

Award Number: W81XWH-17-1-0254 (LC160422)

TITLE: Sputum Biomarkers to Improve CT Screening for the Early Detection of Lung Cancer in Veterans

PRINCIPAL INVESTIGATOR: Feng Jiang

CONTRACTING ORGANIZATION: University of Maryland, Baltimore
Department of Pathology (Research)
Medical School Teaching Facility, 734C
10 South Pine Street
Baltimore, MD 21201-1116

REPORT DATE: JULY 2019

TYPE OF REPORT: Annual report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE JULY 2019		2. REPORT TYPE Annual		3. DATES COVERED 1JUL2018 - 30JUN2019	
4. TITLE AND SUBTITLE Sputum Biomarkers to Improve CT Screening for the Early Detection of Lung Cancer in Veterans				5a. CONTRACT NUMBER W81XWH-17-1-0254	
				5b. GRANT NUMBER GRANT12248530	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) FENG JIANG E-Mail: fjiang@som.umaryland.edu				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Maryland, Baltimore Department of Pathology (Research) Medical School Teaching Facility, 734C 10 South Pine Street Baltimore, MD 21201-1116				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES no					
14. ABSTRACT We propose to develop a model for identifying lung cancer in indeterminate pulmonary nodules (PNs). Specific aims: (1) identify new sputum small noncoding RNA (ncRNA) biomarkers for lung cancer, (2) validate the previously and newly identified biomarkers and develop a prediction model, and (3) blindly validate the prediction model for distinguishing benign from malignant growths in a different cohort. Until 09/2019, we produced 7 peer-reviewed research articles published. We have the important findings as shown below: Demonstrating that aberrant ncRNAs detected in the bronchial epitheliums of sputum could reflects those in malignant PNs (lung tumors). Developing and validating panels of sputum miRNA, snoRNA, and lncRNA biomarkers for lung cancer. Defining a landscape of various types of ncRNAs of bronchial epitheliums of lung cancer patients. Combined analysis of ncRNAs and DNA methylation has a synergistic effect for lung cancer diagnosis. We have developed a prediction model that has more than 90% sensitivity and 95% specificity for lung cancer detection					
15. SUBJECT TERMS Lung cancer, biomarkers, early detection					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
	Unclassified	Unclassified	Unclassified	81	19b. TELEPHONE NUMBER (include area code)

Table of contents,

Front cover	Page 1
Standard Form (SF 298)	Page 2
Table of Contents	Page 3
Introduction	Page 3
Keywords	Page 3
Accomplishments	Pages 4-9
Impact	Page 10
Changes/Problems	Page 11
Products	Page 11
Participants and Other Collaborating Organizations	Page 11
Special Reporting Requirements	Page 11
Appendices	Page 11 and the attachment

Introduction,

To complete the objective, we have three specific aims:

Aim 1. Identify new sputum small ncRNA biomarkers for lung cancer from our previously NGS-defined small ncRNA profile of lung tumors.

Aim 2. Validate the previously and newly identified biomarkers and develop a prediction model by integrating the biomarkers with clinical and imaging variables for identifying NSCLC.

Aim 3. Blindly validate the prediction model for distinguishing benign from malignant growths in a different cohort of patients with LDCT-discovered PNs.

We have achieved the goal of the first two specific aims and will complete the sim 3 before the end of this project. The importance of the findings is described below:

- Demonstrating that aberrant ncRNAs detected in the bronchial epitheliums of sputum could reflects those in malignant PNs (lung tumors)
- Developing and validating panels of sputum miRNA and lncRNA biomarkers for lung cancer by using ddPCR
- Defining a landscape of various types of ncRNAs of bronchial epitheliums of lung cancer patients, providing novel biomarker candidates for malignant PNs
- Demonstrating that ddPCR can sensitively and robustly quantify DNA methylation in clinical specimens
- Combined analysis of ncRNAs and DNA methylation has a synergistic effect for lung cancer diagnosis
- We produced seven peer-reviewed research articles published.
- We have presented these results at national meetings.

Keywords,

Lung cancer, Veterans, small non-coding RNA, biomarkers, diagnosis.

Accomplishments

1. The major goals of the project

Aim #1 is to identify new sputum small ncRNA biomarkers for lung cancer from our previously NGS-defined small ncRNA profile of lung tumors.

1), we have local IRB protocol approved to collected sputum and blood samples and associated clinical information.

2), we have analyzed the changes of the NGS-defined lung tumor tissue-based ncRNAs (68 miRNAs and 29 snoRNAs) in sputum of patients with either malignant or benign PNs using ddPCR. We developed panels of sputum biomarker for NSCLC.

Particularly, we determined expressions of the lncRNAs in sputum of a discovery cohort of 67 lung cancer patients and 65 smokers with benign diseases. Four of the lncRNAs displayed a different level in sputum of the NSCLC patients vs. cancer-free smokers (all $P < 0.001$). From the lncRNAs, 3 lncRNAs (SNHG1, H19, and HOTAIR) are identified as a sputum biomarker panel with an AUC of 0.9 (Figure 1A), producing 82.1% sensitivity and 89.2% specificity for diagnosis of lung cancer. Furthermore, the biomarker panel has a higher sensitivity (82.1% vs. 52.2%, $P=0.02$) compared with sputum cytology (Figure 1B).

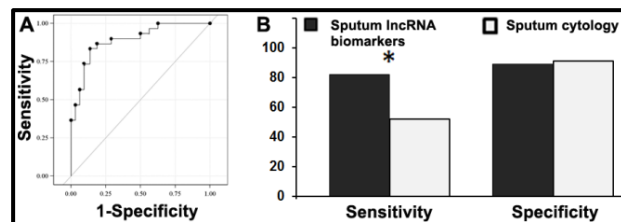


Fig. 3. Three lncRNA were selected as the best sputum biomarkers and incorporated into an algorithm: $P = \exp(U) / [1 + \exp(U)]$, where $U = 2.93 + 2.12 * \log(\text{SNHG1}) - 1.26 * \log(\text{H19}) - 1.1795 * \log(\text{HOTAIR})$. The 3 sputum lncRNAs used in combination by this algorithm produced 0.9 AUC (A). The sputum lncRNA biomarker panel has a higher sensitivity compared with sputum cytology for lung cancer detection (B).

Part of the study was recently published (Sputum long non-coding RNA biomarkers for diagnosis of lung cancer. C Gupta, J Su, M Zhan, SA Stass, F Jiang. 2019. Cancer Biomarkers, 1-9).

3), we have for the first time shown that ddPCR can sensitively and robustly quantify DNA methylation, and further develop a sputum epigenetic classifier for early stage lung cancer.

Conventional qPCR-based platforms, such as methylation-specific PCR (qMSP), have been used for detecting DNA methylation. However, qMSP is an indirect approach, which requires internal controls for data normalization. Furthermore, the sensitivity is poor. We have demonstrated that ddPCR is a straightforward means for precise, direct, and absolute quantification of nucleic acids (e.g., miRNAs, snoRNAs, and lncRNAs), without requiring an internal control gene.

To investigate whether ddPCR could sensitively and robustly quantify DNA methylation, we performed methylation-specific ddPCR (ddMSP) and qMSP in methylated and

unmethylated DNA samples and compared their performance for quantification of DNA methylation. ddMSP had a higher sensitivity for the quantification of DNA methylation compared with qMSP ($p = 0.01$) (Figure 2).

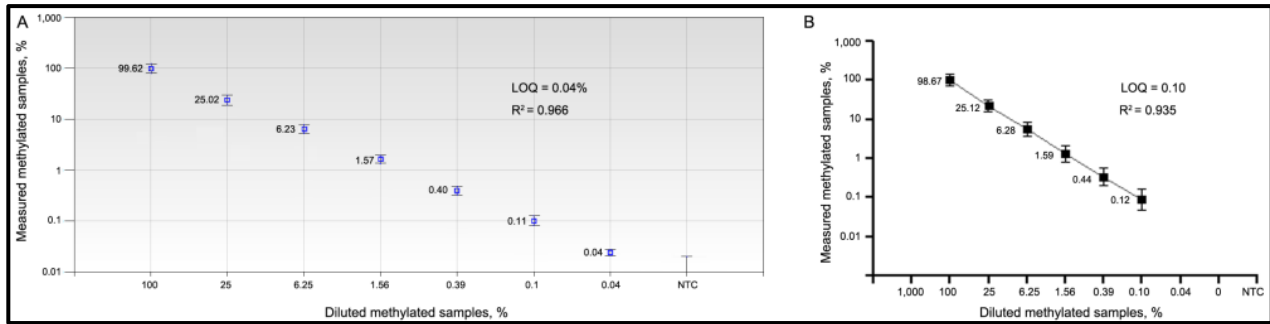


Fig. 2. Comparing the dynamic ranges and sensitivities of ddMSP and qMSP for quantification of DNA methylation. (A), In methylated DNA serially diluted in sputum DNA of a healthy nonsmoker, ddMSP can detect levels of methylated 3OST2 as low as 0.04% (Limit of detection (LOD) = 0.04%). (B), qMSP can detect methylated 3OST2 at 0.10% (LOD = 0.10%) in the same diluted samples ($p = 0.01$).

Furthermore, ddMSP had higher precision and reproducibility than did qMSP (all $p < 0.05$).

To evaluate analytic performance of ddMSP and qMSP in clinical sputum samples, sputum of 20 lung cancer patients and 20 cancer-free controls was tested for *RASSF1A*, whose aberrant methylation level was shown to be elevated in sputum of lung cancer patients. Each well of the samples contained at least 10,000 droplets (Fig. 3A). Therefore, ddMSP analysis of DNA methylation could successfully be performed in clinical sputum specimens. *RASSF1A* analyzed by both the techniques displayed a high methylation level in lung cancer patients vs. controls (all $p < 0.05$). In the ddMSP assay, a specimen with $\geq$ one copy of DNA methylation of *RASSF1A* per microliter was considered to be positive. When the criteria was used, of 20 sputum specimens of lung cancer patients, 11 (55%) had positive methylation of the gene detected by ddMSP. In the qMSP assay, a PMR $\geq 1\%$ was classified as positive for *RASSF1A* in a given sample. When the criteria was used, 9 (45%) were positive for *RASSF1A* by qMSP. The same 4 sputum specimens of control subjects had positive methylation of the gene detected by both ddMSP and qMSP. Therefore, ddMSP analysis of DNA methylation of *RASSF1A* in sputum had a higher sensitivity (55%) than did qMSP (45%) ($p = 0.01$) for distinguishing lung cancer patients from control subjects, while maintaining the same specificity (80%). Furthermore, the CVs of repeated measures by ddMSP on different days by different researchers were approximately twofold lower compared with those generated by qMSP. Altogether, in clinical sputum specimens, ddMSP also exhibited higher sensitivity, accuracy, and reproducibility than did qMSP for quantification of DNA methylation.

We then used ddMSP to analyze 29 potential DNA methylation biomarkers that were previously proposed as potential sputum biomarkers for lung cancer in a set of 127 lung cancer patients and 159 control individuals. All the 29 genes displayed a higher level of methylation in patients vs. controls (all $p < 0.05$). ROC curve and AUC analysis showed that the genes had 29–88% sensitivities and 26–92% specificities in differentiating lung cancer patients from healthy controls. Four genes (*HOXA9*, *RASSF1A*, *SOX17*, and *TAC1*) were identified as the best biomarkers (all $p < 0.001$) and incorporated into a logistic

classifier: Probability of lung cancer = $e^x / (1 + e^x)$, where $x = 1.69 + 1.48 \times \log (HOXA9) - 1.25 \times \log (RASSF1A) + 0.27 \times \log (SOX17) + 0.16 \times \log (TAC1)$. The logistic classifier produced 0.92 AUC for lung cancer detection (Fig. 3B). The logistic classifier had higher accuracy (88.8 vs. 70.63%, $p = 0.004$) and sensitivity (86.6 vs. 44.8%, $p < 0.001$) than did sputum cytology, while maintaining a similar specificity (90.6 vs. 91.2%, $p = 0.34$) (Fig. 3C).

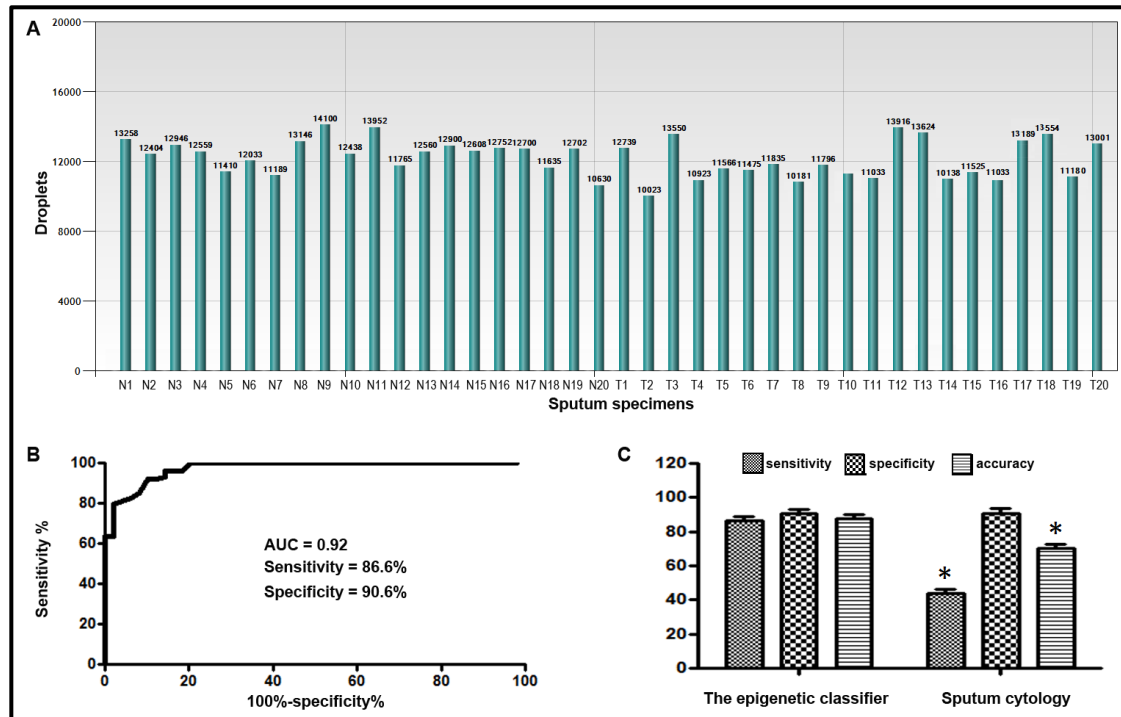


Fig. 3. ddMSP and qMSP analyses of DNA methylation in clinical sputum samples. **A**, ddMSP analysis of sputum samples of 20 cancer-free controls (normal subjects, N) and 20 patients diagnosed with lung tumor (T) for DNA methylation of RASSF1A. **B**, ddMSP analysis of 29 genes in a training set of 127 lung cancer patients and 159 controls developed an epigenetic classifier consisting of four DNA methylation biomarkers. The epigenetic classifier produced 0.92 AUC with 86.6% sensitivity and 90.6% specificity for diagnosis of lung cancer. **C**, the epigenetic classifier had higher accuracy (88.8 vs. 70.63%, $p = 0.004$) and sensitivity (86.6 vs. 44.8%, $p < 0.001$) compared with sputum cytology, whereas keeping a similar specificity (90.6 vs. 91.2%, $p = 0.34$).

Part of the study was recently published (An epigenetic classifier for early stage lung cancer. Y Su, HB Fang, F Jiang. 2018, Clinical epigenetics 10 (1), 68).

4), establishing reference samples as surrogates for clinical specimens to develop plasma biomarkers in CLIA-laboratory.

We worked with a Biomarker Laboratory (BDL) and a Biomarker Reference Laboratory (BRL) that is a CLIA laboratory of the NCI-Early Detection Research Network (EDRN) to establish reference samples that could be used as surrogates for clinical specimens. In this study, both BDL and BRL isolated RNA from the cell lines, split and shared the samples with the other site, and profiled the expression profiles of miRNAs. As shown in Fig.4, miRNA abundance was consistently observed among all possible measurement processes resulting from all combinations of isolation.

Altogether, our developed reference samples have the following properties: 1) express the genes of interest; 2) be re-generable for use over time as a process control; 3) be amenable to identical processing as clinical samples; 4) contain well-established relative differences in the genes; and 5) be commutable between measurement systems.

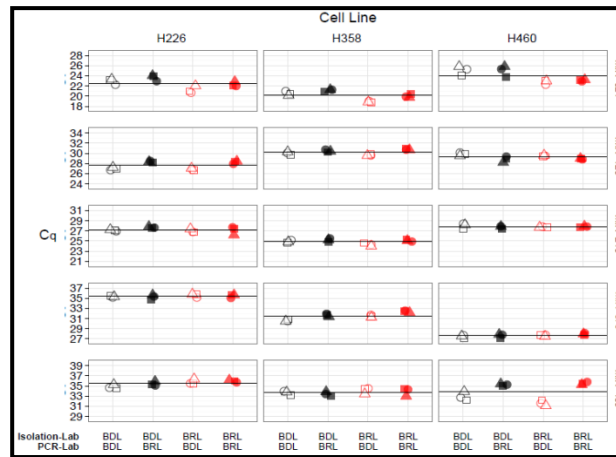


Fig. 4. Profiling of miRNAs in cancer cells. RNAs are isolated in the BDL (black symbols) and BRL (red symbols) on 3 separate occasions (circles, squares, and triangles). qPCR results are from BDL (open symbols) and BRL (filled symbols), where the black line corresponds to the mean Cq for all isolation and PCR combinations^{70, 71}.

Part of the study was recently published (P Scott Pine, Steven P Lund, Sanford A Stass, Debra Kukuruga, Feng Jiang, Lynn Sorbara, Sudhir Srivastava, Marc Salit. BMC biotechnology. 2018).

5), integrating sputum and circulating immunological biomarkers for the early detection of lung cancer

We have demonstrated that assessments of microRNA (miRNA) expressions in circulating peripheral blood mononucleated cell (PBMC) and sputum specimens, respectively, may help diagnose lung cancer. To assess the individual and combined analysis of the miRNAs across the different body fluids for lung cancer early detection, we analyzed a panel of 3 sputum miRNAs (miRs-21, 31, and 210) and a panel of 2 PBMC miRNAs (miRs-19b-3p and 29b-3p) in a discovery cohort of 68 patients with lung cancer and 66 cancer-free smokers.

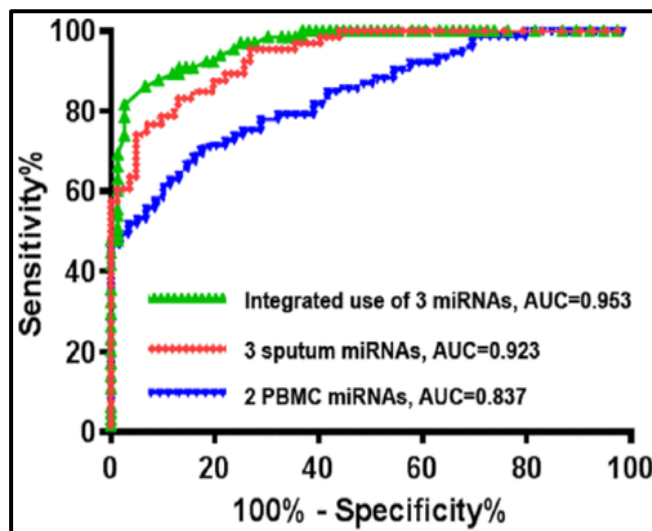


Fig. 5. The comparison of integrated miRNA biomarkers with panels of sputum and PBMC miRNAs in a discovery cohort. A prediction model based on integrated use of 3 miRNA biomarkers (miRs-31, 210, and 19b-3p) across sputum and PBMC specimens was developed for distinguishing lung cancer patients from cancer-free smokers. The ROC curve of the integrated miRNA biomarkers produced a higher AUC (0.953), as compared with the panel of sputum miRNAs (0.923) and the panel of PBMC miRNAs (0.837) (all P's < .05). AUC indicates area under the ROC; miRNAs, microRNAs; PBMC, peripheral blood mononucleated cell.

We find that integrating 2 sputum miRNAs (miRs-31 and 210) and 1 PBMC miRNA (miR-19b-3p) has higher sensitivity (86.8%) and specificity (92.4%) compared with the individual panels (Fig. 4). The synergistic value of the integrated panel of 3 biomarkers

is confirmed in a validation cohort, independent of stage and histological type of lung cancer, and patients' age, sex, and ethnicity. Integrating circulating immunological and sputum biomarkers could improve the early detection of lung cancer.

Part of the study was recently published (Jian Su, Qixin Leng, Yanli Lin, Jie Ma, Fangran Jiang, Cheng-Ju Lee, HongBin Fang, Feng Jiang. Biomarkers in cancer. 2018. 10, 1179299X18759297).

Aim 2 is to validate the newly identified sputum biomarkers, previously identified sputum DNA methylation and protein biomarkers, and PBMC biomarkers, and develop a prediction model by integrating the biomarkers with clinical and imaging variables for identifying NSCLC.

1), validating the previously developed biomarkers for lung cancer.

we have analyzed the new panel of small ncRNA biomarkers, ENO1 and three DNA methylation biomarkers in sputum, and two PBMC miRNA biomarkers in patients with either malignant or benign PNs.

Part of the results are shown in the table below.

Table 1. Successfully validating the biomarkers in the new specimens			
Names	Sample size	Sensitivity	Specificity
3 sputum miRNAs (miR-21-5p, -31-3p, and -210-3p)	117 patients with NSCLC and 116 cancer-free smokers	82.3%	87.9%
2 sputum snoRNAs (snoRD66 and snoRD78)	67 patients with NSCLC and 69 cancer-free smokers	74.6%	84.1%
3 sputum lncRNAs (SNHG1, H19, and HOTAIR)	59 patients with NSCLC and 60 cancer-free smokers	81.4%	88.3%
4 sputum methylation markers (HOXA9, RASSF1A, SOX17, and TAC1)	89 patients with NSCLC and 107 cancer-free smokers	85.4%	88.9%
ENO1	65 patients with NSCLC and 78 cancer-free smokers	82.4%	78.9%
PBMC biomarkers	117 patients with NSCLC and 116 cancer-free smokers	72.1%	81.8%

We validated the miRNAs by using ddPCR in bronchial epithelial cells of the paired sputum samples of lung cancer patients and controls. The miRNAs in sputum displayed changes consistent with those in the matched lung tumor tissues. Therefore, the altered miRNAs detected in the bronchial epitheliums of sputum could reflect those in malignant PNs (lung tumors). Furthermore, from the miRNAs, we optimized a panel of sputum miRNA biomarkers (miR-21, -31, and -210) for distinguishing malignant from benign PNs with 82% sensitivity and 88% specificity in two internal sets of patients with either malignant or benign PNs.

In addition, we validated the three sputum miRNA biomarkers in a different cohort show that the biomarkers have consistent performance as showed in the discovery phase.

Our recent validation of the sputum miRNA biomarkers using an external testing cohort of patients with either malignant (n=117) or benign PNs (n=116) confirmed its diagnostic value (Figure 6).

2), we have developed a prediction model for lung cancer detection with more than 90% sensitivity and 95% specificity

We worked with bio-statistician to develop a prediction model for identifying NSCLC by integrating the multilayered biomarkers across sputum and blood samples, features of PNs, and characteristics of smokers. The prediction model has more than 90% sensitivity and 95% specificity for detection of NSCLC in LDCT-found PNs. the diagnostic performance of this prediction model will be validated in Aim 3.

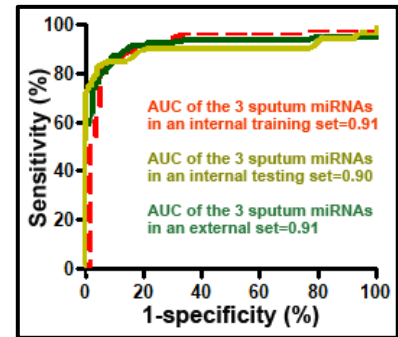


Fig. 6. Comparison of the 3 sputum miRNA biomarkers in internal and external sets of cases and controls shows that the 3 sputum miRNA biomarkers has consistent performance for distinguishing malignant from benign PNs.

In sum, **we have completed 100% goals of Aim 1, and 100% goals of Aim 2.**

Training and professional development opportunities or there is nothing significant to report during this reporting period

"Nothing to Report."

Describe opportunities for training and professional development provided to anyone who worked on the project or anyone who was involved in the activities supported by the project.

During the funded period, with the support, a research-postdoctoral fellow has been trained in the field of development of molecular biomarker for lung cancer.

Workshops, conferences, seminars, study groups, and individual study. Include participation in conferences, workshops, and seminars not listed under major activities.

We have presented our studies in March 2019 EDRN Steering Committee Meeting, Nashville, TN and 11th EDRN Scientific Workshop, September 10-12, 2019 Bethesda, Maryland and

How were the results disseminated to communities of interest?

We have published seven papers in peer-reviewed Journals.

If there is nothing significant to report during this reporting period

"Nothing to Report."

Describe how the results were disseminated to communities of interest. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

"Nothing to Report."

If this is the final report

"Nothing to Report."

Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Our Specific Aim 3 is to blindly validate the prediction model for distinguishing benign from malignant growths in a different cohort of patients with LDCT-discovered PNs.

We will perform the following two major tasks

Subtask 1: Analysis of the small ncRNA, protein, and methylation biomarkers in the specimens from a different set of cases and controls

Subtask 2: Blindly validate the performances of the prediction model for lung cancer early detection

We expect that we will develop the prediction model will have the reproducible ability to distinguish between benign and malignant PNs on the different data. We will also expect that the prediction model will have more than 90% sensitivity and 95% specificity for detection of NSCLC in LDCT-found PNs.

Impact,
No
Changes/problems, No.

Products,

Peer-reviewed articles

1. [Cancer Biomark](#). 2019 Aug 16. doi: 10.3233/CBM-190161. [Epub ahead of print]. **Sputum long non-coding RNA biomarkers for diagnosis of lung cancer.** [Gupta C](#)^{1,1}, [Su J](#)^{1,1}, [Zhan M](#)², [Stass SA](#)¹, [Jiang F](#)¹.
2. [J Cancer Res Clin Oncol](#). 2018 Nov;144(11):2109-2115. doi: 10.1007/s00432-018-2735-0. Epub 2018 Aug 12. **Fucosylation genes as circulating biomarkers for lung cancer.** [Leng Q](#)¹, [Tsou JH](#)¹, [Zhan M](#)², [Jiang F](#)³.
3. [Transl Oncol](#). 2018 Oct;11(5):1225-1231. doi: 10.1016/j.tranon.2018.07.016. Epub 2018 Aug 8. **A Plasma Long Noncoding RNA Signature for Early Detection of Lung Cancer.** [Lin Y](#)¹, [Leng Q](#)², [Zhan M](#)³, [Jiang F](#)⁴.
4. [Oncotarget](#). 2018 May 15;9(37):24684-24692. doi: 10.18632/oncotarget.25227. eCollection 2018 May 15. **An integromic signature for lung cancer early detection.** [Leng Q](#)¹, [Lin Y](#)¹, [Zhan M](#)², [Jiang F](#)¹.
5. [Clin Epigenetics](#). 2018 May 22;10:68. doi: 10.1186/s13148-018-0502-3. eCollection 2018. **An epigenetic classifier for early stage lung cancer.** [Su Y](#)¹, [Fang HB](#)², [Jiang F](#)³.
6. [Transl Oncol](#). 2018 Aug;11(4):883-889. doi: 10.1016/j.tranon.2018.05.001. Epub 2018 May 31. **A Direct Plasma miRNA Assay for Early Detection and Histological Classification of Lung Cancer.** [Leng Q](#)¹, [Wang Y](#)², [Jiang F](#)³.
7. [Biomark Cancer](#). 2018 Feb 13;10:1179299X18759297. doi: 10.1177/1179299X18759297. eCollection 2018. **Integrating Circulating Immunological and Sputum Biomarkers for the Early Detection of Lung Cancer.** [Su J](#)¹, [Leng Q](#)¹, [Lin Y](#)¹, [Ma J](#)¹, [Jiang F](#)¹, [Lee CJ](#)¹, [Fang H](#)¹, [Jiang F](#)¹.

Participants and other collaborating organizations,
The Veterans Affairs Maryland Health Care System
The University of Maryland School of Medicine

Special reporting requirements,
No.

Appendices.
Please see the attached file with the published papers.

Sputum long non-coding RNA biomarkers for diagnosis of lung cancer

Chhavi Gupta^{1†}, Jian Su^{1†}, Min Zhan², and Sanford A Stass¹, and Feng Jiang^{1*}

1. Department of Pathology, University of Maryland School of Medicine, 10 S. Pine St. Baltimore, MD 21201, USA

2. Departments of Epidemiology & Public Health, University of Maryland School of Medicine, 660 W. Redwood St. Baltimore, MD 21201, USA

†. The authors contributed equally to this work.

*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.

Running title: Sputum long non-coding RNAs as biomarkers for lung cancer

Abstract

Background. Analysis of molecular changes in sputum may help diagnose lung cancer. Long non-coding RNAs (lncRNAs) play vital roles in various biological processes, and their dysregulations contribute to the development and progression of lung tumorigenesis. Herein, we determine whether aberrant lncRNAs could be used as potential sputum biomarkers for lung cancer.

Methods. Using reverse transcription PCR, we measure expressions of lung cancer-associated lncRNAs in sputum of a discovery cohort of 67 lung cancer patients and 65 cancer-free smokers with benign diseases and a validation cohort of 59 lung cancer patients and 60 cancer-free smokers with benign diseases.

Results. In the discovery cohort, four of the lncRNAs displayed a significantly different level in sputum of lung cancer patients *vs.* cancer-free smokers with benign diseases (all $P < 0.001$). From the four lncRNAs, three lncRNAs (SNHG1, H19, and HOTAIR) are identified as a biomarker panel, producing 82.09% sensitivity and 89.23% specificity for diagnosis of lung cancer. Furthermore, the biomarker panel has a higher sensitivity (82.09% *vs.* 52.24%, $P=0.02$) and a similar specificity compared with sputum cytology (89.23% *vs.* 90.77%, $P=0.45$). In addition, the lncRNA biomarker panel had a higher sensitivity (87.50% *vs.* 70.07%, $p=0.03$) for diagnosis of squamous cell carcinoma compared with adenocarcinoma of the lung, while maintaining the same specificity (89.23%). The potential of the sputum lncRNA biomarkers for lung cancer detection is confirmed in the validation cohort.

Conclusion. We have for the first time shown that the analysis of lncRNAs in sputum might be a noninvasive approach for diagnosis of lung cancer.

Keywords: diagnosis; lung cancer; sputum; lncRNA; biomarkers.

1 Introduction

Approximately 155,870 Americans will die from lung cancer each year, more than the other three leading cancers combined (breast, prostate, and colorectal cancers). Over 85% lung cancers are non-small cell lung cancers (NSCLC). NSCLC mainly consists of adenocarcinoma (AC) and squamous cell carcinoma (SCC). Tobacco smoking is the major cause of NSCLC. The early detection of lung cancer in a large randomized trial using low-dose CT (LDCT) has revealed a 20% reduction in mortality as compared to chest X-rays[1]. Therefore, lung cancer early detection can increase curability and save lives. LDCT is recommended to be used for lung cancer early detection among smokers in the USA [2; 3]. However, LDCT is associated with over-diagnosis, excessive cost, and radiation exposure [2; 3]. The development of non-invasive biomarkers that can accurately and cost-effectively diagnose early stage lung cancer remains to be clinically imperative [2; 3].

It has been well accepted that lung tumors develop from a field defect characterized by molecular abnormalities resulted from repeated exposure of the entire airway to the tobacco carcinogens [4; 5]. Kadara *et. al.* [6; 7] found that the molecular alterations in the large bronchial airways reflected those in primary lung tumors in the distal lung, regardless of the anatomic location relative to the tumors. Spira *et. al.* [8; 9] demonstrated that molecular changes in epithelial cells collected from the normal-appearing mainstem bronchus of smokers could be developed as biomarkers for NSCLC. Sputum contains bronchial epithelial cells from the lungs and lower respiratory tract. Based on the field cancerizations, examination of the exfoliated bronchial epitheliums of airway in sputum might detect the lung tumor-related molecular alterations, and hence provide a non-invasive and specific means for diagnosis of NSCLC [6; 7; 8; 10] [11] [12].

Long non-coding RNAs (lncRNAs) are transcripts of longer than 200 nucleotides in length, and have important functions in various biological processes [13] [14]. Dysfunctions of lncRNAs play vital roles in the development of lung cancer [15]. Specifically, lncRNAs can regulate different molecular signaling pathways via changing gene expression, and therefore, are implicated in numerous mechanisms of lung carcinogenesis[16; 17] [18] [19] [20]. Since aberrant lncRNAs detected in clinical specimens are associated with lung tumorigenesis, exploration of the molecules for clinical diagnosis is a growing area of research [15; 21; 22]. For examples, Liu *et al.* found that an lncRNA, HOTAIR, was significantly up-regulated in NSCLC tissues and might be a prognostic biomarker for the disease [23]. Furthermore, we have found that expression levels of SNHG1 and RMRP are reliably measured in plasma, and the lncRNAs may provide cell-free circulating biomarkers for lung cancer[24]. In addition, we recently showed that integrated analysis of lncRNAs, miRNAs, and mRNAs in plasma could have a synergistic impact on lung cancer early detection [25]. However, examination of aberrant lncRNA levels in sputum, which is surrogate material for noninvasive and specific diagnosis of lung cancer has not been investigated.

We recently identified six lncRNAs whose aberrant expression level in surgically resected tumor tissues was associated with lung cancer [26; 27]. The objective of this study was to investigate whether the lung cancer-associated lncRNAs could be used as sputum biomarkers for distinguishing lung cancer patients from smokers with benign diseases.

2. Materials and Methods

2.1. Study population

The study protocol was approved by the Institutional Review Board of the University of Maryland Medical Center. Written informed consent forms were obtained from all participants.

The participants were recruited at the point of their referral for suspected lung cancer between the ages of 55–80. Exclusion criteria included pregnancy, current pulmonary infection, surgery within 6 months, radiotherapy within 1 year, and life expectancy of < 1 year. Clinical diagnosis of lung cancer was made using histopathologic examinations of specimens obtained by CT-guided transthoracic needle biopsy, transbronchial biopsy, videotape-assisted thoracoscopic surgery, or surgical resection. The surgical pathologic staging was determined according to the TNM classification of the International Union Against Cancer with the 8th American Joint Committee on Cancer and the International Staging System for Lung Cancer. Histopathological classification was determined according to the World Health Organization classification. A total of 251 subjects including 126 lung cancer patients and 125 cancer-free smokers with benign diseases were recruited. The 126 lung cancer patients were diagnosed with NSCLC consisting of 71 stage I cases, 33 stage II cases, and 22 stage III-IV cases. Seventy-six cases were AC and 50 were SCC of the lungs. All the 125 cancer-free individuals were smokers who served as control group participants. These control participants had granulomatous inflammation (n = 89), nonspecific inflammatory changes (n = 23) or lung infections (n = 13). The cancer-free individuals had been followed for at least 2 years, and none had any evidence of cancer. No difference of age, gender, and smoking status was observed in the lung cancer cases vs. cancer-free smokers (All $p > 0.05$). The lung cases and cancer-free smokers were randomly split into a training cohort and a testing cohort by using a random number generator. The training cohort consisted of 67 lung cancer patients and 65 cancer-free smokers. The testing cohort comprised 59 lung cancer patients and 60 cancer-free smokers. The demographic and clinical characteristics of the two cohorts are presented in Tables 1 and 2.

2.2. Sputum collection, preparation, and sputum cytological study

Sputum was collected and processed from the subjects as previously described [10; 28; 29; 30; 31; 32; 33; 34; 35; 36; 37; 38; 39; 40; 41; 42; 43], before they received any treatment (e.g., surgery, preoperative adjuvant chemotherapy, and radiotherapy). The sputum samples were centrifuged at 800 g for 10 min. The cell pellets were mixed with phosphate-buffered saline solution (PBS) (Sigma-Aldrich, St. Louis, MO). Cytospin slides were prepared from the cell pellets and underwent Papanicolaou staining for evaluating whether the specimens were representative of deep bronchial cells. All sputum samples were of lower respiratory origin as indicated by the presence of macrophages and bronchial epithelial cells. Cytologic diagnosis was made using the classification of Saccomanno [44]. Positive cytology included both carcinoma *in situ* and invasive carcinoma [28; 29]. The remaining sputum cells were stored at -80°C until being tested.

2.3. The analysis of lncRNAs in sputum by reverse transcription-PCR (RT-PCR)

RNA was extracted from cell pellets of sputum as previously described [31; 32; 33; 34; 36; 37]. The purity and concentration of RNA were determined by OD_{260/280} readings using a dual beam UV spectrophotometer (Eppendorf AG, Hamburg, Germany). RNA integrity was determined by capillary electrophoresis using the RNA 6000 Nano Lab-on-a-Chip kit and the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). The expression levels of the six lncRNAs (SNHG1, MALAT1, HOTAIR, H19, MEG3, and RMRP) were determined in sputum by using RT-PCR with Taqman assays (Applied Biosystems, Foster City, CA) as previously described [24; 25]. The cycle threshold (Ct) was defined as the number of cycles required for the fluorescent signal to cross the threshold. An internal control gene, U6, was analyzed in parallel in the specimens. Relative expression of a targeted lncRNA in a given sample was computed using the equation $2^{-\Delta Ct}$, where $\Delta Ct = Ct(\text{targeted lncRNA}) - Ct(U6)$. 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.

2.4. Statistical analysis

We used univariate analysis to identify the lncRNAs whose expression levels were associated with NSCLC. We analyzed the significantly associated lncRNAs by using logistic regression models with constrained parameters as least absolute shrinkage and selection operator (LASSO) to eliminate the less relevant variables[45; 46; 47]. We estimated functions of the combined lncRNA biomarkers by logistic regression with or without adjustment for known risk factors for NSCLC. The performances of the lncRNAs were subjected to receiver-operator characteristic (ROC) curve. We also generated a 95% confidence interval for the difference in the area under the ROCs (AUCs) by using the bootstrap[48]. We established the optimal cut-off value by using the Youden index [49; 50]. The results of the training cohort were blindly validated in the testing cohort by using leave-one-out cross validation[51]. To compare different approaches for lung cancer diagnosis, we computed their AUCs to determine the sensitivity and specificity as previously described [52].

3. Results

3.1. Developing a panel of sputum lncRNA biomarkers for the early detection of lung cancer.

All targeted six lncRNAs had ≤ 32 Ct values in each sputum sample of the training cohort, and therefore were reliably detectable in the specimens. No product was synthesized in the negative control samples. As shown in Table 3, four (SNHG1, H19, and HOTAIR, and MEG3) of the six lncRNAs displayed a significantly different level in sputum of patients with lung cancer vs. cancer-free individuals, producing 0.72-0.80 AUCs (all $P < 0.01$). Of the four lncRNAs, three (SNHG1, H19, and HOTAIR) had a higher, whereas one (MEG3) had a lower expression level in sputum of lung cancer patients vs. cancer-free individuals (All $P < 0.01$) (Table 3). We used logistic regression models with constrained parameters as in LASSO to develop a panel of lncRNA biomarkers for lung cancer. SNHG1, H19, and HOTAIR were selected as the best biomarkers and incorporated into an algorithm, which produced 0.90 AUC for distinguishing lung cancer cases from cancer-free subjects (Supplementary Fig. 1). Furthermore, the estimated correlations among expression levels of the three lncRNAs were low (All $P > 0.05$), implying that the diagnostic values of the lncRNAs were complementary to each other. Using Youden's index, we set up optimal cut-off at 0.36. Based on the cut-off, combined analysis of the three lncRNAs generated 82.03% sensitivity and 89.23% specificity for diagnosis of lung cancer (Table 4). Sputum cytology has 52.24% sensitivity and 90.77% specificity. Therefore, the sputum biomarkers had a higher sensitivity (82.03%) compared with sputum cytology (52.24%) ($P = 0.02$), while maintaining a similar specificity (89.23% vs. 90.77%, $p = 0.45$) (Fig. 1). Interestingly, the panel of the lncRNA biomarkers had a higher sensitivity (87.50%) for SCC compared with AC of the lungs (79.07%), while maintaining the same specificity (89.23 %) (Table 4). Therefore, the sputum biomarker panel could better differentiate SCC cases from cancer-free individuals compared with AC cases. In addition, the three lncRNAs did not display statistical differences of sensitivity and specificity between early and advanced stages (stage I-II vs. stage III-IV) ($P > 0.05$). Moreover, the sputum biomarkers had no special association with the age, gender and ethnicity of the participants (All $p > 0.05$). The expression levels of the sputum lncRNAs were associated with smoking history of lung cancer patients at the edge of significance ($P = 0.05$).

3.2. Validating the sputum lncRNA markers in an independent cohort of lung cancer patients and cancer-free individuals. To evaluate the diagnostic performance of the sputum biomarkers, the three lncRNAs (SNHG1, H19, and HOTAIR) were assessed in

sputum of additional 59 NSCLC patients and 60 cancer-free individuals. The three genes used in combination could differentiate the NSCLC patients from cancer-free individuals with 81.36% sensitivity and 88.33% specificity (Table 4). Furthermore, the panel of the lncRNA biomarkers had a higher sensitivity for SCC (88.46%) compared with AC (75.76%) of the lungs ($P=0.03$) with the same specificity (88.33 %). In addition, the three sputum lncRNA biomarkers didn't show association with the age, gender and ethnicity of the participants (All $p > 0.05$). Taken together, the results generated from the validation cohort confirmed the potential of combined use of the three lncRNAs as a sputum biomarker panel for the early detection of lung cancer.

4. Discussion

As a mirror to lung diseases, sputum is noninvasively obtained and contains bronchial epithelial cells from the lungs and lower respiratory tract, and thus has the advantages as surrogate material for specifically diagnosing lung cancer. Herein, we for the first time report the feasibility of detection of lncRNA in sputum and develop a panel of sputum lncRNA biomarkers for lung cancer. Furthermore, the biomarker panel has a higher sensitivity and a similar specificity compared with sputum cytology, a standard clinical model. Like other sputum molecular biomarkers, the sputum lncRNA biomarkers display a higher diagnostic performance for SCC than AC of the lungs. Interestingly, the sensitivity and specificity of the biomarkers are independent of stage of lung cancer. Therefore, the discovery might be an important characteristic if they are employed for the diagnosis of NSCLC, particularly SCC, at the early stage. Nevertheless, the potential of the sputum lncRNA biomarkers for the early detection of lung cancer need to be prospectively validated in a large cohort study.

The three lncRNAs (SNHG1, H19, and HOTAIR) have diverse and important functions in lung tumorigenesis through regulating different molecular pathways. For instance, elevated expression of SNHG1 was frequently observed in lung cancer tissues [53]. Furthermore, SNHG1 could promote NSCLC progression of lung cancer via miR-101-3p/SOX9/Wnt/ β -catenin regulatory network and miR-145-5p/ MTDH axis[54] [53]. In addition, SNHG1 plays an oncogenic role in lung cancer through ZEB1 signaling pathway by inhibiting TAP63 [55]. H19 is an imprinted oncofetal lncRNA, which is expressed in the embryo, down-regulated at birth and then reappears in tumors[56]. H19 plays important function in cell proliferation and differentiation, epithelial to mesenchymal transition, and mesenchymal to epithelial transition [57]. H19 has also the versatile and functions depending on its context within the process of tumor progression [58]. Interestingly, lung tumor-released H19 could promote gefitinib resistance via packaging into exosomes in lung cancer cells [59]. HOTAIR represses gene expression through recruitment of chromatin modifiers. The increased HOTAIR expression has been founded in lung cancer, and the expression level correlates with metastasis and poor prognosis. Moreover, HOTAIR promotes proliferation, survival, invasion, metastasis, and drug resistance in lung cancer cells through miR-217/DACH1 signaling pathway [60] [61]. Our current study extends the previous findings by developing the lung cancer-associated lncRNA as sputum biomarkers that might be useful in the early detection of lung cancer.

There are some limitations in this present study. 1), the size of the cohorts is small. We will perform a large and prospective study to validate the diagnostic value of the sputum lncRNA biomarkers for the early detection of lung cancer. 2), the panel of the three biomarkers is developed from six lung cancer-associated lncRNAs. Although showing promise, the three biomarkers' diagnostic performance is not high enough to be used in clinical settings. By searching published data, we recently found 21 lncRNAs whose malfunction was characterized in lung tumorigenesis [19; 60; 62; 63; 64; 65; 66; 67; 68; 69; 70; 71; 72; 73; 74; 75; 76; 77; 78; 79; 80; 81; 82; 83; 84; 85; 86]. Our ongoing study is evaluating if the additional lncRNAs might be

added into the biomarker panel to improve the diagnosis of lung cancer. 3), the objective of this present study is to develop sputum biomarkers for differentiating lung cancer from cancer-free smokers who have benign diseases. Toward that end, we collect sputum from lung cancer patients and cancer-free smokers with benign diseases who serve as control individuals. Sputum of healthy controls is not available in this study. Therefore, we are not able to detect differences of lncRNAs in sputum between smokers with benign diseases and healthy controls. In future, it might be important to collect sputum from nonsmokers who don't have any benign disease to determine if there is different lncRNA changes in sputum between the individuals with the smokes with benign diseases and the healthy controls.

In sum, sputum lncRNA biomarkers are developed that could potentially be used as a noninvasive approach for differentiating early stage NSCLC patients from cancer-free smokers. Nevertheless, undertaking a prospective study to further validate the sputum biomarkers in a large cohort is required.

Acknowledgments

This work was supported in part by VA Merit Award Io1 CX000512, the Geaton and JoAnn DeCesaris Family Foundation, NIH/NCI-1R21CA240556, University of Maryland Marlene & Stewart Greenebaum Comprehensive Cancer Center Pilot Grant Program, and DoD-Idea Development Award (F.J.), and NIH/NCI-Early Detection Research Network-5U24CA11509 (S. S.).

Conflict of interest

The authors declare no conflict of interest.

Legends

Table 1. Characteristics of NSCLC patients and cancer-free smokers in a training cohort

Table 2. Characteristics of NSCLC patients and cancer-free smokers in a training cohort

Table 3. Expression levels of sputum lncRNAs in the training cohort of lung cancer patients and cancer-free individuals

Table 4. Diagnostic values of the panel of sputum lncRNA biomarkers in training cohort and testing cohort

Fig. 1. Sensitivity and specificity of sputum lncRNA biomarkers and cytology for diagnosis of lung cancer. The sputum lncRNA biomarker panel has a higher sensitivity and a similar specificity compared with sputum cytology ($P=0.02$). *, $P<0.05$.

Supplementary Fig.1. The receiver-operator characteristic (ROC) curve of sputum lncRNA biomarkers for lung cancer diagnosis.

Table 1. Characteristics of NSCLC patients and cancer-free smokers in a training cohort

	NSCLC cases (n = 67)	Cancer-free individuals (n = 65)	P-value
Age	68.39 (SD 11.36)	66.75 (SD 11.32)	0.38
Sex			0.46
Female	26	23	
Male	41	42	
Race			0.45
White	44	43	
African American	23	22	
Pack-years (median)	35.89	35.47	0.39
Stage			
Stage I	40		
Stage II	17		
Stage III-IV	10		
Histological type			
Adenocarcinoma	43		
Squamous cell carcinoma	24		

Abbreviations: NSCLC, non-small cell lung cancer.

Table 2. Characteristics of NSCLC patients and cancer-free smokers in a training cohort

	NSCLC cases (n = 59)	Cancer-free individuals (n = 60)	P-value
Age	68.57 (SD 12.03)	66.28 (SD 11.60)	0.39
Sex			0.43
Female	23	22	
Male	36	38	
Race			0.47
White	39	40	
African American	20	20	
Pack-years (median)	36.02	35.59	0.40
Stage			
Stage I	31		
Stage II	16		
Stage III-IV	12		
Histological type			
Adenocarcinoma	33		
Squamous cell carcinoma	26		

Abbreviations: NSCLC, non-small cell lung cancer.

Table 3. Expression levels of sputum lncRNAs in the training cohort of lung cancer patients and cancer-free individuals

lncRNAs	Mean (SEM) of level in NSCLC patients	Mean (SEM) of level in cancer-free individuals	P-value	AUC (95% CI)
SNHG1	6.65 (1.08)	5.29 (1.13)	< 0.001	0.80 (0.73 to 0.87)
H19	3.39 (2.63)	2.25 (0.65)	< 0.001	0.77 (0.68 to 0.83)
HOTAIR	1.08 (0.56)	0.29 (0.09)	< 0.001	0.73 (0.66 to 0.83)
MEG3	1.53 (0.92)	3.02 (1.56)	< 0.001	0.72 (0.63 to 0.78)

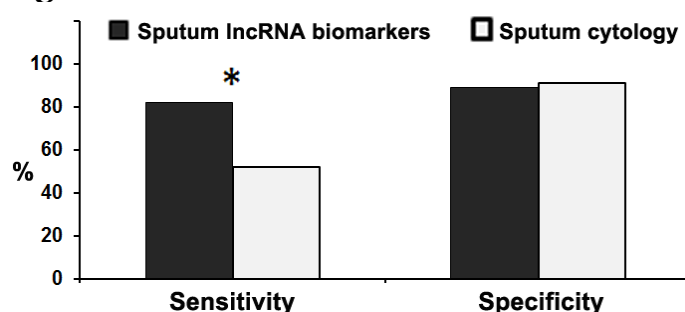
Abbreviations: SEM, the standard error of the mean; AUC, the area under receiver operating characteristic curve; CI, confidence interval.

Table 4. Diagnostic values of the sputum lncRNA biomarkers in training cohort and testing cohort

	The training cohort		The testing cohort	
	Sensitivity (CI)	Specificity (CI)	Sensitivity (CI)	Specificity (CI)
NSCLC	82.09% (70.80% to 90.39%)	89.23 % (79.06% to 95.56%)	81.36% (69.09% to 90.31%)	88.33 % (77.43% to 95.18%)
AC	79.07% (63.96% to 89.96%)	89.23 % (79.06% to 95.56%)	75.76% (57.74% to 88.91%)	88.33 % (77.43% to 95.18%)
SCC	87.50% (67.64% to 97.34%)	89.23 % (79.06% to 95.56%)	88.46% (69.85% to 97.55%)	88.33 % (77.43% to 95.18%)

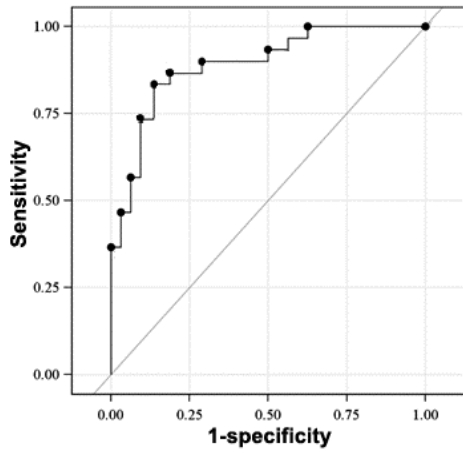
Abbreviations: NSCLC, non-small cell lung cancers; AC, adenocarcinoma; SCC, squamous cell carcinoma, CI, confidence interval.

Fig. 1



Sensitivity and specificity of sputum lncRNA biomarkers and cytology for diagnosis of lung cancer. The sputum lncRNA biomarker panel has a higher sensitivity and a similar specificity compared with sputum cytology (P=0.02). *, P<0.05.

Supplementary Fig.1.



The receiver-operator characteristic (ROC) curve of sputum lncRNA biomarkers for lung cancer diagnosis. Three lncRNAs (SNHG1, H19, and HOTAIR) were selected as the best ones and incorporated into an algorithm: Probability $P = \exp(U) / [1 + \exp(U)]$, where $U = 2.93 + 2.12 * \log(\text{SNHG1}) - 1.26 * \log(\text{H19}) - 1.1795 * \log(\text{HOTAIR})$. The three-lncRNA-algorithm produced the area under the ROC (AUC) of 0.90 for distinguishing lung cancer cases from cancer-free individuals.

References

- [1] D.R. Aberle, A.M. Adams, C.D. Berg, W.C. Black, J.D. Clapp, R.M. Fagerstrom, I.F. Gareen, C. Gatsonis, P.M. Marcus, J.D. Sicks, Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 365 (2011) 395-409.
- [2] E.F. Patz, Jr., P. Pinsky, C. Gatsonis, J.D. Sicks, B.S. Kramer, M.C. Tammemagi, C. Chiles, W.C. Black, D.R. Aberle, Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med* 174 (2014) 269-274.
- [3] D.R. Aberle, C.D. Berg, W.C. Black, T.R. Church, R.M. Fagerstrom, B. Galen, I.F. Gareen, C. Gatsonis, J. Goldin, J.K. Gohagan, B. Hillman, C. Jaffe, B.S. Kramer, D. Lynch, P.M. Marcus, M. Schnall, D.C. Sullivan, D. Sullivan, C.J. Zylak, The National Lung Screening Trial: overview and study design. *Radiology* 258 (2011) 243-253.
- [4] X. Tang, H. Shigematsu, B.N. Bekele, J.A. Roth, J.D. Minna, W.K. Hong, A.F. Gazdar, Wistuba, II, EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. *Cancer Res* 65 (2005) 7568-7572.
- [5] L.M. Solis, C. Behrens, M.G. Raso, H.Y. Lin, H. Kadara, P. Yuan, H. Galindo, X. Tang, J.J. Lee, N. Kalhor, Wistuba, II, C.A. Moran, Histologic patterns and molecular characteristics of lung adenocarcinoma associated with clinical outcome. *Cancer* 118 (2012) 2889-2899.
- [6] H. Kadara, L. Shen, J. Fujimoto, P. Saintigny, C.W. Chow, W. Lang, Z. Chu, M. Garcia, M. Kabbout, Y.H. Fan, C. Behrens, D.A. Liu, L. Mao, J.J. Lee, K.A. Gold, J. Wang, K.R. Coombes, E.S. Kim, W.K. Hong, Wistuba, II, Characterizing the molecular spatial and temporal field of injury in early-stage smoker non-small cell lung cancer patients after definitive surgery by expression profiling. *Cancer Prev Res (Phila)* 6 (2013) 8-17.
- [7] H. Kadara, Wistuba, II, Field cancerization in non-small cell lung cancer: implications in disease pathogenesis. *Proc Am Thorac Soc* 9 (2012) 38-42.
- [8] A. Spira, J.E. Beane, V. Shah, K. Steiling, G. Liu, F. Schembri, S. Gilman, Y.M. Dumas, P. Calner, P. Sebastiani, S. Sridhar, J. Beamis, C. Lamb, T. Anderson, N. Gerry, J. Keane, M.E. Lenburg, J.S. Brody, Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 13 (2007) 361-366.
- [9] G.A. Silvestri, A. Vachani, D. Whitney, M. Elashoff, K. Porta Smith, J.S. Ferguson, E. Parsons, N. Mitra, J. Brody, M.E. Lenburg, A. Spira, A Bronchial Genomic Classifier for the Diagnostic Evaluation of Lung Cancer. *N Engl J Med* 373 (2015) 243-251.
- [10] S.A. Belinsky, K.C. Liechty, F.D. Gentry, H.J. Wolf, J. Rogers, K. Vu, J. Haney, T.C. Kennedy, F.R. Hirsch, Y. Miller, W.A. Franklin, J.G. Herman, S.B. Baylin, P.A. Bunn, T. Byers, Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res* 66 (2006) 3338-3344.
- [11] N. Fusco, C. Fumagalli, E. Guerini-Rocco, Looking for sputum biomarkers in lung cancer secondary prevention: where are we now? *J Thorac Dis* 9 (2017) 4277-4279.
- [12] A.J. Hubers, C.F. Prinsen, G. Sozzi, B.I. Witte, E. Thunnissen, Molecular sputum analysis for the diagnosis of lung cancer. *Br J Cancer* 109 (2013) 530-537.
- [13] L. Ma, V.B. Bajic, Z. Zhang, On the classification of long non-coding RNAs. *RNA Biol* 10 (2013) 925-933.
- [14] R. Spizzo, M.I. Almeida, A. Colombatti, G.A. Calin, Long non-coding RNAs and cancer: a new frontier of translational research? *Oncogene* 31 (2012) 4577-4587.
- [15] J. Yang, J. Lin, T. Liu, T. Chen, S. Pan, W. Huang, S. Li, Analysis of lncRNA expression profiles in non-small cell lung cancers (NSCLC) and their clinical subtypes. *Lung Cancer* 85 (2014) 110-115.
- [16] D. Meseure, K. Drak Alsibai, A. Nicolas, I. Bieche, A. Morillon, Long Noncoding RNAs as New Architects in Cancer Epigenetics, Prognostic Biomarkers, and Potential Therapeutic Targets. *Biomed Res Int* 2015 (2015) 320214.

- [17]J.R. Prensner, A.M. Chinnaiyan, The emergence of lncRNAs in cancer biology. *Cancer Discov* 1 (2011) 391-407.
- [18]M. Zhou, M. Guo, D. He, X. Wang, Y. Cui, H. Yang, D. Hao, J. Sun, A potential signature of eight long non-coding RNAs predicts survival in patients with non-small cell lung cancer. *J Transl Med* 13 (2015) 231.
- [19]M. Li, M. Qiu, Y. Xu, Q. Mao, J. Wang, G. Dong, W. Xia, R. Yin, L. Xu, Differentially expressed protein-coding genes and long noncoding RNA in early-stage lung cancer. *Tumour Biol* 36 (2015) 9969-9978.
- [20]L.H. Schmidt, T. Spieker, S. Koschmieder, S. Schaffers, J. Humberg, D. Jungen, E. Bulk, A. Hascher, D. Wittmer, A. Marra, L. Hillejan, K. Wiebe, W.E. Berdel, R. Wiewrodt, C. Muller-Tidow, The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. *J Thorac Oncol* 6 (2011) 1984-1992.
- [21]J. Chen, R. Wang, K. Zhang, L.B. Chen, Long non-coding RNAs in non-small cell lung cancer as biomarkers and therapeutic targets. *J Cell Mol Med* 18 (2014) 2425-2436.
- [22]H. Tao, J.J. Yang, X. Zhou, Z.Y. Deng, K.H. Shi, J. Li, Emerging role of long noncoding RNAs in lung cancer: Current status and future prospects. *Respir Med* 110 (2016) 12-19.
- [23]X.H. Liu, Z.L. Liu, M. Sun, J. Liu, Z.X. Wang, W. De, The long non-coding RNA HOTAIR indicates a poor prognosis and promotes metastasis in non-small cell lung cancer. *BMC Cancer* 13 (2013) 464.
- [24]Y. Lin, Q. Leng, M. Zhan, F. Jiang, A Plasma Long Noncoding RNA Signature for Early Detection of Lung Cancer. *Transl Oncol* 11 (2018) 1225-1231.
- [25]Q. Leng, Y. Lin, M. Zhan, F. Jiang, An integromic signature for lung cancer early detection. *Oncotarget* 9 (2018) 24684-24692.
- [26]J. Ma, K. Mannoor, L. Gao, A. Tan, M.A. Guarnera, M. Zhan, A. Shetty, S.A. Stass, L. Xing, F. Jiang, Characterization of microRNA transcriptome in lung cancer by next-generation deep sequencing. *Mol Oncol* 8 (2014) 1208-1219.
- [27]L. Gao, J. Ma, K. Mannoor, M.A. Guarnera, A. Shetty, M. Zhan, L. Xing, S.A. Stass, F. Jiang, Genome-wide small nucleolar RNA expression analysis of lung cancer by next-generation deep sequencing. *Int J Cancer* 136 (2015) E623-629.
- [28]M.S. Romeo, I.A. Sokolova, L.E. Morrison, C. Zeng, A.E. Baron, F.R. Hirsch, Y.E. Miller, W.A. Franklin, M. Varella-Garcia, Chromosomal abnormalities in non-small cell lung carcinomas and in bronchial epithelia of high-risk smokers detected by multi-target interphase fluorescence in situ hybridization. *J Mol Diagn* 5 (2003) 103-112.
- [29]M. Varella-Garcia, J. Kittelson, A.P. Schulte, K.O. Vu, H.J. Wolf, C. Zeng, F.R. Hirsch, T. Byers, T. Kennedy, Y.E. Miller, R.L. Keith, W.A. Franklin, Multi-target interphase fluorescence in situ hybridization assay increases sensitivity of sputum cytology as a predictor of lung cancer. *Cancer Detect Prev* 28 (2004) 244-251.
- [30]L. Yu, J. Shen, K. Mannoor, M. Guarnera, F. Jiang, Identification of ENO1 As a Potential Sputum Biomarker for Early-Stage Lung Cancer by Shotgun Proteomics. *Clin Lung Cancer* (2014).
- [31]J. Shen, J. Liao, M.A. Guarnera, H. Fang, L. Cai, S.A. Stass, F. Jiang, Analysis of MicroRNAs in sputum to improve computed tomography for lung cancer diagnosis. *J Thorac Oncol* 9 (2014) 33-40.
- [32]N. Li, J. Ma, M.A. Guarnera, H. Fang, L. Cai, F. Jiang, Digital PCR quantification of miRNAs in sputum for diagnosis of lung cancer. *J Cancer Res Clin Oncol* 140 (2014) 145-150.
- [33]N. Anjuman, N. Li, M. Guarnera, S.A. Stass, F. Jiang, Evaluation of lung flute in sputum samples for molecular analysis of lung cancer. *Clin Transl Med* 2 (2013) 15.

- [34]L. Yu, N.W. Todd, L. Xing, Y. Xie, H. Zhang, Z. Liu, H. Fang, J. Zhang, R.L. Katz, F. Jiang, Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. *Int J Cancer* 127 (2010) 2870-2878.
- [35]F. Jiang, N.W. Todd, R. Li, H. Zhang, H. Fang, S.A. Stass, A panel of sputum-based genomic marker for early detection of lung cancer. *Cancer Prev Res (Phila)* 3 (2010) 1571-1578.
- [36]L. Xing, N.W. Todd, L. Yu, H. Fang, F. Jiang, Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers. *Mod Pathol* 23 (2010) 1157-1164.
- [37]Y. Xie, N.W. Todd, Z. Liu, M. Zhan, H. Fang, H. Peng, M. Alattar, J. Deepak, S.A. Stass, F. Jiang, Altered miRNA expression in sputum for diagnosis of non-small cell lung cancer. *Lung Cancer* 67 (2010) 170-176.
- [38]F. Jiang, N.W. Todd, Q. Qiu, Z. Liu, R.L. Katz, S.A. Stass, Combined genetic analysis of sputum and computed tomography for noninvasive diagnosis of non-small-cell lung cancer. *Lung Cancer* 66 (2009) 58-63.
- [39]R.L. Katz, T.M. Zaidi, R.L. Fernandez, J. Zhang, W. He, C. Acosta, M. Daniely, L. Madi, M.A. Vargas, Q. Dong, X. Gao, F. Jiang, N.P. Caraway, A.A. Vaporciyan, J.A. Roth, M.R. Spitz, Automated detection of genetic abnormalities combined with cytology in sputum is a sensitive predictor of lung cancer. *Mod Pathol* 21 (2008) 950-960.
- [40]Q. Qiu, N.W. Todd, R. Li, H. Peng, Z. Liu, H.G. Yfantis, R.L. Katz, S.A. Stass, F. Jiang, Magnetic enrichment of bronchial epithelial cells from sputum for lung cancer diagnosis. *Cancer* 114 (2008) 275-283.
- [41]R. Li, N.W. Todd, Q. Qiu, T. Fan, R.Y. Zhao, W.H. Rodgers, H.B. Fang, R.L. Katz, S.A. Stass, F. Jiang, Genetic deletions in sputum as diagnostic markers for early detection of stage I non-small cell lung cancer. *Clin Cancer Res* 13 (2007) 482-487.
- [42]W.H. Roa, J.O. Kim, R. Razzak, H. Du, L. Guo, R. Singh, S. Gazala, S. Ghosh, E. Wong, A.A. Joy, J.Z. Xing, E.L. Bedard, Sputum microRNA profiling: a novel approach for the early detection of non-small cell lung cancer. *Clin Invest Med* 35 (2012) E271.
- [43]E.O. Machida, M.V. Brock, C.M. Hooker, J. Nakayama, A. Ishida, J. Amano, M.A. Picchi, S.A. Belinsky, J.G. Herman, S. Taniguchi, S.B. Baylin, Hypermethylation of ASC/TMS1 is a sputum marker for late-stage lung cancer. *Cancer Res* 66 (2006) 6210-6218.
- [44]G. Saccomanno, R.P. Saunders, V.E. Archer, O. Auerbach, M. Kuschner, P.A. Beckler, Cancer of the lung: the cytology of sputum prior to the development of carcinoma. *Acta Cytol* 9 (1965) 413-423.
- [45]Y. Su, M.A. Guarnera, H. Fang, F. Jiang, Small non-coding RNA biomarkers in sputum for lung cancer diagnosis. *Mol Cancer* 15 (2016) 36.
- [46]S.J. Swensen, M.D. Silverstein, D.M. Ilstrup, C.D. Schleck, E.S. Edell, The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med* 157 (1997) 849-855.
- [47]A. McWilliams, M.C. Tammemagi, J.R. Mayo, H. Roberts, G. Liu, K. Soghrati, K. Yasufuku, S. Martel, F. Laberge, M. Gingras, S. Atkar-Khattra, C.D. Berg, K. Evans, R. Finley, J. Yee, J. English, P. Nasute, J. Goffin, S. Puksa, L. Stewart, S. Tsai, M.R. Johnston, D. Manos, G. Nicholas, G.D. Goss, J.M. Seely, K. Amjadi, A. Tremblay, P. Burrowes, P. MacEachern, R. Bhatia, M.S. Tsao, S. Lam, Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 369 (2013) 910-919.
- [48]J.A. Hanley, B.J. McNeil, A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148 (1983) 839-843.
- [49]J. Luo, C. Xiong, Youden index and Associated Cut-points for Three Ordinal Diagnostic Groups. *Commun Stat Simul Comput* 42 (2013) 1213-1234.
- [50]E.F. Schisterman, D. Faraggi, B. Reiser, J. Hu, Youden Index and the optimal threshold for markers with mass at zero. *Stat Med* 27 (2008) 297-315.

- [51]S. Baek, C.A. Tsai, J.J. Chen, Development of biomarker classifiers from high-dimensional data. *Brief Bioinform* 10 (2009) 537-546.
- [52]J. Shen, Z. Liu, N.W. Todd, H. Zhang, J. Liao, L. Yu, M.A. Guarnera, R. Li, L. Cai, M. Zhan, F. Jiang, Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microRNA biomarkers. *BMC Cancer* 11 (2011) 374.
- [53]Y. Cui, F. Zhang, C. Zhu, L. Geng, T. Tian, H. Liu, Upregulated lncRNA SNHG1 contributes to progression of non-small cell lung cancer through inhibition of miR-101-3p and activation of Wnt/beta-catenin signaling pathway. *Oncotarget* 8 (2017) 17785-17794.
- [54]Q. Lu, S. Shan, Y. Li, D. Zhu, W. Jin, T. Ren, Long noncoding RNA SNHG1 promotes non-small cell lung cancer progression by up-regulating MTDH via sponging miR-145-5p. *FASEB J* (2018) fj201701237RR.
- [55]H.Y. Zhang, W. Yang, F.S. Zheng, Y.B. Wang, J.B. Lu, Long non-coding RNA SNHG1 regulates zinc finger E-box binding homeobox 1 expression by interacting with TAp63 and promotes cell metastasis and invasion in Lung squamous cell carcinoma. *Biomed Pharmacother* 90 (2017) 650-658.
- [56]I. Matouk, E. Raveh, P. Ohana, R.A. Lail, E. Gershtain, M. Gilon, N. De Groot, A. Czerniak, A. Hochberg, The increasing complexity of the oncofetal h19 gene locus: functional dissection and therapeutic intervention. *Int J Mol Sci* 14 (2013) 4298-4316.
- [57]L. Liu, S. Lu, lncRNA H19 promotes viability and epithelial-mesenchymal transition of lung adenocarcinoma cells by targeting miR-29b-3p and modifying STAT3. *Int J Oncol* 54 (2019) 929-941.
- [58]E. Raveh, I.J. Matouk, M. Gilon, A. Hochberg, The H19 Long non-coding RNA in cancer initiation, progression and metastasis - a proposed unifying theory. *Mol Cancer* 14 (2015) 184.
- [59]Y. Lei, W. Guo, B. Chen, L. Chen, J. Gong, W. Li, Tumorreleased lncRNA H19 promotes gefitinib resistance via packaging into exosomes in nonsmall cell lung cancer. *Oncol Rep* 40 (2018) 3438-3446.
- [60]G. Loewen, J. Jayawickramarajah, Y. Zhuo, B. Shan, Functions of lncRNA HOTAIR in lung cancer. *J Hematol Oncol* 7 (2014) 90.
- [61]S.S. Chen, M. Peng, G.Z. Zhou, Y.C. Pu, M.C. Yi, Y. Zhu, B. Jiang, Long non-coding RNA HOTAIR regulates the development of non-small cell lung cancer through miR-217/DACH1 signaling pathway. *Eur Rev Med Pharmacol Sci* 23 (2019) 670-678.
- [62]M.M. Wei, Y.C. Zhou, Z.S. Wen, B. Zhou, Y.C. Huang, G.Z. Wang, X.C. Zhao, H.L. Pan, L.W. Qu, J. Zhang, C. Zhang, X. Cheng, G.B. Zhou, Long non-coding RNA stabilizes the Y-box-binding protein 1 and regulates the epidermal growth factor receptor to promote lung carcinogenesis. *Oncotarget* 7 (2016) 59556-59571.
- [63]L. Shen, L. Chen, Y. Wang, X. Jiang, H. Xia, Z. Zhuang, Long noncoding RNA MALAT1 promotes brain metastasis by inducing epithelial-mesenchymal transition in lung cancer. *J Neurooncol* 121 (2015) 101-108.
- [64]P. Li, J. Li, R. Yang, F. Zhang, H. Wang, H. Chu, Y. Lu, S. Dun, Y. Wang, W. Zang, Y. Du, X. Chen, G. Zhao, G. Zhang, Study on expression of lncRNA RGMB-AS1 and repulsive guidance molecule b in non-small cell lung cancer. *Diagn Pathol* 10 (2015) 63.
- [65]Y.R. Yang, S.Z. Zang, C.L. Zhong, Y.X. Li, S.S. Zhao, X.J. Feng, Increased expression of the lncRNA PVT1 promotes tumorigenesis in non-small cell lung cancer. *Int J Clin Exp Pathol* 7 (2014) 6929-6935.
- [66]E.J. Whiteside, I. Seim, J.P. Pauli, A.J. O'Keeffe, P.B. Thomas, S.L. Carter, C.M. Walpole, J.N. Fung, P. Josh, A.C. Herington, L.K. Chopin, Identification of a long non-coding RNA gene, growth hormone secretagogue receptor opposite strand, which stimulates cell migration in non-small cell lung cancer cell lines. *Int J Oncol* 43 (2013) 566-574.

- [67]T. Hu, Y.R. Lu, BCYRN1, a c-MYC-activated long non-coding RNA, regulates cell metastasis of non-small-cell lung cancer. *Cancer Cell Int* 15 (2015) 36.
- [68]J. Li, P. Li, W. Zhao, R. Yang, S. Chen, Y. Bai, S. Dun, X. Chen, Y. Du, Y. Wang, W. Zang, G. Zhao, G. Zhang, Expression of long non-coding RNA DLX6-AS1 in lung adenocarcinoma. *Cancer Cell Int* 15 (2015) 48.
- [69]Z. Zeng, H. Bo, Z. Gong, Y. Lian, X. Li, W. Zhang, H. Deng, M. Zhou, S. Peng, G. Li, W. Xiong, AFAP1-AS1, a long noncoding RNA upregulated in lung cancer and promotes invasion and metastasis. *Tumour Biol* 37 (2016) 729-737.
- [70]Z. Hou, W. Zhao, J. Zhou, L. Shen, P. Zhan, C. Xu, C. Chang, H. Bi, J. Zou, X. Yao, R. Huang, L. Yu, J. Yan, A long noncoding RNA Sox2ot regulates lung cancer cell proliferation and is a prognostic indicator of poor survival. *Int J Biochem Cell Biol* 53 (2014) 380-388.
- [71]J. Luo, L. Tang, J. Zhang, J. Ni, H.P. Zhang, L. Zhang, J.F. Xu, D. Zheng, Long non-coding RNA CARLo-5 is a negative prognostic factor and exhibits tumor pro-oncogenic activity in non-small cell lung cancer. *Tumour Biol* 35 (2014) 11541-11549.
- [72]H. Luo, Y. Sun, G. Wei, J. Luo, X. Yang, W. Liu, M. Guo, R. Chen, Functional Characterization of Long Noncoding RNA Lnc_bc060912 in Human Lung Carcinoma Cells. *Biochemistry* 54 (2015) 2895-2902.
- [73]Y. Wu, H. Liu, X. Shi, Y. Yao, W. Yang, Y. Song, The long non-coding RNA HNF1A-AS1 regulates proliferation and metastasis in lung adenocarcinoma. *Oncotarget* 6 (2015) 9160-9172.
- [74]M. Qiu, Y. Xu, X. Yang, J. Wang, J. Hu, L. Xu, R. Yin, CCAT2 is a lung adenocarcinoma-specific long non-coding RNA and promotes invasion of non-small cell lung cancer. *Tumour Biol* 35 (2014) 5375-5380.
- [75]M. Qiu, Y. Xu, J. Wang, E. Zhang, M. Sun, Y. Zheng, M. Li, W. Xia, D. Feng, R. Yin, L. Xu, A novel lncRNA, LUADT1, promotes lung adenocarcinoma proliferation via the epigenetic suppression of p27. *Cell Death Dis* 6 (2015) e1858.
- [76]L. Zhang, X.F. Zhou, G.F. Pan, J.P. Zhao, Enhanced expression of long non-coding RNA ZXF1 promoted the invasion and metastasis in lung adenocarcinoma. *Biomed Pharmacother* 68 (2014) 401-407.
- [77]F.Q. Nie, M. Sun, J.S. Yang, M. Xie, T.P. Xu, R. Xia, Y.W. Liu, X.H. Liu, E.B. Zhang, K.H. Lu, Y.Q. Shu, Long noncoding RNA ANRIL promotes non-small cell lung cancer cell proliferation and inhibits apoptosis by silencing KLF2 and P21 expression. *Mol Cancer Ther* 14 (2015) 268-277.
- [78]R. Yang, P. Li, G. Zhang, C. Lu, H. Wang, G. Zhao, Long Non-Coding RNA XLOC_008466 Functions as an Oncogene in Human Non-Small Cell Lung Cancer by Targeting miR-874. *Cell Physiol Biochem* 42 (2017) 126-136.
- [79]H. Sang, H. Liu, P. Xiong, M. Zhu, Long non-coding RNA functions in lung cancer. *Tumour Biol* 36 (2015) 4027-4037.
- [80]X. Shi, M. Sun, H. Liu, Y. Yao, R. Kong, F. Chen, Y. Song, A critical role for the long non-coding RNA GAS5 in proliferation and apoptosis in non-small-cell lung cancer. *Mol Carcinog* 54 Suppl 1 (2015) E1-E12.
- [81]L. Han, R. Kong, D.D. Yin, E.B. Zhang, T.P. Xu, W. De, Y.Q. Shu, Low expression of long noncoding RNA GAS6-AS1 predicts a poor prognosis in patients with NSCLC. *Med Oncol* 30 (2013) 694.
- [82]L. Han, E.B. Zhang, D.D. Yin, R. Kong, T.P. Xu, W.M. Chen, R. Xia, Y.Q. Shu, W. De, Low expression of long noncoding RNA PANDAR predicts a poor prognosis of non-small cell lung cancer and affects cell apoptosis by regulating Bcl-2. *Cell Death Dis* 6 (2015) e1665.
- [83]X. Xie, H.T. Liu, J. Mei, F.B. Ding, H.B. Xiao, F.Q. Hu, R. Hu, M.S. Wang, LncRNA HMLincRNA717 is down-regulated in non-small cell lung cancer and associated with poor prognosis. *Int J Clin Exp Pathol* 7 (2014) 8881-8886.

[84]J. Liu, L. Wan, K. Lu, M. Sun, X. Pan, P. Zhang, B. Lu, G. Liu, Z. Wang, The Long Noncoding RNA MEG3 Contributes to Cisplatin Resistance of Human Lung Adenocarcinoma. PLoS One 10 (2015) e0114586.

[85]M. Sun, X.H. Liu, K.M. Wang, F.Q. Nie, R. Kong, J.S. Yang, R. Xia, T.P. Xu, F.Y. Jin, Z.J. Liu, J.F. Chen, E.B. Zhang, W. De, Z.X. Wang, Downregulation of BRAF activated non-coding RNA is associated with poor prognosis for non-small cell lung cancer and promotes metastasis by affecting epithelial-mesenchymal transition. Mol Cancer 13 (2014) 68.

[86]Y. Yang, H. Li, S. Hou, B. Hu, J. Liu, J. Wang, The noncoding RNA expression profile and the effect of lncRNA AK126698 on cisplatin resistance in non-small-cell lung cancer cell. PLoS One 8 (2013) e65309.

Fucosylation genes as circulating biomarkers for lung cancer

Qixin Leng¹, Jen-Hui Tsou¹, and Min Zhan², Feng Jiang^{1*}

1. Department of Pathology, University of Maryland School of Medicine, 10 S. Pine St.
Baltimore, MD 21201, USA

2. Departments of Epidemiology & Public Health, University of Maryland School of Medicine,
660 W. Redwood St. Baltimore, MD 21201, USA

*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.

Short running title: Fucosylation genes as biomarkers for lung cancer

Abstract

Purpose

Fucosyltransferases (FUTs) catalyze fucosylation, which plays a central role in biological processes. Aberrant fucosylation is associated with malignant transformation. Here we investigated whether transcriptional levels of genes coding the FUTs in plasma could provide cell-free circulating biomarkers for lung cancer.

Methods

mRNA expression of all 13 *Futs* (*Fut1-11*, *Pofut1*, and *Pofut2*) was evaluated by PCR assay in 48 lung tumor tissues and the 48 matched noncancerous lung tissues, and plasma of 64 lung cancer patients and 32 cancer-free individuals to develop plasma *Fut* biomarkers. The developed plasma *Fut* biomarkers were validated in an independent cohort of 40 lung cancer patients and 20 controls for their diagnostic performance.

Results

Four of the 13 *Futs* showed a different transcriptional level in 48 lung tumor tissues compared with the 48 matched nonconscious tissues (All <0.05). Two (*Fut8*, and *Pofut1*) of the four *Futs* had a higher plasma level in 64 lung cancer patients compared with 32 control subjects, and consistent with that in lung tissue specimens. Combined analysis of the two *Futs* produced 81% sensitivity and 86% specificity for diagnosis of lung cancer, and was independent of stage and histology of lung tumors. The diagnostic performance of the two plasma biomarkers was successfully validated in the different cohort of 40 lung cancer patients and 20 control individuals.

Conclusion

The fucosylation genes may provide new circulating biomarkers for the early detection of lung cancer.

Keywords: diagnosis; lung cancer; plasma; fucosyltransferases; biomarkers.

1. Introduction

Lung cancer is the leading cause of cancer death among both men and women. More people die of lung cancer than of colon, breast, and prostate cancers combined [1; 2]. Non-small cell lung cancer (NSCLC) accounts for more than 85% of all lung cancers. NSCLC consists of adenocarcinoma (AC), squamous cell carcinoma (SCC), and large cell carcinoma (LC). Tobacco smoking is the major cause of NSCLC. Results of the National Lung Screening Trial (NLST) shows that the early detection of lung cancer using low-dose CT (LDCT) can reduce the mortality[3] [4]. However, LDCT has a specificity of only 61%, producing a high false positive rate or over-diagnosis [5]. As a result, many smokers with benign diseases will have harmful diagnostic and therapeutic procedures. Therefore, the lack of noninvasive approaches for accurately diagnosing early stage lung cancer remains a top clinical challenge.

During tumor development, cancer cells undergo apoptosis and necrosis, and release tumor-associated molecules that can circulate in bloodstream. The tumors-derived molecules in plasma may provide cell-free circulating cancer biomarkers. Although numerous plasma biomarkers have been developed by detecting the circulating cell-free DNA, gene methylated products, proteins, and metabolites for lung cancer early detection, none of them has been well accepted in the clinics [6]. For instance, a blood test (Cancer-SEEK) was recently developed that could detect eight common cancer types by determining circulating proteins and mutations of cell-free DNA [7]. However, the test had about 60% sensitivity for diagnosis of all stages of lung cancer and only 40% sensitivity for the stage I disease[7].

Glycosylation is one of the most abundant protein modifications, and involved in major physiological events, including cell differentiation, proliferation, trafficking, migration and intercellular signaling. Gathering evidences have demonstrated that aberrant glycosylation is the result of alterations in glycosyltransferases that play crucial in the development and

progression of carcinogenesis [8]. Fucosylation is the major type of glycosylation and regulated by fucosyltransferases (FUTs), which catalyze the transfer of the fucose residue from GDP-fucose donor substrate to acceptor substrates present on oligosaccharides, glycoproteins and glycolipids [9]. There are 13 different FUTs, including FUT1 to 11, protein O-fucosyltransferase 1 (POFUT1), and POFUT2. [10; 11]. Abnormal protein expression of FUTs has been proven to associate the development and progression of malignancies, including lung cancer [10]. Furthermore, studying glycans and glycan-binding proteins has shown that changes in fucosylation of glycoproteins were potential cancer biomarkers. For example, increased fucosylated alpha-fetoprotein (AFP) level is observed in sera of patients with hepatocellular carcinoma. AFP is one of the most representative types of glycan-related cancer biomarkers [12; 13]. However, AFP is also elevated in sera of patients with various benign diseases such as accurate and hepatitis [14]. Therefore, the fucosylated glycoproteins-based circulating biomarkers exhibit an insufficient value in the diagnosis of malignancies. In addition, the current proteomic techniques for analysis of the plasma glycobiological biomarkers are labor intensive, expensive, and time consuming.

Since the elevated activity of the 13 FUTs is regulated by their own increased gene expression, we hypothesize that mRNA expression levels of the *Fut* genes in plasma may provide circulating biomarkers for lung cancer. In this study, we first evaluated transcriptional levels of the 13 *Futs* in 48 lung tumor and 48 matched nonconscious lung tissues. We determined the mRNA expression levels in plasma of 64 lung cancer and 32 control subjects and developed a panel of plasma biomarkers, which were validated in additional 40 lung cancer and 20 control subjects. A panel of plasma biomarkers for lung cancer early detection was successfully developed.

2. Materials and Methods

2.1. Patients and clinical specimens. This study was approved by the Institutional Review Boards of University of Maryland Baltimore and Veterans Affairs Maryland Health Care System. Surgically resected tissue specimens were obtained from 48 lung cancer patients who had either a lobectomy or a pneumonectomy. Tumor tissues were intraoperatively dissected from the surrounding lung parenchyma. The paired 48 normal lung tissues were also obtained from the same patients at an area distant from their tumors. Serial cryostat sections from the specimens were prepared and used to confirm the diagnosis based on the WHO classification of tumors of the lung ¹⁰. All 48 cases were diagnosed with histologically confirmed stage I NSCLC, including 25 AC and 23 SCC.

To collect blood samples, we enrolled lung cancer patients and cancer-free smokers by using the inclusion and/or exclusion criteria recommended by U.S. Preventive Services Task Force for lung cancer screening in heavy smokers[15]. We collected blood and prepared plasma using the standard operating protocols (SOPs) developed by The NCI-Early Detection Research Network (EDRN)[16; 17]. Briefly, 10 ml whole blood was collected by venous puncture with butterfly needles into BD Vacutainer spray-coated K2EDTA Tubes (BD, Franklin Lakes, NJ). All the blood samples were processed within 2 hours of collection. To obtain plasma, whole blood samples were centrifuged at 3,000 rpm at room temperature for 15 minutes. The supernatant was immediately removed from the buffy coat, and frozen in 1.8 ml RNase-free cryo-tubes (Thermo Scientific, Rochester, NY), and stored at -80°C until further processing. A total of 104 NSCLC patients and 52 cancer-free smokers were recruited. Among the cancer patients, 26 patients were female and 78 were male. Twenty-five had stage I NSCLC, 17 with stage II, 28 with stage III, 29 with stage IV, and 5 with unknown stage. Of the cancer-free smokers, 13 patients were female and 39 were male. There were no significant differences of age, gender and smoking status between the NSCLC patients and cancer-free

smokers. In this study, the cases and controls were randomly grouped into two cohorts: a development cohort and a validation cohort. The development cohort consisted of 64 lung cancer patients and 32 cancer-free smokers, while the validation cohort comprised 40 lung cancer patients and 20 cancer-free smokers. The demographic and clinical variables of the two sets are shown in Tables 1-2.

2.2. RNA isolation and quantitative reverse transcriptase PCR (qRT-PCR).

RNA was extracted from the specimens by using Trizol LS reagent (Invitrogen Carlsbad, CA) and RNeasy Mini Kit (Qiagen, Hilden, Germany) [18; 19]. RT was carried out to generate cDNA by using a RT Kit (Applied Biosystems, Foster City, CA) as described in our published works [18]. PCR was performed to measure expressions of target genes by using a PCR kit (Applied Biosystems) on a Bio-Red IQ5 Multicolor RT-PCR Detection System (Bio-Red, Hercules, CA). Primers and probes of the targeted 13 *FUTs* genes are shown in the supplementary table 1. Expression levels of the genes were determined using comparative cycle threshold (Ct) method with the equation $2^{-\Delta\Delta Ct}$ by using miR-1228 as an internal control [20]. The targeted genes with Ct values > 35 were considered to be below the detection level of qRT-PCR.

2.3. Droplet Digital PCR (ddPCR).

ddPCR for analysis of expression level of the genes was performed as described in our previous work [21; 22]. Briefly, TaqMan™ reaction mix (Applied Biosystems) containing sample cDNA was partitioned into aqueous droplets in oil via the QX100 Droplet Generator (Bio-Rad), and then transferred to a 96-well PCR plate. A two-step thermocycling protocol (95°C × 10 min; 40 cycles of [94°C × 30 s, 60°C × 60 s], 98°C × 10 min) was undertaken in a Bio-Rad C1000 (Bio-Rad). The PCR plate was loaded on Droplet Reader (Bio-Rad), by which copy number of each gene per µl PCR reaction was directly determined. We used QuantaSoft 1.7.4 analysis software (Bio-Rad) and Poisson

statistics to compute droplet concentrations (copies/ μ L). All assays were done in triplicates, and one no-template control and two interplate controls were carried along in each experiment.

2.4. Statistical analysis. Pearson's correlation analysis was applied to assess relationship between gene expressions and demographic and clinical characteristics of the lung cancer patients and control individuals. The area under receiver operating characteristic (ROC) curve (AUC) analyses were used to determine sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and corresponding cut-off value of each gene [23]. All P values shown were two sided, and a P value of <0.05 was considered statistically significant.

3. Results

3.1. Identifying *Futs* whose abnormal transcriptional levels were associated with lung cancer. The transcription levels of all 13 *Futs* were examined by using qRT-PCR in 48 stage I NSCLC tissues and the matched noncancerous lung tissues. Three (*Futs-7 and 8*, and *pofut1*) displayed a higher, whereas one gen (*Fut-4*) exhibited a lower mRNA expression levels in lung cancer tissues compared with the matched noncancerous lung tissues (All $p<0.05$) (Fig.1). There was no significant difference of the expression of the *Futs* between histological types of lung cancer (All $p>0.05$).

3.2. Developing a panel of plasma *Fut* biomarkers for lung cancer early detection. Transcriptional levels of the 4 *Futs* were first determined in plasma by qRT-PCR in a training set of 64 cases and 32 controls. The amplification curves of the RT-PCR analysis for the genes were not reliably generated by qRT-PCR. In contrast, when ddPCR was used to determine the level of the 4 *Futs* (*Futs-7 and 8*, *pofut1*, and *Fut-4*) in the plasma samples,

each well of the samples contained at least 10,000 droplets. Thus, ddPCR rather than qRT-PCR could quantify the *Fut* genes in plasma.

Of the 4 genes, *Fut8* and *Pofut1* had a high expression level in plasma of lung cancer patients vs. cancer-free controls (All $P < 0.05$) (Fig. 2A). There were significant correlations between the expression levels of the 2 genes in plasma and those in the surgical tissue specimens (All $r \geq 0.92$, all $P \leq 0.05$). Other 2 genes (*Fut4* and *Fut7*) did not show changes in plasma of lung cancer patients vs. controls. Furthermore, *Fut8* and *Pofut1* exhibited AUC values of 0.86 and 0.81, respectively, in distinguishing NSCLC patients from the healthy individuals (Fig. 2B). When optimal cutoffs for the two *FUT*s were set at 1.56 and 1.64, respectively, the individual genes produced 62.50-73.44% sensitivities, 90.63-90.88% specificities, 93.02-94.00% PPVs, and 63.04-71.88% NPVs (Table 3). Combined use of the two genes based on at least one positive result in either *Fut8* or *Pofut1* produced the highest classification accuracy (0.833) compared to any one used alone (all $p < 0.05$) (Table 3). The 2 genes used in combination created 81.76% sensitivity, 86.26% specificity, 92.86% PPV, and 70% NPV for diagnosis of lung cancer, thus significantly improving the cancer detection rate by a single gene with only a 4% decrease in specificity (Table 3). Furthermore, the estimated correlation among the 2 genes was very low ($r=0.187$, $p=0.06$), further supporting that the combined analysis of the 2 *FUT* genes outperformed a single one for diagnosis of NSCLC. In addition, combined analysis of the 2 plasma biomarkers did not show special association with stage and histological type of lung cancer, and patients' age, gender, and smoking status (All $P > 0.05$).

3.3. Validating the sputum markers in an independent set of lung cancer patients and controls. To evaluate the diagnostic performance of the plasma biomarker panel, the 2 genes (*Fut8* and *Pofut1*) were assessed in plasma samples of 40 NSCLC patients

and 20 healthy controls. The 2 genes used in combination could differentiate the NSCLC patients from healthy controls with 82.50% sensitivity and 85.00% specificity. Furthermore, no statistically significant difference was found in the sensitivity and specificity of the markers for stages and histological types of NSCLC (All $p > 0.05$). Moreover, there was no association of expressions of the 2 genes with the age, gender, or smoking status of the lung cancer patients and normal individuals (All $p > 0.05$). Taken together, the results confirm that the panel of 2 plasma biomarkers could be used for the early detection of lung cancer.

4. Discussion

In this present study, we have for the first time demonstrated that mRNA expression of *Futs* could provide cell-free circulating biomarkers for lung cancer. Furthermore, using ddPCR assay, a sensitive and robust technique, we successfully developed 2 *Futs* as small panel biomarkers for effective diagnosis of lung cancer. In addition, the diagnostic performance is independent of stage and histological type of the NSCLC, and age and gender of subjects. Therefore, the plasma biomarkers have an important characteristic if it is employed for more precisely and easily identifying early stage lung cancer.

Chen et al showed that *Fut8* was up-regulated in lung cancer and tissues [24]. A high protein expression of FUT8 was correlated with tumor metastasis, disease recurrence, and poor survival in patients with NSCLC [24; 25]. Downregulation of *Fut8* significantly inhibited the malignant behaviors of lung cancer cells. *Fut8* could globally modify surface antigens, receptors, and adhesion molecules [24]. Dysregulation of *Fut8* was involved in the regulation of dozens of genes associated with the malignancy through multiple mechanisms [24]. The observations in the present study provide further evidence that dysregulation of FUT8 has important role in lung tumorigenesis. Importantly, we demonstrate that a high mRNA

expression level of *Fut8* in plasma may provide useful biomarker for lung cancer early detection. POFUT1 is an important O-glycosyltransferase and has an essential role for extracellular Fringe to function [26; 27]. POFUT1 can active Notch1 in breast cancer [28]. Analysis of POFUT1 may have diagnostic or prognostic value in the patients with cancers [28; 29]. Ma et al. found that POFUT1 overexpression could prompt the binding of Notch ligand Dll1 to Notch1 receptor, and thus activated Notch1 signaling pathway in hepatocellular carcinoma cells [30]. Here we report an elevated mRNA level of *Pofut1* in both lung tumor tissues and plasma specimens of lung cancer patients, suggesting that the gene play an important biological function in lung carcinogenesis. Nevertheless, the possible biological role of aberrant expression of *Fut8* and *Pofut1* in lung cancer development and progression is warranted to be investigated.

There are some weaknesses in this study. i), we evaluate and develop the *Futs* as plasma biomarkers using a retrospective cases and controls, which may produce selection bias and overfitting. Our ongoing project is to prospectively validate the biomarkers in a large cohort of cases and controls. ii), based on the FDA criteria [31], a test directed at a disease with a prevalence of $\leq 5\%$ must have a sensitivity exceeding 95% when the specificity is less than or equal to 95%, and vice versa [31]. The overall sensitivity and specificity of the biomarkers are 81% and 86%, and thus not high enough for clinical diagnosis of NSCLC. The integration of the biomarkers with other classes of biomarkers, such as microRNAs [32; 33; 34; 35; 36; 37; 38; 39; 40; 41] or DNA mutations [42] or methylation [43], is one path to improve the early detection of lung cancer.

In summary, we developed a panel of plasma *Fut* biomarkers for lung cancer early detection. Future use of the biomarkers may provide a potential diagnostic test for the early

detection of lung cancer. Nevertheless, further exploring its value for routine use for the early detection of lung cancer would be required.

Legends

Figure 1. mRNA expression levels of 13 *Futs* in surgical lung tumor tissues and the paired noncancerous lung tissues. Four genes (*Futs-4*, *7* and *8*, and *pofut1*) shows a different level in lung tumor tissues as compared with the paired noncancerous lung tissues (All $p < 0.05$).

Figure 2. mRNA expression levels of *Fut8* and *Pofut1* in plasma samples of 64 lung cancer patients and 32 cancer-free controls. A, *Fut8* and *Pofut1* displayed a higher plasma level in lung cancer patients vs. cancer-free controls (all $p < 0.0001$). B, the receiver operating characteristic (ROC) curves of *Fut8* and *Pofut1* produced an area under the ROC curve (AUC) of 0.86 and 0.81, respectively, in diagnosis of lung cancer.

Table 1. Characteristics of 64 NSCLC patients and 32 cancer-free smokers in a training set.

Table 2. Characteristics of 40 NSCLC patients and 20 cancer-free smokers in a testing set.

Table 3. The comparison of diagnostic performance of one-gene and the two genes used in combination for the detection of lung cancer in a training set.

Supplementary table 1. Primers and probes of 13 *Futs*.

Compliance with Ethical Standards:

Disclosure of potential conflicts of interest: The authors declare no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical

standards. This article does not contain any studies with animals performed by any of the authors.

Funding: This work was supported in part by NCI R21CA205746, VA Merit Award Io1 CX000512, Award from the Geaton and JoAnn DeCesaris Family Foundation, and Maryland Innovation Initiative (MII) Commercialization Program- Phase 1/2 Grant (F.J.)

References:

- [1]American Cancer Society., Cancer facts and figures, in, American Cancer Society., New York,, pp. v.
- [2]X. Yao, Q. Mao, Y. Li, C. Hao, L. Bian, P. Chen, F. Gao, X. Wu, W. Lu, Q. Gao, X. Li, Z. Liang, Poorly neutralizing polyclonal antibody in vitro against coxsackievirus A16 circulating strains can prevent a lethal challenge in vivo. *Hum Vaccin Immunother* (2018) 0.
- [3]Cancer Facts & Figures 2012. American Cancer Society (ACS). *Journal of Consumer Health on the Internet* 16 (2012) 366-367.
- [4]D.R. Aberle, A.M. Adams, C.D. Berg, W.C. Black, J.D. Clapp, R.M. Fagerstrom, I.F. Gareen, C. Gatsonis, P.M. Marcus, J.D. Sicks, Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 365 (2011) 395-409.
- [5]E.F. Patz, Jr., P. Pinsky, C. Gatsonis, J.D. Sicks, B.S. Kramer, M.C. Tammemagi, C. Chiles, W.C. Black, D.R. Aberle, Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med* 174 (2014) 269-274.
- [6]A.J. Hubers, C.F. Prinsen, G. Sozzi, B.I. Witte, E. Thunnissen, Molecular sputum analysis for the diagnosis of lung cancer. *Br J Cancer* 109 (2013) 530-537.
- [7]J.D. Cohen, L. Li, Y. Wang, C. Thoburn, B. Afsari, L. Danilova, C. Douville, A.A. Javed, F. Wong, A. Mattox, R.H. Hruban, C.L. Wolfgang, M.G. Goggins, M. Dal Molin, T.L. Wang, R. Roden, A.P. Klein, J. Ptak, L. Dobbyn, J. Schaefer, N. Silliman, M. Popoli, J.T. Vogelstein, J.D. Browne, R.E. Schoen, R.E. Brand, J. Tie,

- P. Gibbs, H.L. Wong, A.S. Mansfield, J. Jen, S.M. Hanash, M. Falconi, P.J. Allen, S. Zhou, C. Bettegowda, L. Diaz, C. Tomasetti, K.W. Kinzler, B. Vogelstein, A.M. Lennon, N. Papadopoulos, Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* (2018).
- [8] V.B. Chachadi, G. Bhat, P.W. Cheng, Glycosyltransferases involved in the synthesis of MUC-associated metastasis-promoting selectin ligands. *Glycobiology* 25 (2015) 963-975.
- [9] L.H. Wu, B.Z. Shi, Q.L. Zhao, X.Z. Wu, Fucosylated glycan inhibition of human hepatocellular carcinoma cell migration through binding to chemokine receptors. *Glycobiology* 20 (2010) 215-223.
- [10] W. Zhou, H. Ma, G. Deng, L. Tang, J. Lu, X. Chen, Clinical significance and biological function of fucosyltransferase 2 in lung adenocarcinoma. *Oncotarget* 8 (2017) 97246-97259.
- [11] F.X. Sullivan, R. Kumar, R. Kriz, M. Stahl, G.Y. Xu, J. Rouse, X.J. Chang, A. Boodhoo, B. Potvin, D.A. Cumming, Molecular cloning of human GDP-mannose 4,6-dehydratase and reconstitution of GDP-fucose biosynthesis in vitro. *J Biol Chem* 273 (1998) 8193-8202.
- [12] J. Breborowicz, A. Mackiewicz, D. Breborowicz, Microheterogeneity of alpha-fetoprotein in patient serum as demonstrated by lectin affino-electrophoresis. *Scand J Immunol* 14 (1981) 15-20.
- [13] Y. Aoyagi, Y. Suzuki, K. Igarashi, A. Saitoh, M. Oguro, T. Yokota, S. Mori, M. Nomoto, M. Isemura, H. Asakura, The usefulness of simultaneous determinations of glucosaminylation and fucosylation indices of alpha-fetoprotein in the differential diagnosis of neoplastic diseases of the liver. *Cancer* 67 (1991) 2390-2394.
- [14] N. Taniguchi, Y. Kizuka, Glycans and cancer: role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics. *Adv Cancer Res* 126 (2015) 11-51.
- [15] L.L. Humphrey, M. Deffebach, M. Pappas, C. Baumann, K. Artis, J.P. Mitchell, B. Zakher, R. Fu, C.G. Slatore, Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. *Ann Intern Med* 159 (2013) 411-420.
- [16] J.R. Marks, K.S. Anderson, P. Engstrom, A.K. Godwin, L.J. Esserman, G. Longton, E.S. Iversen, A. Mathew, C. Patriotis, M.S. Pepe, Construction and analysis of the NCI-EDRN breast cancer reference set for circulating markers of disease. *Cancer Epidemiol Biomarkers Prev* 24 (2015) 435-441.

- [17] M.K. Tuck, D.W. Chan, D. Chia, A.K. Godwin, W.E. Grizzle, K.E. Krueger, W. Rom, M. Sanda, L. Sorbara, S. Stass, W. Wang, D.E. Brenner, Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. *J Proteome Res* 8 (2009) 113-117.
- [18] J. Ma, A. Jemal, R. Smith, Reply to lung cancer deaths averted by screening should be considered in the context of tobacco control policies. *Cancer* 119 (2013) 3420-3421.
- [19] J. Shen, Z. Liu, N.W. Todd, H. Zhang, J. Liao, L. Yu, M.A. Guarnera, R. Li, L. Cai, M. Zhan, F. Jiang, Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microRNA biomarkers. *BMC Cancer* 11 (2011) 374.
- [20] J. Ma, K. Mannoer, L. Gao, A. Tan, M.A. Guarnera, M. Zhan, A. Shetty, S.A. Stass, L. Xing, F. Jiang, Characterization of microRNA transcriptome in lung cancer by next-generation deep sequencing. *Mol Oncol* 8 (2014) 1208-1219.
- [21] N. Li, J. Ma, M.A. Guarnera, H. Fang, L. Cai, F. Jiang, Digital PCR quantification of miRNAs in sputum for diagnosis of lung cancer. *J Cancer Res Clin Oncol* 140 (2014) 145-150.
- [22] J. Ma, N. Li, M. Guarnera, F. Jiang, Quantification of Plasma miRNAs by Digital PCR for Cancer Diagnosis. *Biomark Insights* 8 (2013) 127-136.
- [23] L.E. Dodd, M.S. Pepe, Partial AUC estimation and regression. *Biometrics* 59 (2003) 614-623.
- [24] C.Y. Chen, Y.H. Jan, Y.H. Juan, C.J. Yang, M.S. Huang, C.J. Yu, P.C. Yang, M. Hsiao, T.L. Hsu, C.H. Wong, Fucosyltransferase 8 as a functional regulator of nonsmall cell lung cancer. *Proc Natl Acad Sci U S A* 110 (2013) 630-635.
- [25] R. Honma, I. Kinoshita, E. Miyoshi, U. Tomaru, Y. Matsuno, Y. Shimizu, S. Takeuchi, Y. Kobayashi, K. Kaga, N. Taniguchi, H. Dosaka-Akita, Expression of fucosyltransferase 8 is associated with an unfavorable clinical outcome in non-small cell lung cancers. *Oncology* 88 (2015) 298-308.
- [26] M. Stahl, K. Uemura, C. Ge, S. Shi, Y. Tashima, P. Stanley, Roles of Pofut1 and O-fucose in mammalian Notch signaling. *J Biol Chem* 283 (2008) 13638-13651.

- [27]C. Lorient, A. Audfray, F. Dupuy, A. Germot, A. Maftah, The two N-glycans present on bovine Pofut1 are differently involved in its solubility and activity. *FEBS J* 274 (2007) 1202-1211.
- [28]G. Wan, L. Tian, Y. Yu, F. Li, X. Wang, C. Li, S. Deng, X. Yu, X. Cai, Z. Zuo, F. Cao, Overexpression of Pofut1 and activated Notch1 may be associated with poor prognosis in breast cancer. *Biochem Biophys Res Commun* 491 (2017) 104-111.
- [29]S. Dong, Z. Wang, B. Huang, J. Zhang, Y. Ge, Q. Fan, Bioinformatics insight into glycosyltransferase gene expression in gastric cancer: POFUT1 is a potential biomarker. *Biochem Biophys Res Commun* 483 (2017) 171-177.
- [30]L. Ma, P. Dong, L. Liu, Q. Gao, M. Duan, S. Zhang, S. Chen, R. Xue, X. Wang, Overexpression of protein O-fucosyltransferase 1 accelerates hepatocellular carcinoma progression via the Notch signaling pathway. *Biochem Biophys Res Commun* 473 (2016) 503-510.
- [31]B.M. Nolen, A. Lomakin, A. Marrangoni, L. Velikokhatnaya, D. Prosser, A.E. Lokshin, Urinary protein biomarkers in the early detection of lung cancer. *Cancer Prev Res (Phila)* 8 (2015) 111-119.
- [32]J. Shen, J. Liao, M.A. Guarnera, H. Fang, L. Cai, S.A. Stass, F. Jiang, Analysis of MicroRNAs in sputum to improve computed tomography for lung cancer diagnosis. *J Thorac Oncol* 9 (2014) 33-40.
- [33]J. Su, J. Liao, L. Gao, J. Shen, M.A. Guarnera, M. Zhan, H. Fang, S.A. Stass, F. Jiang, Analysis of small nucleolar RNAs in sputum for lung cancer diagnosis. *Oncotarget* 7 (2016) 5131-5142.
- [34]L. Xing, J. Su, M.A. Guarnera, H. Zhang, L. Cai, R. Zhou, S.A. Stass, F. Jiang, Sputum microRNA biomarkers for identifying lung cancer in indeterminate solitary pulmonary nodules. *Clin Cancer Res* 21 (2015) 484-489.
- [35]N. Li, J. Ma, M.A. Guarnera, H. Fang, L. Cai, F. Jiang, Digital PCR quantification of miRNAs in sputum for diagnosis of lung cancer. *J Cancer Res Clin Oncol* 140 (2014) 145-150.
- [36]N. Anjuman, N. Li, M. Guarnera, S.A. Stass, F. Jiang, Evaluation of lung flut in sputum samples for molecular analysis of lung cancer. *Clin Transl Med* 2 (2013) 15.
- [37]L. Yu, N.W. Todd, L. Xing, Y. Xie, H. Zhang, Z. Liu, H. Fang, J. Zhang, R.L. Katz, F. Jiang, Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. *Int J Cancer* 127 (2010) 2870-2878.

- [38]F. Jiang, N.W. Todd, R. Li, H. Zhang, H. Fang, S.A. Stass, A panel of sputum-based genomic marker for early detection of lung cancer. *Cancer Prev Res (Phila)* 3 (2010) 1571-1578.
- [39]L. Xing, N.W. Todd, L. Yu, H. Fang, F. Jiang, Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers. *Mod Pathol* 23 (2010) 1157-1164.
- [40]Y. Xie, N.W. Todd, Z. Liu, M. Zhan, H. Fang, H. Peng, M. Alattar, J. Deepak, S.A. Stass, F. Jiang, Altered miRNA expression in sputum for diagnosis of non-small cell lung cancer. *Lung Cancer* 67 (2010) 170-176.
- [41]F. Jiang, N.W. Todd, Q. Qiu, Z. Liu, R.L. Katz, S.A. Stass, Combined genetic analysis of sputum and computed tomography for noninvasive diagnosis of non-small-cell lung cancer. *Lung Cancer* 66 (2009) 58-63.
- [42]A.J. Hubers, D.A. Heideman, Y. Yatabe, M.D. Wood, J. Tull, M. Taron, M.A. Molina, C. Mayo, J. Bertran-Alamillo, G.J. Herder, R. Koning, D. Sie, B. Ylstra, G.A. Meijer, P.J. Snijders, B.I. Witte, P.E. Postmus, E.F. Smit, E. Thunnissen, EGFR mutation analysis in sputum of lung cancer patients: a multitechnique study. *Lung Cancer* 82 (2013) 38-43.
- [43]Y. Su, H. Fang, F. Jiang, Integrating DNA methylation and microRNA biomarkers in sputum for lung cancer detection. *Clin Epigenetics* 8 (2016) 109.

A Plasma Long Noncoding RNA Signature for Early Detection of Lung Cancer



Yanli Lin^{*,†,1}, Qixin Leng^{*,1}, Min Zhan[‡] and Feng Jiang[†]

^{*}Department of Cell Engineering, Beijing Institute of Biotechnology, No. 20 Dongdajie Street, Fengtai District, Beijing 100071, China; [†]Department of Pathology, University of Maryland School of Medicine, 10 S. Pine St. Baltimore, MD 21201, USA; [‡]Departments of Epidemiology & Public Health, University of Maryland School of Medicine, 660 W. Redwood St. Baltimore, MD 21201, USA

Abstract

The early detection of lung cancer is a major clinical challenge. Long noncoding RNAs (lncRNAs) have important functions in tumorigenesis. Plasma lncRNAs directly released from primary tumors or the circulating cancer cells might provide cell-free cancer biomarkers. The objective of this study was to investigate whether the lncRNAs could be used as plasma biomarkers for early-stage lung cancer. By using droplet digital polymerase chain reaction, we determined the diagnostic performance of 26 lung cancer-associated lncRNAs in plasma of a development cohort of 63 lung cancer patients and 33 cancer-free individuals, and a validation cohort of 39 lung cancer patients and 28 controls. In the development cohort, 7 of the 26 lncRNAs were reliably measured in plasma. Two (SNHG1 and RMRP) displayed a considerably high plasma level in lung cancer patients vs. cancer-free controls (all $P < .001$). Combined use of the plasma lncRNAs as a biomarker signature produced 84.13% sensitivity and 87.88% specificity for diagnosis of lung cancer, independent of stage and histological type of lung tumor, and patients' age and sex (all $P > .05$). The diagnostic value of the plasma lncRNA signature for lung cancer early detection was confirmed in the validation cohort. The plasma lncRNA signature may provide a potential blood-based assay for diagnosing lung cancer at the early stage. Nevertheless, a prospective study is warranted to validate its clinical value.

Translational Oncology (2018) 11, 1225–1231

Introduction

Approximately 155,870 Americans will die from lung cancer each year, more than the other 3 leading cancers combined (breast, prostate, and colorectal cancers). Over 85% lung cancers are non-small cell lung cancers (NSCLCs). NSCLC mainly consists of adenocarcinoma (AC) and squamous cell carcinoma (SCC). Tobacco smoking is the major cause of NSCLC. The early detection of lung cancer in a large randomized trial using low-dose CT (LDCT) has revealed a 20% reduction in mortality as compared to chest X-rays [1]. Therefore, LDCT is recently recommended to be used for lung cancer early detection among smokers [2,3]. However, LDCT is associated with overdiagnosis, excessive cost, and radiation exposure [2,3]. The development of noninvasive or circulating biomarkers that can accurately and cost-effectively diagnose early-stage lung cancer is required [4].

Long noncoding RNAs (lncRNAs) have minimum transcript length of 200 bp and play vital roles in various biological processes [5,6]. lncRNAs can regulate different molecular signaling pathways

via changing gene expression, and therefore, their dysregulations are implicated in numerous mechanisms of carcinogenesis [7,8]. Dysregulation of some lncRNAs has been found in relation to oncogenesis and metastasis of lung tumor [9–11]. Importantly, plasma lncRNAs directly released from primary tumors or the circulating cancer cells might provide biomarkers for human

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

¹ The authors contributed equally to this work.

Received 14 May 2018; Revised 19 July 2018; Accepted 24 July 2018

© 2018 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/>). 1936-5233/18

<https://doi.org/10.1016/j.tranon.2018.07.016>

malignancies [12]. To date, several plasma lncRNAs have been identified that show the potential to distinguish lung cancer patients from noncancer subjects [12–20]. Yet none of them has been accepted in the clinical settings for lung cancer diagnosis, mainly due to the low sensitivity and specificity.

Recent studies have characterized 21 lncRNAs whose aberrations are associated with lung cancer [10,21–46]. Furthermore, using whole-genomic next-generation sequencing (NGS) to analyze ncRNA profile of primary lung tumor tissues, we recently identified additional five lung cancer–related lncRNAs [47,48]. These lung tumor–associated lncRNAs may provide a comprehensive list of biomarker candidates for developing circulating lung cancer biomarkers. The objective of this study was to investigate whether the lung cancer–associated lncRNAs could be used as plasma biomarkers for lung cancer.

Materials and Methods

Patients and Clinical Specimens

This study was approved by the Institutional Review Boards of University of Maryland Baltimore and Veterans Affairs Maryland Health Care System. We recruited lung cancer patients and cancer-free smokers by using the inclusion and/or exclusion criteria recommended by the US Preventive Services Task Force for lung cancer screening in heavy smokers [49]. We collected blood in BD Vacutainer spray-coated K2EDTA Tubes (BD, Franklin Lakes, NJ) and prepared plasma using the standard operating protocols developed by The NCI-Early Detection Research Network [50,51]. The specimens were processed within 2 hours of collection by centrifugation at 1300×g for 10 minutes at 4°C. A total of 102 NSCLC patients and 55 cancer-free smokers were recruited. Among the cancer patients, 24 patients were female and 78 were male. Twenty-three had stage I NSCLC, 18 with stage II, 28 with stage III, 28 with stage IV, and 5 with unknown stage. Of the cancer-free smokers, 14 patients were female and 41 were male. There were no significant differences of age, gender, and smoking status between the NSCLC patients and cancer-free smokers. The cases and controls were randomly grouped into two cohorts: a development cohort and a validation cohort. The development cohort consisted of 63 lung cancer patients and 33 cancer-free smokers, while the validation cohort comprised 39 lung cancer patients and 28 cancer-free smokers. The demographic and clinical variables of the two cohorts are shown in Tables 1 and 2.

Table 1. Characteristics of NSCLC Patients and Cancer-Free Smokers in a Development Cohort

	NSCLC Cases (n = 63)	Controls (n = 33)	P Value
Age	67.93 (SD 9.16)	63.79 (SD 16.12)	.18
Sex			.36
Female	15	8	
Male	48	25	
Smoking pack-years (median)	32.1	31.76	.19
Stage			
Stage I	14		
Stage II	10		
Stage III	17		
Stage IV	18		
Unknown	4		
Histological type			
Adenocarcinoma	32		
Squamous cell carcinoma	31		

RNA Isolation and Quantitative Reverse Transcriptase Polymerase Chain Reaction (qRT-PCR)

RNA was extracted from the specimens by using Trizol LS reagent (Invitrogen, Carlsbad, CA) and RNeasy Mini Kit (Qiagen, Hilden, Germany) [52–54]. RT was carried out to generate cDNA by using an RT Kit (Applied Biosystems, Foster City, CA) as described in our published works [52–54]. PCR was performed to measure expressions of target genes by using a PCR kit (Applied Biosystems) on a Bio-Rad IQ5 Multicolor RT-PCR Detection System (Bio-Rad, Hercules, CA). Expression levels of the genes were determined using comparative cycle threshold (C_T) method with miR-1228 as an internal control. The targeted genes with C_T values >35 were considered to be below the detection level of qRT-PCR [55].

Droplet Digital PCR (ddPCR)

ddPCR for analysis of the genes was performed as described in our previous work. Briefly, TaqMan reaction mix (Applied Biosystems) containing sample cDNA was partitioned into aqueous droplets in oil via the QX100 Droplet Generator (Bio-Rad) and then transferred to a 96-well PCR plate. A two-step thermocycling protocol (95°C × 10 minutes; 40 cycles of 94°C × 30 seconds and 60°C × 60 seconds, and 98°C × 10 minutes) was undertaken in a Bio-Rad C1000 (Bio-Rad). The PCR plate was loaded on Droplet Reader (Bio-Rad), by which copy number of each gene per µl PCR reaction was directly determined. We used QuantaSoft 1.7.4 analysis software (Bio-Rad) and Poisson statistics to compute droplet concentrations (copies/µl). Only genes that had at least 10,000 droplets were considered to be robustly detectable by ddPCR in plasma and subsequently underwent further analysis [56,57]. All assays were done in triplicates, and one no-template control and two interplate controls were carried along in each experiment.

Statistical Analysis

Pearson correlation analysis was applied to assess relationship between gene expressions and demographic and clinical characteristics of the lung cancer patients and control individuals. The area under receiver operating characteristic curve (AUC) analyses were used to determine sensitivity, specificity, and corresponding cutoff value of each gene [58]. All P values shown were two sided, and a P value of < .05 was considered statistically significant.

Table 2. Characteristics of NSCLC Patients and Cancer-Free Smokers in a Validation Cohort

	NSCLC Cases (n = 39)	Controls (n = 28)	P Value
Age	66.58 (SD 9.93)	63.68 (SD 13.27)	.25
Sex			.45
Female	9	6	
Male	30	22	
Smoking pack-years (median)	33.39	29.64	.26
Stage			
Stage I	9		
Stage II	8		
Stage III	11		
Stage IV	10		
Unknown	1		
Histological type			
Adenocarcinoma	22		
Squamous cell carcinoma	17		

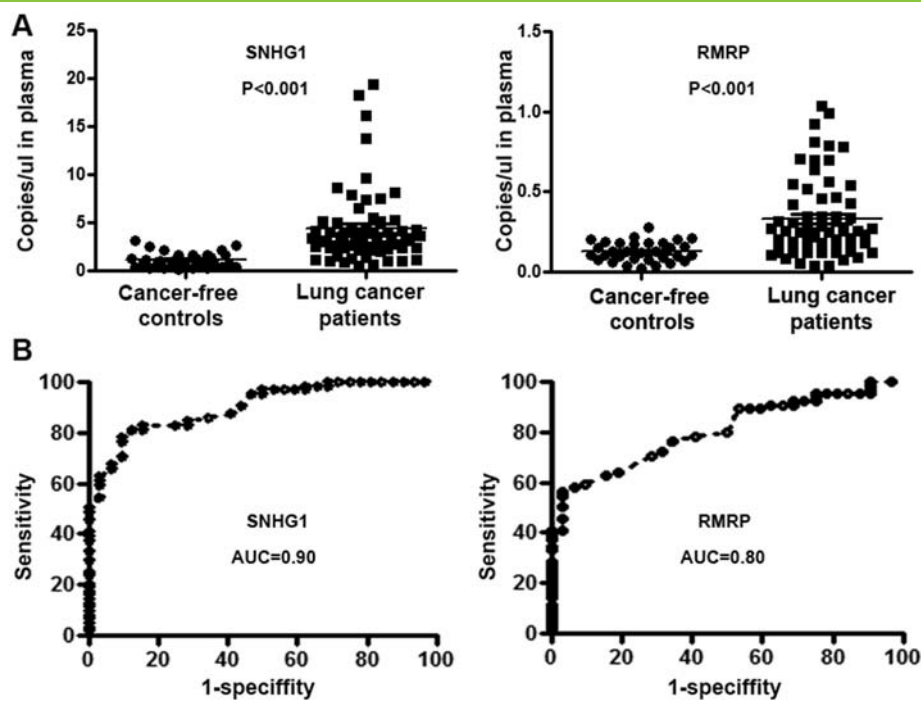


Figure 1. Expression levels of SNHG1 and RMRP in plasma samples of 63 lung cancer patients and 33 cancer-free controls. (A) SNHG1 and RMRP displayed a higher plasma level in lung cancer patients vs. cancer-free controls (all $P < .001$). (B) The receiver operating characteristic curves of SNHG1 and RMRP produced an AUC of 0.90 and 0.80, respectively, in diagnosis of lung cancer.

Results

Developing a Plasma lncRNA Signature for Lung Cancer Early Detection

We first measured expression levels of the 26 lncRNAs (Supplementary Table 1) in plasma by using qRT-PCR in a discovery cohort of 63 cases and 33 controls. The lncRNAs had a C_T value of ≥ 35 in 75% plasma samples. However, the internal control gene, miR-1228, stably displayed a C_T value of 20-22 across the plasma samples. The results suggested that the amplification curves for the lncRNAs were not reliably generated, and their expression levels in plasma were too low to be detectable by qRT-PCR. We have proven that ddPCR is a direct method for absolutely and quantitatively measuring ncRNAs with a higher sensitivity compared with qRT-PCR [56,57,59]. Therefore, we used ddPCR to determine expression level of the lncRNAs in the plasma samples. Seven (26.9%) of the 26 lncRNAs could generated at least 10,000 droplets in each well of the plasma samples. Therefore, the seven lncRNAs could be successfully “read” by ddPCR for the absolute quantification in the plasma samples. The seven genes are SNHG1, MALAT1, HOTAIR, H19, MEG3, MEG8, and RMRP.

Among the seven genes, SNHG1 and RMRP had a higher expression level in plasma of lung cancer patients versus cancer-free

controls (all $P \leq .05$) (Figure 1A). Conversely, other five lncRNAs did not display a different plasma level in lung cancer cases versus cancer-free controls (all $P \geq .05$). Furthermore, the plasma expression levels of the SNHG1 and RMRP were independent of stages and histological types of lung cancer. In addition, SNHG1 and RMRP exhibited AUC values of 0.90 and 0.80, respectively, in distinguishing NSCLC patients from healthy individuals (Figure 1B). Using Youden's index [60], we set up optimal cutoff for the two genes at 1.11 and 0.12, respectively. As a result, the use of the individual genes alone produced 61.00%-78.78% sensitivities and 87.88%-90.91% (Table 3). Combined use of the two genes based on at least one positive result in either SNHG1 or RMRP produced the highest classification accuracy (85.42%) compared to any one used alone (all $P < .05$) (Table 3). The two genes used in combination produced a sensitivity of 84.13% and a specificity of 87.88% for diagnosis of lung cancer, thus considerably improving the detection rate by a single gene with only a 2% decline in specificity (Table 3). Furthermore, the estimated correlation determined by Pearson correlation analysis among levels of the two lncRNAs was very low ($R^2 = 0.011$, $P = .53$), further supporting that the combined analysis of the two genes outperformed a single one. In addition, combined analysis of the two plasma biomarkers did not show special association with stage and histological type of lung cancer and patients' age, gender, and smoking status (all $P > .05$).

Table 3. Diagnostic Performance of One-Gene vs. a Plasma lncRNA Signature for Lung Cancer Diagnosis in a Development Cohort

	Accuracy	Sensitivity (95% CI)	Specificity (95% CI)
SNHG1	81.25% (72.00%-88.49%)	77.78% (65.54%-87.28%)	87.88% (71.80%-96.60%)
RMRP	71.88% (61.78%-80.58%)	61.90% (48.80%-73.85%)	90.91% (75.67%-98.08%)
A plasma lncRNA signature	85.42% (76.74%-91.79%)	84.13% (72.74%-92.12%)	87.50% (71.80%-96.60%)

Abbreviation: CI, confidence interval.

Table 4. Diagnostic Performance of One-Gene vs. a Plasma lncRNA Signature for Lung Cancer Diagnosis in a Validation Set

	Accuracy	Sensitivity (95% CI)	Specificity (95% CI)
<i>SNHG1</i>	80.33% (68.16%-89.40%)	76.92% (60.67%-88.87)	86.36% (65.09%-97.09%)
<i>RMRP</i>	72.13% (59.17%-82.85%)	61.54% (44.62%-76.64%)	90.91% (70.84%-98.88%)
A plasma lncRNA signature	83.62% (71.91%-91.85%)	82.05% (66.47%-92.46%)	86.36% (65.09%-97.09%)

Validating the Plasma lncRNA Marker Signature in an Independent Set of Lung Cancer Patients and Controls

To evaluate the diagnostic performance of the biomarker signature, the two lncRNAs (*SNHG1* and *RMRP*) were assessed by using ddPCR in plasma of additional 39 NSCLC patients and 28 healthy controls. The two genes used in combination could differentiate the NSCLC patients from healthy controls with 82.05% sensitivity and 83.36% specificity (Table 4). Furthermore, no statistically significant difference was found in the sensitivity and specificity of the biomarker signature for stages and histological types of NSCLC (all $P > .05$). Moreover, there was no association of expressions of the two genes with the age, gender, or smoking status of the lung cancer patients and normal individuals (all $P > .05$). Taken together, the results confirm the potential of combined use of the two lncRNAs as a plasma biomarker signature for the early detection of lung cancer.

Discussion

Circulating cell-free lncRNAs biomarkers show promise as biomarkers for cancer diagnosis. However, unlike other ncRNA (e.g., microRNAs), lncRNAs have the lowest levels in plasma among several different RNA species [61], presenting a major challenge in the development of cell-free lncRNAs as cancer biomarkers. For instance, Schlosser et al. recently demonstrated that expressions of lncRNAs were robustly detectable in tissues yet undetectable or sporadically measurable in the matched plasma by using qRT-PCR, a routine platform used for nucleic acid detection [61]. Regular qPCR has some limitations in determining expression of ncRNAs: i) It is an indirect and labor-consuming approach. ii) It requires an internal control gene for normalization. Yet none of the investigated genes has been accepted as a standard control. iii) Its sensitivity for a low copy number of genes is very low. Our current observations are consistent with Schlosser's finding [61]: of the 26 lung cancer-associated lncRNAs, none is reliably measurable in plasma using qRT-PCR. Therefore, conventional qPCR might not be an appropriate tool for the development of lncRNAs as circulating biomarkers given that circulating lncRNAs in body fluids are present in low abundance. We have shown that ddPCR is a direct method for sensitively measuring ncRNAs [56,59] since it depends on limiting partition of the PCR volume, where a positive result of a large number of microreactions indicates the presence of a single molecule in a given reaction [62]. The number of positive reactions, together with Poisson's distribution, produces a straight and high-confidence measurement of the original target concentration. Importantly, ddPCR does not require a reliance on rate-based measurements (C_T values), endogenous controls, and calibration curves and therefore overcome the obstacles linked to the regular qPCR in quantification of genes in plasma. Here we demonstrate that 7 of the 26 lung cancer-associated lncRNAs that are not detectable by qRT-PCR are robustly measurable by ddPCR in plasma. Therefore, ddPCR may address the limitations of the qPCR in quantification of lncRNAs in plasma and hence help develop the genes as cell-free cancer biomarkers.

The previous plasma lncRNA-based assays were mostly developed from the limited number of lung cancer-associated lncRNAs and

only consisted of a single lncRNA gene [12–20]. Since lung tumor is a heterogeneous group of neoplasms and develops from a multitude of molecular changes, a single lncRNA-based assay may not achieve the performance required to move forward for clinically detecting lung cancer. The development of a panel of multiple biomarkers by integrating analysis of multifaceted and diverse lncRNAs would provide a synergistic test for lung cancer diagnosis. By searching published data, we found 21 lncRNAs whose malfunction was well characterized in lung tumorigenesis [10,21–46]. Furthermore, by systematically and comprehensively defining ncRNA changes of NSCLC in surgical lung tumor tissues using whole-transcriptome NGS [47,48], we recently identified additional five lung cancer-associated lncRNAs [47]. Both the published and our NGS-defined lncRNAs of lung tumors may provide a comprehensive set of biomarker candidates for lung cancer. From the 26 lncRNAs, our present study identified and optimized a plasma signature consisting of two lncRNAs that created a higher diagnostic value for lung cancer detection than did individual lncRNAs used alone [12–20]. In addition, the diagnostic performance of the biomarkers was further blindly validated in a different cohort, suggesting that the plasma signature might be a robust assay for lung cancer diagnosis. Moreover, the performance of this plasma lncRNA signature for lung cancer diagnosis was independent of tumor stage and histology. This might be an important characteristic if the plasma lncRNA signature is employed for identifying early-stage lung cancer.

The two lncRNAs (*SNHG1* and *RMRP*) have diverse and important functions in lung tumorigenesis through regulating different molecular pathways. Elevated expression of *SNHG1* was frequently observed in lung cancer tissues and significantly correlated with larger tumor size, advanced stage, lymph node metastasis, and poor overall survival of the patients [63]. Furthermore, *SNHG1* could promote NSCLC progression of lung cancer via miR-101-3p/SOX9/Wnt/ β -catenin regulatory network and miR-145-5p/MTDH axis [63,64]. In addition, *SNHG1* plays an oncogenic role in lung squamous cell carcinoma through ZEB1 signaling pathway by inhibiting TAP63 [65]. *RMRP* is best known for being a component of the nuclear RNase MRP complex, which participates in the processing of ribosomal RNA to generate the short mature 5.8S rRNA [66] and cleaves B-cyclin mRNA, lowering B-cyclin levels during mitosis [67]. In addition, *RMRP* interacts with telomerase to form a complex with RNA-dependent RNA polymerase activity capable of synthesizing dsRNA precursors processed by DICER1 into siRNAs [68]. Moreover, *RMRP* is important for mitochondrial DNA replication and RNA processing [69]. Upregulation of *RMRP* is found in lung adenocarcinoma tissues [70]. *RMRP* might act as an oncogenic lncRNA to promote the expression of KRAS, FMNL2, and SOX9 by inhibiting miR-206 expression in lung cancer [70]. Our current study extends the previous findings by developing them as a biomarker signature that might be clinically useful in the early detection of lung cancer.

There are some limitations in this present study. 1) The size of the cohorts is small. Furthermore, the plasma samples were obtained from the hospital-based patients with clinical diagnosis. The participants might not be representative of high-risk populations (e.g., heavy

smokers) in screening setting for lung cancer. We will perform a prospective and multisite lung cancer screening trial to validate the diagnostic value of the plasma lncRNA signature. 2) The early diagnosis of lung cancer by using LDCT could reduce the mortality [3]. However, LDCT has a low specificity for the early detection of lung cancer, presenting a major clinical challenge [71]. We are evaluating whether the plasma lncRNA signature could improve the specificity of LDCT for the early detection of NSCLC by specifically distinguishing malignant from benign pulmonary growths. 3) The cell-free circulating tumor DNA, microRNAs, or DNA methylation provides other important types of biomarkers for lung cancer [54,72–75,76–78]. Our ongoing efforts are to investigate whether integrating the lncRNA signature with other types of biomarkers might improve the early detection of lung cancer. 4) In the current study, SNHG1 or RMRP expression level in plasma is independent of histological types of lung cancer. Furthermore, the combined use of the two genes is also not associated with histological types of lung cancer. The findings in plasma samples are inconsistent with the two previous studies in tissue specimens that showed that SNHG1 or RMRP was associated with SCC or AC, respectively [65,70]. A new study using paired plasma and tumor tissue specimens from the same lung cancer patients is needed to validate the discrepancy and further understand the mechanism underlying the divergence.

Conclusions

A plasma lncRNA signature was developed that could differentiate early-stage NSCLC patients from cancer-free smokers. Nevertheless, undertaking a prospective study to further validate this plasma biomarker signature in a large cohort is required.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.tranon.2018.07.016>.

Declarations

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

The Institutional Review Boards of University of Maryland Baltimore and Veterans Affairs Maryland Health Care System approved this study. Reference number is H40666.

Competing Interests

The authors declare no conflict of interest.

Funding

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, UMD-UMB Research and Innovation Seed Grant, DoD-Idea Development Award, and Maryland Innovation Initiative (MII) Commercialization Program-Phase 1/2 Grant (F.J.).

Authors' Contributions

Y. L., Q. L., M. Z., and F. J. conducted the experiments and participated in study design, coordination, data interpretation, and

preparing the manuscript. All authors read and approved the final manuscript.

Availability of Data and Materials

Not applicable.

Disclosure of Potential Conflicts of Interest

The authors declare no conflict of interest.

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] 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**(2), 269–274.
- [3] 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.
- [4] Hubers AJ, Prinsen CF, Sozzi G, Witte BI, and Thunnissen E (2013). Molecular sputum analysis for the diagnosis of lung cancer. *Br J Cancer* **109**(3), 530–537.
- [5] Ma L, Bajic VB, and Zhang Z (2013). On the classification of long non-coding RNAs. *RNA Biol* **10**, 925–933.
- [6] Spizzo R, Almeida MI, Colombatti A, and Calin GA (2012). Long non-coding RNAs and cancer: a new frontier of translational research? *Oncogene* **31**, 4577–4587.
- [7] Meseure D, Drak Alsbai K, Nicolas A, Bieche I, and Morillon A (2015). Long noncoding RNAs as new architects in cancer epigenetics, prognostic biomarkers, and potential therapeutic targets. *Biomed Res Int* **2015**, 320214.
- [8] Prensner JR and Chinnaiyan AM (2011). The emergence of lncRNAs in cancer biology. *Cancer Discov* **1**, 391–407.
- [9] Zhou M, Guo M, He D, Wang X, Cui Y, Yang H, Hao D, and Sun J (2015). A potential signature of eight long non-coding RNAs predicts survival in patients with non-small cell lung cancer. *J Transl Med* **13**, 231.
- [10] Li M, Qiu M, Xu Y, Mao Q, Wang J, Dong G, Xia W, Yin R, and Xu L (2015). Differentially expressed protein-coding genes and long noncoding RNA in early-stage lung cancer. *Tumour Biol* **36**, 9969–9978.
- [11] Schmidt LH, Spieker T, Koschmieder S, Schaffers S, Humberg J, Jungen D, Bulk E, Hascher A, Wittmer D, and Marra A, et al (2011). The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. *J Thorac Oncol* **6**, 1984–1992.
- [12] Liang W, Lv T, Shi X, Liu H, Zhu Q, Zeng J, Yang W, Yin J, and Song Y (2016). Circulating long noncoding RNA GAS5 is a novel biomarker for the diagnosis of nonsmall cell lung cancer. *Medicine (Baltimore)* **95**(37)e4608.
- [13] Tantai J, Hu D, Yang Y, and Geng J (2015). Combined identification of long non-coding RNA XIST and HIF1A-AS1 in serum as an effective screening for non-small cell lung cancer. *Int J Clin Exp Pathol* **8**, 7887–7895.
- [14] Li N, Feng XB, Tan Q, Luo P, Jing W, Zhu M, Liang C, Tu J, and Ning Y (2017). Identification of circulating long noncoding RNA Linc00152 as a novel biomarker for diagnosis and monitoring of non-small-cell lung cancer. *Dis Markers* **17**, 743–749.
- [15] Li N, Wang Y, Liu X, Luo P, Jing W, Zhu M, and Tu J (2017). Identification of circulating long noncoding RNA HOTAIR as a novel biomarker for diagnosis and monitoring of non-small cell lung cancer. *Technol Cancer Res Treat* **17**, 1533–1539.
- [16] Wan L, Zhang L, Fan K, and Wang JJ (2017). Diagnostic significance of circulating long noncoding RNA PCAT6 in patients with non-small cell lung cancer. *Oncotargets Ther* **10**, 5695–5702.
- [17] Tan Q, Zuo J, Qiu S, Yu Y, Zhou H, Li N, Wang H, Liang C, Yu M, and Tu J (2017). Identification of circulating long non-coding RNA GAS5 as a potential biomarker for non-small cell lung cancer diagnosis, long non-coding RNA, plasma, GAS5, biomarker. *Int J Oncol* **50**, 1729–1738.
- [18] Zhu Q, Lv T, Wu Y, Shi X, Liu H, and Song Y (2017). Long non-coding RNA 00312 regulated by HOXA5 inhibits tumour proliferation and promotes apoptosis in non-small cell lung cancer. *J Cell Mol Med* **21**, 2184–2198.
- [19] Zhu H, Zhang L, Yan S, Li W, Cui J, Zhu M, Xia N, Yang Y, Yuan J, and Chen X, et al (2017). LncRNA16 is a potential biomarker for diagnosis of early-stage

- lung cancer that promotes cell proliferation by regulating the cell cycle. *Oncotarget* **8**, 7867–7877.
- [20] Wang HM, Lu JH, Chen WY, and Gu AQ (2015). Upregulated lncRNA-UCA1 contributes to progression of lung cancer and is closely related to clinical diagnosis as a predictive biomarker in plasma. *Int J Clin Exp Med* **8**, 11824–11830.
- [21] Wei MM, Zhou YC, Wen ZS, Zhou B, Huang YC, Wang GZ, Zhao XC, Pan HL, Qu LW, and Zhang J, et al (2016). Long non-coding RNA stabilizes the Y-box-binding protein 1 and regulates the epidermal growth factor receptor to promote lung carcinogenesis. *Oncotarget* **7**, 59556–59571.
- [22] Shen L, Chen L, Wang Y, Jiang X, Xia H, and Zhuang Z (2015). Long noncoding RNA MALAT1 promotes brain metastasis by inducing epithelial-mesenchymal transition in lung cancer. *J Neurooncol* **121**, 101–108.
- [23] Loewen G, Jayawickramarajah J, Zhuo Y, and Shan B (2014). Functions of lncRNA HOTAIR in lung cancer. *J Hematol Oncol* **7**, 90–96.
- [24] Li P, Li J, Yang R, Zhang F, Wang H, Chu H, Lu Y, Dun S, Wang Y, and Zang W, et al (2015). Study on expression of lncRNA RGMB-AS1 and repulsive guidance molecule b in non-small cell lung cancer. *Diagn Pathol* **10**, 63–68.
- [25] Yang YR, Zang SZ, Zhong CL, Li YX, Zhao SS, and Feng XJ (2014). Increased expression of the lncRNA PVT1 promotes tumorigenesis in non-small cell lung cancer. *Int J Clin Exp Pathol* **7**, 6929–6935.
- [26] Whiteside EJ, Seim I, Pauli JP, O'Keeffe AJ, Thomas PB, Carter SL, Walpole CM, Fung JN, Josh P, and Herington AC, et al (2013). Identification of a long non-coding RNA gene, growth hormone secretagogue receptor opposite strand, which stimulates cell migration in non-small cell lung cancer cell lines. *Int J Oncol* **43**, 566–574.
- [27] Hu T and Lu YR (2015). BCYRN1, a c-MYC-activated long non-coding RNA, regulates cell metastasis of non-small-cell lung cancer. *Cancer Cell Int* **15**, 36–40.
- [28] Li J, Li P, Zhao W, Yang R, Chen S, Bai Y, Dun S, Chen X, Du Y, and Wang Y, et al (2015). Expression of long non-coding RNA DLX6-AS1 in lung adenocarcinoma. *Cancer Cell Int* **15**, 48–53.
- [29] Zeng Z, Bo H, Gong Z, Lian Y, Li X, Zhang W, Deng H, Zhou M, Peng S, and Li G, et al (2016). AFAP1-AS1, a long noncoding RNA upregulated in lung cancer and promotes invasion and metastasis. *Tumour Biol* **37**, 729–737.
- [30] Hou Z, Zhao W, Zhou J, Shen L, Zhan P, Xu C, Chang C, Bi H, Zou J, and Yao X, et al (2014). A long noncoding RNA Sox2ot regulates lung cancer cell proliferation and is a prognostic indicator of poor survival. *Int J Biochem Cell Biol* **53**, 380–388.
- [31] Luo J, Tang L, Zhang J, Ni J, Zhang HP, Zhang L, Xu JF, and Zheng D (2014). Long non-coding RNA CARLo-5 is a negative prognostic factor and exhibits tumor pro-oncogenic activity in non-small cell lung cancer. *Tumour Biol* **35**, 11541–11549.
- [32] Luo H, Sun Y, Wei G, Luo J, Yang X, Liu W, Guo M, and Chen R (2015). Functional characterization of long noncoding RNA Lnc_bc060912 in human lung carcinoma cells. *Biochemistry* **54**, 2895–2902.
- [33] Wu Y, Liu H, Shi X, Yao Y, Yang W, and Song Y (2015). The long non-coding RNA HNF1A-AS1 regulates proliferation and metastasis in lung adenocarcinoma. *Oncotarget* **6**, 9160–9172.
- [34] Qiu M, Xu Y, Yang X, Wang J, Hu J, Xu L, and Yin R (2014). CCAT2 is a lung adenocarcinoma-specific long non-coding RNA and promotes invasion of non-small cell lung cancer. *Tumour Biol* **35**, 5375–5380.
- [35] Qiu M, Xu Y, Wang J, Zhang E, Sun M, Zheng Y, Li M, Xia W, Feng D, and Yin R, et al (2015). A novel lncRNA, LUADT1, promotes lung adenocarcinoma proliferation via the epigenetic suppression of p27. *Cell Death Dis* **6**, 18–22.
- [36] Zhang L, Zhou XF, Pan GF, and Zhao JP (2014). Enhanced expression of long non-coding RNA ZXF1 promoted the invasion and metastasis in lung adenocarcinoma. *Biomed Pharmacother* **68**, 401–407.
- [37] Nie FQ, Sun M, Yang JS, Xie M, Xu TP, Xia R, Liu YW, Liu XH, Zhang EB, and Lu KH, et al (2015). Long noncoding RNA ANRIL promotes non-small cell lung cancer cell proliferation and inhibits apoptosis by silencing KLF2 and P21 expression. *Mol Cancer Ther* **14**, 268–277.
- [38] Yang R, Li P, Zhang G, Lu C, Wang H, and Zhao G (2017). Long non-coding RNA XLOC_008466 functions as an oncogene in human non-small cell lung cancer by targeting miR-874. *Cell Physiol Biochem* **42**, 126–136.
- [39] Sang H, Liu H, Xiong P, and Zhu M (2015). Long non-coding RNA functions in lung cancer. *Tumour Biol* **36**, 4027–4037.
- [40] Shi X, Sun M, Liu H, Yao Y, Kong R, Chen F, and Song Y (2015). A critical role for the long non-coding RNA GAS5 in proliferation and apoptosis in non-small-cell lung cancer. *Mol Carcinog* **54**, 1–12.
- [41] Han L, Kong R, Yin DD, Zhang EB, Xu TP, De W, and Shu YQ (2013). Low expression of long noncoding RNA GAS6-AS1 predicts a poor prognosis in patients with NSCLC. *Med Oncol* **30**, 694–699.
- [42] Han L, Zhang EB, Yin DD, Kong R, Xu TP, Chen WM, Xia R, Shu YQ, and De W (2015). Low expression of long noncoding RNA PANDAR predicts a poor prognosis of non-small cell lung cancer and affects cell apoptosis by regulating Bcl-2. *Cell Death Dis* **6**, 16–22.
- [43] Xie X, Liu HT, Mei J, Ding FB, Xiao HB, Hu FQ, Hu R, and Wang MS (2014). lncRNA HMLincRNA717 is down-regulated in non-small cell lung cancer and associated with poor prognosis. *Int J Clin Exp Pathol* **7**, 8881–8886.
- [44] Liu J, Wan L, Lu K, Sun M, Pan X, Zhang P, Lu B, Liu G, and Wang Z (2015). The long noncoding RNA MEG3 contributes to cisplatin resistance of human lung adenocarcinoma. *PLoS One* **10**, 11–17.
- [45] Sun M, Liu XH, Wang KM, Nie FQ, Kong R, Yang JS, Xia R, Xu TP, Jin FY, and Liu ZJ, et al (2014). Downregulation of BRAF activated non-coding RNA is associated with poor prognosis for non-small cell lung cancer and promotes metastasis by affecting epithelial-mesenchymal transition. *Mol Cancer* **13**, 68–72.
- [46] Yang Y, Li H, Hou S, Hu B, Liu J, and Wang J (2013). The noncoding RNA expression profile and the effect of lncRNA AK126698 on cisplatin resistance in non-small-cell lung cancer cell. *PLoS One* **8**, 65–60.
- [47] Ma J, Mannoor K, Gao L, Tan A, Guarnera MA, Zhan M, Shetty A, Stass SA, Xing L, and Jiang F (2014). Characterization of microRNA transcriptome in lung cancer by next-generation deep sequencing. *Mol Oncol* **8**, 1208–1219.
- [48] Gao L, Ma J, Mannoor K, Guarnera MA, Shetty A, Zhan M, Xing L, Stass SA, and Jiang F (2015). Genome-wide small nucleolar RNA expression analysis of lung cancer by next-generation deep sequencing. *Int J Cancer* **136**, 623–629.
- [49] Humphrey LL, Deffebach M, Pappas M, Baumann C, Artis K, Mitchell JP, Zakher B, Fu R, and Slatore CG (2013). Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. *Ann Intern Med* **159**, 411–420.
- [50] Marks JR, Anderson KS, Engstrom P, Godwin AK, Esserman LJ, Longton G, Iversen ES, Mathew A, Patriotis C, and Pepe MS (2015). Construction and analysis of the NCI-EDRN breast cancer reference set for circulating markers of disease. *Cancer Epidemiol Biomarkers Prev* **24**, 435–441.
- [51] Tuck MK, Chan DW, Chia D, Godwin AK, Grizzle WE, Krueger KE, Rom W, Sanda M, Sorbara L, and Stass S, et al (2009). Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. *J Proteome Res* **8**, 113–117.
- [52] Ma J, Jemal A, and Smith R (2013). Reply to lung cancer deaths averted by screening should be considered in the context of tobacco control policies. *Cancer* **119**, 3420–3421.
- [53] Shen J, Liu Z, Todd NW, Zhang H, Liao J, Yu L, Guarnera MA, Li R, Cai L, and Zhan M, et al (2011). Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microRNA biomarkers. *BMC Cancer* **11**, 374–379.
- [54] Shen J, Todd NW, Zhang H, Yu L, Lingxiao X, Mei Y, Guarnera M, Liao J, Chou A, and Lu CL, et al (2011). Plasma microRNAs as potential biomarkers for non-small-cell lung cancer. *Lab Invest* **91**(4), 579–587.
- [55] Guthrie JL, Seah C, Brown S, Tang P, Jamieson F, and Drews SJ (2008). Use of Bordetella pertussis BP3385 to establish a cutoff value for an IS481-targeted real-time PCR assay. *J Clin Microbiol* **46**, 3798–3799.
- [56] Li N, Ma J, Guarnera MA, Fang H, Cai L, and Jiang F (2014). Digital PCR quantification of miRNAs in sputum for diagnosis of lung cancer. *J Cancer Res Clin Oncol* **140**, 145–150.
- [57] Ma J, Li N, Guarnera M, and Jiang F (2013). Quantification of plasma miRNAs by digital PCR for cancer diagnosis. *Biomark Insights* **8**, 127–136.
- [58] Dodd LE and Pepe MS (2003). Partial AUC estimation and regression. *Biometrics* **59**, 614–623.
- [59] Li H, Jiang Z, Leng Q, Bai F, Wang J, Ding X, Li Y, Zhang X, Fang H, and Yfantis HG, et al (2017). A prediction model for distinguishing lung squamous cell carcinoma from adenocarcinoma. *Oncotarget* **8**, 50704–50714.
- [60] Schisterman EF, Perkins NJ, Liu A, and Bondell H (2005). Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology* **16**, 73–81.
- [61] Schlosser K, Hanson J, Villeneuve PJ, Dimitrakoulas J, McIntyre L, Pilote L, and Stewart DJ (2016). Assessment of circulating lncRNAs under physiologic and pathologic conditions in humans reveals potential limitations as biomarkers. *Sci Rep* **6**, 36–39.
- [62] Day E, Dear PH, and McCaughan F (2013). Digital PCR strategies in the development and analysis of molecular biomarkers for personalized medicine. *Methods* **59**, 101–107.
- [63] Cui Y, Zhang F, Zhu C, Geng L, Tian T, and Liu H (2017). Upregulated lncRNA SNHG1 contributes to progression of non-small cell lung cancer

- through inhibition of miR-101-3p and activation of Wnt/beta-catenin signaling pathway. *Oncotarget* **8**, 17785–17794.
- [64] Lu Q, Shan S, Li Y, Zhu D, Jin W, and Ren T (2018). Long noncoding RNA SNHG1 promotes non-small cell lung cancer progression by up-regulating MTDH via sponging miR-145-5p. *FASEB J* **20**, 17–23.
- [65] Zhang HY, Yang W, Zheng FS, Wang YB, and Lu JB (2017). Long non-coding RNA SNHG1 regulates zinc finger E-box binding homeobox 1 expression by interacting with TAp63 and promotes cell metastasis and invasion in lung squamous cell carcinoma. *Biomed Pharmacother* **90**, 650–658.
- [66] Schmitt ME and Clayton DA (1993). Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in *Saccharomyces cerevisiae*. *Mol Cell Biol* **13**, 7935–7941.
- [67] Noh JH, Kim KM, Abdelmohsen K, Yoon JH, Panda AC, Munk R, Kim J, Curtis J, Moad CA, and Wohler CM, et al (2016). HuR and GRSF1 modulate the nuclear export and mitochondrial localization of the lncRNA RMRP. *Genes Dev* **30**, 1224–1239.
- [68] Maida Y, Yasukawa M, Furuuchi M, Lassmann T, Possemato R, Okamoto N, Kasim V, Hayashizaki Y, Hahn WC, and Masutomi K (2009). An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. *Nature* **461**, 230–235.
- [69] Chang DD and Clayton DA (1987). A mammalian mitochondrial RNA processing activity contains nucleus-encoded RNA. *Science* **235**, 1178–1184.
- [70] Meng Q, Ren M, Li Y, and Song X (2016). LncRNA-RMRP acts as an oncogene in lung cancer. *PLoS One* **11**, 164–184.
- [71] Marcus PM (2015). Lung cancer screening with low dose computed tomography (LDCT): looking back and moving forward. *Ann Transl Med* **3**, 41–46.
- [72] Leng Q, Lin Y, Jiang F, Lee CJ, Zhan M, Fang H, and Wang Y (2017). A plasma miRNA signature for lung cancer early detection. *Oncotarget* **8**, 111902–111911.
- [73] Lin Y, Leng Q, Jiang Z, Guarnera MA, Zhou Y, Chen X, Wang H, Zhou W, Cai L, and Fang H, et al (2017). A classifier integrating plasma biomarkers and radiological characteristics for distinguishing malignant from benign pulmonary nodules. *Int J Cancer* **141**, 1240–1248.
- [74] Ma J, Li N, Lin Y, Gupta C, and Jiang F (2016). Circulating neutrophil microRNAs as biomarkers for the detection of lung cancer. *Biomark Cancer* **8**, 1–7.
- [75] 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**(10), 1197–1206.
- [76] Shen J and Jiang F (2012). Applications of microRNAs in the diagnosis and prognosis of lung cancer. *Expert Opin Med Diagn* **6**, 197–207.
- [77] Normanno N, Denis MG, Thress KS, Ratcliffe M, and Reck M (2017). Guide to detecting epidermal growth factor receptor (EGFR) mutations in ctDNA of patients with advanced non-small-cell lung cancer. *Oncotarget* **8**, 12501–12516.
- [78] Balgkouranidou I, Chimonidou M, Milaki G, Tsaroucha E, Kakolyris S, Georgoulas V, and Lianidou E (2016). SOX17 promoter methylation in plasma circulating tumor DNA of patients with non-small cell lung cancer. *Clin Chem Lab Med* **54**, 1385–1393.

An integromic signature for lung cancer early detection

Qixin Leng^{1,*}, Yanli Lin^{1,*}, Min Zhan² and Feng Jiang¹

¹Department of Pathology, University of Maryland School of Medicine, Baltimore, MD 21201, USA

²Department of Epidemiology & Public Health, University of Maryland School of Medicine, Baltimore, MD 21201, USA

*These authors have contributed equally to this work

Correspondence to: Feng Jiang, **email:** fjiang@som.umaryland.edu

Keywords: diagnosis; early stage; lung cancer; plasma; biomarkers

Received: March 24, 2018

Accepted: April 07, 2018

Published: May 15, 2018

Copyright: Leng et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

We previously developed three microRNAs (miRs-21, 210, and 486-5p), two long noncoding RNAs (lncRNAs) (SNHG1 and RMRP), and two fucosyltransferase (FUT) genes (FUT8 and POFUT1) as potential plasma biomarkers for lung cancer. However, the diagnostic performance of the individual panels is not sufficient to be used in the clinics. Given the heterogeneity of lung tumors developed from multifactorial molecular aberrations, we determine whether integrating the different classes of molecular biomarkers can improve diagnosis of lung cancer. By using droplet digital PCR, we analyze expression of the seven genes in plasma of a development cohort of 64 lung cancer patients and 33 cancer-free individuals. The panels of three miRNAs (miRs-21, 210, and 486-5p), two lncRNAs (SNHG1 and RMRP), and two FUTs (FUT8 and POFUT1) have a sensitivity of 81-86% and a specificity of 84-87% for diagnosis of lung cancer. From the seven genes, an integromic plasma signature comprising miR-210, SNHG1, and FUT8 is developed that produces higher sensitivity (95.45%) and specificity (96.97%) compared with the individual biomarker panels (all $p < 0.05$). The diagnostic value of the signature was confirmed in a validation cohort of 40 lung cancer patients and 29 controls, independent of stage and histological type of lung tumor, and patients' age, sex, and smoking status (all $p > 0.05$). The integration of the different categories of biomarkers might improve diagnosis of lung cancer.

INTRODUCTION

Over 85% lung cancers are non-small cell lung cancers (NSCLC). NSCLC mainly consists of adenocarcinoma (AC) and squamous cell carcinoma (SCC). Tobacco smoking is the major cause of NSCLC. Since the prognosis for patients with lung cancer is strongly correlated to the tumor stage, diagnosing lung cancer at a curable stage can reduce the mortality [1]. The early detection of lung cancer in a large randomized trial using low-dose CT (LDCT) has revealed a 20% reduction in mortality as compared to chest X-rays [1]. However, LDCT is associated with over-diagnosis, excessive cost, and radiation exposure [2, 3]. The development

of circulating biomarkers that can accurately and cost-effectively identify early stage lung cancer is required [4].

During tumor development, cancer cells undergo apoptosis and necrosis, and release tumor-associated molecules that can circulate in bloodstream. The tumors-derived molecules in plasma provide cell-free circulating cancer biomarkers. Regulatory non-coding RNAs (ncRNAs) can be classified into two major classes based on the transcript size: small ncRNAs (<200 bp) including microRNAs (miRNAs) and long ncRNAs (lncRNAs) (>200 bp) [5, 6]. Through different molecular mechanisms or pathways, the two types of ncRNAs have diverse and critical functions in tumorigenesis [7-10]. Furthermore, plasma miRNAs and lncRNAs directly released from primary lung tumors or the circulating lung cancer cells

might provide cell-free biomarkers for lung cancer [8]. For instance, we recently developed a panel of three plasma miRNA biomarkers with 86% sensitivity and 87% specificity and a panel of two plasma lncRNA biomarkers with 83% sensitivity and 84% specificity for lung cancer early detection [7, 11-15].

Emerging evidences have demonstrated that aberrant glycosylation leads to cancer development and progression [16]. Fucosylation is the major type of glycosylation, and regulated by fucosyltransferases (FUTs) [16, 17]. We recently found that combined use of two plasma FUTs (FUT8 and POFUT1) had 81% sensitivity and 84% specificity for diagnosis of lung cancer, thus providing a new category of cell-free circulating biomarkers for lung cancer.

Since NSCLC is a heterogeneous disease and develops from multifactorial molecular aberrations [18], the analysis of one type of molecular changes may not achieve the performance required to move forward for clinical application. Indeed, although our individual panels of plasma biomarkers show promise for lung cancer diagnosis, their sensitivities (81-86%) and specificities (84-87%) are not sufficient to be used in the laboratory settings.

Because miRNAs, lncRNAs, and FUTs have highly diverse roles that drive the development of lung cancer [7-10, 16], we hypothesize that integrating the different classes of biomarkers may improve the early detection of lung cancer. Here we evaluate the individual and combined applications of the three categories of plasma molecular biomarkers for lung cancer.

RESULTS

The three individual panels of plasma biomarkers displayed a different level in NSCLC patients vs. smokers

Droplet Digital PCR (ddPCR) was used for quantification of the genes (miRs-21, 210, 486-5p, SNHG1, RMRP, FUT8, and POFUT1) in plasma of a development cohort of 64 lung cancer patients and 33 cancer-free individuals. All the seven genes generated at least 10,000 droplets in each well of the plasma samples. Therefore, the seven genes could be successfully “read” by ddPCR for their absolute quantification in plasma. These genes had a significantly different expression level in plasma of the NSCLC patients compared with the control individuals (all $P < 0.05$). As a result, the individual genes resulted in 50.09 to 75.76% sensitivities and 63.64 to 90.91% specificities for detection of NSCLC (Table 1). Furthermore, the panel of three microRNA biomarkers (miRs-21, 210, and 486-5p) had an area under receiver operating characteristic curve (AUC) of 0.92 with 86.36% sensitivity and 87.88% specificity, the panel of two plasma lncRNA biomarkers (SNHG1 and RMRP) displayed 0.89

AUC with 83.33% sensitivity and 84.85% specificity, and the panel of two FUTs (FUT8 and POFUT1) exhibited an AUC of 0.85 with 81.82% sensitivity and 84.85% specificity for diagnosis of lung cancer (Table 2). The individual panels of the genes didn't show special association with stage and histology of the NSCLC, age, gender, and smoking status of the participants (All $p > 0.05$). The seven genes would be potential plasma biomarkers for lung cancer.

An integromic plasma signature for lung cancer early detection

We used logistic regression models with constrained parameters as in least absolute shrinkage and selection operator (LASSO) and AUCs to determine performance of different patterns of combining the genes. From the seven genes, one miRNA (miR-210), one lncRNA (SNHG1), and one FUT (FUT8) were selected as the best biomarkers (all $P < 0.001$). A logisitic regression model with each of the different types of genes was developed as an integromic signature for diagnosing lung cancer: $U = -7.29 + 2.8 \times \log(\text{SNHG1}) + 3.83 \times \log(\text{FUT8}) + 3.36 \times \log(\text{miR-210})$. Combined analysis of the 3 biomarkers by using the logisitic regression model produced a higher AUC (0.97) (Figure 1) than did the individual panels of biomarkers ($p < 0.05$). We used the highest Youden's J index to set up corresponding cut-off value [19]. The optimal cut-off for the integromic signature was $U = 0.79$. Any subject with $U \geq 0.79$ was classified as a lung cancer case. As a result, the integromic plasma signature yielded significantly higher sensitivity (95.45%), specificity (96.97%), and accuracy (95.96%) compared with the individual panels of biomarkers (all $p < 0.05$) (Table 2). Furthermore, combined use of all the seven genes did not produce higher sensitivity and specificity compared with the integromic plasma signature ($p > 0.05$). In addition, Pearson's correlation analysis showed that the relationships among levels of the three genes were very low (All $p > 0.05$), implying that the integration of the different classes of molecular biomarkers has complementary classification. Moreover, the integromic plasma signature had no special association with histological type of the NSCLC, age, gender, and smoking status of the participants (All $p > 0.05$). The integromic signature did not show statistical difference of sensitivity and specificity for different stages of NSCLC (Supplementary Figure 1).

Validating the integromic plasma signature for lung cancer detection

The plasma expression levels of the three genes (miR-210, SNHG1, and FUT8) were assessed by using ddPCR in a validation cohort of additional 40 NSCLC patients and 29 healthy controls. Combined analysis of

Table 1: Diagnostic performance of individual genes for lung cancer in a development cohort

	Sensitivity (95% CI)	Specificity (95% CI)
miRs-21	75.76% (63.64% to 85.46%)	63.64% (45.12% to 79.60%)
miR-210	50.09% (46.29% to 71.05%)	72.73% (54.48% to 86.70%)
miR-486-5p	72.73% (60.36% to 82.97%)	63.64% (45.12% to 79.60%)
SNHG1	75.76% (63.64% to 85.46%)	81.82% (64.54% to 93.02%)
RMRP	62.12% (49.34% to 73.78%)	90.91% (75.67% to 98.08%)
FUT8	71.21 (58.75% to 81.70%)	87.88% (71.80% to 96.60%)
POFUT1	60.61% (47.81% to 72.42%)	90.91% (75.67% to 98.08%)

Abbreviations: CI, confidence interval.

Table 2: The area under receiver operating characteristic curves (AUCs) of the individual panels of biomarkers and the plasma integromic signature in a development cohort

	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy
A panel of 3 miRNAs	0.92 (0.87 to 0.97)	86.36% (75.69% to 93.57%)	87.88% (71.80% to 96.60%)	86.87% (78.59% to 92.82%)
A panel of 2 lncRNAs	0.89 (0.85 to 0.93)	83.33% (72.13% to 91.38%)	84.85% (68.10% to 94.89%)	83.84% (75.09% to 90.47%)
A panel of 2 FUTs	0.85 (0.78 to 0.91)	81.82% (70.39% to 90.24%)	84.85% (68.10% to 94.89%)	82.83% (73.94% to 89.67%)
An integromic signature	0.95 (0.91 to 0.99)	95.45% (87.29% to 99.05%)	96.97% (84.24% to 99.92%)	95.96% (89.98% to 98.89%)

Abbreviations: CI, confidence interval.

the three genes by using the logistic regression model created 0.94 AUC for lung cancer diagnosis. There was no significant difference between the develop cohort and validation cohort with regarding the signature's AUCs (0.95 vs. 0.94, $p=0.46$) (Figure 2). In the validation cohort, the three genes used in combination could differentiate the NSCLC patients from healthy controls with a sensitivity of 95.00% (82.08% to 99.12%) and a specificity of 96.55% (80.37% to 99.82%). In line with the findings in the development cohort, the integromic plasma signature did not show statistical difference of sensitivity and specificity across different stages and subtypes of NSCLC (all $p>0.05$). Moreover, there was no association of expressions of the genes with the age, gender, or smoking status of the individuals (All $p>0.05$).

DISCUSSION

Although showing promise, the use of the individual miRNA, lncRNA, or FUT biomarker panels alone has moderate sensitivities (81-86%) and specificities (84-87%). The miRNAs, lncRNAs, and FUTs have highly different functions in carcinogenesis [20-22]. Given the heterogeneous nature of lung cancer and the numerous

cellular pathways involved, we hypothesize that integrating the different classes of molecular biomarkers may improve the early detection of lung cancer. Intriguingly, the integrated analysis of only one of each type of biomarkers by using a single platform (ddPCR) yields a significantly higher diagnostic performance compared with any panel of one type of genes. Furthermore, the correlations among the changes of the miRNAs, lncRNA, and FUT are very low, supporting that the diagnostic vales of the three classes of molecular alterations could be complementary to each other. Therefore, the observation confirms our hypothesis. Moreover, since the integromic plasma signature shows similar sensitivity and specificity in the early vs. advanced stages of NSCLC, it might be a useful approach for the early detection of lung cancer, a clinically challenging.

miR-210 stimulates a hypoxic phenotype and upsurges radioresistance in NSCLCs [21]. Hypoxia-induced miR-210 can regulate tumor cell susceptibility to cytolytic T-lymphocyte-mediated lysis by a mechanism involving its downstream targets PTPN1, HOXA1, and TP53I11 [23]. SNHG1 could promote NSCLC progression of lung cancer via miR-101-3p/SOX9/Wnt/ β -catenin regulatory network and miR-145-5p/MTDH axis [14, 22]. In addition, SNHG1 plays

an oncogenic role in lung squamous cell carcinoma through ZEB1 signaling pathway by inhibiting TAp63 [24]. FUT8 inhibits the malignant behaviors of lung cancer cells and is involved in the regulation of dozens of genes associated with the malignancy through multiple mechanisms [20]. Upregulation of FUT8 could contribute to epithelial-mesenchymal transition via the transactivation of β -catenin/lymphoid enhancer-binding factor-1 (LEF-1) [20].

The study does have some limitations. 1), the sample size of the cohorts is small. We will perform a new study to prospectively validate the integromic signature for lung cancer early detection using a large population. 2), it is well known that lung cancer-associated molecular genetic changes are also related to chronic obstructive pulmonary disease (COPD) [25]. Many lung cancer patients who are smokers and cancer-free heavy smokers have COPD [25]. COPD could impact molecular genetic profiles in plasma

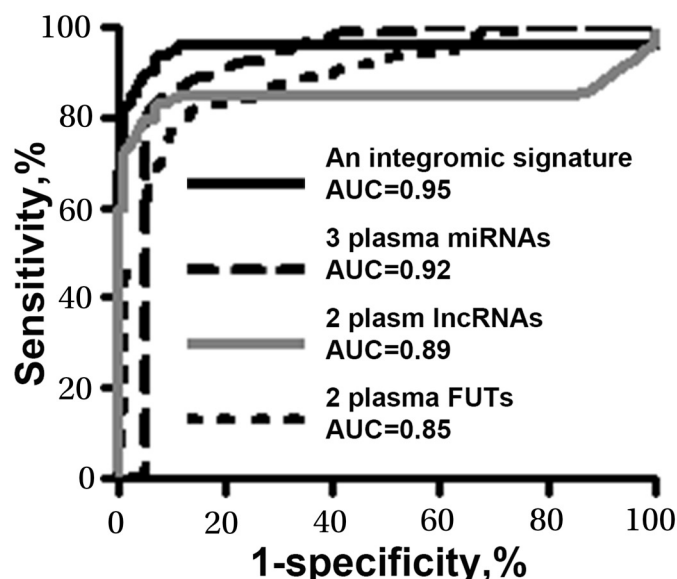


Figure 1: Diagnostic values of the individual panels of biomarkers and integromic plasma signature in a development cohort. The integromic plasma biomarker signature yields a higher area under receiver operating characteristic curve (AUC) than does individual panels of biomarkers (All $p < 0.05$).

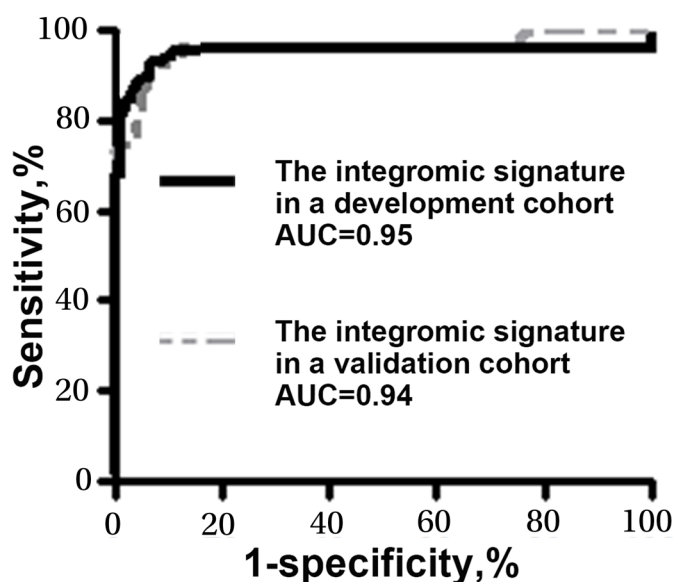


Figure 2: Comparison of AUCs of the integromic plasma signature for diagnosis of lung cancer in different cohorts. The integromic plasma signature has no significant difference of AUCs in the development cohort (black line) vs. validation cohort (gray line) (0.95 vs. 0.94, $p = 0.46$).

Table 3: Characteristics of a development cohort and a validation cohort

	A development cohort		P-value
	NSCLC cases (n = 66)	Controls (n = 33)	
Age	67.68 (SD 9.23)	62.70 (SD 15.33)	0.16
Sex			0.39
Female	17	8	
Male	49	25	
Smoking pack-years (median)	33.6	32.69	0.17
Stage			
Stage I	15		
Stage II	11		
Stage III	17		
Stage IV	19		
Unknown	4		
Histological type			
Adenocarcinoma	34		
Squamous cell carcinoma	32		

	A validation cohort		P-value
	NSCLC cases (n = 40)	Controls (n = 29)	
Age	64.57 (SD 9.28)	63.76 (SD 13.49)	0.28
Sex			0.46
Female	10	8	
Male	30	21	
Smoking pack-years (median)	33.68	31.63	0.23
Stage			
Stage I	9		
Stage II	8		
Stage III	11		
Stage IV	11		
Unknown	1		
Histological type			
Adenocarcinoma	22		
Squamous cell carcinoma	18		

Abbreviations: NSCLC, non-small cell lung cancer.

of both lung cancer patients and cancer-free heavy smokers. In this present project, there is no COPD information of the cases and controls enrolled. Therefore, we are not able to evaluate if the biomarkers identified in the study are associated with COPD. We will recruit lung cancer patients

and cancer-free smokers who have COPD, and determine if COPD is the confounding effect on the molecular changes. 3), the early detection of NSCLC using LDCT followed by appropriate treatments can significantly reduce lung cancer mortality in smokers [1]. LDCT is now recommended for

lung cancer screening in smokers. Yet LDCT has a low specificity for the early detection of lung cancer, presenting a major clinical challenge [1]. The participants enrolled in this project are not representative of the smokers in LDCT screening setting for lung cancer. We will perform a prospective trial to determine if the integromic signature could improve the spesitivity of LDCT for the early deteciton of lung cancer in smokers.

MATERIALS AND METHODS

Patients and clinical specimens

Using a protocol approved by the local Institutional Review Boards Institutional Review Boards, we recruited lung cancer patients and cancer-free smokers according to the inclusion and/or exclusion criteria recommended by U.S. Preventive Services Task Force [26]. Briefly, we enrolled smokers between the ages of 55-80 who had at least a 30 pack-year smoking history and were former smokers (quit within 15 years). Exclusion criteria included pregnancy, current pulmonary infection, surgery within 6 months, radiotherapy within 1 year, and life expectancy of < 1 year. We collected blood in BD Vacutainer spray-coated K2EDTA Tubes (BD, Franklin Lakes, NJ) and prepared plasma using the standard operating protocols developed by The NCI-Early Detection Research Network [27]. The specimens were processed within 2 hours of collection by centrifugation at 1,300 X g for 10 minutes at 4°C. The surgical-pathologic staging of NSCLC was used as the ground truth according to the TNM classification of the International Union Against Cancer (UICC) with the American Joint Committee on Cancer (AJCC) and the International Staging System for Lung Cancer [28, 29]. A total of 106 NSCLC patients and 62 cancer-free smokers were recruited. Among the cancer patients, 27 patients were female and 79 were male. Twenty-four had stage I NSCLC, 19 with stage II, 28 with stage III, 30 with stage IV, and 5 with unknown stage. Fifty-six lung cancer patients were diagnosed with AC, while 40 with SCC. Of the cancer-free smokers, 16 patients were female and 46 were male. There were no significant differences of age, gender and smoking status between the NSCLC patients and cancer-free smokers. The cases and controls were randomly grouped into two cohorts: a development cohort and a validation cohort. The development cohort consisted of 66 lung cancer patients and 33 cancer-free smokers, while the validation cohort comprised 40 lung cancer patients and 29 cancer-free smokers. The demographic and clinical variables of the two cohorts are shown in Table 3.

ddPCR

RNA was extracted from plasma by using Trizol LS reagent (Invitrogen Carlsbad, CA) and RNeasy Mini Kit (Qiagen, Hilden, Germany) [11, 12]. The

qualification and quantification of RNA were assessed by using Biospectrometer (Hutchinson Technology Inc, Hutchinson, MN) and Electrophoresis Bioanalyzer (Agilent Technologies, Foster City, CA). Reverse Transcriptase (RT) was carried out to generate cDNA by using a RT Kit (Applied Biosystems, Foster City, CA) [11, 12]. ddPCR for analysis of expression level of the genes was performed as described in our published works by using a QX200™ Droplet Digital™ PCR System (Bio-Rad, Hercules, CA) [11-15, 30-50]. Briefly, PCR reaction mix containing cDNA was partitioned into aqueous droplets in oil via the QX100 Droplet Generator, and then transferred to a 96-well PCR plate. A two-step thermocycling protocol (95°C ×10min; 40 cycles of [94°C ×30s, 60°C ×60s], 98°C ×10 min) was undertaken in a Bio-Rad C1000 (Bio-Rad, Pleasanton, CA). The PCR plate was then transferred to the QX100 Droplet Reader for automatic reading of samples in all wells. Copy number of each gene per µl PCR reaction was directly determined. Primers and probes of the targeted genes are shown in Supplementary Table 1. We used QuantaSoft 1.7.4 analysis software (Bio-Rad) and Poisson statistics to compute droplet concentrations (copies/µL). Only genes that had at least 10,000 droplets were considered to be robustly detectable by ddPCR in plasma and subsequently underwent further analysis [31]. All assays were done in triplicates, and one no-template control and two interplate controls were carried along in each experiment.

Statistical analysis

To estimate sample size, we set AUC of H0 (the null hypothesis) at 0.5. H1 represented the alternative hypothesis. To have a high reproducibility with adequate precision, we required ≥28 subjects per group. With this sample size, we would have 85% power to detect an AUC of 0.75 at the 2% significance level. Therefore, the sample size in the two cohorts could have enough statistical power. Pearson's correlation analysis was applied to assess relationship between gene expressions and demographic and clinical characteristics of the patients and control individuals. AUCs were used to determine accuracy, sensitivity, and specificity of each gene. We used the highest Youden's J index (sum of sensitivity and specificity—1) to set up corresponding cut-off value [19]. Logistic regression models with constrained parameters as in LASSO were used to eliminate the irrelevant genes, develop composite panels of biomarkers, and optimize a signature with the highest sensitivity and specificity. To compare the signature and our previously developed plasma biomarker panels, we compared their AUCs to determine the sensitivity and specificity as previously described [15].

CONCLUSIONS

Given the heterogeneous nature of NSCLC developed from multifactorial molecular aberrations, we

have for the first time demonstrated that the integration of miRNA, lncRNA, and FUT biomarkers could provide an efficient approach for diagnosis of lung cancer. Nonetheless, a large multi-center clinical project to prospectively validate the full utility of the integromic signature is required.

Author contributions

QL, YL, MZ, and FJ conducted the experiments and participated in study design, coordination, and data interpretation, and preparing the manuscript. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

FUNDING

This work was supported in part by NCI R21CA205746, VA Merit Award I01 CX000512, Award from the Geaton and JoAnn DeCesaris Family Foundation, UMD-UMB Research and Innovation Seed Grant, DoD-Idea Development Award, and Maryland Innovation Initiative (MII) Commercialization Program- Phase 1/2 Grant (F.J.).

REFERENCES

1. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011; 365:395-409.
2. Patz EF Jr, Pinsky P, Gatsonis C, Sicks JD, Kramer BS, Tammemagi MC, Chiles C, Black WC, Aberle DR. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med*. 2014; 174:269-274.
3. Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, Galen B, Gareen IF, Gatsonis C, Goldin J, Gohagan JK, Hillman B, Jaffe C, Kramer BS, et al. The National Lung Screening Trial: overview and study design. *Radiology*. 2011; 258:243-253.
4. Hubers AJ, Prinsen CF, Sozzi G, Witte BI, Thunnissen E. Molecular sputum analysis for the diagnosis of lung cancer. *Br J Cancer*. 2013; 109:530-537.
5. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. *Cell*. 2009; 136:642-655.
6. Orom UA, Shiekhattar R. Long noncoding RNAs usher in a new era in the biology of enhancers. *Cell*. 2013; 154:1190-1193.
7. Shen J, Jiang F. Applications of MicroRNAs in the Diagnosis and Prognosis of Lung Cancer. *Expert Opin Med Diagn*. 2012; 6:197-207.
8. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Brian KC, Allen A, Lin DW, Urban N, Drescher CW, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A*. 2008; 105:10513-10518.
9. Huarte M. The emerging role of lncRNAs in cancer. *Nat Med*. 2015; 21:1253-1261.
10. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens RM, Okamoto A, Yokota J, Tanaka T, Calin GA, Liu CG, Croce CM, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*. 2006; 9:189-198.
11. Shen J, Todd NW, Zhang H, Yu L, Lingxiao X, Mei Y, Guarnera M, Liao J, Chou A, Lu CL, Jiang Z, Fang H, Katz RL, Jiang F. Plasma microRNAs as potential biomarkers for non-small-cell lung cancer. *Lab Invest*. 2011; 91:579-587.
12. Shen J, Liu Z, Todd NW, Zhang H, Liao J, Yu L, Guarnera MA, Li R, Cai L, Zhan M, Jiang F. Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microRNA biomarkers. *BMC Cancer*. 2011; 11:374.
13. Shen J, Liao J, Guarnera MA, Fang H, Cai L, Stass SA, Jiang F. Analysis of MicroRNAs in sputum to improve computed tomography for lung cancer diagnosis. *J Thorac Oncol*. 2014; 9:33-40.
14. Leng Q, Lin Y, Jiang F, Lee CJ, Zhan M, Fang H, Wang Y. A plasma miRNA signature for lung cancer early detection. *Oncotarget*. 2017; 8:111902-111911. <https://doi.org/10.18632/oncotarget.22950>.
15. Lin Y, Leng Q, Jiang Z, Guarnera MA, Zhou Y, Chen X, Wang H, Zhou W, Cai L, Fang H, Li J, Jin H, Wang L, et al. A classifier integrating plasma biomarkers and radiological characteristics for distinguishing malignant from benign pulmonary nodules. *Int J Cancer*. 2017; 141:1240-1248.
16. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer*. 2015; 15:540-555.
17. Kirwan A, Utratna M, O'Dwyer ME, Joshi L, Kilcoyne M. Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. *Biomed Res Int*. 2015; 2015:490531.

18. Kadara H, Wistuba II. Field cancerization in non-small cell lung cancer: implications in disease pathogenesis. *Proc Am Thorac Soc.* 2012; 9:38-42.
19. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950; 3:32-35.
20. Chen CY, Jan YH, Juan YH, Yang CJ, Huang MS, Yu CJ, Yang PC, Hsiao M, Hsu TL, Wong CH. Fucosyltransferase 8 as a functional regulator of nonsmall cell lung cancer. *Proc Natl Acad Sci U S A.* 2013; 110:630-635.
21. Grosso S, Doyen J, Parks SK, Bertero T, Paye A, Cardinaud B, Gounon P, Lacas-Gervais S, Noel A, Pouyssegur J, Barbry P, Mazure NM, Mari B. MiR-210 promotes a hypoxic phenotype and increases radioresistance in human lung cancer cell lines. *Cell Death Dis.* 2013; 4:e544.
22. Cui Y, Zhang F, Zhu C, Geng L, Tian T, Liu H. Upregulated lncRNA SNHG1 contributes to progression of non-small cell lung cancer through inhibition of miR-101-3p and activation of Wnt/beta-catenin signaling pathway. *Oncotarget.* 2017; 8:17785-17794. <https://doi.org/10.18632/oncotarget.14854>.
23. Noman MZ, Buat S, Romero P, Ketari S, Janji B, Mari B, Mami-Chouaib F, Chouaib S. Hypoxia-inducible miR-210 regulates the susceptibility of tumor cells to lysis by cytotoxic T cells. *Cancer Res.* 2012; 72:4629-4641.
24. Zhang HY, Yang W, Zheng FS, Wang YB, Lu JB. Long non-coding RNA SNHG1 regulates zinc finger E-box binding homeobox 1 expression by interacting with TAp63 and promotes cell metastasis and invasion in Lung squamous cell carcinoma. *Biomed Pharmacother.* 2017; 90:650-658.
25. Carr LL, Jacobson S, Lynch DA, Foreman MG, Flanagan EL, Hersh CP, Sciurba FC, Wilson DO, Sieren JC, Mulhall P, Kim V, Kinsey CM, Bowler RP. Features of COPD as Predictors of Lung Cancer. *Chest.* 2018; 55:286-287.
26. Humphrey LL, Deffebach M, Pappas M, Baumann C, Artis K, Mitchell JP, Zakher B, Fu R, Slatore CG. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. *Ann Intern Med.* 2013; 159:411-420.
27. Tuck MK, Chan DW, Chia D, Godwin AK, Grizzle WE, Krueger KE, Rom W, Sanda M, Sorbara L, Stass S, Wang W, Brenner DE. Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. *J Proteome Res.* 2009; 8:113-117.
28. Leong SS, Rocha Lima CM, Sherman CA, Green MR. The 1997 International Staging System for non-small cell lung cancer: have all the issues been addressed? *Chest.* 1999; 115: 242-248.
29. Bulzebruck H, Bopp R, Drings P, Bauer E, Krysa S, Probst G, van Kaick G, Muller KM, Vogt-Moykopf I. New aspects in the staging of lung cancer. Prospective validation of the International Union Against Cancer TNM classification. *Cancer.* 1992; 70:1102-1110.
30. Li N, Ma J, Guarnera MA, Fang H, Cai L, Jiang F. Digital PCR quantification of miRNAs in sputum for diagnosis of lung cancer. *J Cancer Res Clin Oncol.* 2014; 140:145-150.
31. Ma J, Li N, Guarnera M, Jiang F. Quantification of Plasma miRNAs by Digital PCR for Cancer Diagnosis. *Biomark Insights.* 2013; 8:127-136.
32. Su J, Leng Q, Lin Y, Ma J, Jiang F, Lee CJ, Fang H. Integrating Circulating Immunological and Sputum Biomarkers for the Early Detection of Lung Cancer. *Biomark Cancer.* 2018; 10: 299-307.
33. Su Y, Fang H, Jiang F. Integrating DNA methylation and microRNA biomarkers in sputum for lung cancer detection. *Clin Epigenetic.* 2016; 8:109.
34. Su Y, Guarnera MA, Fang H, Jiang F. Small non-coding RNA biomarkers in sputum for lung cancer diagnosis. *Clin Epigenetics.* 2016; 15:36.
35. Su J, Anjuman N, Guarnera MA, Zhang H, Stass SA, Jiang F. Analysis of Lung Flute-collected Sputum for Lung Cancer Diagnosis. *Biomark Insights.* 2015; 10:55-61.
36. Su J, Liao J, Gao L, Shen J, Guarnera MA, Zhan M, Fang H, Stass SA, Jiang F. Analysis of small nucleolar RNAs in sputum for lung cancer diagnosis. *Oncotarget.* 2016; 7:5131-5142. <https://doi.org/10.18632/oncotarget.4219>.
37. Xing L, Su J, Guarnera MA, Zhang H, Cai L, Zhou R, Stass SA, Jiang F. Sputum microRNA biomarkers for identifying lung cancer in indeterminate solitary pulmonary nodules. *Clin Cancer Res.* 2015; 21:484-489.
38. Pine PS, Lund SP, Stass SA, Kukuruga D, Jiang F, Sorbara L, Srivastava S, Salit M. Cell-based reference samples designed with specific differences in microRNA biomarkers. *BMC Biotechnol.* 2018; 18:17-26.
39. Anjuman N, Li N, Guarnera M, Stass SA, Jiang F. Evaluation of lung flute in sputum samples for molecular analysis of lung cancer. *Clin Transl Med.* 2013; 2:15.
40. Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z, Fang H, Zhang J, Katz RL, Jiang F. Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. *Int J Cancer.* 2010; 127:2870-2878.
41. Jiang F, Todd NW, Li R, Zhang H, Fang H, Stass SA. A panel of sputum-based genomic marker for early detection of lung cancer. *Cancer Prev Res (Phila).* 2010; 3:1571-1578.
42. Xing L, Todd NW, Yu L, Fang H, Jiang F. Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers. *Mod Pathol.* 2010; 23:1157-1164.
43. Xie Y, Todd NW, Liu Z, Zhan M, Fang H, Peng H, Alattar M, Deepak J, Stass SA, Jiang F. Altered miRNA expression

- in sputum for diagnosis of non-small cell lung cancer. *Lung Cancer*. 2010; 67:170-176.
44. Jiang F, Todd NW, Qiu Q, Liu Z, Katz RL, Stass SA. Combined genetic analysis of sputum and computed tomography for noninvasive diagnosis of non-small-cell lung cancer. *Lung Cancer*. 2009; 66:58-63.
45. Li H, Jiang Z, Leng Q, Bai F, Wang J, Ding X, Li Y, Zhang X, Fang H, Yfantis HG, Xing L, Jiang F. A prediction model for distinguishing lung squamous cell carcinoma from adenocarcinoma. *Oncotarget*. 2017; 8:50704-50714. <https://doi.org/10.18632/oncotarget.17038>.
46. 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; 10:40-45.
47. Li R, Todd NW, Qiu Q, Fan T, Zhao RY, Rodgers WH, Fang HB, Katz RL, Stass SA, Jiang F. Genetic deletions in sputum as diagnostic markers for early detection of stage I non-small cell lung cancer. *Clin Cancer Res*. 2007; 13:482-487.
48. Ma J, Lin Y, Zhan M, Mann DL, Stass SA, Jiang F. Differential miRNA expressions in peripheral blood mononuclear cells for diagnosis of lung cancer. *Lab Invest*. 2015; 95:1197-1206.
49. Gao L, Ma J, Mannoor K, Guarnera MA, Shetty A, Zhan M, Xing L, Stass SA, Jiang F. Genome-wide small nucleolar RNA expression analysis of lung cancer by next-generation deep sequencing. *Int J Cancer*. 2015; 136:E623-629.
50. Ma J, Mannoor K, Gao L, Tan A, Guarnera MA, Zhan M, Shetty A, Stass SA, Xing L, Jiang F. Characterization of microRNA transcriptome in lung cancer by next-generation deep sequencing. *Mol Oncol*. 2014; 8:1208-1219.

RESEARCH

Open Access



An epigenetic classifier for early stage lung cancer

Yun Su^{1*}, Hong Bin Fang² and Feng Jiang³

Abstract

Background: Methylated genes detected in sputum are promise biomarkers for lung cancer. Yet the current PCR technologies for quantification of DNA methylation and diagnostic value of the sputum biomarkers are not sufficient to be used for lung cancer early detection. The emerging droplet digital PCR (ddPCR) is a straightforward means for precise, direct, and absolute quantification of nucleic acids. Here, we investigate whether ddPCR can sensitively and robustly quantify DNA methylation in sputum for more precise diagnosis of lung cancer.

Results: First, the analytic performance of methylation-specific ddPCR (ddMSP) and quantitative methylation-specific PCR (qMSP) is determined in methylated and unmethylated DNA samples. Second, 29 genes, previously proposed as potential sputum biomarkers for lung cancer, are analyzed by using ddMSP in a training set of 127 lung cancer patients and 159 controls. ddMSP has higher sensitivity, precision, and reproducibility for quantification of methylation compared with qMSP (all $p < 0.05$). A classifier comprising four sputum methylation biomarkers for lung cancer is developed by using ddMSP, producing 86.6% sensitivity and 90.6% specificity, independent of stage and histology of lung cancer (all $p > 0.05$). The classifier has higher accuracy compared with sputum cytology (88.8 vs. 70.6%, $p < 0.01$). The diagnostic performance is confirmed in a testing set of 89 cases and 107 controls.

Conclusions: ddMSP is a robust tool for reliable quantification of DNA methylation in sputum, and the epigenetic classifier could help diagnose lung cancer at the early stage.

Keywords: ddPCR, DNA methylation, Sputum, Diagnosis, Lung cancer

Background

Lung cancer is the leading cause of cancer death among men and women [1]. More than 85% lung tumors are non-small cell lung cancers (NSCLCs), which consist of adenocarcinoma (AC), squamous cell carcinoma (SCC), and large cell carcinoma (LC). Cigarette smoking is the foremost cause of NSCLC [2]. People who smoke cigarettes are nearly 30 times more likely to get lung cancer or die from lung cancer than people who do not smoke. Even smoking a few cigarettes a day or smoking occasionally increases the risk of lung cancer. Individuals who quit smoking have a lower risk of lung cancer than if they had continued to smoke, but their risk is higher than the risk for people who never smoked. The National Lung Screening Trial (NLST) results show that using low-dose CT (LDCT) for the early detection of

lung cancer in smokers can reduce the mortality by 20% as compared to chest X-rays [1]. Therefore, LDCT is recently recommended to be used for lung cancer early detection among smokers [3, 4]. However, LDCT is associated with over-diagnosis, excessive cost, and radiation exposure, limiting its clinical applications [3–5]. The development of noninvasive approaches that can accurately and cost-effectively diagnose early stage lung cancer among smokers remains clinically important [6].

Lung cancer develops from a field defect characterized by an accumulation of molecular abnormalities resulted from repeated exposure of the airway of the smokers to the tobacco-related carcinogens [7–9]. Regardless of the anatomic location relative to the tumors, the molecular alterations observed in the large bronchial airway might reflect the altered changes existed in lung tumors [9–11]. Sputum is defined as secretions from the airways and contains bronchial epithelial cells exfoliated from the airways or lungs [12]. Therefore, the analysis of exfoliated bronchial

* Correspondence: yunsu326@yahoo.com

¹Department of Surgery, Jiangsu Province Hospital of Nanjing University of Chinese Medicine, 138 Xianlin Road, Nanjing 210023, China
Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

epitheliums in sputum for the molecular changes may provide a useful tool for noninvasively and cost-effectively diagnosing lung cancer.

DNA methylations of tumor suppressor genes (TSGs) are early molecular events in lung carcinogenesis and thus show great promise as biomarkers for early stage lung cancer [6, 11, 13–34]. Conventional qPCR-based platforms, particularly, methylation-specific PCR (qMSP), have been used for detecting DNA methylation of TSGs in sputum [13]. However, qMSP has some weaknesses, limiting its use in the clinical settings. For example, qMSP is an indirect approach, which requires internal controls for data normalization [35]. Furthermore, qMSP's sensitivity for analyzing low copy number of genes is poor. This is particularly challenging for quantification of DNA methylation in bronchial epitheliums, as the large excess of non-epithelial cells in sputum could obscure detection of the relative scarcity of methylated DNA from the exfoliated bronchial epitheliums. A more sensitive, precise, and reproducibility method for quantification of methylated DNA in sputum would provide a useful means for noninvasive diagnosis of lung cancer.

Droplet digital PCR (ddPCR) is a direct method for quantitatively measuring nucleic acids [36–45], since it depends on limiting partition of the PCR volume, where a positive result of a large number of microreactions indicates the presence of a single molecule in a given reaction. The number of positive reactions, together with Poisson's distribution, can be used to produce a straight and high-confidence measurement of the original target concentration [43]. Furthermore, ddPCR does not require the reliance on rate-based measurements, endogenous controls, and the use of calibration curves. In addition, previous studies including our own research have demonstrated that ddPCR can quantify low-abundance nucleic acids and has higher sensitivity and precision than does conventional PCR [36, 37, 46]. The objective of this study is to investigate whether methylation-specific ddPCR (ddMSP) could sensitively and robustly quantify DNA methylations in sputum and hence develop a biomarker-based classifier for early stage lung cancer.

Methods

Study population

The study protocol was approved by the local Institutional Review Board. The participants in this study were recruited from the hospital at the point of their referral for suspected lung cancer between the ages of 55–80. Written informed consent was obtained from all enrolled subjects. Exclusion criteria included pregnancy, current pulmonary infection, surgery within 6 months, radiotherapy within 1 year, and life expectancy of < 1 year. Clinical diagnosis of lung cancer was made using

histopathologic examinations of specimens obtained by CT-guided transthoracic needle biopsy, transbronchial biopsy, videotape-assisted thoracoscopic surgery, or surgical resection. The surgical pathologic staging was determined according to the TNM classification of the International Union Against Cancer with the 8th American Joint Committee on Cancer and the International Staging System for Lung Cancer. Histopathological classification was determined according to the World Health Organization classification. A total of 482 subjects including 216 lung cancer patients and 266 cancer-free smokers were recruited. The 216 lung cancer patients were diagnosed with NSCLC consisting of 55 stage I cases, 55 stage II cases, 50 stage III cases, and 56 stage IV cases. One hundred and twelve cases were AC, 91 were SCC, and 13 were LC. The 266 cancer-free patients who were smokers and served as control subjects had granulomatous inflammation ($n = 117$), nonspecific inflammatory changes ($n = 105$) or lung infections ($n = 44$). The cancer-free smokers had been followed for at least 2 years, and none had any evidence of cancer. No difference of age, gender, and smoking status was observed in the lung cancer cases vs. controls (All $p > 0.05$). To refine the biomarkers whose changes specific to NSCLC, the cases were matched to the controls on gender, age, race, and smoking status as a nested case-control study. The cases and controls were then randomly split into a training set and a testing set by using a random number generator. The training set consisted of 127 lung cancer patients and 159 cancer-free controls. The testing set comprised 89 lung cancer patients and 107 cancer-free controls. The demographic and clinical characteristics of the two cohorts are presented in Tables 1 and 2.

Sample collection and sputum cytology

Sputum was collected from the participants as described in previous reports [47–54]. Briefly, to reduce the percentage of oral epithelial cells in the sputum, subjects were asked to blow their nose, rinse their mouth, and swallow water to minimize contamination of squamous cells from postnasal drip and saliva. Sputum samples were then coughed in a sterile container and processed within 2 h. To further minimize oral squamous cell contamination, opaque or dense portions that looked different from saliva under the inverted microscope were selected using blunt forceps from expectorate. The samples were processed on ice in 4 volumes of 0.1% dithiothreitol (Sigma-Aldrich, St. Louis, Mo) followed by 4 volumes of phosphate-buffered saline (PBS) (Sigma-Aldrich). The cell suspension was filtered through 45- μ m nylon gauzes (BNSH Thompson, Scarborough, ON, Canada). Absolute cell numbers and cell viability were quantitated by using a hemacytometer with trypan blue. Two cytocentrifuge slides

Table 1 Characteristics of NSCLC patients and cancer-free smokers in a training set

	NSCLC cases (n = 127)	Controls (n = 159)	p value
Age	65.48 (SD 12.32)	65.63 (SD 11.56)	0.32
Sex			0.35
Female	45	56	
Male	82	103	
Smoking pack-years (median)	35.23	33.79	0.34
Stage			
Stage I	33		
Stage II	32		
Stage III	29		
Stage IV	33		
Histological type			
Adenocarcinoma	63		
Squamous cell carcinoma	57		
Large cell carcinoma	7		

Abbreviations: NSCLC non-small cell lung cancer

were prepared from aliquots of cell suspension by using a cytospin machine (Shandon, Pittsburgh, PA) and were then stained with the Papanicolaou staining technique [12]. A sputum sample was considered adequate if lung macrophages or Curschmann spirals were present on the slides [11, 12]. Cytologic diagnosis was performed on the cytospin slides using the classification of Saccomanno et al. [12]. The remaining cells are stored at -80°C until used.

Table 2 Characteristics of NSCLC patients and cancer-free smokers in a testing set

	NSCLC cases (n = 89)	Controls (n = 107)	p value
Age	65.25 (SD 11.28)	65.36 (SD 11.48)	0.30
Sex			0.36
Female	31	37	
Male	58	70	
Smoking pack-years (median)	35.76	33.29	0.32
Stage			
Stage I	22		
Stage II	23		
Stage III	21		
Stage IV	23		
Histological type			
Adenocarcinoma	49		
Squamous cell carcinoma	34		
Large cell carcinoma	6		

Abbreviations: NSCLC non-small cell lung cancer

DNA isolation and bisulfite conversion

We extracted DNA from the specimens using DNeasy kit (Qiagen, Valencia, CA) as previously described [14]. We eluted DNA with 50 μL of elution buffer (10 mmol/L Tris-Cl, pH 8.5) (Sigma-Aldrich Corporation). DNA was quantified by using the Quantifiler Human DNA Quantification kit (Applied Biosystems, Foster City, CA). Bisulfite conversion was carried out on DNA by using the Zymo EZ DNA Methylation Kit (Zymo Research, Irvine, CA) according to the manufacturer's protocol.

Serially diluted methylated/unmethylated DNA specimens

We purchased 100% methylated and 100% unmethylated control human DNA samples (Zymo Research). We isolated DNA from sputum of a healthy nonsmoker whose sputum DNA did not harbor DNA methylation of TSGs, including *RASSF1A*, *3OST2*, and *PRDM14* [14]. To determine limit of quantification (LOQ) of an assay, we diluted methylated DNA into the sputum DNA sample in the following concentrations: 100, 25, 6.25, 1.56, 0.39, 0.1, 0.04, and 0% methylated DNA. To determine limits of detection (LOD) of an assay, we prepared serially diluted samples containing 5000, 2500, 1250, 625, 313, 156, and 0 pg methylated DNA in H_2O .

Quantification of DNA methylation in sputum by ddMSP

We added bisulfite-treated DNA (2 μL) to ddPCR mixture (18 μL) containing 2 $\times$ ddPCR Supermix for probes (no-dUTP), 750 nmol/L of each primer and 250 nmol/L of the corresponding probe in a final volume of 20 μL . Twenty-nine genes were selected for DNA methylation analysis, since the genes were previously reported as potential sputum methylation biomarkers for lung cancer [6, 13–34]. The 29 genes are *3OST2*, *APC*, *CDH1*, *CDO1*, *CXCL*, *CYGB*, *DAL-1*, *DAPK*, *DCR2*, *FAM19A4*, *FHIT*, *GATA*, *H-cadherin*, *HOXA9*, *JPH3*, *KIFLA*, *MAGE*, *p16*, *PAX5*, *PCDH20*, *PHACTR3*, *PRDM14*, *RAR β* , *RASSF1A*, *SOX17*, *SULF2*, *TAC1*, *TCF2L*, and *ZFP42* (Additional file 1: Table S1). Primers and probes of the targeted genes were designed in the studies [6, 13–34]. A thermocycling protocol ($95^{\circ}\text{C} \times 10$ min; 40 cycles of [$94^{\circ}\text{C} \times 30$ s, $60^{\circ}\text{C} \times 60$ s], $98^{\circ}\text{C} \times 10$ min) was undertaken in a Bio-Rad C1000 (Bio-Rad, Pleasanton, CA). The PCR plate was transferred to the QX100 Droplet Reader (Bio-Rad) for automatic reading of samples in all wells. We used QuantaSoft 1.7.4 analysis software (Bio-Rad) and Poisson statistics to compute droplet concentrations (copies/ μL ; PCR scale). Only tests that had at least 10,000 droplets were used for the ddMSP analysis [36, 37]. All assays were done in triplicates, and one no-template control and two interplate controls were carried along in each experiment.

Quantification of DNA methylation in sputum by qMSP

qMSP was done as previously described [13, 14]. The cycle threshold (Ct) values for each gene were determined. Ct values above 35 were censored according to previous recommendations [13, 14, 55–58]. To determine methylation level of target genes in a given sample, we normalized Ct values of the target genes in relation to that of *myoblast determination protein one* (*MYOD1*) [13, 32]. The percentage of methylated reference (PMR) was defined as target gene/*MYOD1* ratio of the sample divided by target gene/*MYOD1* ratio of the calibrator DNA (methylated control DNA) and multiplying by 100 [14].

Comparison of tolerance of ddMSP and qMSP to PCR inhibitors

To determine tolerance of ddMSP and qMSP to inhibitory substances of PCR, we directly introduced inhibitors, sodium dodecyl sulfate (SDS), and heparin (Sigma-Aldrich Corporation), into the PCR reactions [59, 60]. Differences in the resulting inhibition curves and the half-maximal inhibitory concentrations (IC₅₀) were assessed and compared as described previously [59, 60].

Statistical analysis

We used *t* test to determine significant differences of values of each gene between cases and controls. We used log transformation of the molecular results and applied Pearson's correlation analysis to assess relationship between DNA methylation and demographic characteristics of subjects. We calculated coefficient of variations (CV) to determine the variation between different measurements. We performed the linear regression between different measurements of the assays and the amount of input DNA. We used the receiver-operator characteristic (ROC) curve and area under the curve (AUC) to determine accuracy, sensitivity, and specificity of each gene or the tests. We employed logistic regression models with constrained parameters as in least absolute shrinkage and selection operator (LASSO) based on ROC criterion to eliminate the irrelevant genes and optimize a composite biomarker panel (classifier). The optimal panel of biomarkers was blindly applied to the testing data set to confirm the diagnostic value by comparing the AUC with the goodness-of-fit statistics [61].

Results

ddMSP has higher sensitivity, precision, and reproducibility for quantification of DNA methylation compared with qMSP

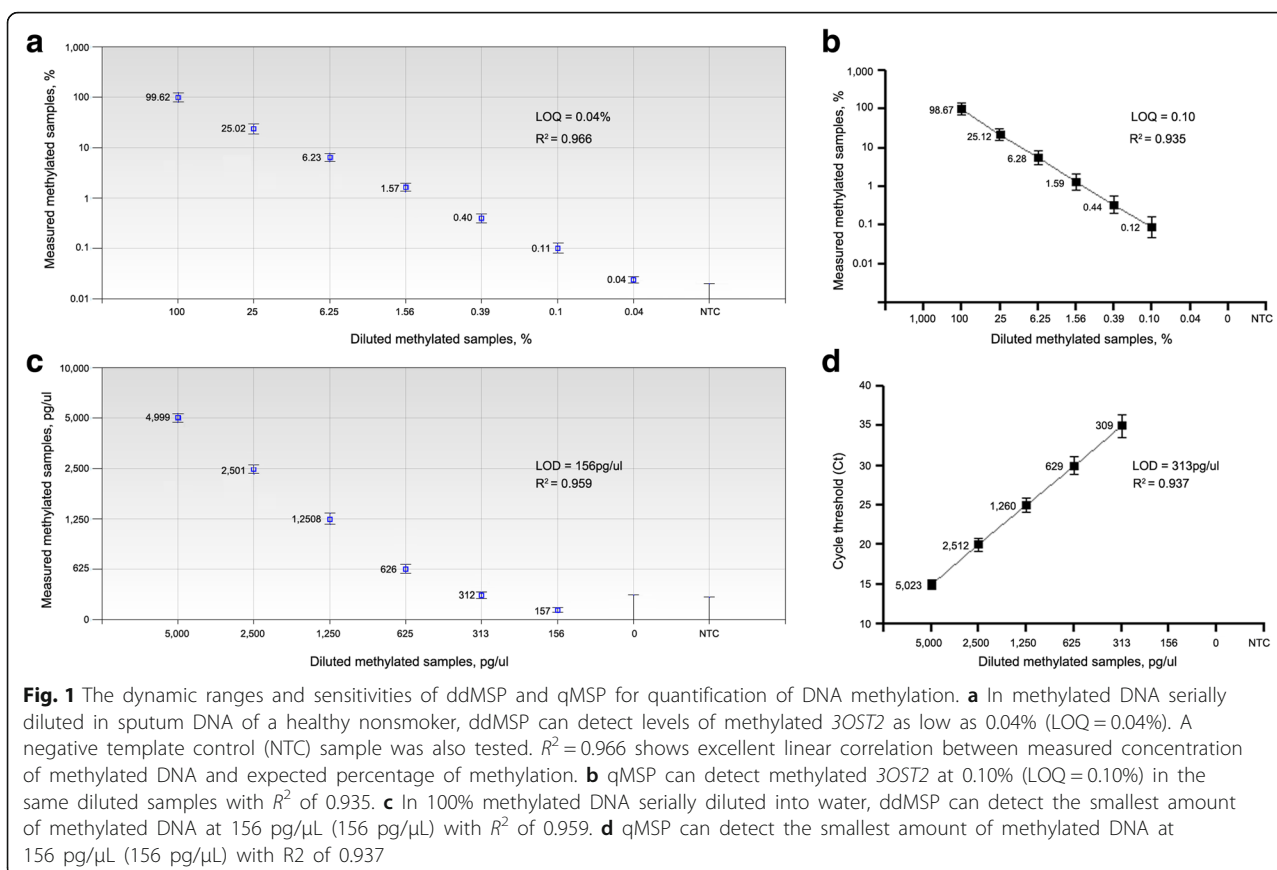
In methylated DNA serially diluted in sputum DNA of a healthy nonsmoker, ddMSP generated at least 10,000 droplets passing through a fluorescence detector. The results suggested that the specimens were successfully

“read” by ddPCR. ddMSP detected methylated genes (*RASSF1A*, *3OST2*, and *PRDM14*) at a concentration of 0.04% (LOQ = 0.04%) ($R^2 = 0.966$) (Fig. 1a), whereas qMSP detected the methylation at a concentration of 0.10% (LOQ = 0.10%) ($R^2 = 0.935$) ($p = 0.008$) (Fig. 1b). There was excellent linearity between the methylated DNA input and values measured by both qMSP and ddMSP (all $R^2 \geq 0.93$). Furthermore, the dispersion of values of the four analyses of the specimen was lower with ddMSP than with qMSP. The repeated measurements by ddMSP had a lower CV value compared with those determined by qMSP ($p = 0.03$) (Additional file 1: Table S2). Therefore, ddMSP had a higher precision for quantification of methylation compared to qMSP ($p = 0.03$) (Additional file 1: Table S2).

To determine the absolute LOD of the two platforms, 100% methylated DNA serially diluted into water and then tested by ddMSP and qMSP. The smallest amount of methylated DNA that can be reliably measured by ddMSP was 156 pg/μL (Fig. 1c), suggesting that ddMSP had a LOD of 156 pg/μL. qMSP produced more than 35 Ct values for the samples that had less than 313 pg methylated DNA per microliter, yielding a LOD of 313 pg/μL (Fig. 1d). Therefore, ddMSP had higher sensitivity as demonstrated by lower LOQ and LOD than did qMSP in the serial dilutions of DNA control samples (all $p < 0.001$).

To determine reproducibility of ddMSP and qMSP, the diluted samples were independently analyzed. The CVs of repeated measures by ddMSP on different days were more than twofold lower compared with those determined by qMSP (Additional file 1: Table S3). Furthermore, the CVs of repeated measures by different research staff using ddMSP were at least twofold lower than did those generated by qMSP (Additional file 1: Table S4). Therefore, ddMSP had a higher reproducibility than did qMSP for quantification of DNA methylation.

To evaluate analytic performance of ddMSP and qMSP in clinical sputum samples, sputum of 20 lung cancer patients and 20 cancer-free controls was tested for *RASSF1A*, whose aberrant methylation level was shown to be elevated in sputum of lung cancer patients [13, 14, 55–58]. Each well of the samples contained at least 10,000 droplets (Fig. 2a). Therefore, ddMSP analysis of DNA methylation could successfully be performed in clinical sputum specimens. *RASSF1A* analyzed by both the techniques displayed a high methylation level in lung cancer patients vs. controls (all $p < 0.05$). In the ddMSP assay, a specimen with $\geq$ one copy of DNA methylation of *RASSF1A*-per microliter was considered to be positive. When the criteria was used, of 20 sputum specimens of lung cancer patients, 11 (55%) had positive methylation of the gene detected by ddMSP. In the qMSP assay, a PMR $\geq 1\%$ was classified as positive for *RASSF1A* in a given sample [62].



When the criteria was used, 9 (45%) were positive for *RASSF1A* by qMSP. The same 4 sputum specimens of control subjects had positive methylation of the gene detected by both ddMSP and qMSP. Therefore, ddMSP analysis of DNA methylation of *RASSF1A* in sputum had a higher sensitivity (55%) than did qMSP (45%) ($p = 0.01$) for distinguishing lung cancer patients from control subjects, while maintaining the same specificity (80%) (Additional file 1: Figure S1). Furthermore, the CVs of repeated measures by ddMSP on different days by different researchers were approximately twofold lower compared with those generated by qMSP. Altogether, in clinical sputum specimens, ddMSP also exhibited higher sensitivity, accuracy, and reproducibility than did qMSP for quantification of DNA methylation.

To compare the tolerance of ddMSP and qMSP to PCR inhibitors, we added SDS and heparin directly into the PCR reactions and then calculated log IC₅₀ values from the resulting inhibition curves. We found greater than a half log increase in IC₅₀ of ddMSP over qMSP for both SDS and heparin (all $p < 0.05$), implying that ddMSP tolerated the presence of the inhibitors better than qMSP.

Diagnostic performance of ddMSP quantified-sputum methylation biomarkers for lung cancer

We first evaluated DNA methylation of 29 genes in the training cohort of 127 NSCLC patients and 159 controls. All the 29 genes displayed a higher level of methylation in patients vs. controls (all $p < 0.05$). ROC curve and AUC analysis showed that the genes had 29–88% sensitivities and 26–92% specificities in differentiating lung cancer patients from healthy controls (Additional file 1: Table S1). Since methylation levels of genes did not follow a normal distribution, we used the log transformation of ddPCR results. We then applied multivariate logistic regression models with stepwise regression based on ROC curve to develop a prediction classifier. Four genes (*HOXA9*, *RASSF1A*, *SOX17*, and *TAC1*) were identified as the best biomarkers (all $p < 0.001$) and incorporated into a logistic classifier: Probability of lung cancer = $e^x / (1 + e^x)$, where $x = 1.69 + 1.48 \times \log(\text{HOXA9}) - 1.25 \times \log(\text{RASSF1A}) + 0.27 \times \log(\text{SOX17}) + 0.16 \times \log(\text{TAC1})$. The logistic classifier produced 0.92 AUC for lung cancer detection (Fig. 2b). Furthermore, Pearson correlation among methylation levels of the four genes was low ($p > 0.05$), implying that their diagnostic values were complementary to each other. Using Youden's index, we set up

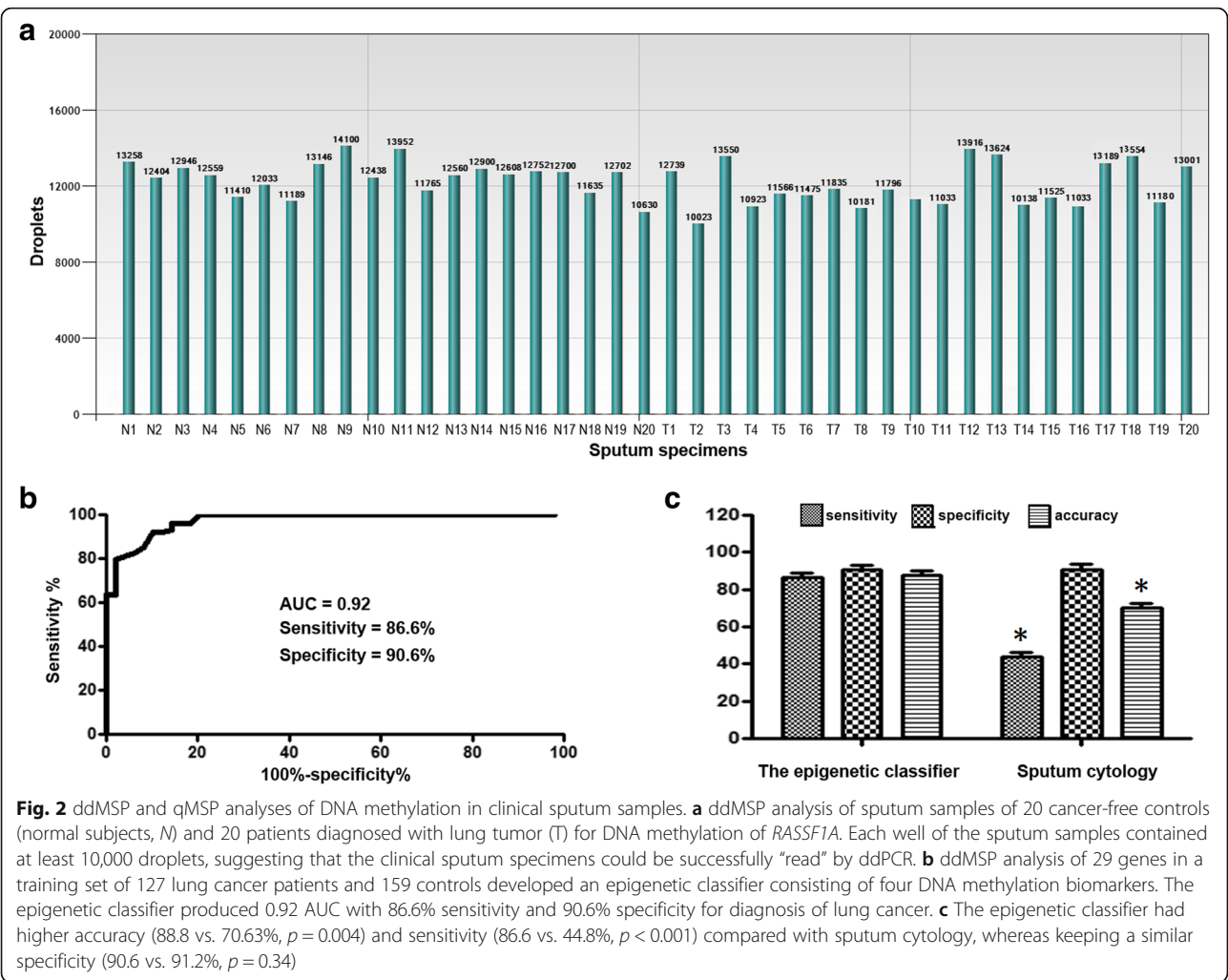


Fig. 2 ddMSP and qMSP analyses of DNA methylation in clinical sputum samples. **a** ddMSP analysis of sputum samples of 20 cancer-free controls (normal subjects, N) and 20 patients diagnosed with lung tumor (T) for DNA methylation of *RASSF1A*. Each well of the sputum samples contained at least 10,000 droplets, suggesting that the clinical sputum specimens could be successfully “read” by ddPCR. **b** ddMSP analysis of 29 genes in a training set of 127 lung cancer patients and 159 controls developed an epigenetic classifier consisting of four DNA methylation biomarkers. The epigenetic classifier produced 0.92 AUC with 86.6% sensitivity and 90.6% specificity for diagnosis of lung cancer. **c** The epigenetic classifier had higher accuracy (88.8 vs. 70.63%, $p = 0.004$) and sensitivity (86.6 vs. 44.8%, $p < 0.001$) compared with sputum cytology, whereas keeping a similar specificity (90.6 vs. 91.2%, $p = 0.34$)

optimal cutoff at 1.28 for the prediction classifier. Subsequently, combined use of the four genes by simply calculating the equation produced 86.6% sensitivity and 90.6% specificity. In addition, including other genes in the prediction classifier did not improve the accuracy for lung cancer diagnosis. The prevalence of the DNA methylation of the four genes was related with pack-years of smoking ($p = 0.03$). Since the cases and controls were matched 1:1 by age, gender, and smoking status as a nested case-control study, we adjusted the parameters during model building. The logistic classifier did not show special association with stage and histological type of lung cancer, and patients’ age, gender, and smoking status (all $p > 0.05$).

Moreover, the logistic classifier had higher accuracy (88.8 vs. 70.63%, $p = 0.004$) and sensitivity (86.6 vs. 44.8%, $p < 0.001$) than did sputum cytology, while maintaining a similar specificity (90.6 vs. 91.2%, $p = 0.34$) (Fig. 2c). The integrated use of the biomarkers and sputum cytology did not significantly increase the diagnostic value.

Validating the panel of ddMSP-quantified methylation biomarkers in a testing cohort

In the testing cohort, the panel of the four genes had 85.4% sensitivity and 91.6% specificity in differentiating lung cancer patients from controls (Table 3). In line with findings in the training set, the logistic classifier was not

Table 3 The diagnostic performance of the epigenetic classifier and sputum cytology in a testing set

	Accuracy (%) (95% CI)*	Sensitivity (%) (95% CI)*	Specificity (%) (95% CI)
The epigenetic classifier	88.78 (83.50 to 92.83)	85.39 (76.32 to 91.99)	91.59 (84.63 to 96.08)
Sputum cytology	71.43 (64.56 to 77.64)	46.07 (35.44 to 56.96)	92.52 (85.80 to 96.72)

Abbreviations: NSCLC non-small cell lung cancer, CI confidence interval
* $p < 0.05$

associated with patient's age, gender, and smoking status, as well as histological type and stage of NSCLC (all $p > 0.05$). Moreover, the logistic classifier yielded higher accuracy (88.8 vs. 71.4%, $p = 0.002$) and sensitivity (85.4 vs. 46.1%, $p < 0.001$), while keeping a similar specificity (91.6 vs. 92.5%, $p = 0.45$), as compared with sputum cytology (Table 3). Taken together, the validation data confirmed the potential of the ddMSP-quantified sputum biomarkers as a sensitive classifier for the early detection of lung cancer.

Discussion

This current study presents the earliest assessment of ddPCR, an emerging technique, for quantitative detection of DNA methylation in sputum. We find that ddMSP can absolutely and robustly quantify DNA methylation in sputum without requiring external references. Therefore, determination of DNA methylation in sputum by ddMSP is highly efficient, and data handling is forthright. Furthermore, our head-to-head comparison of ddMSP and qMSP reveals that ddMSP displays higher precision and reproducibility in measuring copy number of DNA methylation in both control DNA samples and clinical sputum specimens. The sensitivity of conventional qMSP for analyzing low-abundance methylated DNA is poor. This is particularly challenging for the determination of DNA methylation in bronchial epitheliums, since the large excess of normal cells in sputum could obscure detection of the relative scarcity of methylated DNA. We find that ddMSP has a higher sensitivity to quantify cancer-specific methylation and thus could overcome the obstacle of qMSP. In addition, the total time required for ddMSP is about twofold shorter than did qMSP and might be further reduced when an automated system is used [63, 64]. Moreover, ddMSP is not expensive and tolerates the PCR inhibitors better compared with conventional qMSP. Altogether, ddMSP is a straightforward and robust approach for accurate quantification of DNA methylation in sputum.

Importantly, using ddMSP, we develop and validate a DNA methylation-based classifier that has higher accuracy and sensitivity compared with sputum cytology, the clinical gold standard. Furthermore, ddMSP analysis of the four genes by simply calculating the equation would be a convenient tool to be used in the clinics. In addition, the diagnostic performance of the logistic classifier is independent of stage and histological type of the NSCLC, as well as age, gender, and smoking status of subjects. Therefore, the classifier has an important characteristic if it is employed for more precisely and easily identifying early stage lung cancer among smokers.

However, some limitations do exist in this present study: (i) we evaluate ddMSP for quantification of DNA methylation using a retrospective cases and controls,

which may produce selection bias and overfitting. Furthermore, the cases and controls are hospital-based patients, and not representative of the smokers in a screening setting for lung cancer early detection. We will perform a large trial to prospectively validate if the logistic classifier could help identify lung cancer at the early stage in a screening setting among smokers. (ii) the overall sensitivity and specificity of the DNA methylation-based classifier are 86.6 and 90.6%, which are not high enough for clinical diagnosis of NSCLC. The integration of the methylation biomarkers with other classes of biomarkers, such as microRNAs [36, 47–52, 65–67] or DNA mutations [33], is one path to improve the early detection of lung cancer [14]. (iii) since there is no sample left from the patients of the training and cohorts, we are not able to test the classification performance of qMSP in all the samples of the training and testing cohorts. We are consenting new cases and controls and collecting specimens and will then compare performance of qMSP and ddMSP in the same samples. (iv) like qMSP, ddMSP has the same limitations as do other PCR-based platforms. For instance, large experiments can become quite labor intensive to perform, when multiple genes are targeted. In this present study, we use 96-well PCR plates. Given the four genes to be analyzed with duplication, we only test 12 clinical samples at one time. In the future, we will use 484-well PCR plates, in which, we could simultaneously test 60 samples.

Conclusion

ddMSP could be a sensitive and robust tool for reliable quantitation of sputum methylation biomarkers. The ddMSP-quantified methylation classifier may provide a potential diagnostic test for the early detection of lung cancer and thus reduce the deaths and costs associated with the disease. Nevertheless, the continued development of this new technology, and further exploring its value for routine use for the early detection of lung cancer would be required.

Additional file

Additional file 1: Table S1. Twenty-nine genes displayed a higher level of methylation in sputum of lung cancer patients vs. controls. **Table S2.** CVs of the four analyses by ddMSP and qMSP on the same specimens. **Table S3.** CVs of repeated measures by ddMSP and qMSP on different times. **Table S4.** CVs of repeated measures by different researches using ddMSP and qMSP. **Figure S1.** Sensitivities and specificities of measuring DNA methylation of *RASSF1A* by ddMSP and qMSP for diagnosis of lung cancer in sputum samples of 20 cancer-free controls (normal subjects, N) and 20 patients diagnosed with lung tumor (T). ddMSP analysis of DNA methylation of *RASSF1A* had a higher sensitivity (55%) than did qMSP (45%) ($p = 0.01$) for distinguishing lung cancer cases from controls, while maintaining the same specificity (80%). *, $p = 0.01$. (PDF 235 kb)

Acknowledgements

This study was supported in part by a grant for cancer research from Jiangsu Province Hospital of Nanjing University of Chinese Medicine (Y. S.)

Authors' contributions

YS, HF, and FJ conducted the experiments and participated in study design, coordination, and data interpretation, and preparing the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

This study was supported in part by a grant for cancer research from the Jiangsu Province Hospital of Nanjing University of Chinese Medicine. The study was approved by the Nanjing University of Chinese Medicine Institutional Review Board and adhered to the Declaration of Helsinki, and all patients provided written informed consent.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Surgery, Jiangsu Province Hospital of Nanjing University of Chinese Medicine, 138 Xianlin Road, Nanjing 210023, China. ²Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University Medical Center, 4000 Reservoir Road, N.W, Washington D.C. 20057, USA. ³Department of Pathology, University of Maryland School of Medicine, Baltimore, MD, USA.

Received: 24 February 2018 Accepted: 14 May 2018

Published online: 22 May 2018

References

- Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395–409.
- Blumer G. Cigarette smoking and cancer of the lung. *Ill Med J*. 1951;100:98–9.
- Patz EF Jr, Pinsky P, Gatsonis C, Sicks JD, Kramer BS, Tammemagi MC. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med*. 2014;174:269–74.
- Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, Galen B. The national lung screening trial: overview and study design. *Radiology*. 2011; 258:243–53.
- Sant M, Allemani C, Santaquilani M, Knijn A, Marchesi F, Capocaccia R. EUROcare-4. Survival of cancer patients diagnosed in 1995–1999. Results and commentary. *Eur J Cancer*. 2009;45:931–91.
- Hubers AJ, Prinsen CF, Sozzi G, Witte BJ, Thunnissen E. Molecular sputum analysis for the diagnosis of lung cancer. *Br J Cancer*. 2013;109:530–7.
- Belinsky SA. Gene-promoter hypermethylation as a biomarker in lung cancer. *Nat Rev Cancer*. 2004;4:707–17.
- Grossman DC, Curry SJ, Owens DK, Barry MJ, Davidson KW, Doubeni CA. Screening for adolescent idiopathic scoliosis: US preventive services task force recommendation statement. *JAMA*. 2018;319:165–72.
- Brody JS, Spira A. State of the art. Chronic obstructive pulmonary disease, inflammation, and lung cancer. *Proc Am Thorac Soc*. 2006;3:535–7.
- Kadara H, Wistuba II. Field cancerization in non-small cell lung cancer: implications in disease pathogenesis. *Proc Am Thorac Soc*. 2012;9:38–42.
- Belinsky SA, Liechty KC, Gentry FD, Wolf HJ, Rogers J, Vu K. Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res*. 2006;66:3338–44.
- Saccomanno G, Saunders RP, Archer VE, Auerbach O, Kuschner M, Beckler PA. Cancer of the lung: the cytology of sputum prior to the development of carcinoma. *Acta Cytol*. 1965;9:413–23.
- Hubers AJ, Heideman DA, Burgers SA, Herder GJ, Sterk PJ, Rhodius RJ. DNA hypermethylation analysis in sputum for the diagnosis of lung cancer: training validation set approach. *Br J Cancer*. 2015;112:1105–13.
- Su Y, Fang H, Jiang F. Integrating DNA methylation and microRNA biomarkers in sputum for lung cancer detection. *Clin Epigenetics*. 2016;8:109.
- Liu D, Peng H, Sun Q, Zhao Z, Yu X, Ge S. The indirect efficacy comparison of DNA methylation in sputum for early screening and auxiliary detection of lung cancer: a meta-analysis. *Int J Environ Res Public Health*. 2017; 23:14–9.
- Hsu HS, Chen TP, Wen CK, Hung CH, Chen CY, Chen JT. Multiple genetic and epigenetic biomarkers for lung cancer detection in cytologically negative sputum and a nested case-control study for risk assessment. *J Pathol*. 2007;213:412–9.
- Guzman L, Depix MS, Salinas AM, Roldan R, Aguayo F, Silva A. Analysis of aberrant methylation on promoter sequences of tumor suppressor genes and total DNA in sputum samples: a promising tool for early detection of COPD and lung cancer in smokers. *Diagn Pathol*. 2012;7:87.
- Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A*. 1996;93:9821–6.
- Wang X, Cao A, Peng M, Hu C, Liu D, Gu T. The value of chest CT scan and tumor markers detection in sputum for early diagnosis of peripheral lung cancer. *Zhongguo Fei Ai Za Zhi*. 2004;7:58–63.
- Leng S, Wu G, Klinge DM, Thomas CL, Casas E, Picchi MA. Gene methylation biomarkers in sputum as a classifier for lung cancer risk. *Oncotarget*. 2017;8: 63978–85.
- Belinsky SA, Leng S, Wu G, Thomas CL, Picchi MA, Lee SJ. Gene methylation biomarkers in sputum and plasma as predictors for lung cancer recurrence. *Cancer Prev Res (Phila)*. 2017;10:635–40.
- Miyake M, Gomes Giacoia E, Aguilar Palacios D, Rosser CJ. Lung cancer risk assessment for smokers: gene promoter methylation signature in sputum. *Biomark Med*. 2012;6:512.
- Leng S, Do K, Yingling CM, Picchi MA, Wolf HJ, Kennedy TC. Defining a gene promoter methylation signature in sputum for lung cancer risk assessment. *Clin Cancer Res*. 2012;18:3387–95.
- Hwang SH, Kim KU, Kim JE, Kim HH, Lee MK, Lee CH. Detection of HOXA9 gene methylation in tumor tissues and induced sputum samples from primary lung cancer patients. *Clin Chem Lab Med*. 2011;49:699–704.
- Liu Y, Lan Q, Shen M, Jin J, Mumford J, Ren D. Aberrant gene promoter methylation in sputum from individuals exposed to smoky coal emissions. *Anticancer Res*. 2008;28:2061–6.
- Belinsky SA, Grimes MJ, Casas E, Stidley CA, Franklin WA, Bocklage TJ. Predicting gene promoter methylation in non-small-cell lung cancer by evaluating sputum and serum. *Br J Cancer*. 2007;96:1278–83.
- Su SB, Yang LJ, Zhang W, Jin YL, Nie JH, Tong J. p16 and MGMT gene methylation in sputum cells of uranium workers. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi*. 2006;24:92–5.
- Belinsky SA, Klinge DM, Dekker JD, Smith MW, Bocklage TJ, Gilliland FD. Gene promoter methylation in plasma and sputum increases with lung cancer risk. *Clin Cancer Res*. 2005;11:6505–11.
- Olaussen KA, Soria JC, Park YW, Kim HJ, Kim SH, Ro JY. Assessing abnormal gene promoter methylation in paraffin-embedded sputum from patients with NSCLC. *Eur J Cancer*. 2005;41:2112–9.
- Hubers AJ, Heideman DA, Duin S, Witte BJ, de Koning HJ, Groen HJ. DNA hypermethylation analysis in sputum of asymptomatic subjects at risk for lung cancer participating in the NELSON trial: argument for maximum screening interval of 2 years. *J Clin Pathol*. 2017;7:250–4.
- Hubers AJ, Brinkman P, Boksem RJ, Rhodius RJ, Witte BJ, Zwinderman AH. Combined sputum hypermethylation and eNose analysis for lung cancer diagnosis. *J Clin Pathol*. 2014;67:707–11.
- Hubers AJ, van der Drift MA, Prinsen CF, Witte BJ, Wang Y, Shivapurkar N. Methylation analysis in spontaneous sputum for lung cancer diagnosis. *Lung Cancer*. 2014;84:127–33.
- Hubers AJ, Heideman DA, Yatabe Y, Wood MD, Tull J, Taron M, et al. EGFR mutation analysis in sputum of lung cancer patients: a multitechnique study. *Lung Cancer*. 2013;82(1):38–43.
- Hubers AJ, Heideman DA, Herder GJ, Burgers SA, Sterk PJ, Kunst PW. Prolonged sampling of spontaneous sputum improves sensitivity of hypermethylation analysis for lung cancer. *J Clin Pathol*. 2012;65:541–5.
- Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ. Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat Methods*. 2013;10:1003–5.
- Li N, Ma J, Guarnera MA, Fang H, Cai L, Jiang F. Digital PCR quantification of miRNAs in sputum for diagnosis of lung cancer. *J Cancer Res Clin Oncol*. 2014;140:145–50.
- Ma J, Li N, Guarnera M, Jiang F. Quantification of plasma miRNAs by digital PCR for cancer diagnosis. *Biomark Insights*. 2013;8:127–36.
- Bhat S, Herrmann J, Armishaw P, Corbisier P, Emslie KR. Single molecule detection in nanofluidic digital array enables accurate measurement of DNA copy number. *Anal Bioanal Chem*. 2009;394:457–67.

39. Kiss MM, Ortoleva-Donnelly L, Beer NR, Warner J, Bailey CG, Colston BW. High-throughput quantitative polymerase chain reaction in picoliter droplets. *Anal Chem*. 2008;80:8975–81.
40. Kreutz JE, Munson T, Huynh T, Shen F, Du W, Ismagilov RF. Theoretical design and analysis of multivolume digital assays with wide dynamic range validated experimentally with microfluidic digital PCR. *Anal Chem*. 2011;83: 8158–68.
41. Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, Bhat S. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal Chem*. 2012;84:1003–11.
42. Pohl G, Shih IM. Principle and applications of digital PCR. *Expert Rev Mol Diagn*. 2004;4:41–7.
43. Vogelstein B, Kinzler KW. Digital PCR. *Proc Natl Acad Sci U S A*. 1999;96: 9236–41.
44. Hayden RT, Gu Z, Ingersoll J, Abdul-Ali D, Shi L, Pounds S, et al. Comparison of droplet digital PCR to real-time PCR for quantitative detection of cytomegalovirus. *J Clin Microbiol*. 2013;51(2):540–6.
45. Day E, Dear PH, McCaughan F. Digital PCR strategies in the development and analysis of molecular biomarkers for personalized medicine. *Methods*. 2013;59:101–7.
46. Li H, Jiang Z, Leng Q, Bai F, Wang J, Ding X. A prediction model for distinguishing lung squamous cell carcinoma from adenocarcinoma. *Oncotarget*. 2017;8:50704–14.
47. Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z. Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. *Int J Cancer*. 2010;127:2870–8.
48. Xing L, Todd NW, Yu L, Fang H, Jiang F. Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers. *Mod Pathol*. 2010; 23:1157–64.
49. Xie Y, Todd NW, Liu Z, Zhan M, Fang H, Peng H. Altered miRNA expression in sputum for diagnosis of non-small cell lung cancer. *Lung Cancer*. 2010; 67:170–6.
50. Anjuman N, Li N, Guarnera M, Stass SA, Jiang F. Evaluation of lung flute in sputum samples for molecular analysis of lung cancer. *Clin Transl Med*. 2013;2:15.
51. Jiang F, Todd NW, Li R, Zhang H, Fang H, Stass SA. A panel of sputum-based genomic marker for early detection of lung cancer. *Cancer Prev Res (Phila)*. 2010;3:1571–8.
52. Jiang F, Todd NW, Qiu Q, Liu Z, Katz RL, Stass SA. Combined genetic analysis of sputum and computed tomography for noninvasive diagnosis of non-small-cell lung cancer. *Lung Cancer*. 2009;66:58–63.
53. Qiu Q, Todd NW, Li R, Peng H, Liu Z, Yfantis HG. Magnetic enrichment of bronchial epithelial cells from sputum for lung cancer diagnosis. *Cancer*. 2008;114:275–83.
54. Li R, Todd NW, Qiu Q, Fan T, Zhao RY, Rodgers WH. Genetic deletions in sputum as diagnostic markers for early detection of stage I non-small cell lung cancer. *Clin Cancer Res*. 2007;13:482–7.
55. Li W, Deng J, Jiang P, Zeng X, Hu S, Tang J. Methylation of the RASSF1A and RARB genes as a candidate biomarker for lung cancer. *Exp Ther Med*. 2012;3:1067–71.
56. Lee SM, Lee WK, Kim DS, Park JY. Quantitative promoter hypermethylation analysis of RASSF1A in lung cancer: comparison with methylation-specific PCR technique and clinical significance. *Mol Med Rep*. 2012;5:239–44.
57. Fischer JR, Ohnmacht U, Rieger N, Zemaitis M, Stoffregen C, Manegold C. Prognostic significance of RASSF1A promoter methylation on survival of non-small cell lung cancer patients treated with gemcitabine. *Lung Cancer*. 2007;56:115–23.
58. Chen H, Suzuki M, Nakamura Y, Ohira M, Ando S, Iida T. Aberrant methylation of RASGRF2 and RASSF1A in human non-small cell lung cancer. *Oncol Rep*. 2006;15:1281–5.
59. Dingle TC, Sedlak RH, Cook L, Jerome KR. Tolerance of droplet-digital PCR vs real-time quantitative PCR to inhibitory substances. *Clin Chem*. 2013;59: 1670–2.
60. Wilson IG. Inhibition and facilitation of nucleic acid amplification. *Appl Environ Microbiol*. 1997;63:3741–51.
61. Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol*. 1982;115: 92–106.
62. Grote HJ, Schmiemann V, Gedert H, Bocking A, Kappes R, Gabbert HE. Methylation of RAS association domain family protein 1A as a biomarker of lung cancer. *Cancer*. 2006;108:129–34.
63. Kurimoto K, Hayashi M, Guerrero-Preston R, Koike M, Kanda M, Hirabayashi S. PAX5 gene as a novel methylation marker that predicts both clinical outcome and cisplatin sensitivity in esophageal squamous cell carcinoma. *Epigenetics*. 2017;12: 865–74.
64. Campomenosi P, Gini E, Noonan DM, Poli A, D'Antona P, Rotolo N. A comparison between quantitative PCR and droplet digital PCR technologies for circulating microRNA quantification in human lung cancer. *BMC Biotechnol*. 2016;16:60.
65. Shen J, Liao J, Guarnera MA, Fang H, Cai L, Stass SA. Analysis of MicroRNAs in sputum to improve computed tomography for lung cancer diagnosis. *J Thorac Oncol*. 2014;9:33–40.
66. Su J, Liao J, Gao L, Shen J, Guarnera MA, Zhan M. Analysis of small nucleolar RNAs in sputum for lung cancer diagnosis. *Oncotarget*. 2016;7:5131–42.
67. Xing L, Su J, Guarnera MA, Zhang H, Cai L, Zhou R. Sputum microRNA biomarkers for identifying lung cancer in indeterminate solitary pulmonary nodules. *Clin Cancer Res*. 2015;21:484–9.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



A Direct Plasma miRNA Assay for Early Detection and Histological Classification of Lung Cancer



Qixin Leng^{*}, Yue Wang[†] and Feng Jiang^{*}

^{*}Department of Pathology, University of Maryland School of Medicine, Baltimore, MD 21201, USA; [†]Customer Value Partners, Towson, MD 21286, USA

Abstract

Cell-free microRNAs in plasma provide circulating biomarkers for lung cancer. Most techniques for analysis of miRNAs require a large plasma volume to purify a sufficient RNA yield followed by complicated downstream processing. Small differences in the multiple procedures often cause large analytical variations and poor diagnostic values of the plasma biomarkers. Here we investigate whether directly quantifying plasma miRNAs without RNA purification could diagnose lung cancer. FirePlex assay was directly applied to 20 μ l plasma of 56 lung cancer patients and 28 cancer free controls for quantifying 11 lung tumor-associated miRNAs. FirePlex assay is easier, less expensive and time-consuming for quantification of plasma miRNAs compared with conventional reverse transcription PCR with an equivalent analytic performance. From the lung tumor-associated miRNAs, a prediction model based on two miRNAs (miRs-205-5p and -210-3p) was developed, producing 78.6% sensitivity and 89.3% specificity for identifying lung cancer. The diagnostic value was independent of stage of lung tumor, and patients' age and sex (all $P > 0.05$). Furthermore, based on the same two miRNAs, additional prediction models were developed with 75.0% sensitivity and 89.3% specificity for diagnosis of lung squamous cell carcinoma, and 82.2% sensitivity and 89.3% specificity for lung adenocarcinoma. The direct plasma assay can improve the efficacy of miRNA assessment in a small plasma volume by reducing multiple procedure-associated analytical variables. The developed plasma miRNA biomarkers might be useful for the early detection and histological classification of lung cancer.

Translational Oncology (2018) 11, 883–7

Introduction

Non-small-cell lung cancer (NSCLC) accounts for approximately 85% of lung cancer that is the number one cancer killer [1]. NSCLC mainly consists of two major histological subtypes: squamous cell carcinoma (SCC) and adenocarcinoma (AC) [2]. The 5-year survival rate for stage IV NSCLC is only 10%, whereas nearly 85% for stage IA NSCLC [1]. These statistics provide the primary rationale to improve the early detection of NSCLC. Furthermore, histological subtype is predictive of a differential response rate, overall survival, or toxicity profile from certain therapies. Precisely classifying subtypes of lung cancer is also clinically important for the personalized treatment of the malignancy [3–5]. MicroRNAs (miRNAs) are small molecules that have important functions in diverse biological processes, including cell proliferation, differentiation, and apoptosis [6–8]. Furthermore, dysfunction of miRNAs contributes to the development and progression of human malignancies, including lung cancer [9]. In addition, miRNA expression profiles represent molecular

signatures for the classification, diagnosis, and progression of tumor, and thus could be developed as cancer biomarkers [8,10]. Blood is one of the most easily and noninvasively accessible body fluids. The use of blood plasma for assessment of circulating miRNAs originating from primary tumors would provide a useful tool for the early detection and histological classification of lung cancer [11–13].

The most common techniques used in analysis of plasma miRNAs mainly include quantitative reverse transcription polymerase chain

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

Received 23 April 2018; Revised 30 April 2018; Accepted 1 May 2018

© 2018 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/>).

1936-5233/18

<https://doi.org/10.1016/j.tranon.2018.05.001>

reaction (qRT-PCR) and microarray [14]. Being sensitive and quantitative, qRT-PCR is relatively inexpensive and flexible, and thus the preferred choice for assessment of limited number of miRNAs in relatively small experiments. Compared with qRT-PCR, microarrays can target more miRNAs for larger studies, however, have a low detection sensitivity. Since the cancer-associated miRNAs are derived from primary tumor and could be ‘diluted’ in a background of normal miRNAs, the circulating miRNAs are present in very low abundance in plasma [11,15,16]. In order to profile the low-abundance miRNAs in plasma, these current techniques require a large plasma volume to purify a sufficient RNA yield followed by complex downstream workflow, including reverse transcription [17]. However, small differences in the complex processes often result in large analytical variations and poor diagnostic performances of the plasma biomarkers [14,17].

FirePlex technology is a new approach for quantification of miRNAs [18]. This technique is dependent on hydrogel particles that contain custom selected probes against target miRNAs. miRNAs bind to these probes and are then ligated to adaptor sequences for pre-detection amplification. Particles are optimized for use with common bench-top flow cytometers, allowing detection without specialized laboratory equipment. Furthermore, data analysis is relatively straightforward and does not require advanced bioinformatics skills. This technique can quantify multiple miRNAs across a range of samples without the labor-intensive workflow or large sample requirement for other techniques. Importantly, it can be used directly with crude body fluids without RNA extraction and reverse transcription. In addition, miRNA profiling by the FirePlex technology can be achieved from input of as little as 10 µL of plasma or 100 pg purified RNA [18]. Therefore, FirePlex technology for multiplex miRNA profiling may overcome the obstacles of the current techniques by eliminating the errors induced during sample purification and insufficient methods for analytical performance. The objective of the present study was to use FirePlex miRNA assay to develop plasma miRNA biomarkers for lung cancer early detection and histological classification.

Materials and Methods

Patients and Clinical Specimens

The Institutional Review Boards of University of Maryland Baltimore and Veterans Affairs Maryland Health Care System

approved this study. We recruited lung cancer patients and cancer-free smokers by using the inclusion and/or exclusion criteria recommended by U.S. Preventive Services Task Force for lung cancer screening in heavy smokers [19]. We collected blood in BD Vacutainer spray-coated K2EDTA Tubes (BD, Franklin Lakes, NJ) and prepared plasma using the standard operating protocols developed by The NCI-Early Detection Research Network within 2 hours of the collection [20]. A total of 56 NSCLC patients and 28 cancer-free smokers were recruited. Among the cancer patients, 9 patients were female and 47 were male. Twelve had stage I NSCLCs, 6 with stage II, 12 with stage III, 18 with stage IV, and 7 with unknown stage. Histologically, 28 lung cancer patients were diagnosed with AC, while 28 with SCC. Of the cancer-free smokers, 2 patients were female and 28 were male. The cancer-free patients who were smokers and served as control subjects had granulomatous inflammation (n = 14), nonspecific inflammatory changes (n = 12) or lung infections (n = 2). The cancer-free smokers had been followed for at least 2 years, and none had any evidence of cancer. The demographic and clinical variables of the cohort are shown in Table 1.

miRNA Quantification by the FirePlex Assay

Using next-generation deep sequencing to comprehensively characterize miRNA profiles in eight lung tumor tissues, we identified miRNAs whose dysregulation was associated with NSCLC [21]. From the miRNAs, we found that 11 miRNAs might be potential plasma biomarkers for lung cancer [12,13,21–26]. The 11 miRNAs were analyzed in this study, including miRs-205-5p, -145, -422a, -34a-5p, -93-5p, -223-3p, -210-3p, and 628-3p, and lets-7d-5p, -7g-5p, and -7i-5p.

Plasma 20 µL was mixed with 36 µL Digest Buffer, 20 µL water and 4 µL Protease Mix and incubated at 60°C for 45 minutes with shaking. For each sample run, FirePlex Particles (35 µL) were added to a well of a 96-well filter plate and filtered. Twenty-five µL Hybe Buffer was added to each well. One ng total RNA was loaded. The plate was incubated at 37°C for 60 minutes with shaking. After rinsing twice with 1× Rinse A, 75 µL of 1× Labeling Buffer was added to each well. The plate was incubated at room temperature for 60 minutes with shaking. After two rinses with 1× Rinse B followed by one rinse with 1× Rinse A, a catch plate was added to the vacuum manifold and the filter plate put under constant vacuum. 65 µL of 95°C RNase-free water was added twice to each well to elute the ligated sample. 30 µL of this melt-off was added to a clean PCR plate and mixed with 20 µL PCR master mix. The mixture underwent 32 cycles of PCR amplification. 60 µL of Hybe Buffer was added back to each well of the original particles followed by 20 µL of the PCR product, and the plate was incubated at 37°C for 30 minutes with shaking. After rinsing twice with 1× Rinse B followed by one rinse with 1× Rinse A, 75 µL of 1× Reporting Buffer was added to each well and the plate incubated at room temperature for 15 minutes with shaking. After rinsing twice with 1× Rinse A, 175 µL of Run Buffer was added to each well. The samples were then scanned on an EMD Millipore Guava 6HT flow cytometer (MilliporeSigma, Germany). Flow cytometry quantification data was analyzed with the FirePlex™ Analysis Workbench software as previously described [18].

RNA Isolation and qRT-PCR

Plasma 200 µL was used for RNA extraction by Trizol LS reagent (Invitrogen Carlsbad, CA) and RNeasy Mini Kit (Qiagen, Hilden, Germany) [22,26,27]. RT was carried out to generate cDNA by using

Table 1. Characteristics of NSCLC Patients and Cancer-Free Smokers

	NSCLC Cases (n = 56)	Controls (n = 28)	P-Value
Age	65.68 (SD 10.09)	62.37 (SD 9.26)	0.262
Sex			
Female	9	2	0.2
Male	47	26	
Pack-years (median)	52.88	48.75	0.37
Stage			
Stage I	12		
Stage II	6		
Stage III	13		
Stage IV	18		
Unknown stage	7		
Histological type			
Adenocarcinoma	28		
Squamous cell carcinoma	28		

Abbreviations: NSCLC, non-small cell lung cancer; SD, standard deviation.

a RT Kit (Applied Biosystems, Foster City, CA) as described in our published works [22,26,27]. PCR was performed to measure expressions of target genes by using a PCR kit (Applied Biosystems) on a Bio-Red IQ5 Multi-color RT-PCR Detection System (Bio-Red, Hercules, CA). Expression levels of the genes were determined using comparative cycle threshold (C_T) method with miR-1228 as an internal control. The targeted genes with C_T values >35 were considered to be below the detection level of qRT-PCR [28]. All reactions were run in triplicate. C_T values were converted into absolute copy numbers using a standard curve from a synthetic miRNA as previously described [22,26,27].

Statistical Analysis

We used the area under receiver operating characteristic (ROC) curve (AUC) with the null hypothesis (H_0) of 0.5 to determine sample size. 28 subjects were required in each category of cases and controls to show a minimum difference of interest between an AUC of 0.75 versus an AUC of 0.5 with 80% power at the 5% significance level. To determine if a particular miRNA signal was detectable, assay background was subtracted from the miRNA signal measured in the samples and the result was compared to a limit of detection. The detection threshold was set based on comparison of fluorescence intensity of sample wells to negative (water) wells by calculating $[(3S_{ww})^2 + (4S_{pp})^2]^{1/2}$. S_{ww} was the well-to-well standard deviation of the mean signals, and S_{pp} was the mean of the particle-particle standard errors. The detection threshold was set at 2.00 as previously established [18]. We generated the heatmap based on a grid of raw and background subtracted miRNA signal intensities indicated by color. We used an ANOVA to determine if the groups exhibit significantly different expression levels for a particular miRNA, and the ANOVA results were reported as an unadjusted P -value. Differential expression with P -values less than 0.05 indicated that the two groups were different with at least a 95% confidence level. Pearson's correlation analysis was utilized to assess relationship between plasma miRNA expressions and demographic characteristics of the patients and cancer-free controls. Correlation between FirePlex versus qRT-PCR output analyses was tested by using linear regression models. The precision of miRNA measurements was estimated by using coefficient of variation [$CV = (SD/mean) \times 100$] of quadruplicate measures for each sample. The ROC and AUC analyses were applied to determine sensitivity and specificity of each miRNA. We used logistic regression models with constrained parameters as in least absolute shrinkage and selection operator (LASSO) to eliminate the irrelevant genes and develop prediction models with the highest sensitivity and specificity.

Results

Directly Targeting Plasma Samples for Quantification of miRNAs

The releases of miRNAs in plasma by hemolysis of blood cells often produce nonspecific and low reproducible results [16,29–32]. To determine if the plasma had hemolysis of blood cells, we included hemolysis-associated miRNA markers (miRs-451a and -486-5) in the FirePlex miRNA Assay. All the specimens had negative results of the hemolysis miRNA markers, suggesting no hemolysis in the plasma samples. All the 11 miRNAs, including miRs-205-5p, -145, -422a, -34a-5p, -93-5p, -223-3p, -210-3p, and 628-3p, and lets-7d-5p, -7g-5p, and -7i-5p, had background-subtracted signal

above the detection threshold (Figure 1). Therefore, the FirePlex miRNA Assay could quantify the 11 miRNAs in plasma.

To determine the robustness of the FirePlex assay in miRNA quantification, we performed qRT-PCR in plasma samples of 20 cancer-free controls and 20 lung cancer patients for quantitation of a miRNA, miRs-34a-5p. There was no statistical difference of the CVs of miRNA levels determined by FirePlex vs. qRT-PCR in the same samples ($P = 0.26$) (Supplementary Figure 1), implying that FirePlex had comparable precision and repeatability compared with qRT-PCR for quantification of plasma miRNA. Furthermore, there was a close correlations of expression levels of the miRNA determined by the two platforms (Pearson $r = 0.9961$, $P < 0.0001$) (Supplementary Figure 2). In addition, the total time required for the FirePlex assay is 3–5 hours, whereas 24 hours for qRT-PCR. Altogether, the FirePlex assay is a less time-consuming technique for quantification of plasma miRNAs with an equivalent analytic performance compared with qRT-PCR.

Plasma Biomarker Panels for the Early Detection and Histological Classification of Lung Cancer

Of the 11 miRNAs that robustly detectable in the plasma samples, 10 miRNAs (miRs-93-5p, 628-3p, -422a, -34a-5p, -223-3p, -210-3p, and -205-5p, and lets-7d-5p, -7g-5p, and -7i-5p) had a different expression level in lung cancer patients vs. cancer-free individuals (All $P < 0.05$) (Figure 2) (Supplementary Figure 3). We used ROC to evaluate the diagnostic performance of the miRNAs for lung cancer. The 10 miRNAs showed 0.663–0.77 AUC values with 58.93–73.21% sensitivity and 60.25–82.14% specificity (Table 2), implying that the miRNAs held promise as plasma biomarkers for lung cancer.

We used logistic regression models with constrained parameters as in LASSO and AUCs to optimize a marker panel for lung cancer. From the 10 genes, a logistic regression model with 2 miRNAs (miRs-205-5p and -210-3p) was developed for diagnosing lung cancer: $U = -1.44 + 0.05 \times \text{miR205-5p} + 0.07 \times \text{miR-210-3p}$. Incorporated analysis of the 2 biomarkers by using the logistic regression model produced a higher AUC (0.85) than did the individual biomarkers (Figure 3A) (Table 2) ($P < 0.05$). We calculated the distance to the perfect point (0, 1) with varying cut-offs for U , and the cut-off corresponding to the shortest distance in the AUC analysis was considered the optimal one. The optimal cut-off for the logistic regression model was $U = 0.62$. Any subject with $U \geq 0.62$ was classified as a lung cancer case. As a result, combined analysis of the 2 miRNAs by using the logistic regression model yielded significantly higher sensitivity (78.6%) and specificity (89.3%) compared with the individual biomarkers (all $P < 0.05$) (Table 2). Furthermore, based on the 2 miRNAs, a second logistic regression model was developed for diagnosis of AC with AUC of 0.82: $U = -1.69 + 0.05 \times \text{miR-205-5p} + 0.03 \times \text{miR-210-3p}$ (Figure 3B). When cut-off was set at $U = 0.53$, this logistic regression model produced 75.0% sensitivity and 89.3% specificity for diagnosis of AC. In addition, using the same 2 miRNAs, a third logistic regression model was established for diagnosis of SCC with AUC of 0.88: $U = -2.62 + 0.043 \times \text{miR-205-5p} + 0.1 \times \text{miR-210-3p}$ (Figure 3C). When cut-off was set at $U = 0.35$, this logistic regression model produced 82.1% sensitivity and 89.3% specificity for diagnosis of SCC.

Combined use of all the 10 miRNA genes did not produce higher sensitivity and specificity compared with the two miRNAs (miRs-205-5p and -210-3p) used in combination ($P > 0.05$).

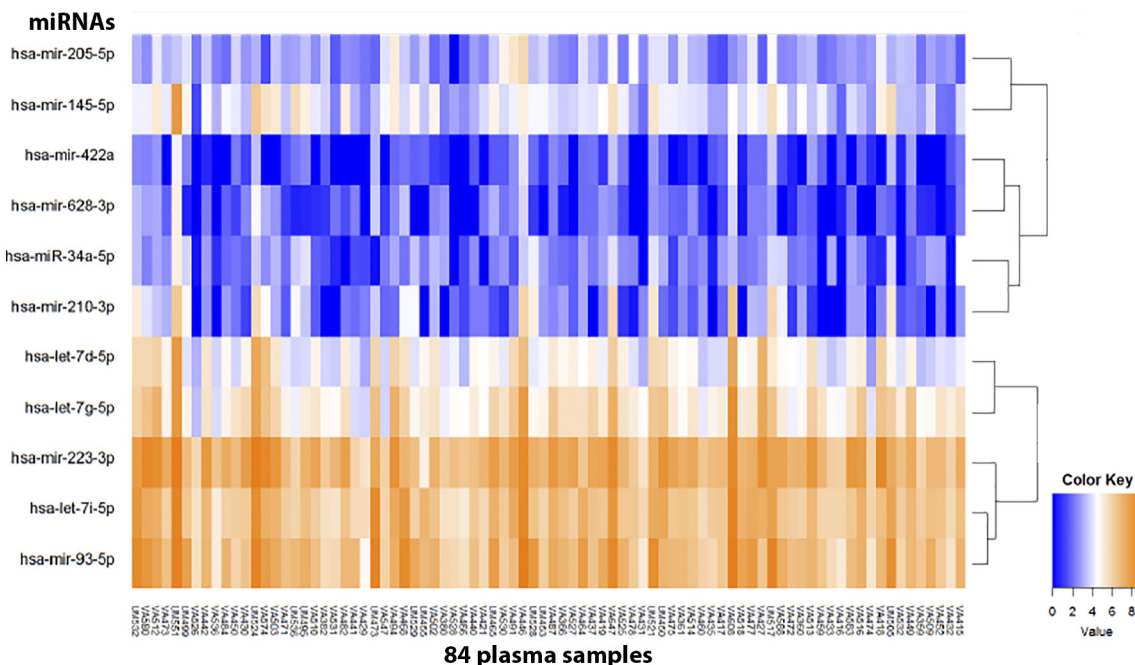


Figure 1. Directly targeting 84 plasma samples for quantifying 11 miRNAs. The heatmap is a grid of raw, background subtracted miRNA signal intensities indicated by color. The 11 genes have background-subtracted signal above the detection threshold (2.00) and thus are reliably quantified in plasma.

Furthermore, the estimated correlations among levels of the 2 genes were very low (Pearson $r = -0.0219$, $P = 0.7720$) (Supplementary Figure 4), implying that the integration of the two molecular biomarkers has complementary classification. In addition, the logistic regression models had no special association with age, gender, and smoking status of the participants (All $P > 0.05$). Moreover, the three logistic regression models with the same two miRNAs had similar sensitivity and specificity for diagnosis of different stages of NSCLC (Supplementary Figure 5).

Discussion

The current techniques for the assessment of plasma miRNAs require a large plasma volume followed by the complicated downstream handling. Small differences in the processing will lead to poor analytical and diagnostic values of plasma biomarkers. We propose to avoid the multiple procedures by direct analyzing plasma using the

FirePlex technology for quantifying multiple miRNAs. Our head-to-head comparison of the FirePlex Assay and qRT-PCR demonstrates that the FirePlex can eliminate intra-assay variables and improve reliability of miRNA quantification by directly targeting crude plasma and minimizing the complicated workflow. Furthermore, unlike qRT-PCR requiring a large sample volume (200 μL), the FirePlex assay can simultaneously quantify multiple miRNAs from input of as little as 20 μL of plasma. In addition, without the expensive procedures including RNA isolation, RNA quantification, and RT, the FirePlex technology can be used at less cost compared with qRT-PCR. The FirePlex assay only takes 3–5 hours from sample to data, and thus requires a shorter turnaround time compared with qRT-PCR demanding 24 hours. Moreover, the data handling for the FirePlex technology is straightforward for fast analysis and interpretation, as the results are obtained by using conventional flow cytometry with simple software. Therefore, the FirePlex assay might

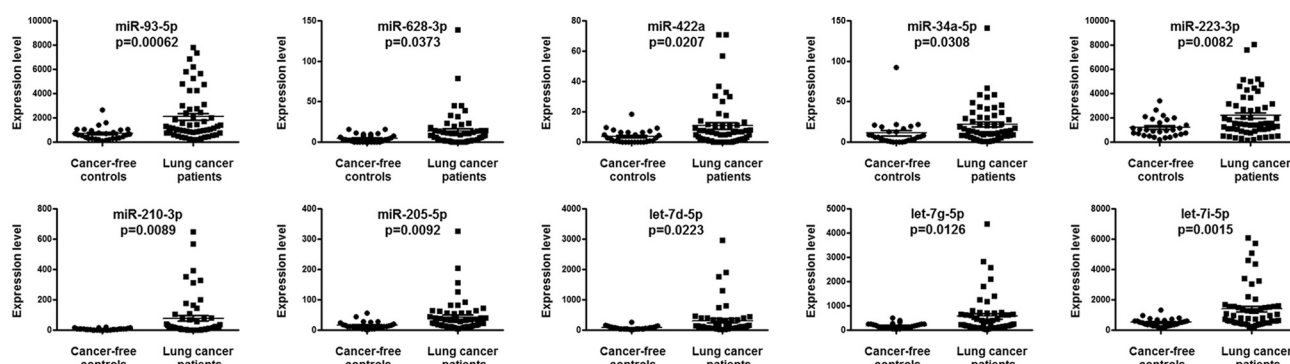


Figure 2. Of the 11 miRNAs tested in plasma samples of cancer-free smokers vs. lung cancer patients, 10 miRNAs have a significantly different level in the lung cancer patients vs. cancer-free smokers (all $P < 0.05$). The inside line denotes the median.

Table 2. miRNAs with a Different Expression Level in Lung Cancer Patients vs. Cancer-Free Smokers

miRNAs	AUC	Standard Deviation	P Value	Sensitivity (%)	Specificity (%)
mir-93-5p	0.7691	0.0518	<0.0001	71.43	71.43
miR-628-3p	0.6633	0.0600	0.0152	58.93	60.71
mir-422a	0.6773	0.0599	0.0084	62.5	60.71
miR-34a-5p	0.7165	0.0597	0.0013	67.86	67.86
mir-223-3p	0.6709	0.0592	0.0110	73.21	60.71
mir-210-3p	0.7545	0.0518	0.0002	64.29	82.14
miR-205-5p	0.7519	0.0543	0.0002	69.64	67.86
lets-7d-5p	0.6823	0.0368	0.0013	63.92	60.25
let-7i-5p	0.6678	0.0156	0.0002	59.66	61.26
let-7g-5p	0.7126	0.0387	0.0036	68.23	67.53
A panel of 2 miRNAs	0.8510	0.0162	0.0001	78.57	89.29

Abbreviations: AUC, area under the receiver-operator characteristic curve.

provide an efficient means for quantification of the plasma miRNAs whose changes are the hallmark of lung cancer.

Importantly, we develop 2 miRNA biomarkers, which used in combination could diagnose lung cancer. Since the biomarkers show similar sensitivity and specificity in the early vs. advanced stages of NSCLC, it might be a useful approach for the early detection of lung cancer, a clinically challenging. Furthermore, the use of the 2 miRNA biomarkers would potentially improve personalized treatment of lung cancer by precisely classifying the subtypes. In addition, our analysis of the plasma miRNAs by simply calculating the equation with the established cut-off could be a convenient tool in the clinics. Therefore, the developed biomarkers with the easy-to-use prediction models may represent promising noninvasive diagnostic tools not only for early detection of lung cancer, but improving the efficacy of therapeutics of NSCLC.

Increased miR-205-5p expression contributes to the development and progression of NSCLC [21,33]. We have shown that miR-205-5p is one of the miRNAs that could be used as biomarkers for the early detection of lung cancer [21,24,34]. miR-210-3p stimulates a hypoxic phenotype and upsurges radioresistance in NSCLCs [35]. Hypoxia-induced miR-210-3p can regulate tumor cell susceptibility to cytolytic T-lymphocyte-mediated lysis by a mechanism involving its downstream targets PTPN1, HOXA1, and

TP53I11 [36]. The results created from our present study support these previous findings, and further provide evidence of using the miRNAs as potential biomarkers for diagnosis of lung cancer.

There are limitations in the present study. 1), the sample size is small. Furthermore, this single and retrospective cohort of cases and controls may produce selection bias. To diminish the bias, we will perform pivotal evaluation of the diagnostic assay in a large cohort by using prospective-specimen collection, retrospective-blinded-evaluation design developed by NCI [37]. 2), in the present study, we only assess 11 miRNAs whose changes are associated with lung cancer, from which 2 miRNA biomarkers are optimized. Although showing promise, the 2 miRNAs have moderate sensitivity and specificity that are not sufficient in routine laboratory settings. We will use this platform to evaluate more lung tumor-specific miRNAs to identify additional plasma miRNA biomarkers that can be added to the current ones so that the diagnostic efficacy of the plasma assays could be improved. Furthermore, the assessments of cell-free circulating tumor DNA or DNA methylation status of gene promoters have attracted increasing attention as potential liquid biopsy tests for lung cancer [38–40]. Our ongoing efforts are to compare the diagnostic efficiency of the plasma miRNA signature with those of the cell-free DNA biomarkers in the early detection of lung cancer. 3), given that the changes of the miRNAs are not specific for lung cancer, the patents with positive results might have other type of cancers. In future clinical trials, the patients who have positive results with the molecular analysis should undertake clinical examinations and tests to have final clinical diagnosis.

Conclusions

This study demonstrates that an approach for directly targeting less blood and skipping the multipole procedures that are associated with large analytic variations would improve the efficacy of miRNA assessment. The developed plasma miRNA biomarkers may help diagnose lung cancer at the early stage and classify the subtypes. Nevertheless, the continued development of this technology and validation of the biomarkers are required.

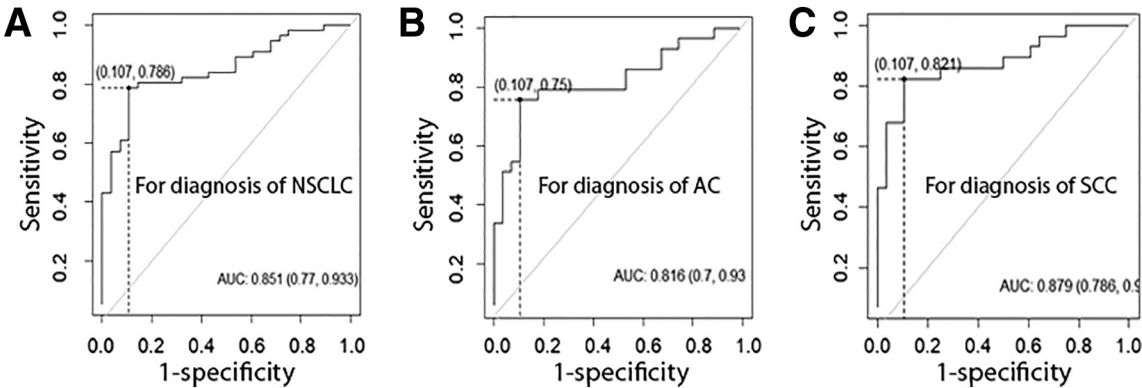


Figure 3. Receiver-operator characteristic (ROC) curves of 2 plasma miRNAs (miRs-205-5p and -210-3p). The area under the ROC curve (AUC) for the biomarkers conveys the accuracy. (A), a logistic regression model based on the 2 miRNAs creates an AUC value of 0.85 with 78.6% sensitivity and 89.3% specificity for diagnosis of non-small-cell lung cancer (NSCLC). (B), a logistic regression model based on the 2 miRNAs produces 0.82 AUC with 75.0% sensitivity and 89.3 % specificity for diagnosis of lung adenocarcinoma (AC). (C), a logistic regression model based on the 2 miRNAs generates 0.88 AUC with 82.1% sensitivity and 89.3% specificity for diagnosis of lung squamous cell carcinoma (SCC).

Declarations

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

The Institutional Review Boards of University of Maryland Baltimore and Veterans Affairs Maryland Health Care System approved this study. reference number is H40666.

Competing Interests

The authors declare no conflict of interest.

Funding

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, UMD-UMB Research and Innovation Seed Grant, DoD-Idea Development Award, and Maryland Innovation Initiative (MII) Commercialization Program- Phase 1/2 Grant (F.J.).

Authors' Contributions

QL, YW, and FJ conducted the experiments and participated in study design, coordination, and data interpretation, and preparing the manuscript. All authors read and approved the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.tranon.2018.05.001>.

References

- [1] Siegel R, Naishadham D, and Jemal A (2013). Cancer statistics, 2013. *CA Cancer J Clin* **63**, 11–30.
- [2] Travis WD, Brambilla E, and Geisinger KR (2016). Histological grading in lung cancer: one system for all or separate systems for each histological type? *Eur Respir J* **47**, 720–723.
- [3] Patnaik S, Mallick R, Kannisto E, Sharma R, Bshara W, Yendamuri S, and Dhillon SS (2015). MiR-205 and MiR-375 microRNA assays to distinguish squamous cell carcinoma from adenocarcinoma in lung cancer biopsies. *J Thorac Oncol* **10**, 446–453.
- [4] Gressett SM and Shah SR (2009). Intricacies of bevacizumab-induced toxicities and their management. *Ann Pharmacother* **43**, 490–501.
- [5] Kazandjian D, Suzman DL, Blumenthal G, Mushti S, He K, Libeg M, Keegan P, and Pazdur R (2016). FDA Approval Summary: Nivolumab for the Treatment of Metastatic Non-Small Cell Lung Cancer With Progression On or After Platinum-Based Chemotherapy. *Oncologist* **21**, 634–642.
- [6] Ambros V (2003). MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell* **113**, 673–676.
- [7] Galasso M, Sana ME, and Volinia S (2010). Non-coding RNAs: a key to future personalized molecular therapy? *Genome Med* **2**, 12–16.
- [8] Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, and Ferrando AA, et al (2005). MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838.
- [9] Iorio MV and Croce CM (2012). microRNA involvement in human cancer. *Carcinogenesis* **33**, 1126–1133.
- [10] Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens RM, Okamoto A, Yokota J, and Tanaka T, et al (2006). Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9**, 189–198.
- [11] Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, Tait JF, and Tewari M (2012). Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer Prev Res (Phila)* **5**, 492–497.
- [12] Shen J and Jiang F (2012). Applications of MicroRNAs in the Diagnosis and Prognosis of Lung Cancer. *Expert Opin Med Diagn* **6**, 197–207.
- [13] Shen J, Stass SA, and Jiang F (2013). MicroRNAs as potential biomarkers in human solid tumors. *Cancer Lett* **329**, 125–136.
- [14] Levy B, Hu ZI, Cordova KN, Close S, Lee K, and Becker D (2016). Clinical Utility of Liquid Diagnostic Platforms in Non-Small Cell Lung Cancer. *Oncologist* **21**, 1121–1130.
- [15] Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, Guo J, Zhang Y, Chen J, and Guo X, et al (2008). Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res* **18**, 997–1006.
- [16] Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Brian KC, and Allen A, et al (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* **105**, 10513–10518.
- [17] Tiberio P, Callari M, Angeloni V, Daidone MG, and Appierto V (2015). Challenges in using circulating miRNAs as cancer biomarkers. *Biomed Res Int* **2015**, 731–738.
- [18] Tackett MR and Diwan I (2017). Using FirePlex() Particle Technology for Multiplex MicroRNA Profiling Without RNA Purification Methods. *Mol Biol* **1654**, 209–219.
- [19] Humphrey LL, Deffebach M, Pappas M, Baumann C, Artis K, Mitchell JP, Zakher B, Fu R, and Slatore CG (2013). Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. *Ann Intern Med* **159**, 411–420.
- [20] Tuck MK, Chan DW, Chia D, Godwin AK, Grizzle WE, Krueger KE, Rom W, Sanda M, Sorbara L, and Stass S, et al (2009). Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. *J Proteome Res* **8**, 113–117.
- [21] Ma J, Mannoor K, Gao L, Tan A, Guarnera MA, Zhan M, Shetty A, Stass SA, Xing L, and Jiang F (2014). Characterization of microRNA transcriptome in lung cancer by next-generation deep sequencing. *Mol Oncol* **8**, 1208–1219.
- [22] Shen J, Todd NW, Zhang H, Yu L, Lingxiao X, Mei Y, Guarnera M, Liao J, Chou A, and Lu CL, et al (2011). Plasma microRNAs as potential biomarkers for non-small-cell lung cancer. *Lab Invest* **91**, 579–587.
- [23] Leng Q, Lin Y, Jiang F, Lee CJ, Zhan M, Fang H, and Wang Y (2017). A plasma miRNA signature for lung cancer early detection. *Oncotarget* **8**, 111902–111911.
- [24] Li H, Jiang Z, Leng Q, Bai F, Wang J, Ding X, Li Y, Zhang X, Fang H, and Yfantis HG, et al (2017). A prediction model for distinguishing lung squamous cell carcinoma from adenocarcinoma. *Oncotarget* **8**, 50704–50714.
- [25] Ma J, Guarnera MA, Zhou W, Fang H, and Jiang F (2017). A Prediction Model Based on Biomarkers and Clinical Characteristics for Detection of Lung Cancer in Pulmonary Nodules. *Transl Oncol* **10**, 40–45.
- [26] Shen J, Liu Z, Todd NW, Zhang H, Liao J, Yu L, Guarnera MA, Li R, Cai L, and Zhan M, et al (2011). Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microRNA biomarkers. *BMC Cancer* **11**, 374–379.
- [27] Ma J, Jemal A, and Smith R (2013). Reply to lung cancer deaths averted by screening should be considered in the context of tobacco control policies. *Cancer* **119**, 3420–3421.
- [28] Guthrie JL, Seah C, Brown S, Tang P, Jamieson F, and Drews SJ (2008). Use of Bordetella pertussis BP3385 to establish a cutoff value for an IS481-targeted real-time PCR assay. *J Clin Microbiol* **46**, 3798–3799.
- [29] Jarry J, Schadenbord D, Greenwood C, Spatz A, and van Kempen LC (2014). The validity of circulating microRNAs in oncology: five years of challenges and contradictions. *Mol Oncol* **8**, 819–829.
- [30] Hunter MP, Ismail N, Zhang X, Aguda BD, Lee EJ, Yu L, Xiao T, Schafer J, Lee ML, and Schmittgen TD, et al (2008). Detection of microRNA expression in human peripheral blood microvesicles. *PLoS One* **3**, e3694.
- [31] Moldovan L, Batte KE, Trgovcich J, Wisler J, Marsh CB, and Piper M (2014). Methodological challenges in utilizing miRNAs as circulating biomarkers. *J Cell Mol Med* **18**, 371–390.
- [32] Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, Mitchell PS, Bennett CF, Pogosova-Agadjanyan EL, and Stirewalt DL, et al (2011). Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc Natl Acad Sci U S A* **108**, 5003–5008.

- [33] Lebanony D, Benjamin H, Gilad S, Ezagouri M, Dov A, Ashkenazi K, Gefen N, Izraeli S, Rechavi G, and Pass H, et al (2009). Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from nonsquamous non-small-cell lung carcinoma. *J Clin Oncol* **27**, 2030–2037.
- [34] Xing L, Todd NW, Yu L, Fang H, and Jiang F (2010). Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers. *Mod Pathol* **23**, 1157–1164.
- [35] Grosso S, Doyen J, Parks SK, Bertero T, Paye A, Cardinaud B, Gounon P, Lacas-Gervais S, Noel A, and Pouyssegur J, et al (2013). MiR-210 promotes a hypoxic phenotype and increases radioresistance in human lung cancer cell lines. *Cell Death Dis* **4**e544.
- [36] Noman MZ, Buart S, Romero P, Ketari S, Janji B, Mari B, Mami-Chouaib F, and Chouaib S (2012). Hypoxia-inducible miR-210 regulates the susceptibility of tumor cells to lysis by cytotoxic T cells. *Cancer Res* **72**, 4629–4641.
- [37] Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, and Yasui Y (2001). Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* **93**, 1054–1061.
- [38] Nie K, Jia Y, and Zhang X (2015). Cell-free circulating tumor DNA in plasma/serum of non-small cell lung cancer. *Tumour Biol* **36**, 7–19.
- [39] Balgouranidou I, Chimonidou M, Milaki G, Tsaroucha E, Kakolyris S, Georgoulas V, and Lianidou E (2016). SOX17 promoter methylation in plasma circulating tumor DNA of patients with non-small cell lung cancer. *Clin Chem Lab Med* **54**, 1385–1393.
- [40] Normanno N, Denis MG, Thress KS, Ratcliffe M, and Reck M (2017). Guide to detecting epidermal growth factor receptor (EGFR) mutations in ctDNA of patients with advanced non-small-cell lung cancer. *Oncotarget* **8**(7), 12501–12516.

Integrating Circulating Immunological and Sputum Biomarkers for the Early Detection of Lung Cancer

Jian Su, Qixin Leng, Yanli Lin, Jie Ma, Fangran Jiang, Cheng-Ju Lee, HongBin Fang and Feng Jiang

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

Biomarkers in Cancer
Volume 10: 1–6
© The Author(s) 2018
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1179299X18759297



ABSTRACT: We have demonstrated that assessments of microRNA (miRNA) expressions in circulating peripheral blood mononucleated cell (PBMC) and sputum specimens, respectively, may help diagnose lung cancer. To assess the individual and combined analysis of the miRNAs across the different body fluids for lung cancer early detection, we analyse a panel of 3 sputum miRNAs (miRs-21, 31, and 210) and a panel of 2 PBMC miRNAs (miRs-19b-3p and 29b-3p) in a discovery cohort of 68 patients with lung cancer and 66 cancer-free smokers. We find that integrating 2 sputum miRNAs (miRs-31 and 210) and 1 PBMC miRNA (miR-19b-3p) has higher sensitivity (86.8%) and specificity (92.4%) compared with the individual panels. The synergistic value of the integrated panel of 3 biomarkers is confirmed in a validation cohort, independent of stage and histological type of lung cancer, and patients' age, sex, and ethnicity. Integrating circulating immunological and sputum biomarkers could improve the early detection of lung cancer.

KEYWORDS: Lung cancer, early detection, biomarkers

RECEIVED: November 19, 2017. **ACCEPTED:** January 23, 2018.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by NC1 R21CA205746, VA Merit Award I01 CX000512, Award from the Geaton and JoAnn DeCesaris Family Foundation, UMD-UMB Research and Innovation Seed Grant, DoD-Idea Development Award, and Maryland Innovation Initiative (MII) Commercialization Program-Phase 1/2 grant (F.J.)

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Feng Jiang, Department of Pathology, University of Maryland School of Medicine, 10 S. Pine St. Baltimore, MD 21201, USA. Email: fjiang@som.umaryland.edu

Introduction

Lung cancer is the biggest cancer killer in both men and women. More than 85% lung cancers are non-small-cell lung cancers (NSCLCs), which mainly consist of adenocarcinoma (AC) and squamous cell carcinoma (SCC). Because the early detection of the disease can reduce the mortality, non-invasive biomarkers that can accurately and cost-effectively diagnose early-stage lung cancer are urgently needed.¹

Cigarette smoking is the leading cause of NSCLC. Smoking can cause damages across the entire respiratory epithelial cells. The process is called field cancerization. The field cancerization could be observed in histologically normal-appearing lung epithelium, which may anchor the same molecular aberrations as the ones in primary lung tumours.^{2,3} Examination of the exfoliated bronchial epithelia of airway in sputum can identify the lung tumour-related molecular alterations and hence provide a non-invasive and specific means for diagnosis of NSCLC.^{4–7} MicroRNAs (miRNAs) are small molecules and the aberrations contribute to carcinogenesis.⁸ We have found that aberrant miRNA expression detected in sputum biologically reflects those in primary lung cancer.^{9–14} We have also developed a 3-miRNA biomarker panel in sputum for diagnosis of NSCLC with 83% sensitivity and 88% specificity.^{15–17}

Immune evasion occurs as a primary and important event in tumorigenesis.^{18–20} Peripheral blood mononuclear cells (PBMCs) are the front-line defence against tumorigenesis within immune system.²¹ Furthermore, miRNAs have important effects on the immune cells' function, and the abnormalities contribute to the evasion proceedings of cancer cells in

immunological surveillance.^{22–24} Therefore, miRNA changes of PBMCs would provide novel biomarkers for diagnosis of cancer.^{18,19,21,25} In addition, miRNA biomarkers of PBMCs may overwhelm the difficulties of cell-free (eg, plasma or serum)-based biomarkers because of the following reasons: (1) circulating PBMCs are a rich and readily accessible cell source in blood and thus produce large RNA quantity that has high quality for consistent analysis,¹⁹ and (2) miRNAs are purified from the PBMCs and thus do not have other sources of extracellular miRNAs (RBC-, protein-, exosome-, and vesicle-related miRNAs). We recently identified 2 PBMC miRNA (miRs-19b-3p and 29b-3p) as circulating immunologic biomarkers for NSCLC with 73% sensitivity and 83% specificity.²⁶

Given that the miRNAs in sputum and peripheral PBMC have diverse functions in lung carcinogenesis through various cellular paths; here, we investigated whether integrating sputum and circulating immunological miRNA biomarkers would have a synergistic influence on the early detection of lung cancer.

Materials and Methods

Patients

The Institutional Review Boards at University of Maryland Baltimore and Veterans Affairs Maryland Health Care System reviewed and approved this study. Using the inclusion and/or exclusion criteria recommended by US Preventive Services Task Force for lung cancer screening in heavy smokers,²⁷ we enrolled subjects between the ages of 55 to 80 years who had not less than a 30 pack-year smoking



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Table 1. Characteristics of patients with NSCLC and cancer-free smokers in a discovery cohort.

	NSCLC CASES (N=68)	CONTROLS (N=66)	P VALUE
Age	64.46 (SD: 11.37)	64.68 (SD: 11.32)	.31
Sex			.32
Female	28	28	
Male	40	38	
Race			.40
White	51	50	
African American	17	16	
Pack-years (median)	34.56	33.27	.47
Stage			
Stage I	22		
Stage II	20		
Stage III-IV	26		
Histological type			
Adenocarcinoma	37		
Squamous cell carcinoma	31		

Abbreviation: NSCLC, non-small-cell lung cancer.

history and were former smokers (quit within 15 years). Exclusion criteria included pregnancy, current pulmonary infection, surgery within 6 months, radiotherapy within 1 year, and life expectancy of <1 year. A total of 117 patients with NSCLC and 116 cancer-free smokers were recruited. No difference of age, race, sex, and smoking status was observed in the lung cancer cases vs cancer-free smokers. The cases and controls were randomly grouped into 2 cohorts: a discovery cohort and a validation cohort. The discovery cohort comprised 68 patients with lung cancer and 66 cancer-free smokers. The discovery cohort was used to simultaneously evaluate sputum miRNAs and PBMC miRNAs in the same set of clinical specimens and develop an integrated panel of miRNA biomarkers for NSCLC. The validation cohort included 49 patients with lung cancer and 50 cancer-free smokers and was used to confirm the synergistic effect of the integrated panel of miRNAs on the early detection of NSCLC. Detailed demographic and clinical characteristics of the 2 cohorts are shown in Tables 1 and 2.

Sample collection and preparation

From the participants before they had any treatment, sputum and blood samples were collected as previously described.^{9-12,28-36} The preparation of airway bronchial epithelia from the sputum specimens was performed using a protocol developed in our previous studies.^{7,9-12,28-36} Peripheral blood was collected in BD Vacutainer spray-coated K₂EDTA Tubes

(BD, Franklin Lakes, NJ, USA), and PBMCs were isolated using Ficoll gradient centrifugation as previously described.²⁶

Assessing expressions of the miRNAs using quantitative reverse transcription polymerase chain reaction

RNA was extracted from the samples using our established protocols.^{9-12,26,30,31} The expressions of 3 miRNAs (miRs-21, 31, and 210) in sputum, and 2 miRNAs (miRs-19b-3p and 29b-3p) in PBMC were determined by using TaqMan miRNA assays (Applied Biosystems, Foster City, CA, USA).^{9-12,26,30,31} The levels of miRNAs were determined using a threshold cycle method.^{9-12,15,16} We included controls in each experiment: RNA extracted from a H358 NSCLC cell line (a positive control for RT [reverse transcription], preamplification, and ddPCR [Droplet Digital polymerase chain reaction]), genomic DNA (a positive control for genomic DNA detection), and nuclease-free water (control for contamination). All experiments were repeated 3 times.

Statistical analysis

We used univariate analysis to identify the miRNAs whose expression levels being related to NSCLC. The significantly associated factors were then analysed using multivariate logistic regression models with stepwise regression based on receiver operator characteristic (ROC) curve to select an optimal prediction model for NSCLC. We also generated a 95% confidence

Table 2. Characteristics of patients with NSCLC and cancer-free smokers in a validation cohort.

	NSCLC CASES (N = 49)	CONTROLS (N = 50)	P VALUE
Age	65.23 (SD: 10.48)	64.70 (SD: 11.15)	.30
Sex			.35
Female	19	20	
Male	30	30	
Race			.36
White	37	38	
African American	12	12	
Pack-years (median)	35.52	35.16	.43
Stage			
Stage I	15		
Stage II	15		
Stage III-IV	19		
Histological type			
Adenocarcinoma	26		
Squamous cell carcinoma	23		

Abbreviation: NSCLC, non-small-cell lung cancer.

interval for the difference in the area under the ROCs (AUCs) using the bootstrap.³⁷ The optimal cut-off value was generated using the Youden index.^{38,39} To compare different miRNA biomarker panels, we computed their AUCs to determine the sensitivity and specificity as previously described.⁴⁰

Results

The individual sputum miRNAs and PBMC miRNAs could distinguish patients with lung cancer from cancer-free controls

In the discovery cohort, the 3 sputum miRNAs (miRs-21, -31, and -210) and 2 PBMC miRNAs (miRs-19b-3p and 29b-3p) exhibited a different level in patients with NSCLC vs cancer-free smokers (all P 's < .05). The analysis of the 3 sputum miRNAs created 0.923 AUC (Figure 1). Successively, the examination of the 3 sputum miRNAs produced 82.3% sensitivity and 87.9% specificity for diagnosis of lung cancer (Table 3). The level of miR-21 in sputum was related to AC (P < .05); however, miR-210 was correlated to SCC (P < .05). Overall, the 3 sputum miRNA panel did not have specific relationship with stage and histological type of lung cancer, and patients' age, race, and sex (all P 's > .05), but smoking history (P < .05).

The analysis of the 2 PBMC miRNAs had an AUC of 0.837 (Figure 1), producing 72.1% sensitivity and 81.8% specificity for diagnosis of NSCLC (Table 3). Furthermore, the 2-PBMC miRNA biomarker panel created 80.7% sensitivity and 89.4% specificity for SCC, and 75.7% sensitivity

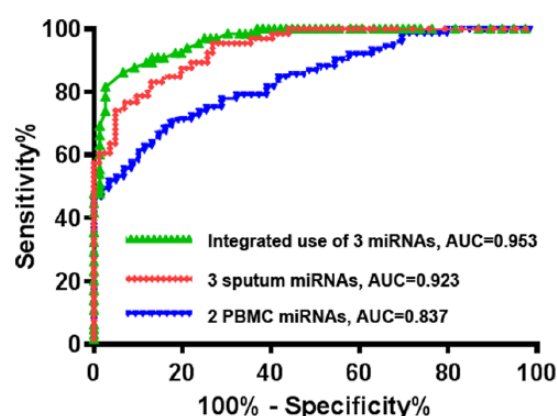


Figure 1. The comparison of integrated miRNA biomarkers with panels of sputum and PBMC miRNAs in a discovery cohort. A prediction model based on integrated use of 3 miRNA biomarkers (miRs-31, 210, and 19b-3p) across sputum and PBMC specimens was developed for distinguishing lung cancer patients from cancer-free smokers. The ROC curve of the integrated miRNA biomarkers produced a higher AUC (0.953), as compared with the panel of sputum miRNAs (0.923) and the panel of PBMC miRNAs (0.837) (all P 's < .05). AUC indicates area under the ROC; miRNAs, microRNAs; PBMC, peripheral blood mononucleated cell.

and 68.2% specificity for AC. As a result, the 2-PBMC miRNA biomarker panel had a higher diagnostic value for SCC compared with AC. The 2 PBMC miRNA biomarkers did not have relationship with stage of NSCLC, and patients' age, sex, and ethnicity of the participants (all P 's ≥ .05), except the smoking history (P = .03). Altogether, the results from

Table 3. The comparison of integrated panel of 2 sputum and 1 PBMC miRNA biomarkers with individual panels of sputum and PBMC miRNA biomarkers in a discovery cohort and a validation cohort.

	A DISCOVERY COHORT		A VALIDATION COHORT	
	SENSITIVITY (95% CI)	SPECIFICITY (95% CI)	SENSITIVITY (95% CI)	SPECIFICITY (95% CI)
3 sputum miRNAs	82.35 (71.20-90.53)	87.88 (77.51-94.62)	81.63 (67.98-91.24)	86.00 (73.26-94.18)
2 PBMC miRNAs	72.06 (59.85-82.27)	81.82 (70.39-90.24)	71.43 (56.74-83.42)	80.00 (66.28-89.97)
Integrated 2 sputum and 1 PBMC miRNAs	86.76 (76.36-93.77)	92.42 (83.20-97.49)	85.71 (72.76-94.06)	92.00 (80.77-97.78)

Abbreviations: CI, confidence interval; miRNAs, microRNAs; PBMC, peripheral blood mononucleated cell.

this study confirmed our prior discoveries^{15,26} that both the sputum and PBMC miRNAs could have the potential as biomarkers for NSCLC.

Incorporated examination of 2 miRNA and 1 PBMC miRNA biomarkers has a synergistic influence on the early detection of NSCLC

In the discovery cohort, logistic regression model was applied to define performance of different patterns of the 3 sputum and 2 PBMC miRNA biomarkers. From the 5 genes, 2 sputum miRNAs (miR-31 and 210) and 1 PBMC miRNA (miR-19b-3p) were selected as the best markers (all P 's < .001) and integrated into a logistic model. Integrated use of the 3 biomarkers created a greater AUC (0.953) than did the 3-sputum biomarker panel (0.923) or the 2-PBMC biomarker panel (0.837) (P < .05) (Figure 1). Using Youden index, we set up optimal cut-off at 0.64. As a result, integrated use of the 2 sputum miRNAs and 1 PBMC miRNA produced higher sensitivity (86.8%) and specificity (92.4%) than did the 3-sputum biomarkers and the 2-PBMC biomarkers (all P 's < .05) (Table 3). Furthermore, combined use of all the 5 biomarkers (3 miRNA and 2 PBMC biomarkers) did not produce higher sensitivity and specificity compared with the integrated panel of the only 3 biomarkers (P > .05). Moreover, integrated analysis of the 3 biomarkers across the different body fluids did not show special association with stage and histology of lung cancer, and patients' age, race, and sex (all P 's > .05). Importantly, the estimated Pearson correlation among levels of the 3 miRNAs was very low (all P 's > .05), suggesting that the combination had complementary classification for diagnosis of lung cancer.

Validating the synergistic effect of the integrated panel of 3 miRNAs across the body fluids for detection of NSCLC

In the validation cohort, the integrated panel of 3 biomarkers was blindly confirmed in the specimens of 49 patients with lung cancer and 50 cancer-free smokers. The integration of the 3 biomarkers across sputum and blood specimens showed higher sensitivity (85.7% vs 81.6% and 71.4%) and specificity (92.0% vs 86.0% and 80.0%) than did the 3-sputum miRNA

panel and 2-PBMC miRNA panel (all P 's < .05) (Table 3). Inconsistent with the findings in the discovery cohort, integrated analysis of the 3 biomarkers had no special relationship with stage and histological type of lung cancer, and patients' age, race, and sex (all P 's > .05). Furthermore, the Pearson correlation between the 3 miRNAs' expression levels was very low (all P 's > .05), supporting the synergistic value of this integrated panel of the 3 biomarkers for the early detection of NSCLC.

Discussion

The development of sputum biomarkers is based on the concept that lung tumorigenesis developed in a field defect characterized by molecular aberrations in the airways. According to the field defect concept, the alterations of miRNAs in bronchial airways reflect those in primary lung tumours in the distal lungs.⁶ The analysis of exfoliated airway bronchial epithelia in sputum for the altered miRNA expressions might indicate the lung tumour-related molecular alterations and hence identify NSCLC. On the contrary, the development of PBMC miRNAs as circulating tumour markers is based on the concept that PBMCs act as the primary defence line against cancer, and the abnormal miRNA expressions of PBMC contributes to the escaping events in immunological surveillance.^{18,19,21,22,25} Based on the different concepts, miRNA expression profile in PBMCs, unlike miRNAs detected in sputum cells, may not signify the abnormal miRNA changes in the lung tumours or mimic that in the primary cancer tissues.⁴¹⁻⁴³ Therefore, there was no overlap between the PBMC miRNA vs sputum miRNA profiles of patients with lung cancer.^{10,11,15,26,44} Given that altered miRNAs in sputum and PBMCs contribute to lung tumorigenesis via the different mechanisms,^{5,9,10,22,26} we hypothesize that integrating the miRNAs across the different body fluids may have a synergistic effect and thus improve the diagnosis of lung cancer. Indeed, the combined analysis of the 2-sputum miRNA and 1-PBMC miRNA generated a better performance than did a single category of the markers. Also, there was no association among the miRNA aberrations in the 2 surrogate clinical materials, further supporting that the performances of the 2 categories of miRNA biomarkers might have a synergistic value. In addition, unlike PBMC miRNA markers that are

more specific to SCC, the integration of sputum and PBMC biomarkers was not related to histology of lung tumour and hence substantiates the utility for predicting lung cancer. Moreover, integrated use of the 3 miRNAs had a comparable diagnostic performance for lung tumour at the early vs late stages. Altogether, the findings would be an important feature if the assay is used for identifying early-stage lung cancer.

Dysregulations of miRs-31, 210, and 19b-3p have been found to associate with a variety of tumours. For example, miR-31 downregulation in breast cancer is related with tumour metastases, whereas elevated miR-31 expression in colorectal cancer may associate with late stage of the tumours.^{45,46} MiR-31 dysregulation endorses KRAS mutation-driven NSCLC and hence plays a key role in lung tumorigenesis.⁴⁷ MiR-210 stimulates a hypoxic phenotype and upsurges radioresistance in NSCLCs.⁴⁸ The analysis of miR-210 in serum can diagnose several types of malignancies, including NSCLC.^{49–51} Abnormal miR-19b-3p expression has not been described in the immune system. Our ongoing project is to investigate the mechanisms, by which miR-19b-3p may involve in lung cancer cells' evasion of immunological surveillance.

This study may have some limitations. (1) Clinically useful biomarkers should have very high sensitivity and specificity. Integrated use of 3 miRNAs has a sensitivity of 88.2% and a specificity of 92.4%, which may not be sufficient in the laboratory settings. We will find new biomarkers that can be used together with the current ones to improve clinical performance of the assay. Furthermore, an ongoing project in our laboratory is to simultaneously evaluate cell-free miRNAs (eg, plasma- and serum-derived miRNAs) and PBMC miRNAs using the same clinical specimens to compare the diagnostic values of the 2 different miRNA biomarkers and investigate whether there is synergistic diagnosis between them. (2) The clinical samples were obtained from the patients who already have clinical diagnosis. The subjects might not represent heavy smokers in a lung cancer screening programme. We will conduct a large lung cancer screening trial to confirm the performance of the integrated panel of miRNA biomarkers. (3) Using low-dose computed tomography (LDCT) for the early detection of lung cancer can reduce the mortality.⁵² However, LDCT always produces over-diagnosis or a false-positive rate, presenting a major clinical obstacle for lung cancer early detection.⁵² It would be interesting to evaluate whether the 3 miRNA biomarkers could complement LDCT for the early detection of lung cancer by specifically reducing its over-diagnosis.

Conclusions

We have for the first time demonstrated that integrating circulating and sputum markers has a synergistic value and presents a potential approach for the early detection of lung cancer. Nevertheless, a large multi-centre trial is required to confirm the integrated method before it could be translated in laboratory settings.

Author Contributions

JS, QL, YL, JM, FJ, and C-JL conducted the experiments and participated in data interpretation. HF and FJ participated in study design, coordination, and data analysis and interpretation, and prepared the manuscript. All authors read and approved the final manuscript.

Disclosures and Ethics

The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material.

REFERENCES

1. American Cancer Society. *Cancer Facts & Figures for African Americans*. Atlanta, GA: American Cancer Society.
2. Tang X, Shigematsu H, Bekele BN, et al. EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. *Cancer Res*. 2005;65:7568–7572.
3. Solis LM, Behrens C, Raso MG, et al. Histologic patterns and molecular characteristics of lung adenocarcinoma associated with clinical outcome. *Cancer*. 2012;118:2889–2899.
4. Kadara H, Shen L, Fujimoto J, et al. Characterizing the molecular spatial and temporal field of injury in early-stage smoker non-small cell lung cancer patients after definitive surgery by expression profiling. *Cancer Prev Res (Phila)*. 2013;6:8–17.
5. Kadara H, Wistuba II. Field cancerization in non-small cell lung cancer: implications in disease pathogenesis. *Proc Am Thorac Soc*. 2012;9:38–42.
6. Spira A, Beane JE, Shah V, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med*. 2007;13:361–366.
7. Belinsky SA, Liechty KC, Gentry FD, et al. Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res*. 2006;66:3338–3344.
8. Croce CM, Calin GA. miRNAs, cancer, and stem cell division. *Cell*. 2005;122:6–7.
9. 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.
10. Xie Y, Todd NW, Liu Z, et al. Altered miRNA expression in sputum for diagnosis of non-small cell lung cancer. *Lung Cancer*. 2010;67:170–176.
11. Xing L, Todd NW, Yu L, Fang H, Jiang F. Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers. *Mod Pathol*. 2010;23:1157–1164.
12. 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;127:2870–2878.
13. Pang W, Tian X, Bai F, et al. Pim-1 kinase is a target of miR-486-5p and eukaryotic translation initiation factor 4E, and plays a critical role in lung cancer. *Mol Cancer*. 2014;13:240.
14. Wang J, Tian X, Han R, et al. Downregulation of miR-486-5p contributes to tumor progression and metastasis by targeting protumorigenic ARHGAP5 in lung cancer. *Oncogene*. 2014;33:1181–1189.
15. Xing L, Su J, Guarnera MA, et al. Sputum microRNA biomarkers for identifying lung cancer in indeterminate solitary pulmonary nodules. *Clin Cancer Res*. 2015;21:484–489.
16. Su J, Liao J, Gao L, et al. Analysis of small nucleolar RNAs in sputum for lung cancer diagnosis. *Oncotarget*. 2016;7:5131–5142.
17. Su Y, Guarnera MA, Fang H, Jiang F. Small non-coding RNA biomarkers in sputum for lung cancer diagnosis. *Mol Cancer*. 2016;15:36.
18. Bluth M, Lin YY, Zhang H, Viterbo D, Zenilman M. Use of gene expression profiles in cells of peripheral blood to identify new molecular markers of acute pancreatitis. *Arch Surg*. 2008;143:227–233 (discussion 233–234).
19. Baine MJ, Chakraborty S, Smith LM, et al. Transcriptional profiling of peripheral blood mononuclear cells in pancreatic cancer patients identifies novel genes with potential diagnostic utility. *PLoS ONE*. 2011;6:e17014.
20. Melief CJ, Finn OJ. Cancer immunology. *Curr Opin Immunol*. 2011;23:234–236.

21. Twine NC, Stover JA, Marshall B, et al. Disease-associated expression profiles in peripheral blood mononuclear cells from patients with advanced renal cell carcinoma. *Cancer Res.* 2003;63:6069–6075.
22. Sasaki K, Kohanbash G, Hoji A, et al. miR-17-92 expression in differentiated T cells: implications for cancer immunotherapy. *J Transl Med.* 2010;8:17.
23. Bhatia A, Kumar Y. Cellular and molecular mechanisms in cancer immune escape: a comprehensive review. *Expert Rev Clin Immunol.* 2014;10:41–62.
24. Kohlhaup FJ, Mitra AK, Lengyel E, Peter ME. MicroRNAs as mediators and communicators between cancer cells and the tumor microenvironment. *Oncogene.* 2015;34:5857–5868.
25. Whitney AR, Diehn M, Popper SJ, et al. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci U S A.* 2003;100:1896–1901.
26. Ma J, Lin Y, Zhan M, Mann DL, Stass SA, Jiang F. Differential miRNA expressions in peripheral blood mononuclear cells for diagnosis of lung cancer. *Lab Invest.* 2015;95:1197–1206.
27. Humphrey LL, Deffebach M, Pappas M, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US preventive services task force recommendation. *Ann Intern Med.* 2013;159:411–420.
28. Romeo MS, Sokolova IA, Morrison LE, et al. Chromosomal abnormalities in non-small cell lung carcinomas and in bronchial epithelia of high-risk smokers detected by multi-target interphase fluorescence in situ hybridization. *J Mol Diagn.* 2003;5:103–112.
29. Yu L, Shen J, Mannoor K, Guarnera M, Jiang F. Identification of ENO1 as a potential sputum biomarker for early-stage lung cancer by shotgun proteomics. *Clin Lung Cancer.* 2014;15:372–378.
30. Li N, Ma J, Guarnera MA, Fang H, Cai L, Jiang F. Digital PCR quantification of miRNAs in sputum for diagnosis of lung cancer. *J Cancer Res Clin Oncol.* 2014;140:145–150.
31. Anjuman N, Li N, Guarnera M, Stass SA, Jiang F. Evaluation of lung flut in sputum samples for molecular analysis of lung cancer. *Clin Transl Med.* 2013;2:15.
32. Jiang F, Todd NW, Li R, Zhang H, Fang H, Stass SA. A panel of sputum-based genomic marker for early detection of lung cancer. *Cancer Prev Res (Phila).* 2010;3:1571–1578.
33. Jiang F, Todd NW, Qiu Q, Liu Z, Katz RL, Stass SA. Combined genetic analysis of sputum and computed tomography for noninvasive diagnosis of non-small-cell lung cancer. *Lung Cancer.* 2009;66:58–63.
34. Katz RL, Zaidi TM, Fernandez RL, et al. Automated detection of genetic abnormalities combined with cytology in sputum is a sensitive predictor of lung cancer. *Mod Pathol.* 2008;21:950–960.
35. Qiu Q, Todd NW, Li R, et al. Magnetic enrichment of bronchial epithelial cells from sputum for lung cancer diagnosis. *Cancer.* 2008;114:275–283.
36. Li R, Todd NW, Qiu Q, et al. Genetic deletions in sputum as diagnostic markers for early detection of stage I non-small cell lung cancer. *Clin Cancer Res.* 2007;13:482–487.
37. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology.* 1983;148:839–843.
38. Luo J, Xiong C. Youden index and associated cut-points for three ordinal diagnostic groups. *Commun Stat Simul Comput.* 2013;42:1213–1234.
39. Schisterman EF, Faraggi D, Reiser B, Hu J. Youden index and the optimal threshold for markers with mass at zero. *Stat Med.* 2008;27:297–315.
40. Samawi HM, Yin J, Rochani H, Panchal V. Notes on the overlap measure as an alternative to the Youden index: how are they related? *Stat Med.* 2017;36:4230–4240.
41. 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.
42. Sozzi G, Boeri M, Rossi M, et al. Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: a correlative MILD trial study. *J Clin Oncol.* 2014;32:768–773.
43. Jarry J, Schadendorf D, Greenwood C, Spatz A, van Kempen LC. The validity of circulating microRNAs in oncology: five years of challenges and contradictions. *Mol Oncol.* 2014;8:819–829.
44. 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;10:40–45.
45. Wang S, Hu J, Zhang D, Li J, Fei Q, Sun Y. Prognostic role of microRNA-31 in various cancers: a meta-analysis. *Tumour Biol.* 2014;35:11639–11645.
46. Laurila EM, Kallioniemi A. The diverse role of miR-31 in regulating cancer associated phenotypes. *Gene Chromos Cancer.* 2013;52:1103–1113.
47. Edmonds MD, Boyd KL, Moyo T, et al. MicroRNA-31 initiates lung tumorigenesis and promotes mutant KRAS-driven lung cancer. *J Clin Invest.* 2016;126:349–364.
48. Grosso S, Doyen J, Parks SK, et al. MiR-210 promotes a hypoxic phenotype and increases radioresistance in human lung cancer cell lines. *Cell Death Dis.* 2013;4:e544.
49. Chang W, Lee CY, Park JH, et al. Survival of hypoxic human mesenchymal stem cells is enhanced by a positive feedback loop involving miR-210 and hypoxia-inducible factor 1. *J Vet Sci.* 2013;14:69–76.
50. Lawrie CH, Gal S, Dunlop HM, et al. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *Br J Haematol.* 2008;141:672–675.
51. Papaconstantinou IG, Manta A, Gazouli M, et al. Expression of microRNAs in patients with pancreatic cancer and its prognostic significance. *Pancreas.* 2013;42:67–71.
52. Patz EF Jr, Pinsky P, Gatsonis C, et al; NLST Overdiagnosis Manuscript Writing Team. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med.* 2014;174:269–274.