current smoker	33	33
Former smoker	23	22
Pack-years	45.36 (SD 12.57)	24.23 (SD 11.56)
Years quit	13.46 (SD 8.36)	11.27 (SD 8.34)
History of cancer	6	2
Stage of non-small cell cancer		
Stage I	15	
Stage II	16	
Stage III-VI	19	
Histological type		
AC	23	
SCC	22	
LC	5	
SCLC	6	
Radiological		
Nodule size (mm)	20.48 (SD 11.57)	12.36 (SD 8.48)
Nodule location		
Left lower lobe	7	11
Left upper lobe	21	15
Right lower lobe	12	15
Right middle lobe	3	5
Right upper lobe	12	7
Nodule type (number)		0
Nonsolid or ground-glass opacity	16	17
Perifissural	7	8
Part-solid	7	6
Solid	6	5
Spiculation	18	2

Furthermore, the prediction model produced a higher specificity (94.8% vs 89.4%,  $P = .03$ ) for the detection of SCC compared with the diagnosis of all types of lung cancer while maintaining a similar sensitivity (80.4% vs 80.3%,  $P = .46$ ) (Supplementary Table 2). The prediction model did not exhibit statistical differences of sensitivity and specificity between stages of lung cancer ( $P > .05$ ).

**Validating the Prediction Model for Differentiating Malignant from Benign PNs in a Testing Cohort**

The molecular and clinical data and radiographic features of PNs in the validation set were used to test the performance of the model by comparing the calculated results with final clinical diagnosis and the

**Table 3.** Univariate and Multivariate Analyses of Potential Predictors of Malignant PNs

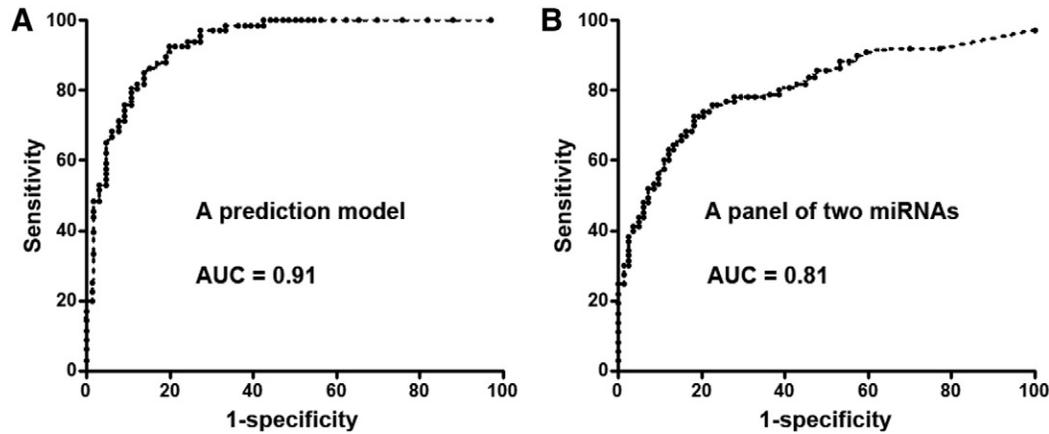
Variable	A Training Set			A Testing Set		
	OR	95% CI	P	OR	95% CI	P
Univariate analysis						
Smoking pack-year	1.22	1.01-1.26	.002	1.26	1.19-1.32	.003
History of cancer	1.35	1.30-1.40	.021	1.34	1.28-1.43	.023
Nodule size	1.16	1.12-1.22	.001	1.17	1.11-1.22	.010
Nodule Location	2.79	2.70-3.16	.036	2.66	2.56-3.15	.020
Spiculation	5.76	5.62-6.23	<.001	5.68	5.59-6.58	<.001
Multivariate analysis						
Smoking pack-year	1.26	1.12-1.30	.001	1.29	1.22-1.36	.006
Nodule size	2.23	1.88-2.86	.022	2.46	1.75-2.87	.020
Spiculation	2.68	2.03-3.69	.002	2.95	2.12-3.99	.001

AUC value. The prediction model created an AUC of 0.91 (Supplementary Table 3), suggesting its reproducibility in distinguishing between benign and malignant PNs. Furthermore, the prediction model had 80.4% sensitivity and 89.1% specificity for the identification of malignant PNs. In addition, the prediction model had higher sensitivity and specificity for detection of malignant PNs compared with the biomarkers (80.4% vs 73.2%, 89.1% vs 81.2%, all  $P < .05$ ) (Supplementary Table 3). Moreover, the prediction model had a higher specificity (94.6% vs 89.1%,  $P = .03$ ) for the detection of SCC compared with all types of lung cancer while maintaining a similar sensitivity (81.8% vs 80.4%,  $P = .39$ ) (Supplementary Table 2). The results generated from the testing cohort confirmed the potential of using the prediction model as an assay for distinguishing malignant from benign PNs.

**Discussion**

Each year, about 2.6 million subjects in the United States will be screened by using LDCT [22]. More than 700,000 heavy smokers will have a positive diagnosis without a clear indication of cancer [22]. Therefore, precisely identifying lung cancer in the subjects with indeterminate PNs is a primary medical concern. In a retrospective cohort of 629 patients with PNs, Swensen et al. identified 6 independent predictors of malignancy: older age, a history of smoking, a history of an extrathoracic cancer more than 5 years before nodule detection, larger nodule diameter, upper lobe location, and spiculated margins [14]. A prediction model was developed using the six predictors that had 0.83 AUC in the identification of the malignancy in radiologically indeterminate PNs. Furthermore, in a retrospective cohort of 375 veterans with PNs, Gould et al. derived another model including older age, a history of smoking, larger nodule diameter, and shorter time since quitting smoking, producing 0.78 AUC for estimating the clinical probability of lung cancer in patients with PNs [15,16]. In addition, using data from two separate LDCT screening cohorts, McWilliams et al. recently developed prediction models that could estimate the probability of lung nodules that might be malignant ones [17]. Although the clinical and radiological characteristics-based models show promising, the sensitivity and specificity need to be improved for lung cancer detection in PNs.

Our present study confirms the previous findings that the clinical and radiological variables could be predictors for malignant PNs, including the history of cancer; smoking pack-years of the subjects; and the diameter, spiculation, and upper lobe location of the PNs. Furthermore, this study validates the diagnostic values of our previously identified PBMC-miRNA biomarkers for lung cancer [12]. Importantly, from the heterogeneous predictors, using logistic regression models with stepwise regression, we optimize a new prediction model for distinguishing malignant from benign PNs. This parsimonious model comprises two biomarkers, one radiological variable of PNs, and one clinical factor of smokers. The prediction model produced higher sensitivity and specificity for prediction of malignant PNs compared with the biomarkers used alone. In addition, the performance of this model could identify SCC with a higher specificity compared with other types of lung cancer while keeping a similar sensitivity. Given that radiographic imaging has limited capability to differentiate between benign and malignant lesions, particularly centrally located SCCs [1,23], future use of the prediction model would help make decisions about the management of CT-detected abnormalities of undetermined importance.



**Figure 1.** ROC curve analysis of a prediction model and a panel of two PBMC miRNA biomarkers (miRs-19b-3p and -29b-3p) for distinguishing between malignant and benign PNs in a training set of patients. The AUC for each approach conveys its accuracy for diagnosis of malignant PNs. The prediction model produces a higher AUC value for identifying malignant PNs (A) compared with the panel of the two PBMC miRNA biomarkers (B) ( $P = .02$ ).

The study has some limitations. 1) Based on the Food and Drug Administration criteria [24], any screening test directed at a disease with a prevalence of 5% or less must detect preclinical disease with a sensitivity exceeding 95% when the specificity is less than or equal to 95%, and vice versa [24]. The prevalence of lung cancer in high-risk groups is at 1% to 3%, whereas LDCT has about 90% sensitivity and 61% specificity for lung cancer early detection. Ideal prediction modalities should have 95% sensitivity and 95% specificity for predicting malignant PNs and thus augment the performance of CT for targeted screening for lung cancer. However, the prediction model, although promising, does not possess the required diagnostic performance (95% sensitivity and 95% specificity) for routine clinical application. In the future, we should identify additional biomarkers and clinical and radiological characteristics that can be added to the current prediction model so that the diagnostic efficacy of the approach could be improved. 2) The sample size is small. Furthermore, cases and controls used in this study were recruited from the hospital-based patients with PNs. The prevalence of lung cancer in the hospital-based population with PNs was 19.5%. The subjects might not be representative of the individuals in LDCT screening setting for lung cancer. We will perform a prospective trial to determine if the analysis of the prediction model could be an effective high-throughput screening for specifically identifying NSCLC in a large population-based LDCT screening positive setting among heavy smokers.

## Conclusion

We have for the first time developed a parsimonious prediction model by integrating biomarkers and clinical and radiological characteristics of smokers that could identify lung cancer among indeterminate PNs. Future use of the prediction model would have the following important benefits [14]: 1) surgery for early-stage lung cancer could immediately be expedited without the risk of the tumor metastasizing during a long follow-up duration; 2) invasive and expensive bronchoscopy and transthoracic needle aspiration biopsy procedures could be used more appropriately; 3) multiple and expensive noninvasive techniques that might lead to unnecessary procedures, radiation exposure, anxiety, cost, and low accuracy for subjects with benign lesions could be avoided; and 4) the expense and risk of

surgery for benign PNs could be reduced. Nonetheless, undertaking external and prospective validation study of the prediction model for lung cancer is required.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgements

This work was supported in part by NCI R21CA1456782-01, VA Merit Award I01 CX000512, LUNGevity/Upstage Foundation Early Detection Award, University of Maryland Cancer Epidemiology Alliance Seed Grant, and UMD-UMB Research and Innovation Seed Grant (F.J.).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.tranon.2016.11.001>.

## References

- [1] Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, and Sicks JD (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* **365**, 395–409.
- [2] Moyer VA (2014). Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* **160**, 330–338.
- [3] Diederich S and Das M (2006). Solitary pulmonary nodule: detection and management. *Cancer Imaging* **6**, S42–S46.
- [4] Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, Galen B, Gareen IF, Gatsonis C, Goldin J, and Gohagan JK, et al (2011). The National Lung Screening Trial: overview and study design. *Radiology* **258**, 243–253.
- [5] Patz Jr EF, Pinsky P, Gatsonis C, Sicks JD, Kramer BS, Tammemagi MC, Chiles C, Black WC, and Aberle DR (2014). Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med* **174**, 269–274.
- [6] Melief CJ and Finn OJ (2011). Cancer immunology. *Curr Opin Immunol* **23**, 234–236.
- [7] Bluth M, Lin YY, Zhang H, Viterbo D, and Zenilman M (2008). Use of gene expression profiles in cells of peripheral blood to identify new molecular markers of acute pancreatitis. *Arch Surg* **143**, 227–233.
- [8] Baine MJ, Chakraborty S, Smith LM, Mallya K, Sasson AR, Brand RE, and Batra SK (2011). Transcriptional profiling of peripheral blood mononuclear cells in pancreatic cancer patients identifies novel genes with potential diagnostic utility. *PLoS One* **6**, e17014.

- [9] Twine NC, Stover JA, Marshall B, Dukart G, Hidalgo M, Stadler W, Logan T, Dutcher J, Hudes G, and Dorner AJ, et al (2003). Disease-associated expression profiles in peripheral blood mononuclear cells from patients with advanced renal cell carcinoma. *Cancer Res* **63**, 6069–6075.
- [10] Croce CM and Calin GA (2005). miRNAs, cancer, and stem cell division. *Cell* **122**, 6–7.
- [11] Sasaki K, Kohanbash G, Hoji A, Ueda R, McDonald HA, Reinhart TA, Martinson J, Lotze MT, Marincola FM, and Wang E, et al (2010). miR-17-92 expression in differentiated T cells — implications for cancer immunotherapy. *J Transl Med* **8**, 17.
- [12] Ma J, Lin Y, Zhan M, Mann DL, Stass SA, and Jiang F (2015). Differential miRNA expressions in peripheral blood mononuclear cells for diagnosis of lung cancer. *Lab Invest* **95**, 1197–1206.
- [13] Wang WS, Liu LX, Li GP, Chen Y, Li CY, Jin DY, and Wang XL (2013). Combined serum CA19-9 and miR-27a-3p in peripheral blood mononuclear cells to diagnose pancreatic cancer. *Cancer Prev Res (Phila)* **6**, 331–338.
- [14] Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, and Edell ES (1997). The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med* **157**, 849–855.
- [15] Gould MK, Ananth L, and Barnett PG (2007). A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest* **131**, 383–388.
- [16] Schultz EM, Sanders GD, Trotter PR, Patz Jr EF, Silvestri GA, Owens DK, and Gould MK (2008). Validation of two models to estimate the probability of malignancy in patients with solitary pulmonary nodules. *Thorax* **63**, 335–341.
- [17] McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, Yasufuku K, Martel S, Laberge F, and Gingras M, et al (2013). Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* **369**, 910–919.
- [18] Li Y, Chen KZ, and Wang J (2011). Development and validation of a clinical prediction model to estimate the probability of malignancy in solitary pulmonary nodules in Chinese people. *Clin Lung Cancer* **12**, 313–319.
- [19] Lin H, Zhong WZ, Yang XN, Zhang XC, Yang JJ, Zhou Q, Yan HH, Liao RQ, Nie Q, and Dong S, et al (2012). A clinical model to estimate the pretest probability of lung cancer, based on 1198 pedigrees in China. *J Thorac Oncol* **7**, 1534–1540.
- [20] MacMahon H, Austin JH, Gamsu G, Herold CJ, Jett JR, Naidich DP, Patz Jr EF, and Swensen SJ (2005). Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology* **237**, 395–400.
- [21] Hanley JA and McNeil BJ (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **148**, 839–843.
- [22] Goulart BH and Ramsey SD (2013). Lung cancer screening with low-dose computed tomography. *J Natl Compr Cancer Netw* **11**, 366–367.
- [23] Toloza EM, Harpole L, and McCrory DC (2003). Noninvasive staging of non-small cell lung cancer: a review of the current evidence. *Chest* **123**, 137S–146S.
- [24] Nolen BM, Lomakin A, Marrangoni A, Velikokhatnaya L, Prosser D, and Lokshin AE (2015). Urinary protein biomarkers in the early detection of lung cancer. *Cancer Prev Res (Phila)* **8**, 111–119.