Published in final edited form as:

Int J Cancer. 2017 September 15; 141(6): 1240–1248. doi:10.1002/ijc.30822.

A classifier integrating plasma biomarkers and radiological characteristics for distinguishing malignant from benign pulmonary nodules

Yanli Lin¹, Qixin Leng¹, Zhengran Jiang^{1,2}, Maria A. Guarnera¹, Yun Zhou³, Xueqi Chen⁴, Heping Wang⁵, Wenxian Zhou⁵, Ling Cai⁵, HongBin Fang⁵, Jie Li⁶, Hairong Jin⁶, Linghui Wang⁶, Shaoqiong Yi⁶, Wei Lu⁷, David Evers⁸, Carol B Fowle⁸, Yun Su⁹, and Feng Jiang^{1,8,*}

¹Department of Pathology, University of Maryland School of Medicine, Baltimore. MD. USA

²The F. Edward Hébert School of Medicine at the Uniformed Services University of the Health Sciences, Bethesda, MD. USA

³The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore. MD. USA

⁴Department of Nuclear Medicine, Peking University First Hospital, Beijing. China

⁵Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University Medical Center, Washington D.C. USA

⁶Department of thoracic surgery, the general hospital of PLA, Beijing. China

⁷Department of Medical Physics, Memorial Sloan Kettering Cancer Center, NY. USA

⁸VA Maryland Health Care System, Baltimore VA Medical Center, Baltimore, MD. USA

⁹Department of Surgery, Jiangsu Province Hospital of Traditional Chinese Medicine (TCM), Affiliated Hospital of Nanjing University of TCM. Nanjing. China

Abstract

Lung cancer is primarily caused by cigarette smoking and the leading cancer killer in the USA and across the world. Early detection of lung cancer by low-dose CT (LDCT) can reduce the mortality. However, LDCT dramatically increases the number of indeterminate pulmonary nodules (PNs), leading to overdiagnosis. Having a definitive preoperative diagnosis of malignant PNs is clinically important. Using microarray and droplet digital PCR to directly profile plasma miRNA expressions of 135 patients with PNs, we identified 11 plasma miRNAs that displayed a significant difference between patients with malignant versus benign PNs. Using multivariate logistic regression analysis of the molecular results and clinical/radiological characteristics, we

*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. fjiang@som.umaryland.edu.

Conflict of interest: The authors declare no conflict of interest.

Author contributions statement

YL, QL, and ZJ conducted the experiments and participated in data interpretation. MAG participated in coordination and acquisition of data. YZ, XC, HW, WZ, LC, HF, WL, YS participated in data analysis. FJ participated in study design, coordination, and data analysis and interpretation, and prepared the manuscript. All authors read and approved the final manuscript. The contents do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

developed an integrated classifier comprising two miRNA biomarkers and one radiological characteristic for distinguishing malignant from benign PNs. The classifier had 89.9% sensitivity and 90.9% specificity, being significantly higher compared with the biomarkers or clinical/radiological characteristics alone (All $P < 0.05$). The classifier was validated in two independent sets of patients. We have for the first time shown that the integration of plasma biomarkers and radiological characteristics could more accurately identify lung cancer among indeterminate PNs. Future use of the classifier could spare individuals with benign growths from the harmful diagnostic procedures, while allowing effective treatments to be immediately initiated for lung cancer, thereby reduces the mortality and cost. Nevertheless, further prospective validation of this classifier is warranted.

Keywords

lung cancer; miRNA; biomarkers; CT; pulmonary nodules

Lung cancer is the leading cancer killer in the USA and worldwide, of which more than 85% are non-small cell lung cancer (NSCLC). Tobacco smoking is the major cause of NSCLC. Histologically, NSCLC comprises adenocarcinoma (AC), squamous cell carcinoma (SCC), and large cell carcinoma (LCC). A NCI-National Lung Screening Trial (NLST) showed that the early detection of lung cancer using low-dose CT (LDCT) can significantly reduce the mortality¹. LDCT is now used for lung cancer screening in smokers². However, the CT scan has dramatically increased the number of indeterminate pulmonary nodules (PNs) in asymptomatic individuals. 24.2% of heavy smokers had indeterminate PNs detected by LDCT, whereas 96.4% of these PNs were ultimately confirmed as benign growths³. Furthermore, the CT screening trial by LDCT yielded a sensitivity of more than 90%, but a specificity of only 61%¹, producing a high false positive rate or overdiagnosis³. To manage the indeterminate PNs, a work-up consisting of noninvasive and invasive techniques is proposed⁴. 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 often lead to radiation exposure, anxiety, cost, and low accuracy for subjects with benign lesions. Invasive techniques comprise CT-guided transthoracic needle and transbronchial biopsies. The invasive techniques might have a risk of pneumothorax, hemorrhage, and a false negative result⁵. Thus, it is clinically imperative to develop new approaches that can accurately distinguish malignant from benign PNs in a safe and cost-effective manner to prevent individuals with benign growths from the expensive and harmful diagnostic procedures, while allowing effective treatments to be immediately initiated for NSCLC.

Analysis of lung tumor-associated molecular changes in body fluids provides a liquid biopsy for safely and cost-effectively detecting NSCLC. MicroRNAs (miRNAs) are small non-coding RNA molecules (containing about 22 nucleotides) that function as posttranscriptional regulators of gene expression⁶. Dysregulation of miRNAs plays a crucial role in lung tumorigenesis^{6, 7, 8, 9}. Due to their small size and relative resistance to nucleases, miRNAs are highly stable in peripheral plasma which is an easily accessible and rich biological fluid⁹. Plasma miRNAs that are directly released from primary lung tumors or the

circulating lung cancer cells might provide circulating biomarkers for lung cancer^{9, 10}. For example, we were one of the first to demonstrate that testing the lung tumor-associated miRNAs in plasma could discriminate NSCLC patients from healthy individuals^{8, 11–13}. Numerous miRNAs have been identified as diverse panels of biomarkers, which, however, produce widespread inconsistent results in lung cancer diagnosis¹¹. Furthermore, sensitivities and specificities of these plasma miRNA biomarkers are not high enough to be used in the clinical settings for predicting malignancy among indeterminate PNs⁸.

Analysis of patients' demographical characteristics and radiological features of PNs on CT images could also discriminate lung cancer from benign growths^{14–18}. For instance, Swensen, et al. developed a Mayo Clinic model based on patients' age, smoking status, cancer history, and location, size and shape of nodules, which could identify malignant PNs with an area under the curve (AUC) of 0.83¹⁴. Similarly, Gould, et al. established a different prediction model based on age, smoking history, nodule diameter, and time since quitting smoking, which yielded 0.78 AUC^{15, 16}. McWilliams, et al. recently developed another two prediction models with 0.89–0.91 AUCs¹⁷. However, the clinical/radiological features-based models also suffered from moderate sensitivity and specificity^{14–19}.

Given the intratumoral heterogeneity and clonal selection throughout the course of lung tumorigenesis, the use of either circulating biomarkers or clinical/radiological factors alone might not provide sufficient power for diagnosis of lung cancer. We aimed to determine if integrating plasma miRNA biomarkers with clinical/radiological characteristics could more accurately identify NSCLC among the indeterminate PNs.

Materials and Methods

Patient cohorts and research design

The study protocols were approved by the Institutional Review Boards of the University of Maryland Medical Center (UMMC), the Baltimore VA Medical Center (BVAMC), and Jiangsu Province Hospital of Traditional Chinese Medicine (JPHTCM). Inclusion criteria were current and former smokers who had CT-detected PNs and were between the ages of 55–74. Exclusion criteria included pregnancy or lactation, current pulmonary infection, thoracic surgery within 6 months, radiotherapy to the chest within 1 year, and life expectancy of < 1 year. A PN was defined as a solitary, round, or oval lesion in the lung parenchyma in the absence of adenopathy, atelectasis, or pneumonia. 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). We also obtained radiographic characteristics of the PNs on CT images, including the maximum transverse size, the visually determined type (nonsolid or ground-glass opacity, part-solid, solid, perifissural, and spiculation), and the location in the lungs. A definitive malignant diagnosis was established and verified based on pathologic examination of tissues obtained via surgery or biopsy. A definitive benign diagnosis was established when a specific benign etiology was confirmed pathologically, or the PNs were clinically and radiographically stable after a 2-year follow-up with multiple examinations^{20, 21}. 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 Cancer²². Histopathologic classification was determined according to the World Health Organization classification²³. Altogether, we recruited 135, 126, and 98 patients with PNs from UMMC, BVAMC, and JPHTCM, respectively. Of the UMMC cohort, 69 had malignant PNs and were diagnosed with NSCLC, and 66 had benign PNs (Table 1). The 66 subjects with benign PNs were diagnosed with granulomatous inflammation (n = 34), nonspecific inflammatory changes (n = 23), or lung infections (n = 9). From the UMMC cohort, we randomly selected 18 individuals with malignant PNs and 18 individuals with benign PNs, from whom, the plasma samples were analyzed by using a microarray to identify miRNA biomarker candidates for lung cancer. The identified miRNAs were then validated in all 135 plasma samples of the UMMC cohort by using droplet digital PCR (ddPCR). The resulted molecular data and clinical/radiological characteristics of the UMMC cohort of patients with PNs were analyzed to identify an optimal panel of biomarkers and then construct a classifier for identifying malignant PNs. 126 patients with PNs recruited from BVAMC were used as an independent testing cohort, while 98 patients with PNs recruited from JPHTCM were used an external testing cohort to confirm the classifier for the differentiation of malignant from benign PNs. The BVAMC cohort consisted of 63 patients with malignant PNs (NSCLC) and 63 patients with benign PNs (Table 2). The 63 subjects with benign PNs were diagnosed with granulomatous inflammation (n = 30), nonspecific inflammatory changes (n = 19), or lung infections (n = 14). In the JPHTCM cohort, 49 had malignant PNs (NSCLC) and 49 had benign PNs (Supplementary table 1). The 49 subjects with benign PNs were diagnosed with granulomatous inflammation (n = 26), nonspecific inflammatory changes (n = 17), or lung infections (n = 6). The demographic and clinical parameters, including information about nodules size, of the three cohorts are shown in Tables 1–2 and Supplementary table 1, respectively.

Blood collection, plasma preparation, and RNA isolation

The blood samples were collected before any treatment regimen. Variability in the blood collection and preparation might have confounding effects on the molecular analysis of the body fluid specimens. Furthermore, qualities of RNA samples are crucial for the accurate and robust measurement of plasma miRNAs. To reduce the variability and bias linked to sampling methods, storage or purification, in the three medical centers we collected blood and prepared plasma using the standard operating protocols (SOPs) developed by The National Cancer Institute Early Detection Research Network (EDRN)^{24, 25}. Furthermore, the release of contaminating miRNAs in plasma by hemolysis of blood cells such as red blood cells (RBCs) could yield nonspecific results. To avoid the contamination, we prepared plasma from blood within 2 hours after the collection as previously described^{11, 26, 27}. Moreover, we used RBC lysis solution to maximally reduce the possible contamination from RBCs in plasma^{13, 44–49}. We extracted RNA from plasma by using a protocol with miRNeasy Mini Kit spin column as described in our published work^{8, 11–13, 26–41}. In addition, we analyzed expression levels of RBC-related miRNA (mir-451), myeloid-related miRNA (miR-223), and lymphoid-associated miRNA (miR-150) in all the RNA samples. The samples that were positive to these blood cells-related miRNAs were excluded from the study. RNA was immediately stored at –80 in a barcoded cryotube until use.

Microarray analysis

The plasma RNA specimens were analyzed for miRNA expressions by using “Exiqon Services” (Exiqon, Denmark) with an established protocol as described in our previous reports^{41, 42}. Briefly, 6 μ l RNA was reversely transcribed in 30 μ l reactions using the miRCURY LNATM Universal RT miRNA PCR, Polyadenylation and cDNA synthesis kit (Exiqon). cDNA was diluted 100 x and assayed in 10 μ l PCR reactions according to the protocol for miRCURY LNATM Universal RT miRNA PCR. Each miRNA was assayed by qPCR on the miRNA Ready-to-Use PCR, Hama Panels I using ExiLent SYBR® Green master mix. Negative controls excluding template from the reverse transcription reaction were performed and profiled like the samples. The amplification was performed in a LightCycler® 480 Real-Time PCR System (Roche, San Francisco, CA) in 384 well plates. The amplification curves were made by using quantification cycle (Cq), which was used as a relative value for further quantification of the tested genes. We normalized the resulted data by using the average of assays detected in the samples (average-assay Cq).

Droplet digital PCR (ddPCR)

ddPCR analysis for quantification of miRNAs was done as described in our published studies^{26, 33, 43}. Briefly, 1 μ l RNA per sample was obtained for RT to produce cDNA by TaqMan miRNA RT Kit (Applied Biosystems, Foster City, CA) and specific primers for each gene. 20 μ l reaction mixture containing 5 μ l of cDNA solution, 10 μ l Supermix, 1 μ l of Taqman primer/probe mix was loaded into a cartridge with droplet Generation oil (Bio-Rad, Hercules, CA) and then placed into the QX100 Droplet Generator (Bio-Rad). The generated droplets were transferred to a 96-well PCR plate. PCR amplification was carried on a T100 thermal cycler (Bio-Rad). ddPCR was a direct method for quantitatively measuring nucleic acids³³. The number of positive reactions, together with Poisson’s distribution, were used to produce a straight and high-confidence measurement of the original target concentration⁴⁴. Therefore, ddPCR could absolutely quantify targeted nucleic acid sequences without requiring external calibrators or endogenous controls (genes)^{26, 44, 45}. The plate was loaded on Droplet Reader (Bio-Rad), by which copy number of each miRNA per μ l PCR reaction mixture was directly determined. All assays were performed in triplicates. Furthermore, two interplate controls and one no-template control were carried along in each experiment. The no template control for RT was RNease free water instead of RNA sample input, and no template control for PCR was RNease free water instead of RT products input.

Statistical analysis

To identify plasma miRNAs that were differentially expressed in patients with malignant versus benign PNs, we expected the acceptable number of false positives to be 1.0, fold difference between cases and controls at 2.0, standard deviation of the gene measurements on the base-two logarithmic scale at 0.7, and desired power at 80%. Given 375 miRNAs included in the array, at least 15 specimens for each type of the patients were required to achieve the statistical criteria. Furthermore, based on one-sample with binomially distributed outcomes, we needed 45 cases from each group at 5% significant level with 80% power to discover and validate a panel of biomarkers or classifier for predicting malignant PNs. For analysis of microarray data, we performed an unpaired unequal variance t test with

Benjamini-Hochberg correction⁴⁶ to identify differentially expressed miRNAs in plasma of patients with malignant versus benign PNs. We used univariate analysis to determine which of plasma miRNAs and clinical and radiological variables were associated with malignant PNs. The significantly associated factors were then analyzed by using multivariate logistic regression models with constrained parameters as in least absolute shrinkage and selection operator (LASSO) based on receiver-operator characteristic (ROC) curve to identify an optimal panel of miRNA biomarkers and construct a classifier for malignant PNs. The optimal cutoff value was generated using the Youden index⁴⁷. The 95% confidence intervals in the ROC plot for proportions were estimated. To compare the performance of the classifier with that of the plasma biomarkers and the Mayo Clinic model¹⁴, we used the method of Hanley and McNeil with the area under an ROC curve (AUC) analysis⁴⁸. The classifier was blindly validated in two additional sets of patients by comparing the calculated results with final clinical diagnosis and the AUCs.

Results

Identifying differentially-expressed miRNAs in plasma of patients with malignant versus benign PNs

Of the 375 miRNAs embodied on the miRNA array, 282 (75.2%) showed a <35 Cq value in all plasma specimens of 36 patients with either malignant (18) or benign (18) PNs. Furthermore, the miRNA expression levels measured by using the microarray in the replicates of each sample were highly correlated (all $p < 0.0001$). Therefore, the 282 miRNAs were reliably measurable in plasma of the patients with PNs. Among the miRNAs, 11 (miRs-21-5p, -103a-3p, -126-3p, -135a-5p, -145-5p, -141-3p, -193b-3p, -200b-3p, -205-5p, -210, and -301b) exhibited more than 2.0 fold-changes with a $p < 0.05$ in plasma of patients with malignant versus benign PNs (Supplementary table 2). Of the 11 miRNAs, nine (miRs-21-5p, -103a-3p, -126, 141-3p, -193b-3p, -205-5p, 210, and -301b) had a higher expression level, whereas three (miRs-135a-5p, 145-5p, and -200b-3p) displayed a lower level in plasma of patients with malignant versus benign PNs (All $P < 0.05$).

Validating the changes of the plasma miRNAs, and developing miRNA biomarkers for malignant PNs

The changes of 11 malignant PN-related plasma miRNAs identified by the microarray should be validated using a different and reliable technique¹³. We demonstrated that ddPCR was a more sensitive technique with greater precision and reproducibility to detect expression of miRNAs in plasma than did the conventional reverse transcription PCR (RT-PCR)^{26, 33}. Moreover, ddPCR could absolutely and precisely determine copy number of miRNAs without the need of an internal control gene, such as U6. Therefore, we used ddPCR to assess changes of the 11 miRNAs in 135 plasma samples of the UMMC cohort. Each well of the plasma samples contained at least 10,000 droplets. By contrast, no product was synthesized in the negative control samples. Thus, the plasma samples were successfully “read” for the absolute quantification of the 11 miRNAs by using a reliable and accurate assay. ddPCR analysis showed that all the 11 miRNAs displayed a significantly different level in plasma samples of patients with malignant PNs versus individuals with benign diseases (all $P < 0.05$). Furthermore, the miRNAs had consistent changes detected by

ddPCR in the same direction as in the microarray analysis: nine displayed a higher expression level, whereas three exhibited a lower level in plasma of subjects with malignant versus benign PNs (Table 3).

To determine the diagnostic values of the plasma miRNAs, ROC and the AUCs were calculated by using the copy number of each miRNA in the UMMC cohort of 135 patients. The individual miRNAs exhibited AUC values of 0.57–0.82 in distinguishing malignant from benign PNs (Table 3). We used logistic regression models with constrained parameters as in LASSO based on ROC criterion to identify and optimize a panel of miRNA biomarkers. The three miRNAs, miRs-126, 210, and 205-5p are selected as the best biomarkers (all $P < 0.001$). Combined use of the three miRNAs produced 0.87 AUC for distinguishing malignant from benign PNs. Furthermore, Pearson correlation analysis indicated that the estimated correlation among expression levels of the three miRNAs in plasma was low (All $P > 0.05$), implying that the diagnostic values of the miRNAs were complementary to each other. Subsequently, combined use of the three miRNAs generated a sensitivity of 81.2% and a specificity of 86.4% (Figure 1A). Moreover, including other miRNAs in the model did not improve the accuracy for the identification of malignant PNs. The logistic model had no statistically significant association with histological type and stage of the NSCLC, and patients' age, gender, ethnicity, and smoking history (All $p > 0.05$).

Developing a classifier by integrating the biomarkers and radiographic features of PNs for identifying malignant PNs

Although showing promise, the 81.2% sensitivity and 86.4% specificity of the three miRNA biomarkers used together are not sufficient in the clinic for distinguishing malignant from benign PNs. We used univariate analysis to determine which of clinical and radiological variables were associated with malignant PNs in 135 patients of the UMMC cohort. History of cancer and smoking pack-years of the patients, the diameter, spiculation, and upper lobe location of the PNs were associated with malignant PNs (Table 4). We then used logistic regression models with constrained parameters as in LASSO based on ROC criterion to eliminate the large number of the parameters to construct a classifier including: miRs-205-5p and 126, and diameter of PN. The classifier had 0.95 AUC in distinguishing malignant from benign PNs (Figure 1B). Adding other miRNAs and imaging/clinical variables in the classifier did not improve the performance for predicting malignant PNs. Furthermore, prediction models based on parameters of PN on CT images and clinical characteristics of smokers have been developed for predicting the probability of malignant PNs^{14–17}, of which, the Mayo Clinic model is a commonly used one. We applied the Mayo Clinic model¹⁴ in the same set of 135 patients for predicting lung cancer by using the equation. The Mayo Clinic model produced an AUC of 0.82 (Figure 1C), a similar value as shown in the previous reports^{14–16}. The direct comparison showed that the classifier had a significantly higher AUC value (0.95) compared with the Mayo Clinic model (0.82) and the panel of three biomarkers (0.87) used alone (All $P < 0.05$) in the same set of patients (Figure 1A–C). As a result, our classifier yielded 89.9% sensitivity and 90.9% specificity for diagnosis of malignant PNs, which were also significantly higher compared with those of the Mayo Clinic model (75.4% sensitivity and 80.3% specificity) and the biomarkers (81.2% sensitivity and 86.4% specificity) (all $P < 0.05$) (Figure 1A–C). Moreover, the classifier did

not exhibit statistical differences of sensitivity and specificity between histological types and stages of NSCLC ($P>0.05$).

Validating the classifier for differentiating malignant from benign PNs in two independent cohorts of patients with PNs

The three miRNAs (miRs-126, 210, and 205-5p) were analyzed for the expression levels in plasma of both BVAMC and JPHTCM cohorts. The three miRNAs showed a similar change pattern in the two independent cohorts as in the UMMC cohort, providing additional evidence that the plasma miRNAs could be reproducibly measured. The classifier produced 0.94 AUC with a sensitivity of 88.9% and a specificity of 90.5% for diagnosis of malignant PNs in BVAMC cohort (Table 5). The classifier produced higher sensitivity and specificity than did the panel of the biomarkers (80.9% sensitivity and 85.7% specificity) and the Mayo Clinic model (74.6% sensitivity and 79.4% specificity) (All $P<0.05$) (Table 5). The classifier was further validated in an external set of 98 patients with PNs (the JPHTCM cohort) recruited in China for the diagnostic value in a blinded fashion. The classifier created an AUC of 0.94 with a sensitivity of 87.8% and a specificity of 89.8% for detection of malignant PNs. Furthermore, the classifier had higher sensitivity and specificity (87.8% sensitivity and 89.8% specificity) for detection of malignant PNs than did the panel of the biomarkers (81.6% sensitivity and 85.7% specificity) and the Mayo Clinic model (73.5% sensitivity and 75.5% specificity) (Table 5) (All $P<0.05$). Taken together, the results created from the extensive validations confirmed the potential of the classifier for estimating the probability of lung cancer among indeterminate PNs.

Discussion

Numerous plasma miRNA biomarkers have been searched by detecting circulating miRNAs directly released from primary tumors or the circulating lung cancer cells, but have limited success, due to several challenges⁹: 1), the release of contaminating miRNAs in plasma by hemolysis of blood cells always produces a low specificity with inconsistent results for cancer diagnosis. 2), since the amount of miRNAs directly derived from primary tumors in plasma is very low and further 'diluted' in a background of normal miRNAs^{9, 10}, some cancer cell derived-miRNAs presenting at very low abundance in plasma are undetectable by RT-PCR, producing a very poor sensitivity for cancer detection⁴⁴. 3), no standard endogenous control exists in plasma for normalizing circulating miRNAs, resulting in poor reproducibility and robustness among different studies. To address the challenges, first, in this study we use the EDRN-established SOPs for collecting and preparing blood specimens to reduce bias related to sampling methods, storage or purification, and to diminish the contamination of the blood cells-related miRNAs in plasma. Indeed, expression levels of blood cells-related miRNAs in our samples are negative, suggesting that there is no contaminated miRNAs from hemolysis of the blood cells. Second, we have demonstrated that ddPCR could directly and reliably quantify low abundance miRNAs in clinical samples with a higher sensitivity than conventional RT-PCR^{26, 33}. Furthermore, ddPCR can absolutely quantify copy number of miRNAs. In addition, ddPCR does not require external calibrators or endogenous control genes. Moreover, it is relatively resistant to PCR inhibitors. In this study we use ddPCR to analyze miRNAs in plasma. Our results show that

miR-205-5p, a key lung tumor-specific miRNA presenting at a very low level unreliably detectable by RT-PCR in plasma¹¹, is robustly and reproducibly quantified by ddPCR. Since ddPCR could reliably and sensitively measure the vital miRNAs that were not previously detectable in plasma using RT-PCR, our newly developed three miRNA biomarkers are not the same as the previous ones developed by using RT-PCR^{8, 11, 13, 27}. However, although using only three miRNAs, the logistic model has a higher specificity compared with a circulating miRNA signature composed by reciprocal ratios among 24 miRNAs (87% vs. 81%) for identifying lung cancer⁴⁹.

Previously, some models based on the clinical/radiological variables have shown the potential for predicting malignant PNs^{14, 16, 17}. Our present study confirms the previous findings that the clinical and radiological variables could be predictors for malignant PNs. However, the moderate sensitivity and specificity of these models limits the application in the clinical settings^{14, 16}. We reason that integrating plasma biomarkers with radiological/clinical characteristics might have a synergistic effect for identifying NSCLC among the indeterminate PNs. Indeed, by integrating the plasma biomarkers with radiological characteristics, we develop a classifier for distinguishing malignant from benign PNs. Although this simple classifier comprises only two biomarkers and one radiological variable of PNs, it has higher sensitivity and specificity compared with the panel of biomarkers or the Mayo Clinic model used alone. Furthermore, the performance of the classifier developed in White Americans and African Americans is confirmed in a geographically independent cohort (Chinese population), further implying the usefulness for detection of NSCLC among indeterminate PNs. In addition, the classifier with a simple equation and a single cut-off value would offer a convenient analytic means in the laboratory settings for the classification of malignant from benign PNs.

The areas for improvement do exist in the current study. A screening assay directed at a malignancy with an incidence 5% should have a sensitivity exceeding 95% when the specificity is 95%, and vice versa. The incidence of lung cancer in heavy smokers is less than 3%, while LDCT has 90% sensitivity and only 61% specificity, producing a high false positive rate. Only the approaches with more than 95% specificity and appropriate sensitivity for identifying malignant PNs could supplement LDCT lung cancer screening. Although showing promise, our developed classifier with 89% sensitivity and 90% specificity does not hold the required performance. The two plasma miRNA biomarkers in the classifier were developed from the limited number of miRNAs identified by microarrays, however, by which other important lung cancer-associated miRNAs might not been included. We are using high-throughput next-generation sequencing to directly analyze plasma samples of patients with either malignant or benign PNs to identify new miRNA biomarkers for lung cancer. The performance of this classifier could be further improved by adding the new plasma miRNA biomarkers that are more specific to malignant PNs. Furthermore, the radiological features used in the study are obtained through the conventional image analysis that is based on subjective observation and limited to the measurements of nodule size in one dimension. Radiomics, an emerging technique extracting a large number of quantitative features from medical images automatically, provides a more detailed quantification of tumor phenotypic characteristics that have diagnostic value. In the future, we intend to incorporate the molecular biomarkers, radiomic

features of nodules, and clinical characteristics of smokers to develop a classifier that could more accurately and conveniently identify lung cancer among the indeterminate PNs.

In summary, we have for the first time developed a simple classifier by integrating plasma biomarkers with radiological characteristics that could identify lung cancer among indeterminate PNs. Future use of the classifier by sparing individuals with benign growths from the harmful diagnostic procedures, while allowing effective treatments to be immediately initiated for NSCLC, would complement LDCT for the early detection of lung cancer. Nevertheless, undertaking a prospective study to further validate the classifier for lung cancer in a large population-based LDCT screening positive setting among heavy smokers is required.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant support: This work was supported in part by NCI R21CA205746, VA Merit Award I01 CX000512, Award from the Geaton and JoAnn DeCesaris Family Foundation, and DoD-Idea Development Award (F.J.), and NCI R01CA172638 and NCI P30 CA008748 (W.L.).

References

1. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011; 365:395–409. [PubMed: 21714641]
2. Moyer VA. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014; 160:330–8. [PubMed: 24378917]
3. 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. [PubMed: 24322569]
4. Diederich S, Das M. Solitary pulmonary nodule: detection and management. *Cancer Imaging*. 2006; 6:S42–6. [PubMed: 17114077]
5. Aberle DR, Berg CD, Black WC, et al. The National Lung Screening Trial: overview and study design. *Radiology*. 2011; 258:243–53. [PubMed: 21045183]
6. Costa FF. Non-coding RNAs: new players in eukaryotic biology. *Gene*. 2005; 357:83–94. [PubMed: 16111837]
7. Yanaihara N, Caplen N, Bowman E, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*. 2006; 9:189–98. [PubMed: 16530703]
8. Shen J, Jiang F. Applications of MicroRNAs in the Diagnosis and Prognosis of Lung Cancer. *Expert Opin Med Diagn*. 2012; 6:197–207. [PubMed: 22615714]
9. Mitchell PS, Parkin RK, Kroh EM, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A*. 2008; 105:10513–8. [PubMed: 18663219]
10. Chen X, Ba Y, Ma L, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res*. 2008; 18:997–1006. [PubMed: 18766170]
11. Shen J, Todd NW, Zhang H, et al. Plasma microRNAs as potential biomarkers for non-small-cell lung cancer. *Lab Invest*. 2011; 91:579–87. [PubMed: 21116241]
12. Shen J, Liao J, Guarnera MA, et al. Analysis of MicroRNAs in sputum to improve computed tomography for lung cancer diagnosis. *J Thorac Oncol*. 2014; 9:33–40. [PubMed: 24305007]
13. Shen J, Stass SA, Jiang F. MicroRNAs as potential biomarkers in human solid tumors. *Cancer Lett*. 2013; 329:125–36. [PubMed: 23196059]

14. Swensen SJ, Silverstein MD, Ilstrup DM, et al. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med*. 1997; 157:849–55. [PubMed: 9129544]
15. 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; 131:383–8. [PubMed: 17296637]
16. 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; 63:335–41. [PubMed: 17965070]
17. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med*. 2013; 369:910–9. [PubMed: 24004118]
18. Li Y, Chen KZ, Wang J. Development and validation of a clinical prediction model to estimate the probability of malignancy in solitary pulmonary nodules in Chinese people. *Clin Lung Cancer*. 2011; 12:313–9. [PubMed: 21889113]
19. Lin H, Zhong WZ, Yang XN, et al. A clinical model to estimate the pretest probability of lung cancer, based on 1198 pedigrees in China. *J Thorac Oncol*. 2012; 7:1534–40. [PubMed: 22982654]
20. 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; 237:395–400. [PubMed: 16244247]
21. Moyer VA. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2012; 157:120–34. [PubMed: 22801674]
22. Ohori M, Wheeler TM, Scardino PT. The New American Joint Committee on Cancer and International Union Against Cancer TNM classification of prostate cancer Clinicopathologic correlations. *Cancer*. 1994; 74:104–14. [PubMed: 7516262]
23. 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; 10:1243–60. [PubMed: 26291008]
24. Marks JR, Anderson KS, Engstrom P, et al. Construction and analysis of the NCI-EDRN breast cancer reference set for circulating markers of disease. *Cancer Epidemiol Biomarkers Prev*. 2015; 24:435–41. [PubMed: 25471344]
25. Tuck MK, Chan DW, Chia D, et al. Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. *J Proteome Res*. 2009; 8:113–7. [PubMed: 19072545]
26. Ma J, Li N, Guarnera M, et al. Quantification of Plasma miRNAs by Digital PCR for Cancer Diagnosis. *Biomark Insights*. 2013; 8:127–36. [PubMed: 24277982]
27. Shen J, Liu Z, Todd NW, et al. Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microRNA biomarkers. *BMC Cancer*. 2011; 11:374. [PubMed: 21864403]
28. Su Y, Fang H, Jiang F. Integrating DNA methylation and microRNA biomarkers in sputum for lung cancer detection. *Clin Epigenetics*. 2016; 8:109. [PubMed: 27777637]
29. Su Y, Guarnera MA, Fang H, et al. Small non-coding RNA biomarkers in sputum for lung cancer diagnosis. *Mol Cancer*. 2016; 15:36. [PubMed: 27176474]
30. Su J, Anjuman N, Guarnera MA, et al. Analysis of Lung Flute-collected Sputum for Lung Cancer Diagnosis. *Biomark Insights*. 2015; 10:55–61.
31. Su J, Liao J, Gao L, et al. Analysis of small nucleolar RNAs in sputum for lung cancer diagnosis. *Oncotarget*. 2016; 7:5131–42. [PubMed: 26246471]
32. Xing L, Su J, Guarnera MA, et al. Sputum microRNA biomarkers for identifying lung cancer in indeterminate solitary pulmonary nodules. *Clin Cancer Res*. 2015; 21:484–9. [PubMed: 25593345]
33. Li N, Ma J, Guarnera MA, et al. Digital PCR quantification of miRNAs in sputum for diagnosis of lung cancer. *J Cancer Res Clin Oncol*. 2014; 140:145–50. [PubMed: 24281335]
34. Anjuman N, Li N, Guarnera M, et al. Evaluation of lung flute in sputum samples for molecular analysis of lung cancer. *Clin Transl Med*. 2013; 2:15. [PubMed: 24053570]
35. Yu L, Todd NW, Xing L, et al. Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. *Int J Cancer*. 2010; 127:2870–8. [PubMed: 21351266]

36. Xing L, Todd NW, Yu L, et al. Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers. *Mod Pathol*. 2010; 23:1157–64. [PubMed: 20526284]
37. Xie Y, Todd NW, Liu Z, et al. Altered miRNA expression in sputum for diagnosis of non-small cell lung cancer. *Lung Cancer*. 2010; 67:170–6. [PubMed: 19446359]
38. Cao X, Wu Z, Jiang F, et al. Identification of chilling stress-responsive tomato microRNAs and their target genes by high-throughput sequencing and degradome analysis. *BMC Genomics*. 2014; 15:1130. [PubMed: 25519760]
39. Li P, Zhang Q, Wu X, et al. Circulating microRNAs serve as novel biological markers for intracranial aneurysms. *J Am Heart Assoc*. 2014; 3:e000972. [PubMed: 25249297]
40. Huang Y, Yang S, Zhang J, et al. MicroRNAs as promising biomarkers for diagnosing human cancer. *Cancer Invest*. 2010; 28:670–1. [PubMed: 20210518]
41. Ma J, Li N, Lin Y, et al. Circulating Neutrophil MicroRNAs as Biomarkers for the Detection of Lung Cancer. *Biomark Cancer*. 2016; 8:1–7.
42. Ma J, Lin Y, Zhan M, et al. Differential miRNA expressions in peripheral blood mononuclear cells for diagnosis of lung cancer. *Lab Invest*. 2015; 95:1197–206. [PubMed: 26146958]
43. Ma J, Mannoor K, Gao L, et al. Characterization of microRNA transcriptome in lung cancer by next-generation deep sequencing. *Mol Oncol*. 2014; 8:1208–19. [PubMed: 24785186]
44. Whale AS, Huggett JF, Cowen S, et al. Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. *Nucleic Acids Res*. 2012; 40:e82. [PubMed: 22373922]
45. Li H, Jiang Z, Leng Q, et al. A prediction model for distinguishing lung squamous cell carcinoma from adenocarcinoma. *Oncotarget*. 2017; 11:226–32.
46. Liu H, Zhu L, Liu B, et al. Genome-wide microRNA profiles identify miR-378 as a serum biomarker for early detection of gastric cancer. *Cancer Lett*. 2012; 316:196–203. [PubMed: 22169097]
47. Schisterman EF, Perkins NJ, Liu A, et al. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*. 2005; 16:73–81. [PubMed: 15613948]
48. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983; 148:839–43. [PubMed: 6878708]
49. Sozzi G, Conte D, Leon M, et al. Quantification of free circulating DNA as a diagnostic marker in lung cancer. *J Clin Oncol*. 2003; 21:3902–8. [PubMed: 14507943]

Novelty & Impact Statements

Early detection of lung cancer by low-dose CT (LDCT) can reduce the mortality. However, LDCT dramatically increases the number of indeterminate pulmonary nodules (PNs), leading to overdiagnosis. Having a definitive preoperative diagnosis of malignant PNs is clinically important. Using multivariate logistic regression analysis of the molecular results and clinical/radiological characteristics, we have for the first time developed a simple classifier comprising two miRNA biomarkers and one radiological characteristic for accurately distinguishing malignant from benign PNs.

Figure 1.

Receiver-operator characteristic (ROC) curve analysis of the classifier, panel of the three plasma miRNA biomarkers, and Mayo Clinic model for distinguishing malignant from benign PNs in a training set of patients. The area under the ROC curve (AUC) for each approach conveys its accuracy for diagnosis of malignant PNs. The classifier produces a higher AUC value (B) for identifying malignant PNs, compared with the panel of the three plasma miRNA biomarkers (A) and Mayo Clinic model (C) (All $P < 0.05$).

Table 1

Characteristics of patients recruited in the University of Maryland Medical Center

	Patients with malignant PNs (n = 69)	Patients with benign PNs (n = 66)
Clinical		
Age	68.22 (SD 9.90)	65.27 (SD 8.26)
Sex		
Male	46	45
Female	23	21
Race		
African American	21	20
White	48	46
Smoking history		
Current smoker	41	40
Former smoker	28	26
Pack-years	43.26 (SD 13.12)	23.69 (SD 12.28)
Years quit	7.16 (SD 4.69)	12.69 (SD 8.27)
History of cancer	7	2
Stage of non-small cell cancer		
Stage I	18	
Stage II	18	
Stage III–VI	23	
Histological type		
AC	33	
SCC	29	
LC	7	
Radiological		
Nodule size (mm)	19.89 (SD 12.16)	10.18 (SD 5.55)
Nodule Location		
Left lower lobe	9	13
Left upper lobe	25	18
Right lower lobe	15	20
Right middle lobe	4	7
Right upper lobe	15	8
Nodule type (number)		
Nonsolid or ground-glass opacity	18	20
Perifissural	7	9
Part-solid	9	11
Solid	13	14
Spiculation	22	12

Abbreviations: PN, pulmonary nodule; SD, standard deviation; AC, adenocarcinoma; SCC, squamous cell carcinoma; LCC, large cell carcinoma.

Table 2

Characteristics of patients recruited in the Baltimore VA Medical Center.

Characteristics	Patients with malignant PNs (n = 63)	Patients with benign PNs (n = 63)
Clinical		
Age	67.38 (SD 9.16)	64.48 (SD 9.01)
Sex		
Male	43	42
Female	20	21
Race		
African American	19	20
White	44	43
Smoking history		
Current smoker	40	39
Former smoker	23	24
Pack-years	44.67 (SD 12.19)	25.78 (SD 13.19)
Years quit	6.78 (SD 8.38)	11.15 (SD 7.79)
History of cancer	4	1
Stage of non-small cell cancer		
Stage I	20	
Stage II	19	
Stage III–VI	24	
Histological type		
AC	31	
SCC	28	
LC	4	
Radiological		
Nodule size (mm)	18.34 (SD 13.02)	10.39 (SD 6.02)
Nodule Location		
Left lower lobe	8	11
Left upper lobe	21	17
Right lower lobe	15	20
Right middle lobe	5	7
Right upper lobe	14	8
Nodule type (number)		
Nonsolid or ground-glass opacity	17	19
Perifissural	7	9
Part-solid	7	10
Solid	12	13
Spiculation	20	12
		63

Abbreviations: PN, pulmonary nodule; SD, standard deviation; AC, adenocarcinoma; SCC, squamous cell carcinoma; LCC, large cell carcinoma.

Table 3

Expression levels of 11 plasma miRNAs and their diagnostic significance in 135 patients with malignant versus benign PNs

miRNAs	Mean (SD) in patients with benign PNs	Mean (SD) in patients with malignant PNs	P-value	AUC (95% confidence interval)
miR-21-5p	8.9893 (5.7922)	18.9694 (23.4001)	0.0458	0.6135 (0.5102 to 0.7168)
miR-103a-3p	1.5987 (1.9266)	7.0492 (14.2826)	0.0386	0.6160 (0.5155 to 0.7166)
miR-126-3p	4.6113 (4.5983)	20.1927 (40.4664)	0.0301	0.6075 (0.4977 to 0.7173)
miR-135a-5p	0.0237 (0.0305)	0.0123 (0.0180)	0.0068	0.5696 (0.4784 to 0.6608)
miR-145-5p	0.9343 (1.4553)	0.7990 (1.4933)	0.0004	0.6761 (0.5868 to 0.7655)
miR-141-3p	0.0239 (0.0663)	0.0481 (0.1143)	0.0480	0.5962 (0.5064 to 0.6861)
miR-193b-3p	0.0523 (0.0525)	0.0772 (0.0810)	0.0203	0.6318 (0.5468 to 0.7168)
miR-200b-3p	0.0319 (0.0453)	0.0151 (0.0190)	0.0010	0.6199 (0.5330 to 0.7069)
miR-205-5p	0.0114 (0.0191)	0.0672 (0.1054)	<0.001	0.8187 (0.7503 to 0.8871)
miR-210	0.3579 (0.6268)	0.8951 (1.2892)	0.0054	0.6525 (0.5527 to 0.7523)
miR-301b	0.2491 (0.2926)	0.9524 (1.8623)	0.0130	0.6115 (0.5201 to 0.7029)

Abbreviations: PN, pulmonary nodules; SD, standard deviation; AUC, the area under receiver operating characteristic curve.

Table 4

Association of clinical and radiological variables with malignant PNs

Variables	p-value
Age (year)	0.1729
Sex	0.7388
Race	0.9648
Current smoker	0.2295
Former smoker	0.6548
Smoking package-year	<0.0001
Cancer history	<0.0001
Nodule diameter on CT	<0.0001
Nodule spiculation on CT	<0.0001
Year after quit	0.0126
Upper lobe locations of PNs	0.0065

Abbreviations: PN, pulmonary nodule.

Table 5

Comparison of the classifier, panel of the three plasma miRNA biomarkers, and Mayo Clinic model for distinguishing malignant from benign PNs in two independent cohorts of patients *

Approaches	BVAMC patients		JPHTCM	
	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
The classifier	88.89% (78.44% to 95.41%)	90.48 % (80.41% to 96.42%)	87.76% (75.23% to 95.37%)	89.80 % (77.77% to 96.60%)
The biomarker panel	80.95% (69.09% to 89.75%)	85.71 % (74.61% to 93.25%)	81.63% (67.98% to 91.24%)	85.71 % (72.76% to 94.06%)
The Mayo Clinic model	74.60% (62.06% to 84.73%)	79.37% (67.30% to 88.53%)	73.47% (58.92% to 85.05%)	75.51% (61.13% to 86.66%)

Abbreviations: BVAMC, Baltimore VA Medical Center patients; JPHTCM, Jiangsu Province Hospital of Traditional Chinese Medicine; CI, confidence interval.

* All P values <0.05.