ORIGINAL RESEARCH

A NOVEL COMPOSITE BIOMARKER PANEL FOR DETECTION OF EARLY STAGE NON-SMALL CELL LUNG CANCER

VOL 44 NO 1 MARCH 2021

Eric L.R. Bédard, MD^{1,3*}, Aswin G. Abraham, MD, DPhil^{2,3}, Anil A. Joy, MD^{3,5}, Sunita Ghosh, PhD^{3,5}, Xiaoyu Wang, MSc², Andrew Lim, MD^{2,3}, Dongyu Shao, MSc², Raimar Loebenberg, PhD⁴, Wilson H. Roa, MD^{2,3*}

¹ Division of Thoracic Surgery, Department of Surgery, University of Alberta, Edmonton, AB, Canada

² Division of Radiation Oncology, Department of Oncology, University of Alberta, Edmonton, AB, Canada

³ Alberta Health Services, Edmonton, AB, Canada

⁴ Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, AB, Canada

⁵ Division of Medical Oncology, Department of Oncology, University of Alberta, Edmonton, AB, Canada

* Both authors contributed equally

Submitted 9th September 2020, Accepted 31st January, 2021

Clin Invest Med 2021;44(1):E15-24; DOI: 10.25011/cim.v44i1.36016

ABSTRACT

Purpose: To investigate a novel composite methodology of using targeted serum microRNAs (micro ribonucleic acid; miRNA) and urine metabolites for the accurate detection of early stage non-small cell lung cancer (NSCLC).

Methods: Consecutively consenting NSCLC patients and matched control subjects were recruited to provide samples of serum for miRNA and/or urine for metabolite analyses. Serum miRNA levels were measured using quantitative real-time reverse-transcription with exogenous control, and the comparative delta cycle threshold (ΔC_T) method was used to calculate relative miRNA expression of two targeted miRNAs (miR-21 and miR-223). The concentrations of six targeted urinary metabolites in patients and healthy controls were measured using proton nuclear magnetic resonance (¹H NMR) spectroscopy. A composite methodology of using the 35 accruals with both serum and urine biomarkers was then established with binary logistic regression, receiver operating characteristic (ROC) models with or without artificial intelligence (AI).

Results: The ROC analysis of miRNA expression yielded a sensitivity of 96.4% and a specificity of 88.2% for the detection of early stage NSCLC, with area under the curve (AUC) = 0.91 (CI 95%: 0.80-1.0). Relative urinary concentrations of 4-methoxyphenylacetic acid (4MPA) were significantly different between NSCLC and healthy control (p=0.008). The ROC analysis of 4MPA yielded a sensitivity of 82.1% and a specificity of 88.2%, with AUC = 0.85. The composite process combining miRNA and metabolite expression demonstrated a sensitivity and specificity of nearly 100% and AUC=1.

Conclusions: A highly specific, sensitive and non-invasive detection method for NSCLC was developed. Pending validation, this can potentially improve the early detection and, hence, the treatment and survival outcomes of patients.

Correspondence to: Wilson Roa **Email:** wilson.roa@albertahealthservices.ca

INTRODUCTION

Non-small cell lung cancer (NSCLC) is the leading cause of cancer-related mortality worldwide [1]. Overall survival remains poor at approximately 19% [2] with most patients presenting with advanced or metastatic disease. Outcomes in stage I disease, however, may be as high as 88% at 10 years [3], and the early detection of lung cancer through screening programs may improve survival by

way of early treatment [6-9]. The National Lung Cancer Screening Trial (NLST) showed a 20% reduction in lung cancer mortality with low-dose computed tomography (CT) screening, albeit with a high false-positive rate [4]. The subsequent NELSON study did show improved sensitivity and specificity for low dose CT [5] and there was also a decrease in mortality by 26% in high-risk men and up to 61% in high-risk women over a 10-year period. Unfortunately, patients do need to have invasive

interventions [10-12] to verify the imaging findings. Additionally, cumulative radiation exposure from repeated scans may increase the risk of developing cancers [13]. Complimentary screening tools to improve the specificity and cost-efficiency of low dose CT screening would be highly valuable.

MicroRNAs (micro ribonucleic acids; miRNAs) are small noncoding single-stranded RNAs involved in cellular proliferation, differentiation, apoptosis and oncogenesis [14-16]. Their role in lung cancer is an area of active research [17-18], and they rank among the top biomarker candidates [19-20]. Our previous studies investigated the utility of sputum miRNAs for lung cancer detection and identified a panel of miRNAs that could differentiate early stage lung cancer patients from non-cancer patients with matched risk factors [21-25]. Serum and plasma samples are alternative bio-specimens for miRNA profiling that avoid the difficulty of poor sample quality which can commonly occur with sputum sampling [26-31]. Several studies have investigated large miRNA panels for lung cancer diagnosis with varying accuracy [26-29]. An ongoing challenge has been to reduce the size of the miRNA panel without compromising sensitivity or specificity. Wang et al. [30] and Nadal et al. [31] used a panel of just five and four serum miRNAs to differentiate NSCLC and healthy controls with significant accuracy. Unfortunately, the relatively large number of serum miRNAs used even in these studies, results in high costs and practical challenges for their application to clinical diagnosis and screening programs. Based on evidence from our own studies and those in the literature, our group identified a panel of four miRNAs that was then tested on retrospectively collected samples of plasma from NSCLC patients and controls from the Lung Cancer Biospecimen Resource Network (LCBRN) bio-repository [32,33].

Along with the use of miRNAs as biomarkers, metabolomics is a promising approach in systems biology. It involves the comprehensive analysis of low-molecular weight metabolites in biological samples [34-36]. Altered metabolism is considered an emerging cancer hallmark, with cancer cells exhibiting a distinct metabolic behavior. Therefore, assessing metabolic changes in biofluids such as plasma/serum and urine can relate to tumor onset and progression and has been used as a tool in cancer detection and screening [37-44]. Although most NMR metabolomics studies investigate metabolite profiling in plasma/serum, Carrola et al. [40] showed that the NMR metabolomics profiles of urine consistently differed between lung cancer patients and healthy subjects. Our pilot study indicated that a targeted metabolite, 4-methoxyphenylacetic acid (4MPLA), together with citrate, creatine ribosome, creatinine, choline, N-acetylneuraminic acid (NANA), could be selected as

potential reference biomarkers in urine to discriminate between lung cancer patients and healthy controls [37-40]. Nuclear magnetic resonance (NMR) spectroscopy has the advantages of exhibiting high reproducible metabolite measurement and requiring minimum sample preparation as compared with other techniques such as mass spectrometry [44-45]. Metabolomics, using ^1H NMR spectroscopy, has been used for the identification of different biomarkers in lung [38-41], breast [42] gastric, prostate [43] and breast [44] cancers, and may provide a faster and cheaper diagnosis platform of lung cancer than miRNA profiling [44-45].

Based on our previous work [32,33], we have used two miRNA candidates for targeted serum miRNA profiling and one urine metabolite in the present study to establish a composite methodology that can discriminate early stage non-small cell lung cancer accurately.

MATERIALS AND METHODS

Serum samples and urine samples of consecutively consenting early stage NSCLC patients and controls subjects were collected. Ethics approval was granted from the institutional and provincial ethics committees.

Inclusion criteria

Patients with histologically proven (by biopsy or cytology) stage I and II NSCLC patients aged ≥ 18 years old and ≤ 75 years old, with an Eastern Cooperative Oncology Group performance status $\leq$ two, a life expectancy $\geq$ three months, and who were willing to complete a questionnaire about personal medical history were included in this study. Control individuals without lung cancer were recruited from pulmonology clinics, and included patients attending the clinics for management of non-cancer lung diseases.

Serum MiRNA analysis

For quality control, one NSCLC serum sample with a hemolysis index (HI) > 200 was excluded. Data on samples available separately upon request. miR-39 (cel-miR-39) was added to each 150 μL serum sample to serve as synthetic exogenous control. RNA was extracted and purified using miRNeasy Serum/Plasma Kit (Qiagen, Canada) according to manufacturer protocols. RNA was converted to complementary DNA by reverse transcription (RT) using TaqMan MicroRNA RT Kit and TaqMan MicroRNA Assay's specific stem-loop primers (Applied Biosystems, USA). Reverse transcription reactions were mixed according to manufacturer protocols. Expression levels of miRNAs (miR-21, miR223 and the control, cel-miR-39) were quantified using quantitative real-time reverse-transcription polymerase chain reaction (RT-qPCR) analysis. Cycle threshold (C_t), defined as the number of

PCR cycles required for a fluorescent signal to be higher than baseline variability. Using cel-miR-39 as the control, relative expression levels of serum miRNAs were calculated and represented by $2^{-\Delta C_T}$, where $\Delta C_T = C_T$ (miRNA of interest) – C_T (cel-miR-39).

Metabolomics analysis

For NMR analysis, 400 μ L of clear urine was mixed with 200 μ L of buffer solution (KH_2PO_4 1.5 M in D_2O ; containing 0.1% TSP-d4 as a chemical shift reference). NMR [39, 45] spectra were acquired on a Bruker Avance DRX-600 spectrometer according to standard practise. The chemical shift (peak position) and intensity were automatically determined by TopSpin (V3.2, Bruker, Germany). The intensity of the strongest peak of the targeted metabolite was used as the apparent and relative concentration. This concentration of each sample was then used as the variable for the case and control.

Conventional statistical analysis

Mean and standard deviations were reported for normally distributed continuous variables and median (range) was reported for non-normally distributed continuous variables. Frequency and proportions were reported for categorical variables. Two categorical variables were compared using Chi-square tests. Fisher's exact tests were reported when the cell frequencies were less than five. Mann-Whitney test was used to compare the continuous variables between two groups. Binary logistic regression (BLR) and receiver-operator curve (ROC) was used to discriminate lung cancer patients and healthy controls. The cut-off point obtained from the ROC curve was selected based on a point that maximizes sensitivity and specificity.

Statistical analysis was conducted using SPSS version 22 software (IBM SPSS, USA). A p -value <0.05 was used for all statistical comparison and two-tailed tests were used.

Artificial intelligence tool development

An algorithm was generated using the TensorFlow™ open-source software library. The artificial neural network was given a training set of data from combinations of miRNA expression levels and urine metabolite concentrations data from both controls and cancer patients for a randomly selected subset of 16 individuals, and an AUC set at one to predict the outcome of interest (lung cancer subjects versus healthy controls). A 3 X 3 neural network with a rectified linear activation unit activation function was used. Hyperparameter optimization to configure the convolution layers was explored using the Bayesian method. The machine learning was

then verified using an additional validation study set of seven other individuals. Lastly, a separate test data set including data from 11 individuals was submitted for evaluation.

RESULTS

Patient population

Forty-six subjects with matched serum samples and fifty-six subjects with matched urine samples were included in the study. A total of thirty-five subjects provided matched samples of both serum and urine. Table 1 displays the baseline disease characteristics of the NSCLC patients. The NSCLC and control groups did not differ significantly in the group of subjects that gave serum samples and the group that had both serum and urine samples in terms of smoking history, gender and pulmonary disease.

Evaluation of exogenous control

Cel-miR-39 was detectable by RT-qPCR for all 45 subjects, and 200 copies was the lower limit of detection ($C_T = 35.29 \pm 0.36$). There was no significant difference ($p=0.59$) in C_T between the NSCLC group ($C_T=28.38 \pm 0.725$) and the control group ($C_T=28.51 \pm 0.78$), and thus spiked-in cel-miR-39 was validated as a suitable exogenous control.

Expression levels of miRNA

Both miR-21, and miR-223 were detectable by RT-qPCR in all 45 subjects. Table 2 and box plot of the same (Figure 1A) shows the miR21 and miR223 expressions in NSCLC subjects and cancer-free controls.

The expression levels of miR-21 and miR-223 were analysed using BLR and ROC. The combined effect of miR-21 and miR-223 allowed for a model of risk threshold determination with a weighted combination of miR-21 and miR-223 based on the opposite bio-functional contributions, which correlate with the regulatory mechanisms of miR-21 and miR-223 for cancer.

A genomic variable from miRNA binary logistic regression analysis (genomVAR) was thus generated to express a subject's disease status as follows:

$$\text{genomVAR} = 0.840 + 2.389\text{RQmiR21} - 2.70\text{RQmiR223}$$

where RQ represents RNA quantification data from RT-qPCR and the coefficients indicate the weighted contributions of miR-21 and miR-223.

ROC analysis of genomVAR was used to differentiate NSCLC and cancer-free controls (Figure 1B, right) and showed an AUC of 0.912 (0.792-1.0). A cut-point of 0.4944 can differentiate NSCLC and controls with 96.4% sensitivity (27/28) and 88.2% specificity (15/17).

TABLE 1.
PARTICIPANT BASELINE CHARACTERISTICS AND TUMOR CHARACTERISTICS.

Participant baseline characteristics		NSCLC patients	Controls	P-value
45 patients with serum samples included for analysis				
Number of participants		n = 28	n = 17	-
Age, years; median [range]		67 [51-76]	57 [46-74]	0.022
Gender, n [%]	Male	12 [42.81]	9 [52.9]	0.511
	Female	16 [57.2]	8 [47.1]	
Current smoker, n [%]	Yes	12 [42.9]	8 [47.1]	0.783
	No	16 [57.1]	9 [52.9]	
Smoking history, pack-years; mean [SD]		42 [21.1]	33.8 [21.3]	0.233
COPD, n [%]	Yes	7 [25]	1 [5.9]	0.132
	No	21 [75]	16 [94.1]	
Tumor histology, n [%]				
Adenocarcinoma		19 [67.8]	NA	-
Squamous cell carcinoma		7 [25]		
Large cell carcinoma		1 [3.6]		
Sarcomatoid carcinoma		1 [3.6]		
TNM stage (AJCC 7), n [%]				
T1aN0M0		7 [25]	NA	-
T1bN0M0		5 [17.9]		
T2aN0M0		9 [32.1]		
T2bN0M0		1 [3.6]		
T3N0M0		2 [7.1]		
T1bN1M0		1 [3.6]		
T2aN1M0		2 [7.1]		
T2bN1M0		1 [3.6]		
Maximum tumor diameter, cm				
Mean [SD]		2.8 [1.7]	NA	-
35 patients with both serum and urine samples included for analysis				
Number of participants		n = 21	n = 14	-
Age, years; median [range]		64 [51-75]	60 [46-74]	-
Gender, n [%]	Male	12 [57.1]	8 [57.1]	0.999
	Female	9 [42.9]	6 [42.9]	
Current smoker, n [%]	Yes	11 [52.4]	6 [42.9]	0.581
	No	10 [47.6]	8 [57.1]	
Smoking history, pack-years Mean [SD]		42 [18.1]	34 [22.2]	0.170
COPD, n [%]	Yes	7 [33.3]	1 [7.1]	0.162
	No	14 [66.6]	13 [92.9]	
Tumor histology, n [%]				
Adenocarcinoma		16 [76.2]	NA	-
Squamous cell carcinoma		4 [19]		
Large cell carcinoma		1 [4.8]		
TNM stage (AJCC 7), n [%]				
T1aN0M0		7 [20]	NA	-
T1bN0M0		3 [8.6]		
T2aN0M0		6 [17.1]		
T2bN0M0		1 [2.9]		
T3N0M0		1 [2.9]		
T1bN1M0		1 [2.9]		
T2aN1M0		1 [2.9]		
T2bN1M0		1 [2.9]		
Maximum tumor diameter, cm				
Mean [SD]		2.8 [1.7]	NA	-

Abbreviations: AJCC, American Joint Committee on Cancer; COPD, chronic obstructive pulmonary disease; n, number; NA, not applicable; NSCLC, non-small cell lung cancer; TNM, tumor node metastases

TABLE 2.
EXPRESSION LEVEL AND RECEIVER OPERATING CHARACTERISTIC ANALYSIS OF TWO MICRO-RNAS IN NON-SMALL CELL LUNG CANCER PATIENTS AND CONTROLS.

miRNA	NSCLC patients (n=28)		Controls (n=17)		AUC in ROC
	Mean	SD	Mean	SD	
miR-21	1.591	3.679	0.524	0.665	0.63 (95% CI: 0.47-0.80)
miR-223	0.795	2.020	2.312	3.300	0.79 (95% CI: 0.64-0.95)

Abbreviations: AUC, area under the curve; NSCLC, non-small cell lung cancer; n, number; ROC, receiver operating characteristics; SD, standard deviation

A ratio variable genomic ratio of miRNAs (genomRAT) of miR-21 to miR-223 for detection of NSCLC was expressed as follows:

$$\text{genomRAT} = \text{RQmiR21}/\text{RQmiR223}$$

where RQ represents RNA quantification data from quantitative real-time reverse-transcription polymerase

chain reaction for miR-21 and miR-223. The exogenous control cel-miR-39 did not affect genomRAT as a ratio. The ROC analysis of genomRAT (Figure 1B, left) was used to differentiate NSCLC and cancer-free controls. The simplified variable reached the same level of performance as genomVAR, with an AUC of 0.876 (95%: 0.723-1.0). A cut-point of 0.8060 can differentiate NSCLC and controls with 96.4% sensitivity (27/28) and 88.2% specificity (15/17). Figure 1B demonstrates ROC curve for genomRAT.

Given the relatively small sample size, internal validation of miRNA data was conducted using k-fold cross-validation. The data were divided into three parts (k=3). The cut points were determined on the training set by repeating the ROC five times on a randomly chosen training set. The results were averaged to yield the final values. The cut point was determined on the training set and validated using the test set. The AUC for the training set was 86.46% with a sensitivity of 94.4% and specificity of 92.0%. The AUCs for the training set and the test set with the cut point (0.91297) were 89.26% and 92.98%, respectively.

FIGURE 1A.
BOX PLOT OF MIR21 AND MIR223 FOR LUNG CANCER (LC) AND HEALTHY CONTROL (HV) SHOWING MEDIAN EXPRESSION OF MIR-21 AND MIR223 WITH THE RESPECTIVE 95% CI.

FIGURE 1B.
ROC FOR genomVAR (RIGHT) AND genomRAT (LEFT) FOR MIRNA PROFILING OF N=45 SUBJECTS (N=28 NSCLC AND N=17 HEALTHY CONTROLS) SHOWED AN AUC OF 0.912 AND 0.876, RESPECTIVELY.

TABLE 3.
STATISTICAL RESULTS OF SIX TARGETED METABOLITES FOR LUNG CANCER AND HEALTHY CONTROL.

Metabolite	Lung cancer mean $\pm$ std	Healthy control mean $\pm$ std	P-value	AUC
4MPLA	0.380 $\pm$ 0.429	1.349 $\pm$ 1.539	0.001	0.83 (95% CI: 0.71-0.94)
Citrate	0.863 $\pm$ 1.383	0.303 $\pm$ 0.431	0.061	0.73 (95% CI: 0.59-0.86)
Creatine ribosome	0.271 $\pm$ 0.566	0.662 $\pm$ 1.208	0.112	0.68 (95% CI: 0.53-0.82)
Creatinine	0.137 $\pm$ 0.157	0.170 $\pm$ 0.180	0.458	0.70 (95% CI: 0.57-0.84)
Choline	0.208 $\pm$ 0.131	0.250 $\pm$ 0.169	0.298	0.57 (95% CI: 0.42-0.73)
NANA	0.193 $\pm$ 0.597	0.066 $\pm$ 0.024	0.304	0.59 (95% CI: 0.43-0.74)

Abbreviations: 4MPLA, 4-methoxyphenylacetic acid; AUC, area under the curve; NANA, N-acetylneuraminic acid

FIGURE 2A.
BOX PLOT OF 4MPLA (LEFT) AND CITRATE (RIGHT) FOR LUNG CANCER (LC) AND HEALTHY CONTROL (HV) SHOWING MEDIAN EXPRESSION OF 4MPLA AND CITRATE WITH THE RESPECTIVE 95% CI.

FIGURE 2B.
ROC OF URINE METABOLITE PROFILING OF metabVAR (LEFT) AND metabRAT (RIGHT) FOR N=56 SUBJECTS (N=32 NSCLC AND N=24 HEALTHY CONTROLS) SHOWED AN AUC OF 0.827 AND 0.854, RESPECTIVELY.

NMR metabolomics

For metabolites, only 4MPLA demonstrated significant difference ($p < 0.05$) (Table 3). The mean concentration of 4MPLA for lung cancer was significantly lower than healthy control ($p = 0.001$), and the mean concentration of citrate for lung cancer patient was higher than healthy control ($p = 0.062$) (Figure 2A). The ROC of 4MPLA demonstrated a sensitivity of 82.1% and a specificity of 88.2%, and the ROC of citrate demonstrated a sensitivity of 67.9% sensitivity and a specificity of 70.6%. 4MPLA and citrate were used as covariates for binary logistic regression owing to their statistical significance (Table 3). The metabolite model process was as follows:

$$\text{Metabolomics variable (metabVAR)} = -0.828 + 1.588C_{4\text{MPLA}} - 1.103C_{\text{Citrate}}$$

For the BLR model, a cut-point of -0.5661 for the metabVAR demonstrated a sensitivity of 87.5% and a specificity of 71.9% to differentiate NSCLC from control samples and the AUC of the ROC was noted to be 0.827 (95%: 0.713-1.00). We further created a metabolomics ratio (metabRAT) of 4MPLA and citrate as a simple ratio to discriminate NSCLC:

$$\text{metabRAT} = C_{4\text{MPLA}} / C_{\text{Citrate}}$$

A cut-point of 1.3641 for the metabRAT demonstrated a sensitivity of 87.5% and a specificity of 75% to differentiate NSCLC from controls, and the AUC of ROC is 0.854 (95%: 0.751-0.958). Figure 2B demonstrates the ROC curves for metabVAR and metabRAT.

In a manner similar to the internal validation of the miRNA data, internal validation of the metabolomics data was conducted using k-fold cross-validation. The cut point determined on the training set for 4MPLA was 0.405476 and validated using the test set. The AUC for the training set was 0.826 (95% CI: 0.713-0.938), with a sensitivity of 87.5% and specificity of 78.1%. The AUCs for the training set and the test set with the above cut point were 79.2% and 87.5%, respectively.

Combining miRNA profiling and metabolomics

Although the miRNA profiling and metabolomics were measured on different analytical platforms, miRNA expressions and metabolite concentrations are continuous variables that correspond to the biological state of a human body, including its oncological related mechanisms. We established a mathematical model from the combined miRNA profiling of miR-21 and miR-223, and metabolomics data using the single metabolite 4MPLA to further improve detection. Binary logistic regression was used to test miRNA profiling plus metabolomics, resulting in a combined variable from miRNA (comboVAR) expressed as follows:

$$\text{ComboVAR} = 87.32 + 346.29\text{RQmiR21} - 412.60\text{RQmiR223} - 53.72C_{4\text{MPLA}}$$

ROC of comboVAR identified a cut-point of 1.361 that could accurately differentiate NSCLC from controls with

FIGURE 3.
ROC CURVE FOR COMBINATION MODELS OF COMBOVAR FOR N=35 SUBJECTS SHOWED AN AUC OF 1.00 (N=21 NSCLC AND N=14 HEALTHY CONTROLS).

a sensitivity of 100% and a specificity of 100%, and an AUC of 1.00. Figure 3 shows the ROC curve for comboVAR. Addition of a further metabolite beyond 4MPLA to the analysis was not deemed to be of any additional benefit.

Internal validation of combining miRNA expression levels and urine metabolite concentrations was done using k-fold cross-validation. The cut point, which was determined on the training set, was -1.24675 and validated using the test set. The AUC for the training set was 92.62 with a sensitivity of 100% and specificity of 86.7%. The AUCs for the training set and the test set with the cut point were 92.9% and 85.7%, respectively.

Artificial intelligence tool development

The artificial neural network yielded a sensitivity of 100% and a specificity of 100%, and an AUC of 1.00 for predicting the presence of cancer, all using the same selected combinations of miRNA expression levels and urine metabolite concentrations.

DISCUSSION

A novel non-invasive process using only body fluid samples for the early detection of lung cancer could dramatically alter the current conventional clinical practice. The results of this study demonstrate that using a targeted panel of two miRNAs from serum samples and one targeted urine metabolite can distinguish between patients with early stage NSCLC and controls with a sensitivity and specificity between 90-100%. In our study, the use of two miRNAs alone gave a sensitivity of 96.4% and a specificity of 88.2% to detect NSCLC which is similar to or better than previously reported larger miRNA panels [28-33]. A panel of two miRNAs is more cost-effective

than larger miRNA panels and would facilitate screening programs on a population level. The findings from this study have been registered for patent protection.

Sputum has been used as a non-invasive bio-specimen, but often requires multiple collections and patients often have difficulty producing a quality sample. Our early work in NSCLC used sputum miRNA profiling and demonstrated sensitivities and specificities ranging between 67–85% [23–25], which has recently been confirmed by others [26–27]. On the other hand, utilizing serum as a bio-specimen ensures ease of sample collection and reproducibility. Subsequent work from our group on retrospectively collected serum samples from the Lung Cancer Biospecimen Resource Network (LCBRN) bio-repository was able to verify a panel of miRNAs that could identify early NSCLC with high accuracy [32, 33]. This information was used to identify the two miRNA panels used for the analyses in this study.

The ROC analyses of the individual miRNAs were conducted to assess the individual ability to discriminate between NSCLC and control subjects. The AUCs for miRNA-21 and miRNA-223 ranged from 0.63 and 0.79, respectively (Table 2). To determine the diagnostic ability of miR21 and miR223 in combination, we conducted a risk score analysis and the AUC improved to 0.91, with sensitivity of 96.4% and specificity of 88.2%.

We utilized synthetic spiked-in cel-miR-39 as miRNA exogenous control, and thus ensured effective miRNA extraction, complementary DNA synthesis, and PCR amplification. Cel-miR-39 avoids the issue of frequent hemolysis that can occur with miR-16 as an endogenous control [47–48], and the problem of extensive RNase-mediated degradation with using U6 snRNA [49].

Although metabolomics has attracted attention owing to its close relevance to phenotype [50], it is also challenging due to the sheer number of different molecules present in biofluids like urine [51], and at present only about 220 metabolites have been identified and quantitated by NMR or liquid chromatography-mass spectrometry [52]. In cancer diagnosis by urine NMR metabolomics, only a limited number of metabolites can be quantified by untargeted metabolomics [40, 42] due to the significant overlap of the NMR spectra of the different biomolecules, which makes untargeted metabolomics a complex challenge. Hence, targeted metabolomics is the way forward, and our study identified that the relative urinary concentrations of 4MPLA and citrate were significantly different between NSCLC cases and matched controls, making them potential targets for use in detection of NSCLC.

A composite methodology showed unique results: the miRNA expression alone yielded a sensitivity of 96.4% and a specificity of 88.2% for the detection of early stage

NSCLC, with AUC = 0.91 (CI 95%: 0.80–1.0), whereas ROC analysis of the metabolite 4MPLA alone yielded a sensitivity of 82.1% and a specificity of 88.2%. The composite methodology of miRNA/4MPLA with internal validation yielded a sensitivity of 100.0% and a specificity of 86.7%. Evaluation of the data using machine-learning further improved our findings of detecting early stage NSCLC, with sensitivity and specificity of close to 100%.

Naturally, our findings may be skewed because of the relatively small data set used in the study. To verify these results, we are currently conducting a large prospective validation trial, which should be completed in a year. If the findings support the current results, then it may be a landmark step towards early non-invasive detection of lung cancers. The low cost of running the tests and the ease of scaling down the metabolomics testing to point-of-care devices are other potential areas that can have a significant impact on routine uptake of this panel.

CONCLUSION

We have described a composite methodology using serum miRNAs and urine metabolites to detect early stage NSCLC with sensitivity and specificity of close to 100%. Pending validation, this is likely to improve the early detection of early lung cancers and will have a clinical impact by increasing the likelihood of providing curative treatment and prolonged survival.

ACKNOWLEDGEMENTS

We would like to thank Dr. Mary Hitt, Division of Experimental Oncology, 5-142H Katz Group Centre for Pharmacy and Health Research, University of Alberta, Edmonton, Alberta for the laboratory support. We would like to thank Dr. Linghong Guo and Dr. Rene Razzak for technical support.

FUNDING

This study was partly funded by the Alberta Cancer Foundation.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359–386.
2. SEER Stat Fact Sheets: Lung and Bronchus Cancer. Surveillance epidemiology and end results database. <http://seer.cancer.gov/statfacts/html/lungb.html>. Published 2016. Accessed 3 July, 2016.
3. Henschke CI, Yankelevitz DF, Libby DM, et al. Survival of patients with stage I lung cancer detected on CT screening. *New Engl J Med*. 2006;355(17):1763–1771.

4. Aberle DR, Adams AM, Berg CD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New Engl J Med*. 2011;365(5):395-409.
5. Horeweg N, Scholten ET, de Jong PA, et al. Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers. *Lancet Oncol*. 2014;15(12):1342-1350.
6. Manser R, Lethaby A, Irving LB, et al. Screening for lung cancer. *Cochrane Database Syst Rev*. 2013(6):CD001991.
7. Kaneko M, Eguchi K, Ohmatsu H, et al. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. *Radiology*. 1996;201(3):798-802.
8. Sone S, Takashima S, Li F, et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. *Lancet*. 1998;351(9111):1242-1245.
9. Sobue T, Moriyama N, Kaneko M, et al. Screening for lung cancer with low-dose helical computed tomography: anti-lung cancer association project. *J Clin Oncol*. 2002;20(4):911-920.
10. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet*. 1999;354(9173):99-105.
11. Aberle DR, Abtin F, Brown K. Computed Tomography Screening for Lung Cancer: Has It Finally Arrived? Implications of the National Lung Screening Trial. *J Clin Oncol*. 2013;31(8):1002-1008.
12. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and Harms of CT Screening for Lung Cancer A Systematic Review. *JAMA*. 2012;307(22):2418-2429.
13. Sarma A, Heilbrun ME, Conner KE, Stevens SM, Woller SC, Elliott CG. Radiation and Chest CT Scan Examinations What Do We Know? *Chest*. 2012;142(3):750-760.
14. Gregory RI, Shiekhhattar R. MicroRNA biogenesis and cancer. *Cancer Res*. 2005;65(9):3509-3512.
15. Lu J, Getz G, Miska EA, et al. MicroRNA expression profiles classify human cancers. *Nature*. 2005;435(7043):834-838.
16. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nature Reviews Cancer*. 2006;6(11):857-866.
17. Weiss GJ, Bemis LT, Nakajima E, et al. EGFR regulation by microRNA in lung cancer: correlation with clinical response and survival to gefitinib and EGFR expression in cell lines. *Annals of Oncology*. 2008;19(6):1053-1059.
18. Eder M, Scherr M. MicroRNA and lung cancer. *New Engl J Med*. 2005;352(23):2446-2448.
19. Schultz NA, Dehlendorff C, Jensen BV, et al. MicroRNA Biomarkers in Whole Blood for Detection of Pancreatic Cancer. *JAMA*. 2014;311(4):392-404.
20. Matamala N, Vargas MT, Gonzalez-Campora R, et al. Tumor MicroRNA Expression Profiling Identifies Circulating MicroRNAs for Early Breast Cancer Detection. *Clin Chem*. 2015;61(8):1098-1106.
21. Roa WH, Kim JO, Razzak R, et al. Sputum MicroRNA Profiling: A Novel Approach for the Early Detection of Non-Small Cell Lung Cancer. *Clin Invest Med*. 2012;35(5):E271-E281.
22. Kim JO, Gazala S, Razzak R, et al. Non-small Cell Lung Cancer Detection Using MicroRNA Expression Profiling of Bronchoalveolar Lavage Fluid and Sputum. *Anticancer Res*. 2015;35(4):1873-1880.
23. Razzak R, Bedard ELR, Kim JO, et al. MicroRNA expression profiling of sputum for the detection of early and locally advanced non-small-cell lung cancer: a prospective case-control study. *Current Oncology*. 2016;23(2):E86-E94.
24. Shen J, Liao JP, Guarnera MA, et al. Analysis of MicroRNAs in Sputum to Improve Computed Tomography for Lung Cancer Diagnosis. *J Thorac Oncol*. 2014;9(1):33-40.
25. Xing LX, Su J, Guarnera MA, et al. Sputum microRNA Biomarkers for Identifying Lung Cancer in Indeterminate Solitary Pulmonary Nodules. *Clin Cancer Res*. 2015;21(2):484-489.
26. Bianchi F, Nicassio F, Marzi M, et al. A serum circulating miRNA diagnostic test to identify asymptomatic high-risk individuals with early stage lung cancer. *Embo Mol Med*. 2011;3(8):495-503.
27. Wozniak MB, Scelo G, Muller DC, Mukeria A, Zaridze D, Brennan P. Circulating MicroRNAs as Non-Invasive Biomarkers for Early Detection of Non-Small-Cell Lung Cancer. *PLoS ONE*. 2015;10(5).
28. Boeri M, Verri C, Conte D, et al. MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. *Proc Natl Acad Sci U S A*. 2011;108(9):3713-3718.
29. 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(8):768-773.
30. Wang C, Ding M, Xia MD, et al. A Five-miRNA Panel Identified From a Multicentric Case-control Study Serves as a Novel Diagnostic Tool for Ethnically Diverse Non-small-cell Lung Cancer Patients. *Ebiomedicine*. 2015;2(10):1377-1385.
31. Nadal E, Truini A, Nakata A, et al. A Novel Serum 4-MicroRNA Signature for Lung Cancer Detection. *J Thorac Oncol*. 2015;10(9):S304-S304.
32. Gyoba J, Roa W, Guo L, Ghosh S, Bedard E. Lung Cancer Risk Score Analysis Using Plasma microRNA Profiles. *J Thorac Oncol*. 2017;12(11):S1857-S1857.
33. Gyoba J. Translational Application of microRNA Profiling to Detect Non-Small Cell Lung Cancers: Department of Surgery, University of Alberta; 2017.
34. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Bio*. 2016;17(7):451-459.
35. Liu XJ, Locasale JW. Metabolomics: A Primer. *Trends Biochem Sci*. 2017;42(4):274-284.
36. Ritchie MD, Holzinger ER, Li RW, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet*. 2015;16(2):85-97.
37. Hwang VJ, Weiss RH. Metabolomic profiling for early cancer detection: current status and future prospects. *Expert Opin Drug Met*. 2016;12(11):1263-1265.

38. Klupczynska A, Dereziński P, Garrett TJ, et al. Study of early stage non-small-cell lung cancer using Orbitrap-based global serum metabolomics. *J Cancer Res Clin.* 2017;143(4):649-659.
39. Mazzone PJ, Wang XF, Beukemann M, et al. Metabolite Profiles of the Serum of Patients with Non-Small Cell Carcinoma. *J Thorac Oncol.* 2016;11(1):72-78.
40. Carrola J, Rocha CM, Barros AS, et al. Metabolic Signatures of Lung Cancer in Biofluids: NMR-Based Metabonomics of Urine. *J Proteome Res.* 2011;10(1):221-230.
41. Rocha CM, Carrola J, Barros AS, et al. Metabolic Signatures of Lung Cancer in Biofluids: NMR-Based Metabonomics of Blood Plasma. *J Proteome Res.* 2011;10(9):4314-4324.
42. Chan AW, Mercier P, Schiller D, et al. H-1-NMR urinary metabolomic profiling for diagnosis of gastric cancer. *Br J Cancer.* 2016;114(1):59-62.
43. Gkotsos G, Virgiliou C, Lagoudaki I, et al. The Role of Sarcosine, Uracil, and Kynurenine Acid Metabolism in Urine for Diagnosis and Progression Monitoring of Prostate Cancer. *Metabolites.* 2017;7(1).
44. Zhou J, Wang Y, Zhang X. Metabonomics studies on serum and urine of patients with breast cancer using 1 H-NMR spectroscopy. *Oncotarget.* 2017.
45. Markley JL, Bruschweiler R, Edison AS, et al. The future of NMR-based metabolomics. *Curr Opin Biotech.* 2017;43:34-40.
46. Beckonert O, Keun HC, Ebbels TMD, et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc.* 2007;2(11):2692-2703.
47. Navarro A, Diaz T, Gallardo E, et al. Prognostic Implications of miR-16 Expression Levels in Resected Non-Small-Cell Lung Cancer. *J Surg Oncol.* 2011;103(5):411-415.
48. Kirschner MB, Kao SC, Edelman JJ, et al. Haemolysis during Sample Preparation Alters microRNA Content of Plasma. *PLoS ONE.* 2011;6(9).
49. Xiang M, Zeng Y, Yang R, et al. U6 is not a suitable endogenous control for the quantification of circulating microRNAs. *Biochem Biophys Res Commun.* 2014;454(1):210-214.
50. Emwas AHM, Salek RM, Griffin JL, Merzaban J. NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. *Metabolomics.* 2013;9(5):1048-1072.
51. Moldovan L, Batte KE, Trgovcich J, Wisler J, Marsh CB, Piper M. Methodological challenges in utilizing miRNAs as circulating biomarkers. *J Cell Mol Med.* 2014;18(3):371-390.
52. Bouatra S, Aziat F, Mandal R, et al. The Human Urine Metabolome. *PLoS ONE.* 2013;8(9).