

Published in final edited form as:

Lung Cancer. 2010 February ; 67(2): 170–176. doi:10.1016/j.lungcan.2009.04.004.

## Altered miRNA Expression in Sputum for Diagnosis of Non-small Cell Lung Cancer

Ying Xie<sup>1</sup>, Nevins W. Todd<sup>2</sup>, Zhenqiu Liu<sup>3</sup>, Min Zhan<sup>4</sup>, HongBin Fang<sup>3</sup>, Hong Peng<sup>1</sup>, Mohammed Alattar<sup>2</sup>, Janaki Deepak<sup>2</sup>, Sanford A. Stass<sup>1</sup>, and Feng Jiang<sup>1,\*</sup>

<sup>1</sup> Department of Pathology, University of Maryland School of Medicine, Baltimore, MD

<sup>2</sup> Department of Medicine, University of Maryland School of Medicine, Baltimore, MD

<sup>3</sup> Division of Biostatistics of The University of Maryland Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, MD

<sup>4</sup> Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Baltimore, MD

### SUMMARY

Analysis of molecular genetic markers in biological fluids has been proposed as a useful tool for cancer diagnosis. MicroRNAs (miRNAs) are small regulatory RNAs that are frequently dysregulated in lung cancer and have shown promise as tissue-based markers for its prognostication. The aim of this study was to determine whether aberrant miRNA expression can be used as a marker in sputum specimen for the diagnosis of non-small cell lung cancer (NSCLC). Experimental Design: Expressions of mature miRNAs, *mir-21* and *mir-155*, were examined by real-time reverse transcription polymerase chain reaction (RT-PCR) and normalized to that of control miRNA, *U6B*, in sputum of 23 patients with NSCLC and 17 cancer-free subjects. The data was compared with conventional sputum cytology for the diagnosis of lung cancer. All endogenous miRNAs were present in sputum in a remarkably stable form and sensitively and specifically detected by real-time RT-PCR. *Mir-21* expression in the sputum specimens was significantly higher in cancer patients ( $76.32 \pm 9.79$ ) than cancer-free individuals ( $62.24 \pm 3.82$ ) ( $p < 0.0001$ ). Furthermore, overexpression of *mir-21* showed highly discriminative receiver-operator characteristic (ROC) curve profile, clearly distinguishing cancer patients from cancer-free subjects with areas under the ROC curve at  $0.902 \pm 0.054$ . Detection of *mir-21* expression produced 69.66% sensitivity and 100.00% specificity in diagnosis of lung cancer, as compared with 47.82% sensitivity and 100.00% specificity by sputum cytology. The measurement of altered miRNA expression in sputum could be a useful noninvasive approach for the diagnosis of lung cancer.

### Key words

MicroRNA; sputum; lung cancer; real-time RT-PCR; diagnosis

---

\*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.

#### A Conflict of Interest Statement

All authors have no potential conflict of interest relevant to this manuscript was reported.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## INTRODUCTION

Non-small cell lung cancer (NSCLC) is the leading cause of cancer death in the United States and worldwide (1). Five-year survival rate of NSCLC patients is only 13%. An important reason for poor outcomes of the patients is that NSCLC is often detected after it has spread beyond the primary site at the time of diagnosis (2). Therefore, finding NSCLC early is a realistic approach to reduce the mortality associated with lung cancer (2). Although computed tomography appears promising as it can identify lung cancer at a smaller size compared with a chest X-ray, the improved sensitivity is related to a high false-positive rate (3–4). The fluorescence bronchoscopy excels at detecting centrally occurring lung tumor. However, it is invasive technique (3–4). The development of highly reliable and noninvasive diagnostic tools would facilitate the early detection of lung cancer, and is thus clinically important.

Lung carcinogenesis develops from respiratory airway of the patients (5). Because sputum contains exfoliated airway epithelial cells and is the most easily accessible biological fluid, cytological analysis of sputum has been used for lung cancer diagnosis (6–7). However, sputum cytology depends on the skills required for identifying subtle cellular morphological abnormalities in bronchial epithelial cells, thus the sensitivity is very low (6–7). Molecular genetic study of sputum can identify the cells bearing tumor-related molecular genetic aberrations, therefore, might be more sensitive than cytology in identifying neoplastic cells (6–7). For instance, we demonstrated that assessment of deletions of tumor suppressor genes, *HYAL2* and *FHIT*, could detect abnormal cells not only in all the cytological positive sputum, but also in 55% cytologically negative sputum from stage I NSCLC patients (8–9).

MicroRNAs (miRNAs) are a class of small non-protein-coding RNAs that can posttranscriptionally regulate the expression of hundreds of their target genes, thereby controlling a wide range of biological functions such as cellular proliferation, differentiation, and apoptosis (10). Furthermore, miRNAs may function as tumor suppressors or oncogenes, and dysregulated miRNA expressions participate in cancer development and progression (11–2). Therefore, miRNAs can potentially be useful in the diagnosis and classification of human malignancies (11). For example, overexpressions of miRNAs, *mir-21* and *mir-155*, in surgically resected lung tumor tissues were negative prognostic factors for overall survival of the patients (13–4). However, the feasibility of examining aberrant miRNA expressions in sputum for noninvasive diagnosis of lung cancer has not been investigated.

The objective of this study was to determine whether altered miRNA expression detected in sputum could be a useful biomarker for the diagnosis of NSCLC. Real-time quantitative reverse transcription polymerase chain reaction (RT-PCR) was performed with the following two goals of exploring whether 1) miRNA expression could robustly and reliably be measured in sputum and 2) analysis of miRNA expression in sputum might diagnose lung cancer. We showed that endogenous miRNAs stably existed and could be sensitively and specifically measured in sputum. Overexpression of *mir-21* in sputum specimens would constitute a diagnostic marker for lung cancer.

## MATERIALS AND METHODS

### Patients

Twenty-three NSCLC patients who had flexible bronchoscopy at the University of Maryland Medical Center and Baltimore VA Medical Center participated in the study. In the same institutions, 17 cancer-free subjects were also recruited from patients who had received primary care and underwent fiberoptic bronchoscopy for other diseases than cancer. The research was approved by an institutional review board. The NSCLC patients included 13 adenocarcinomas and 10 squamous cell carcinomas, and 3 stage I, 5 stage II, 7 stage III, and 8 stage IV, as

determined according to WHO classification and the International Union against Cancer staging system. Of the lung cancer patients, 9 cancer patients had central tumors and 14 had peripheral tumors in the lungs (Table 1).

### Collection and processing of sputum

Sputum was collected from the participants as described in our recent reports (8–9). Two cytocentrifuge slides were made from each sputum sample and stained with Papanicolaou stain for cytologic diagnosis using the classification of Saccomanno (15). Positive cytology included both carcinoma *in situ* and invasive carcinoma. The main variables used to select sputum specimens for the study were high cellularity, good nuclear morphology, and lack of purulence, debris, and residual cytoplasm (8–9).

### RNA isolation

RNA was isolated from sputum by using mirVana™ miRNA Isolation Kit (Ambion, Austin, TX), according to the manufacturer's protocol. The qualification and quantification of RNA were assessed by using both Biospectrometer (Hutchinson Technology Inc, Hutchinson, MN) and Electrophoresis Bioanalyzer (Agilent Technologies, Foster City, CA).

### Reverse transcription (RT) with miRNA-specific looped primer

RNA was applied for RT by using the Applied Biosystems 9700 Thermocycler (Applied Biosystems) and TaqMan® MicroRNA RT Kit (Applied Biosystems), according to the manufacturer's instructions. The reaction includes 50 nM stem-loop RT primer, 1x RT buffer, 0.25 mM each of dNTPs, and 3.33 U/μl MultiScribe reverse transcriptase in a total volume of 15 μL.

### Quantification of mature miRNAs by Real-time PCR

Real-time PCR was performed to measure expressions of target miRNAs by using TaqMan® PCR kit (Applied Biosystems) on a Bio-Red IQ5 Multicolor Real-time PCR Detection System (Bio-Red, Hercules, CA). The 20 μl PCR reaction included RT product, 1x TaqMan® Universal PCR Master Mix (Applied Biosystems), and the corresponding primers and Taqman probe for the target genes. The reactions were incubated in a 96-well plate at 95°C for 15 min, followed by 45 cycles of 95°C for 15 s and 60°C for 1 min. The threshold cycle (Ct) was defined as the fractional cycle number at which the fluorescence passed the fixed threshold. Expression of target miRNAs, *mir-21* and *mir-155*, was normalized in relation to expression of small nuclear *U6B RNA*, a commonly used internal control for miRNA quantification assay (16). All assays were performed in triplicates.

### Characterization of stability of miRNAs in sputum

Synthesized *Caenorhabditis elegans* miRNA, *cel-miR-238*, that did not exist in the human genomes (17) was obtained from Integrated DNA Technologies (Integrated DNA Technologies, Inc, Coralville, IA). One sputum specimen was split in two parts. In the first part, 2X Denaturing Solution (Ambion) was first added to inhibit RNase activity in the sputum sample, and the synthesized *C. elegans* miRNA, *cel-miR-238*, was subsequently added with a concentration of 100 fmol/μl. Conversely, in the second part of the specimen, *cel-miR-238* was added and, after 5 minutes the Denaturing Solution was added in the sputum sample. RNA was isolated from the sputum samples, respectively. Quantification of the *cel-miR-238* and human endogenous miRNAs (*mir-21* and *mir-155*) was performed by using real-time RT-PCR in the samples, respectively.

To further evaluate stability of the miRNAs in sputum, six sputum specimens were randomly selected, and each one was divided into 4 parts. The first aliquot from each sputum specimen

was processed immediately for isolating RNA, while others were stored in 4°C and processed on day 1, 3 and 7. Expression of endogenous miRNAs (*mi-21*, *miR-155*, and *U6B*) was measured by using real-time RT-PCR at the same time in these specimens that were processed from the different time points. All experiments were performed at least three times.

### Determination of specificity, sensitivity, and dynamic range of miRNAs quantification in sputum

To determine specificity of miRNA quantification in sputum by real-time RT-PCR assay, the synthesized *cel-miR-238* was diluted in human RNA isolated from a denatured sputum specimen by ten orders of magnitude, including  $10^{10}$ ,  $10^9$ ,  $10^8$ ,  $10^7$ ,  $10^6$ ,  $10^5$ ,  $10^4$ ,  $10^3$ ,  $10^2$ , and  $10^1$  copies of *cel-miR-238*/μg human RNA. 10 ng human RNA was used as a negative control. The real-time RT-PCR assay with primers specific to *cel-miR-238* was performed in the samples. To determine the sensitivity and dynamic range of miRNA quantification in sputum, RNA was extracted from six sputum specimens and then diluted by ten orders of magnitude in DEPC water, ranging from 0.86 pg to 860 ng/ml. Expressions of human miRNAs, *mir-21*, *mir-155* and *U6B*, were assessed by using real-time RT-PCR in the samples. The result data was analyzed to determine the sensitivity and dynamic range of the quantification of the miRNAs as previously described (16). All tests were performed in triplicates.

### Statistical analysis

Statistical analyses were done using Statistical Analysis System software version 6.12 (SAS Institute, Cary, NC). All P values shown were two sided, and a P value of <0.05 was considered statistically significant. Furthermore, receiver-operator characteristic (ROC) curve analysis was undertaken using expression level for each gene in the sputum specimens from cancer patients and cancer-free controls by Analyse-it software (Analyse-it Software, Leeds, UK). Using this approach, areas under the ROC curve (AUROCs) identified optimal sensitivity and specificity levels at which to distinguish normal individuals from cancer patients, and corresponding thresholds were calculated for each miRNA. In addition, contingency table and logistic regression analysis were applied to determine the associations between the genetic aberrations and clinicopathologic and demographic characteristics of the cases and controls, including cytology, gender, smoking status, location of tumor, and stage and histology classification.

## RESULTS

### Stability of miRNAs in sputum

Given that sputum might contain high levels of RNase activity, we determined whether miRNAs could stably be detected in the specimen. First, we mixed synthetic miRNA corresponding to *Caenorhabditis elegans* miRNA, *cel-miR-238*, into sputum either before or after the addition of a denaturing solution that inhibits RNase activity. The *cel-miR-238* was selected because of the absence of homologous sequences in humans (17). Quantification of the *cel-miR-238* expression by real-time RT-PCR showed that the synthetic miRNA was not detectable when added to the sputum specimen that had RNase activity. However, the *cel-miR-238* expression was clearly observed in the specimen that had no RNase activity, because Denaturing Solution was added before the synthetic *cel-miR-238* was spiked into sputum. Furthermore, the human endogenous miRNAs (*mir-21* and *mir-155*) were clearly detected in the same samples. These results indicate that endogenous sputum miRNAs rather than naked exogenous miRNAs exist in a form that is resistant to RNase activity.

Given that sputum is the most easily accessible biological fluid and can be used as a surrogate diagnostic material for lung cancer in clinic settings. We next sought to investigate the stability of endogenous miRNAs in archived sputum. Aliquots of six sputum specimens were stored at

4°C for 1 to 7 days, respectively. Bioanalyzer showed that there was increasing degradation of total RNA illustrated by gradually decreased heights of 18S and 28S peaks from day 1 to day 7 (Figure 1A). However, there was no effect on expression levels of *mir-21* or *mir-155* as measured by real-time RT-PCR in the same specimens (Figure 1B). The data imply that even there was apparent degradation of total RNA, endogenous miRNAs could robustly and reliably be examined in sputum over the storage time points.

### Sensitivity, specificity, and dynamic range of miRNA quantification in sputum

To define the dynamic range, sensitivity, and specificity of miRNA quantification in sputum specimen, the synthesized *cel-miR-238* miRNA was mixed with the human total RNA isolated from the denatured sputum specimen by ten orders of magnitude. Although the *cel-miR-238* was detected in the mixed samples, it was not observed in the total human RNA that was not spiked with the *cel-miR-238* miRNA, implying that the target miRNA can be specifically detected in sputum. Furthermore, as shown in Figure 2A and B, the real-time RT-PCR assay demonstrated excellent linearity between the log of the *cel-miR-238* input and Ct value, suggesting that the assay had a dynamic range of at least 8 logs and was thus capable of detecting as few as 1,000 copies of the *cel-miR-238* in one  $\mu$ g human RNA, and the correlation coefficient was 0.9981.

To further determine the sensitivity of detection miRNA in sputum specimens, the total RNA was isolated from cellular pellets of six sputum specimens and then diluted in DEPC water by ten orders of magnitude, respectively. The serially diluted RNAs served as experimental samples for measuring expression of *mir-21*, *mir-155*, and *U6B*. The results showed excellent linearity between the RNA input and the Ct values for the real-time RT-PCR assay. Furthermore, the assay had a dynamic range of at least six orders of magnitude ( $R^2=0.9986$ ), and was capable of detecting as little as 0.86 pg of RNA and 10 copies of the target genes (Fig. 2C and D).

### Genetic changes in sputum of lung cancer patients and cancer-free individuals

To determine clinical significance of dysregulated expressions of miRNAs in sputum for diagnosis of lung cancer, we evaluated expressions of two cancer-associated miRNAs, *mir-21* and *mir-155*, in sputum obtained from a cohort of 23 NSCLC patients and 17 cancer-free subjects. All assays were successfully performed in the sputum specimens. Means of expression for *mir-21* in the specimens of cancer patients and cancer-free individuals were 76.32 (SD 9.79) and 62.24 (SD 3.82), respectively. Means of expression for *mir-155* in the specimens of cancer patients and cancer-free individuals were 88.64 (SD 9.01) and 85.18 (SD 6.54), respectively. Expression of *mir-21* was significantly higher in cancer patients than in cancer free controls ( $p=0.0001$ ), while there was not statistically significant difference of *mir-155* expression in the specimens between the two groups ( $P=0.1669$ ). ROC curves with corresponding AUROCs for *mir-21* and *mir-155* expressions in sputum from cancer vs control cases are shown in Fig. 3.

The cutoff normalized miRNA expression value (72.97) for *mir-21* was further chosen from the ROC curve in order to maximize sensitivity and specificity. Based on the cutoff, *mir-21* overexpression was found in 16 of 23 and 0 of 17 sputum samples from cancer patients and cancer-free subjects, resulting in 69.66% (95% CI, 0.46–0.86) sensitivity and 100.00% (95% CI, 0.77–1.00) specificity in the diagnosis of lung cancer. Furthermore, the prevalence of *mir-21* expressions detected in sputum was not associated with pack-years of smoking, patient age, gender, histological tumor type and stage, and location of the tumors (all  $P > 0.05$ ) (Table 2). The data suggested that measuring expressions of *mir-21* could provide a potential noninvasive diagnostic tool for lung cancer. Interestingly, of the three lung cancer patients diagnosed with stage I NSCLC, two patients' sputum had high expression level of *mi-21*,

indicating that the elevated *mir-21* expression in sputum may be a useful biomarker for the early detection of lung cancer.

### Comparison of cytology and measurement of altered miRNA expression in sputum for the diagnosis of lung cancer

As shown in Table 3, 11 of 23 sputum specimens from lung cancer patients were diagnosed as positive by cytological examination. All cytologically positive sputum specimens also had increased *mir-21* expression. However, of the 12 cytologically negative sputum samples from lung cancer patients, 6 had *mir-21* overexpression, two of which were obtained from stage I lung cancer patients. Overall, the sensitivity and specificity of sputum cytology were 47.82% (95% CI, 0.27–0.69) and 100.00% (95% CI, 0.77–1.00), whereas the sensitivity and specificity of measuring *mir-21* expressions by real-time RT-PCR in sputum were 69.66% (95% CI, 0.46–0.86) and 100.00% (95% CI, 0.77–1.00), respectively. Therefore, examination of *mir-21* expression had higher sensitivity than that of sputum cytology in the diagnosis and early detection of lung cancer patients ( $P < 0.05$ ).

## DISCUSSION

Our current study clearly shows that miRNAs are present in sputum in a very stable form. Expression of miRNAs could be readily and robustly measured by real-time RT-PCR in the easily accessible biological fluid with high sensitivity and specificity. Furthermore, examination of *mir-21* overexpression yields higher sensitivity than that of sputum cytology, suggesting that elevated *mir-21* expression would provide a useful biomarker for noninvasive diagnosis of lung cancer.

Sputum has the advantages as potential surrogate material for molecular genetic diagnosis of lung cancer, because of its non-invasive procurement, and the fact that it contains bronchial epithelial cells from the lungs and lower respiratory tract. Therefore, there is a long history of identifying and developing molecule genetic changes that can be tested in sputum as biomarkers (7). Although miRNAs have been proposed as tissue-based markers for predicting outcomes of lung cancer, analysis of miRNA expression in sputum poses several potential challenges. First, the high concentrations of RNase activity in sputum will degenerate RNA. Second, in clinical practice, there might be a time gap between collecting sputum from patients and processing the sample for analysis, during which, the likelihood of RNA degradation increases. Thus, it is imperative to investigate whether microRNA is stable in sputum and feasibility of detection of miRNA expression in sputum for diagnosis of lung cancer. Using real-time quantitative RT-PCR, we examined the expressions of endogenous and exogenous miRNAs, and found that endogenous miRNAs were resistant to RNase activity and present in sputum in a remarkably stable form. Furthermore, the miRNAs could reproducibly be detected after sputum was stored for up to 7 days, although there was significant degradation of total RNA. To our knowledge, this is the first report of feasibility of detection of miRNA in sputum. The results presented here will lay the basis and rationale to prompt future global investigations of miRNAs as sputum-based biomarkers for lung cancer.

*Mir-21* is located in chromosome 17q23.1 that is site of translocation breakpoints or amplifications in various cancers (18). Furthermore, *mir-21* has been found to specifically target and downregulate tumor suppressor genes, *PTEN* and *TPM1*, and therefore, might play an oncogenic role in tumorigenesis (19–20). Chen et al. recently found that *mir-21* was highly overexpressed in glioblastoma tumors compared with normal tissues (18). Moreover, inhibition of *mir-21* in glioblastoma cell lines triggered caspase production and led to increased programmed cell death. In the current study, we show that the real-time RT-PCR assay for the assessment of *mir-21* overexpression can detect abnormal specimens that escape cytologic examination, and provide a potential useful diagnostic biomarker for lung cancer. Furthermore,

2 of 3 stage I lung cancer patients are identified by measuring *mir-21* expression in sputum indicate that *mir-21* might serve as a biomarker for the early detection of lung cancer.

Elevated *mir-21* expression occurs in sputum of NSCLC patients with equal frequency among all histologic types and locations of lung tumor, suggesting that the genetic change might not be specific to histologic type. Detection of the abnormality may be useful in detecting all histologic types of NSCLC, including peripheral tumors, predominantly, adenocarcinomas of lung. This is particularly important because adenocarcinomas, which arise from the smaller airways, are difficult to be detected by bronchoscopy or sputum cytology and have become more prevalent than squamous cell lung cancers worldwide (1).

Furthermore, most of the previously identified lung cancer associated molecular genetic changes were related to the smoking status. Some of the changes can be found in healthy smokers who never develop lung cancer (7). The use of such molecular genetic alterations as biomarkers could produce high false positive diagnostic rate or over-diagnosis, impeding their application in clinical settings, especially in screening or early detection of lung cancer. We here find that the increased *mir-21* expression is not statistically significantly associated with number of smoking pack-years in both cancer patients and controls, suggesting that dysregulation of *mir-21* in lung cancer might not be tobacco smoking-related damage. Therefore, *mir-21* expression in sputum would provide a specific marker for the early detection of lung cancer with high accuracy.

*Mir-155* can target 3-prime UTR of *TP53INP1*, which is also target gene of p53 (22). Nikiforova et al. recently found that measuring altered *mir-155* expression can detect thyroid cancer with high accuracy in the surgical and preoperative fine-needle aspiration samples (22). Furthermore, high levels of *mir-155* in surgical lung tumor tissues correlated with shorter survival of the patients after resection (11); implying that elevated *mir-155* expression might be a prognostic biomarker of lung cancer. However, results from our current study showed that detection of *mir-155* expression in sputum could not distinguish the lung cancer patients from cancer-free controls. Possible explanations for the observation might be the followings. First, the primer and probe we utilized in the current study is different from ones used in the previous reports from others. The different primers might cause contradictory results. Second, dysregulation of *mir-155* expression might be associated with late-stage lung cancer and the recurrence of tumors (23). Although the detection of the *mir-155* expression in tumor tissues could predict relapse of cancer and poor outcomes of the patients, measuring the advanced stage lung tumor-related genetic change in sputum might not provide high accuracy for the early detection of lung cancer. Third, the sample size in the current study is small, additional studies in large case-control population will be needed to explore the difference.

Although the detection of *mir-21* expression in sputum seems to be more sensitive than cytology in the detection of lung cancer, the 69.99% sensitivity is still not yet efficient for routine clinical application. The result might indicate the need to develop a strategy for simultaneous assessment of a panel of tumor-specific miRNA biomarkers in sputum for highly sensitive and specific diagnosis of lung cancer. To that end, we are currently performing global miRNA characterization of bronchial epithelial cells from lung cancer patients to identify a panel of lung tumor specific miRNAs for diagnosis of lung cancer with high sensitivity and specificity.

In summary, although the sample size used in the research is small, the proof of principle study yields 1) interesting results on miRNA stability that provide a firm grounding for further investigation of this class of molecules as sputum-based biomarkers for lung cancer, and 2) measuring *mir-21* expression in sputum can distinguish patients with lung cancer from cancer-free controls with higher sensitivity, as compared with conventional sputum cytology.

Nevertheless, further validation of *mir-21* status in large populations and identification of additional miRNAs whose aberrations can be detected in sputum and specific to lung cancer are required.

## Acknowledgments

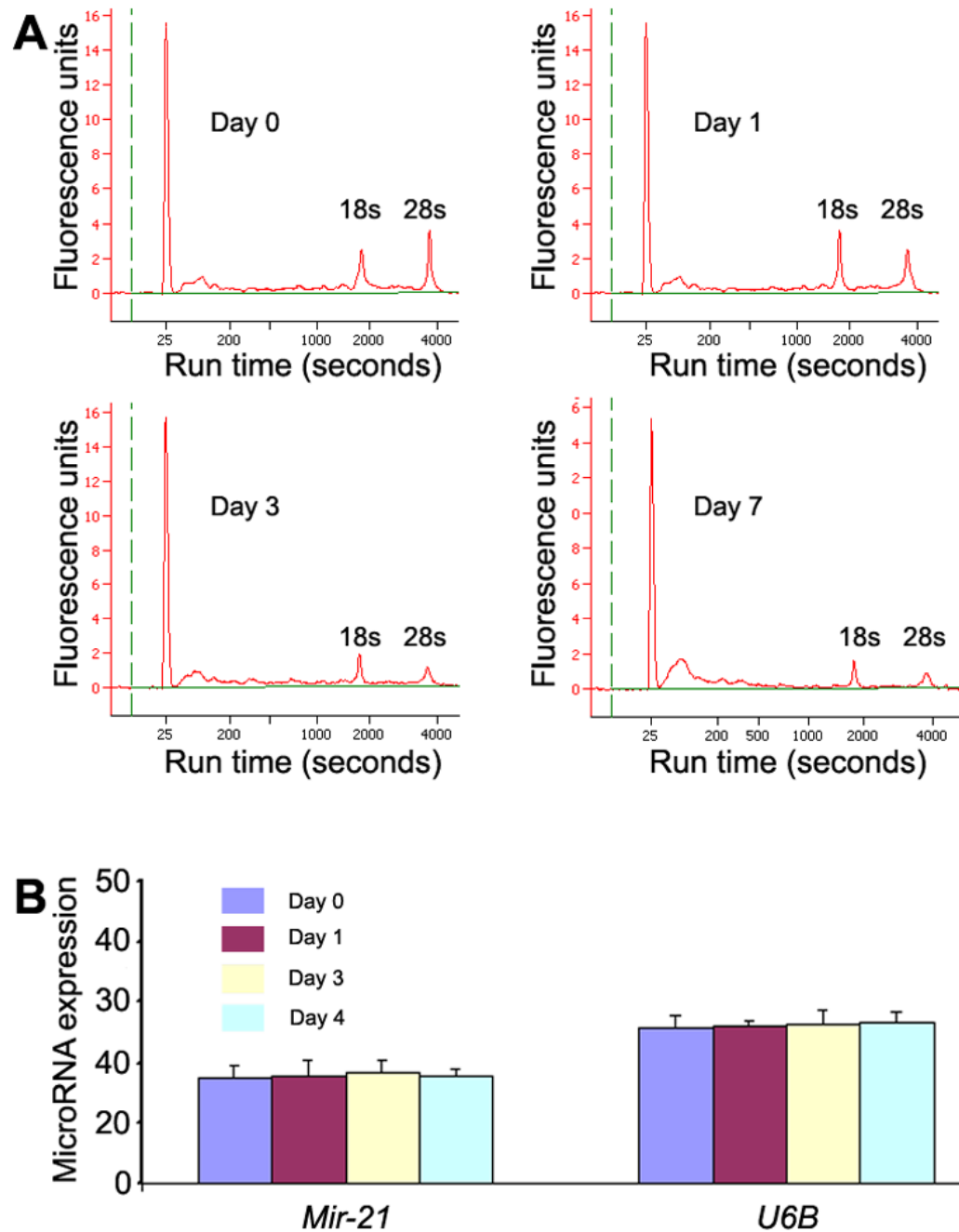
Grant Support: This work was supported in part by National Cancer Institute (NCI) grants CA-135382, CA-137742, and CA-133956, Wendy Will Case Cancer Award, an award from The University of Maryland Statewide Health Network, American Cancer Society Research Scholar Grant, Associate Member Award from NCI-The Early Detection Research Network, an clinical innovator award from Flight Attendant Medical Research Institute, and a scholar career development award from NIH K12RR023250-University of Maryland Multidisciplinary Research Career Development Program (F. J.).

## References

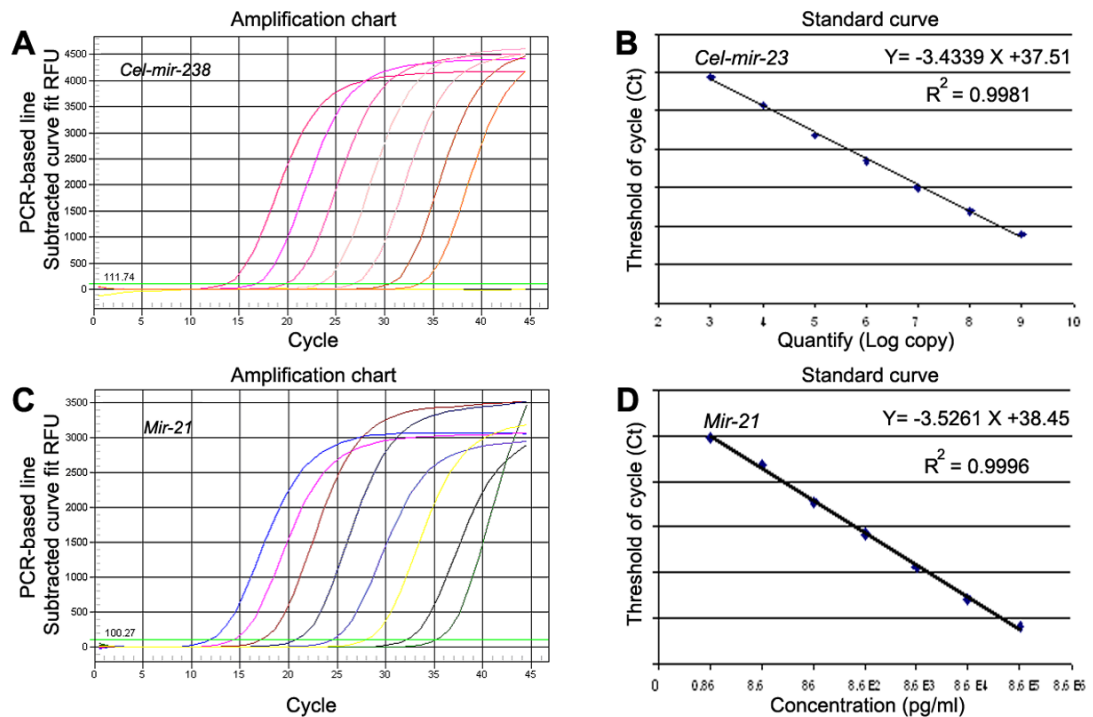
- Greenlee RT, Hill-Harmon MB, Murray T, Thun M. Cancer statistics, 2001. *CA Cancer J Clin* 2001;51:15–36. [PubMed: 11577478]
- Flehinger BJ, Melamed MR, Zaman MB, Heelan RT, Perchick WB, Martini N. Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Johns Hopkins study. *Am Rev Respir Dis* 1984;130:549–554. [PubMed: 6091505]
- Hirsch FR, Franklin WA, Gazdar AF, Bunn PA Jr. Early detection of lung cancer: clinical perspectives of recent advances in biology and radiology. *Clin Cancer Res* 2001;7:5–22. [PubMed: 11205917]
- Tolozza EM, Harpole L, McCrory DC. Noninvasive staging of non-small cell lung cancer: a review of the current evidence. *Chest* 2003;123:137S–146S. [PubMed: 12527573]
- Minna JD, Roth JA, Gazdar AF. Focus on lung cancer. *Cancer Cell* 2002;1:49–52. [PubMed: 12086887]
- Sacomanno G, Archer VE, Auerbach O, Saunders RP, Brennan LM. Development of carcinoma of the lung as reflected in exfoliated cells. *Cancer* 1974;33:256–70. [PubMed: 4810100]
- Thunnissen FB. Sputum examination for early detection of lung cancer. *J Clin Pathol* 2003;11:805–10. [PubMed: 14600122]
- Li R, Todd NW, Qiu Q, Fan T, Zhao RY, Rodgers WH, et al. Genetic deletions in sputum as diagnostic markers for early detection of stage I non-small cell lung cancer. *Clin Cancer Res* 2007;13:482–7. [PubMed: 17255269]
- Qiu Q, Todd NW, Li R, Peng H, Liu Z, Yfantis HG, Katz RL, Stass SA, Jiang F. Magnetic enrichment of bronchial epithelial cells from sputum for lung cancer diagnosis. *Cancer* 2008;114:275–83. [PubMed: 18484646]
- Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. *Science* 2001;294:853–8. [PubMed: 11679670]
- Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* 2006;9:189–98. [PubMed: 16530703]
- Ortholan C, Puissegur MP, Ilie M, Barbry P, Mari B, Hofman P. MicroRNAs and Lung Cancer: New Oncogenes and Tumor Suppressors, New Prognostic Factors and Potential Therapeutic Targets. *Curr Med Chem* 2009;16:1047–61. [PubMed: 19275611]
- Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res* 2004;64:3753–6. [PubMed: 15172979]
- Markou A, Tsaroucha EG, Kaklamanis L, Fotinou M, Georgoulas V, Lianidou ES. Prognostic value of mature microRNA-21 and microRNA-205 overexpression in non-small cell lung cancer by quantitative real-time RT-PCR. *Clin Chem* 2008;54:1696–704. [PubMed: 18719201]
- Sacomanno G, Archer VE, Auerbach O, Saunders RP, Brennan LM. Development of carcinoma of the lung as reflected in exfoliated cells. *Cancer* 1974;33:256–70. [PubMed: 4810100]
- Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, et al. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res* 2005;33:e179. [PubMed: 16314309]



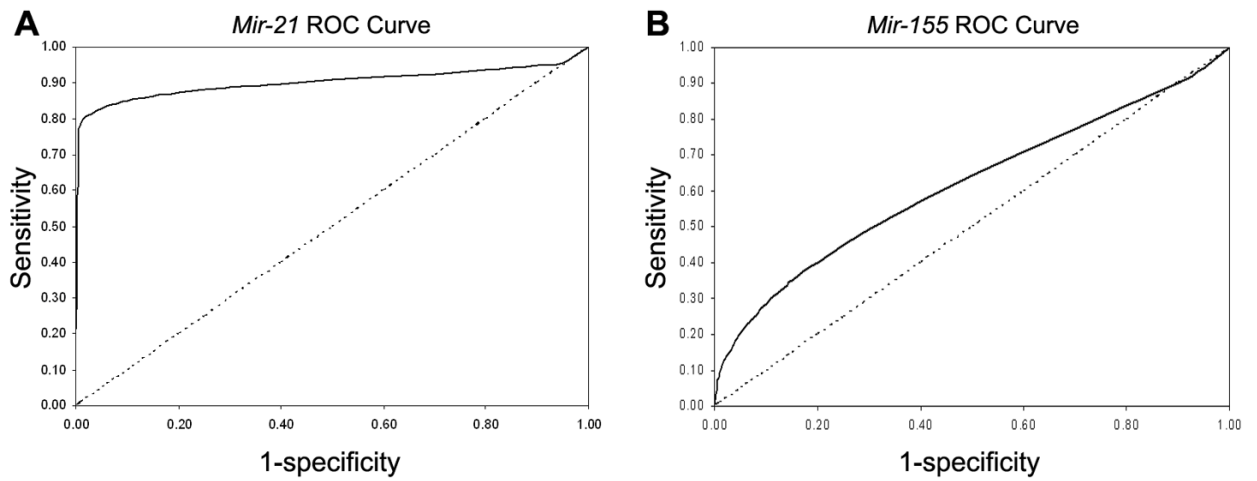
17. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Nat Acad Sci USA* 2008;105:10513–8. [PubMed: 18663219]
18. Chan JA, Krichevsky AM, Kosik KS. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res* 2005;65:6029–33. [PubMed: 16024602]
19. Pezzolesi MG, Platzer P, Waite KA, Eng C. Differential expression of PTEN-targeting microRNAs miR-19a and miR-21 in Cowden syndrome. *Am J Hum Genet* 2008;82:1141–9. [PubMed: 18460397]
20. Zhu S, Si ML, Wu H, Mo YY. MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). *J Biol Chem* 3007 282:14328–36.
21. Gironella M, Seux M, Xie MJ, Cano C, Tomasini R, Gommeaux J, et al. Tumor protein 53-induced nuclear protein 1 expression is repressed by miR-155, and its restoration inhibits pancreatic tumor development. *Proc Nat Acad Sci USA* 2007;104:16170–5. [PubMed: 17911264]
22. Nikiforova MN, Tseng GC, Steward D, Diorio D, Nikiforov YE. MicroRNA expression profiling of thyroid tumors: biological significance and diagnostic utility. *J Clin Endocr Metab* 2008;93:1600–1608. [PubMed: 18270258]



**Fig. 1.** Endogenous miRNAs could robustly and reliably be examined in sputum over seven-day storage. RNA assessed by capillary electrophoresis (A) show great degradation in sputum on day 7 compared with days 0, 1, and 3. However, the expression levels of each miRNA measured by real-time RT-PCR in the specimens do not change after seven-day storage (B). All assays were performed at least three times.

**Fig. 2.**

Sensitivity and dynamic range of microRNA quantification in sputum. A. Amplification curve of synthesized cel-miR-238 miRNA in RNA extracted from a denatured sputum sample. The assay showed high sensitivity and broad dynamic range between signal of 103 copies and 1010 copies input template and no-template-control background; B. standard curve of cel-miR-238 ( $R^2 = 0.9981$ , slope =  $-3.43$ ). C. Amplification plot of total RNA from sputum specimen at eight orders of magnitude for detection of mir-21. The total RNA input ranged from 0.86 pg/ml to 860ng/ml in RT-PCR reaction; D: Correlation of total RNA input with the threshold of cycle (Ct) values for mir-21 assay ( $R^2 = 0.9986$ , slope =  $-3.53$ ). All experiments were done in triplicates.



**Fig. 3.** Receiver-operator characteristic (ROC) curve analysis of expression of miRNAs in sputum. A, ROC curve with corresponding the area under the ROC curve (AUROC) for mir-21 expression in sputum from NSCLC patients vs cancer-free individuals. B. ROC curve with corresponding AUROC for mir-155 expression in sputum specimens from NSCLC patients vs cancer-free individuals. AUROC for each gene expression conveys its accuracy in distinguishing cancer-free subjects from cancer patients in terms of sensitivity and specificity.

**Table 1**

## Demographics in cases and controls

	NSCLC (n = 23)	Controls (n = 17)	P-value
Age	68.1 (SD 11.3)	45.5 (SD 16.7)	0.012
Sex			
Female	5	5	
Male	18	12	0.580
Smoking status			
Pack-years	40.5 (SD 24.1)	16.2 (SD 20.6)	0.0027
Location of tumor			
Central	9		
Peripheral	14		0.639
Stage			
Stage I	3		
Stage II	5		
Stage III	7		
Stage IV	8		
Histology			
AC of lung	13		
SC of lung	10		

Abbreviations: NSCLC, non-small-cell lung cancer; AC, adenocarcinoma; SC, squamous cell carcinoma.

**Table 2**Prevalence of *mir-21* expression in sputum of NSCLC patients

	No. of cases	Mean of <i>mi-21</i> expression (SD)	P-value*
Smoking status (Pack-year)			0.1885
<20	5	73.11 (6.96)	
21–40	7	71.54 (11.48)	
41–60	8	76.89 (9.84)	
61–100	3	74.62 (6.72)	
Location of tumor			0.4186
Central	9	78.43 (40.75)	
Peripheral	14	74.96 (9.27)	
Stage			0.1274
Stage I	3	78.51 (9.94)	
Stage II	5	68.86 (7.60)	
Stage III	7	72.41(10.88)	
Stage IV	8	83.58 (4.58)	
Histology			0.4633
AC of lung	13	78.89 (4.58)	
SC of lung	10	75.12 (4.19)	

Abbreviations: NSCLC, non–small-cell lung cancer; AC, adenocarcinoma; SC, squamous cell carcinoma.

\* Student's t test.

**Table 3**

Comparison of cytology and measurement of altered miRNA expression for the detection of lung cancer in sputum specimen

	<b>23 sputum from lung cancer patients</b>	
	<b>11 cytology positive</b>	<b>12 cytology negative</b>
No. sputum that are positive by detection of <i>mi-21</i> expression	11	6