

RESEARCH ARTICLE

Editorial Process: Submission:06/26/2023 Acceptance:10/16/2023

Small RNA Deep Sequencing of Circulating RNAs Discovers a Unique Panel of microRNAs as Feasible and Reliable Biomarkers of Non-Small Cell Lung Cancers in Northern Thailand

Moe Thi Thi Han¹, Sakorn Pornprasert¹, Somcharoen Saeteng², Apichat Tantraworasin², Sophon Siwachat², Pradchaya Thuropathum², Busayamas Chewaskulyong³, Ratchada Cressey^{1,4*}

Abstract

Objective: This study aimed to assess the practicality and reliability of utilizing microRNAs (miRNAs) as a potential screening and diagnosing tool for non-small cell lung cancers (NSCLCs) in Northern Thailand. **Methods:** Small RNA sequencing and a literature review was performed to obtain a list of serum miRNA candidates. Serum levels of these selected miRNA candidates were measured in patients with NSCLC and healthy volunteers by real-time RT-PCR and receiver operating characteristic curve (ROC) were used to assess diagnostic performance. **Results:** Sequencing data revealed 148 known miRNAs and 230 novel putative miRNAs in serum samples; 19 serum miRNAs were significantly downregulated and 242 were upregulated. Seven miRNAs selected according to sequencing data and 11 miRNAs according to previous reports were evaluated in training cohort (45 lung cancer patients, 26 controls) and 6 miRNAs were found differentially expressed ($p < 0.05$, Mann Whitney U test) and associated ($p < 0.05$, Chi-square test) with NSCLC development. Further analysis and verification identified an optimal combination of 4 miRNAs composed of hsa-miR23a, hsa-miR26b, hsa-miR4488 and novel-130 to provide the optimal AUC of 0.901 ± 0.034 . Detection of serum miRNA by real-time RT-PCR showed good reproducibility with the coefficient of variation (CV) $\leq 4\%$. The optimal screening miRNAs panel was primarily identified through sequencing data of local patient population, thus indicating that the etiology of NSCLCs may differ from one population to other and thus require a unique panel of miRNAs for their identification. **Conclusion:** Circulating miRNA is a feasible screening tool for NSCLCs. Nevertheless, populations with different lung cancer etiology may need to identify their own most suitable miRNA panel.

Keywords: Cancer biomarkers- microRNA- serum- small RNA sequencing- lung cancer

Asian Pac J Cancer Prev, 24 (10), 3585-3598

Introduction

Lung cancer, which accounts for 29% of cancer deaths in men and 26% in women, is the leading cause of cancer-related mortality globally (Jemal et al., 2009). Two main types of lung cancer are non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), which make up around 85% and 15% of all lung cancers, respectively. Lung cancer has a poor prognosis despite recent improvements in detection and treatment methods, with a 5-year overall survival rate of only about 16% Alberg et al. (2007). The absence of obvious symptoms and insufficiently sensitive biomarkers in early stages of lung cancer limits their early diagnosis (Yu et al., 2019).

The American College of Chest Physicians (ACCP) has originally issued a guideline recommended lung cancer screening using low-dose computed tomography (LDCT) for individuals with high-risk factors including current and former smokers aged 55 to 74 years with a 30 or more pack-year history of smoking. Nevertheless, this eligibility was expanded to individuals with less smoking exposure of 20 or more pack-years in March 2021 (Force et al., 2021). Reports have shown that early detection of lung cancer was achievable through LDCT (National Lung Screening Trial Research et al., 2011), however, this technique possesses the drawbacks of increased cancer risk associated with repeated ionizing radiation exposure and high cost (de Koning et al., 2020). Especially in

¹Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Thailand. ²Department of Surgery, Faculty of Medicine, Chiang Mai University, Thailand. ³Department of Internal Medicine, Faculty of Medicine, Chiang Mai University, Thailand. ⁴Cancer Research Unit, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Thailand. *For Correspondence: Ratchada.cr@cmu.ac.th

Thailand, LDCT for lung cancer screening is not covered by the national healthcare system, thus individuals subjected to LDCT screening need to pay for their own expense. Consequently, it is urgent to identify novel, more sensitive, low cost and easy-to-detect biomarkers which can be used in diagnosis and prognosis prediction of lung cancers for Thai population (Yu et al., 2019).

Recent evidence has suggested that miRNAs may serve as ideal biomarkers in screening of lung cancer, as they can be secreted into blood circulation and existed in a remarkable stable form (Mitchell et al., 2008). miRNAs are a type of endogenous, single-stranded, short noncoding RNA (22 nucleotides) that play role as important regulators of gene expression and define part of the epigenetic signature (Vrijens et al., 2015). miRNAs are involved in the response of cells to environmental stressors; thus, representing ideal markers to reflect on etiology fingerprint of the disease. Development of several diseases, including cancer, have been attributed to altered miRNA expression pattern (Kadkhoda et al., 2022a; Kadkhoda et al., 2022b). The altered miRNA expression patterns were reported to be observed before the obvious clinical symptoms or clear biopsy and image examination became evidenced (Zhang et al., 2020). An investigation in animal model has also demonstrated the significant change of miRNA expression during lung tumorigenesis induced by vinyl carbamate (VC) treatment, indicating that miRNAs are promising biomarkers for screening and diagnosing of lung cancer (Melkamu et al., 2010). Serum miRNA has been used to detect lung cancer in an asymptomatic high-risk group with 80% accuracy (Bianchi et al., 2011). miRNA signatures in tissues and plasma could also be used to predict the development and prognosis of computed tomography detected pre-cancerous lung cancer (Boeri et al., 2011). A number of high impact investigations have demonstrated circulating miRNA as a supersensitive biomarker for early-stage (stage 0-I) lung cancer screening (Inagaki et al., 2023) (Ying et al., 2020).

Therefore, the altered expression pattern of miRNA might represent a novel approach that could be introduced to routine examination for monitoring and early diagnosis of lung cancers. In this study, a comprehensive list of circulating miRNAs profiles differentially expressed in lung cancer patients compared to normal individuals was identified by small RNA sequencing and literature review. The clinical usefulness of the selected miRNA candidates was then validated in a larger group of samples by realtime RT-PCR. Our results suggest that identifying unique clinically useful miRNAs according to cancer etiology for each population is critically important for detection of lung cancer and realtime RT-PCR is a robust and reliable technique of miRNA detection.

Materials and Methods

Study population

Serum samples were obtained from patients diagnosed with non-small cell lung cancer (NSCLC) at Maharaj Nakorn Chiang Mai hospital and healthy volunteers who came for annual health check up at the Clinical Service

Center of the Faculty of Associated Medical Sciences, Chiang Mai University. All lung cancer patients were newly diagnosed, previously untreated (chemotherapy or radiotherapy), and histologically confirmed. At the time of blood sample collection, healthy volunteers had no history or symptoms of lung cancer, such as cough, shortness of breath, wheezing, weight loss, weakness, and fatigue. All cancer patients and healthy controls were resident in one of the 17 provinces of northern Thailand. This study was approved by the human ethics committee at the Faculty of Medicine, Chiang Mai University, Thailand (Study code : NONE-2564-08207)).

Sample processing

After written inform consent were obtained from the subjects, blood specimens (5-10 mL) were collected before the initiation of chemotherapy or radiotherapy, and isolated serums were stored at -80°C. Total miRNA was isolated from serum samples using NucleoSpin miRNA Plasma, Mini kit for circulating miRNA (MACHEREY-NAGEL) according to the protocol of the manufactures.

Small RNA sequencing

Total miRNA isolated from serum of 6 subjects (3 NSCLCs and 3 controls) were submitted to the Vishuo Biomedical Pte. Ltd. to perform small RNA sequencing and data analysis. About 10ng small RNA was used for library preparation. 3'SR Adaptor and 5'SR Adaptor for Illumina were ligated to the small RNA using ligation enzymes and first strand cDNA was synthesized using ProtoScript II Reverse Transcriptase. Each sample was then amplified by PCR using P5 and P7 primers, the PCR product were purified, cleaned, and validated by electrophoresis using an Agilent 2100 Bioanalyzer. Then libraries with different indexes were multiplexed. The qualified libraries were sequenced pair end PE150 on the illumina HiseqXten/Novaseq / MGI2000 System. To obtain high quality clean data, raw data were processed by Trimmomatic (v0.30) (Martin, 2011) to remove the adapter sequences, bases with average quality score below 20 and reads that were less than 18 base-pair long after trimming. Filtered high quality data was then analyzed by miRDeep2 software to identify both known and novel miRNA from deep sequencing as well as their expression data. Differential expression analysis was performed using DESeq (Gill and Dhillon, 2022)/DESeq2 Bioconductor package.

Literature search for candidate miRNA previously reported to be differentially expressed in NSCLCs

A list of candidate miRNAs was obtained from searching previous reports investigated in a large group of population or identified through meta-analysis to be significantly differentially expressed in NSCLCs compared to healthy volunteers. The chosen miRNAs were as followed: let-7a-5p, miR-1-3p (Ying et al., 2020), miR-21 (Bica-Pop et al., 2018), miR-23a (Hsu et al., 2017), miR-26b (Lu et al., 2018), miR-126 (Grimolizzi et al., 2017), miR-146b (Yang et al., 2019), miR-205 (Duan et al., 2017), miR-375-5p (Ying et al., 2020), miR-1268b (Asakura et al., 2020) and miR-6075 (Asakura et al.,

2020).

Quantitative RT-PCR (RT-qPCR) and Quality Control (QC) of the assay

Serum level of miRNAs of interest was verified by realtime RT-PCR utilizing polyadenylation-based approach (Shi and Chiang, 2005) with SYBR Green fluorescent-based method (Biorad). In brief, poly-A tailing, and adapter ligation was performed using 10ng of total small RNA isolated from serum and the single-stranded complementary DNA (cDNA) was synthesized using a Tetro cDNA Synthesis Kit (Meridian Bioscience) and quantified by quantitative Polymerase Chain Reaction (qPCR) utilizing SYBR Green fluorescent-based method (iTaQ™Universal SYBR® Green supermix; Biorad) with a CFX96 Touch™ Real-Time PCR Detection System (Biorad). miR-484 (Liu et al., 2020) previously reported to have stable expression across cancer and control subjects were used as endogenous references. The relative miRNA expression was calculated using the equation $2^{-\Delta CT}$, in which ΔCT = cycle threshold (CT) of cancer patients or healthy control - average CT of endogenous reference miRNAs.

To ensure reliability and reproducibility of the miRNA detection by real-time qPCR procedure, a pooled of total miRNA extracted from healthy volunteer was used as control material to subject to the detection along with miRNA extracted from tested samples. This control material was used to set up OCV (optimal condition variance) by performing 10 replicative assays for each specific miRNA and the Ct value of the control material that felt within 2-fold of the mean \pm 2SD of the OCV would be consider acceptable.

Statistical Analysis

All statistical analyses will be performed with the SPSS program. The relationship between miRNA expression and diagnosis will be assessed using Chi-square tests. The Mann-Whitney U test was used to analyze differences in miRNA expression levels between the case and control groups. A two-sided p value less than 0.05 was defined as statistically significant. Receiver-operating curve (ROC) analysis was performed to evaluate the diagnostic performance of serum miRNAs and the area under the ROC curve (AUC) of each miRNA and their combinations were calculated for direct comparison.

Results

Identification of differentially expressed circulating miRNAs using small RNA deep sequencing

Deep sequencing data analysis revealed a total of 148 known miRNA and 230 novel putative miRNAs in serum samples. Further analysis to determine miRNA with significant differential expression according to the criteria of fold change greater than 2 and p value less than 0.05 showed that 19 miRNAs were significantly downregulated, and 242 miRNAs were significantly upregulated in cancer patients compared to non-cancer volunteers (Figure 1A-B). Results of novel and known top 10 circulating miRNA DEGs in cancer patients compared

non-cancer control are presented in Table 1, 2. Cluster analysis of differentially expressed miRNA is shown in Figure 1C. Clustering analysis groups miRNAs based on similarity. The regions of different colors represent different cluster groups. miRNAs with similar expression patterns in the same group are likely to have similar functions or participate in the same biological process. Figure 1C shows that the majority of circulating miRNA significantly increases in lung cancer patients were involved in the cellular component (CC) and molecular function (MF) whereas miRNA involved in the biological process (BP) appeared to be reduced in comparison to non-cancer control.

The biological processes of differently expressed miRNA were assessed using GO analysis, which is shown in the histogram with the specification of the relevant BP, CC and MF. The major mRNA gene targets were enriched in CC (Figure 2A), which the majority were identified to be involved in every part of the cell, i.e., cytosol, nucleus, membrane, and extracellular space (Figure 2B).

Differential miRNA target genes were then annotated with KEGG pathway. Rich factors were used to measure the degree of pathway enrichment. Rich factor refers to the ratio of the number of differential miRNA target genes in the pathway to the total number of genes in the pathway. The top 30 pathways of most significantly enriched for the analysis are shown in the Figure 3A-B. It can be noted that viral carcinogenesis, transcriptional mis-regulation in cancer, regulation of actin cytoskeletal, MAPK signaling pathway and calcium signaling are amongst the most significantly enriched pathways in NSCLC serum samples when compared with serum of non-cancer individuals (Figure 3A-B).

Verification of candidate miRNAs in the training set of clinical samples by realtime RT-PCR and quality control of the assay

Considering the fold changes of miRNAs of deep

Table 1. Characteristic of Patients with NSCLC and Non-Cancer Volunteers

Characteristic	Lung cancer (n=45)	Controls (n=26)	P value
Gender			
Male (n (%))	30 (66.7%)	14 (53.8%)	0.284 ^a
Female (n (%))	15 (33.3%)	12 (46.2%)	
Age			
Mean \pm SD	61.1 \pm 11.04	55.9 \pm 11.37	0.092 ^b
\leq 55 (n (%))	14 (31.1%)	12 (46.2%)	0.205 ^a
56 up (n (%))	31 (68.9%)	14 (53.8%)	
Smoking status			
Yes (n (%))	33 (73.3%)	11 (42.3%)	0.009 ^a
No (n (%))	12 (26.7%)	15 (57.7%)	
Stages (NSCLC)			
Early (stage I + II)	11 (24.4%)	-	
Late (stage III+IV)	30 (66.7%)	-	
Unidentified	4 (8.9%)	-	

^a, Chi Square; ^b, Mann Whitney U test

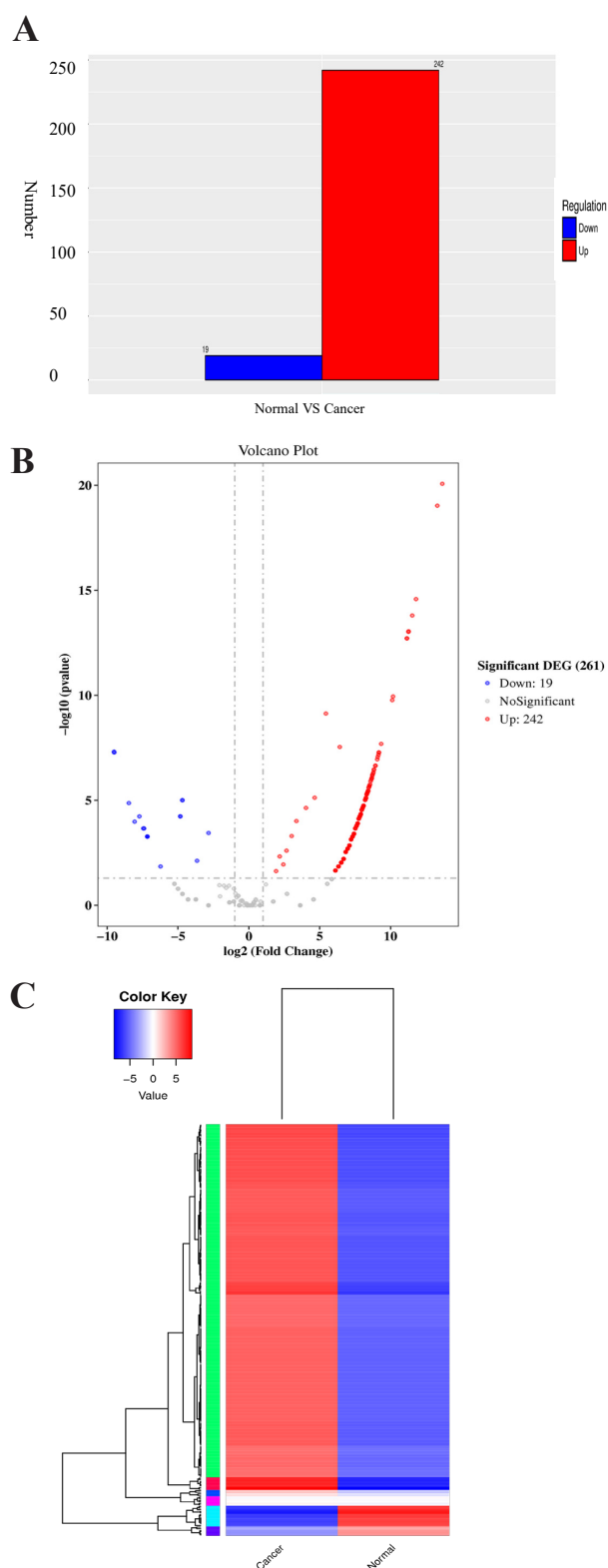


Figure 1 Differentially Expressed miRNAs in the Serum of Patients with NSCLC Identified through Small RNA Sequencing. miRNA up- or down-regulation across Samples (A). Volcano plot of miRNA differential expression (B). Red dots represent up-regulated miRNAs and blue dots down-regulated. X axis represents the fold change of miRNA expression in different samples. Y axis represents the statistical significance of the change in expression. Differential miRNA cluster analysis result (C). miRNA of high expressed are in red, and low expression in blue. The gradient from blue to red represents the increase of miRNA expression.

sequencing results (Table 2) compared between non-cancer volunteers and lung cancer patients, 5 miRNAs significantly upregulated (hsa-miR-483-5p, novel-78, hsa-miR-4488, novel-133, novel-130) and 2 miRNAs significantly downregulated (hsa-miR-200b-3p and novel-145) were chosen for verification with larger clinical sample by realtime RT-PCR together with 11 miRNAs chosen from previous reports (hsa-Let-7a-5p, hsa-miR-1-3p, hsa-miR-21, hsa-miR-23a, hsa-miR-26b, hsa-miR-126, hsa-miR-146b, hsa-miR-205, hsa-miR-375-5p, hsa-miR-1268b, hsa-miR-6075).

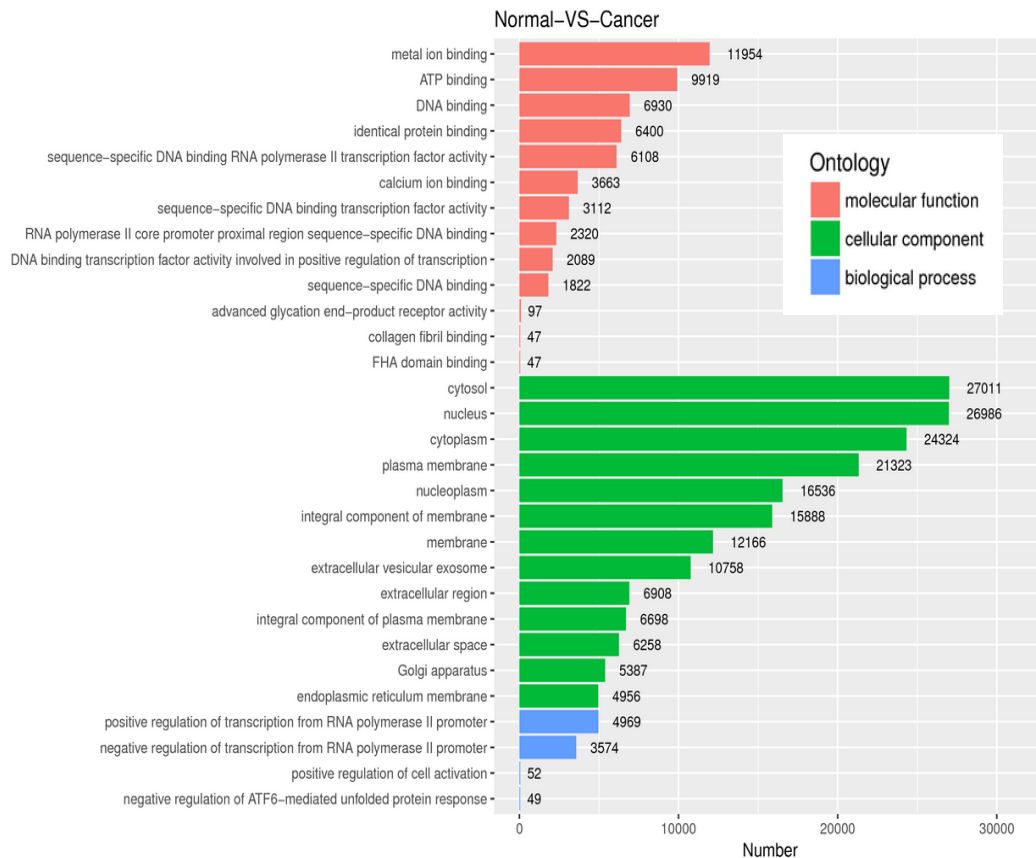
To ensure reproducibility of the detection method of miRNA by real-time RT-PCR, a control material was prepared and subjected to the detection along with clinical samples. Some of the representative quality control charts are shown in Figure 4. The result showed that the Ct value of the control material fell within 2-fold of the mean \pm 2SD of the OCV with the CV less than 5% (ranging from 1-4%, which indicated a very good reproducibility of the assay. The results show that out of 7 circulating miRNAs identified by small RNA sequencing, expression level of 4 miRNAs (novel-78, novel-130, hsa-miR 483-5p and hsa-miR4488) were found to be significantly increased ($p < 0.05$, Mann Whitney U test, Table 3, Figure 5A) in patients with NSCLC. On the other hand, out of 11 circulating miRNAs chosen according to previous reports, only 1 miRNA (hsa-miR 23a) was found to be differently expressed in lung cancer patients. Further analysis into different subtypes showed that adenocarcinoma (ADC), which were the majority of NSCLCs significantly showed an increase level of hsa-miR23a, hsa-miR26b, novel-130, novel-78, hsa-miR 483-5p, hsa-miR4488 (Figure 5B). Although, it's not statistically significant, which may be due to a lower number of samples, other sub-types of NSCLCs including squamous cell carcinoma (SCC) and large cell carcinoma (LCC) also showed similar trends.

Upper bound of 95% CI for mean of relative miRNA serum levels ($2^{-\Delta\text{Ct}}$) in control group was set as the cutoff value and the association of each miRNA with lung cancer development was further analyzed using Chi-square and diagnostic odd ratio calculation. According to these cutoff values, an increase serum level of hsa-miR23a, hsa-miR26b, novel-130, hsa-miR 483-5p, hsa-miR4488 was significantly associated ($p < 0.05$, Chi square) with NSCLC development with diagnostic odd ratio ranging from 3.339-11.290 (Table 3).

Panel building and validation

ROC (Receiver Operating Characteristic) curve analyses were performed on the serum miRNAs significantly associated or significantly increased in serum of lung cancer patients, which included hsa-miR23a, hsa-miR 26b, novel-130, novel-78, hsa-miR 483-5p and hsa-miR4488, to determine their diagnostic performance. The results demonstrated that the six miRNAs exhibited similar performance with AUC values between 0.597–0.664 (Table 4) for the discrimination of every stage of NSCLCs from controls, which were lower than the AUC of cytokeratin 19 fragment (CYRFA 21-1) of 0.774. Interestingly, various combinations of these 6-serum miRNA increased the AUC values significantly

A



B

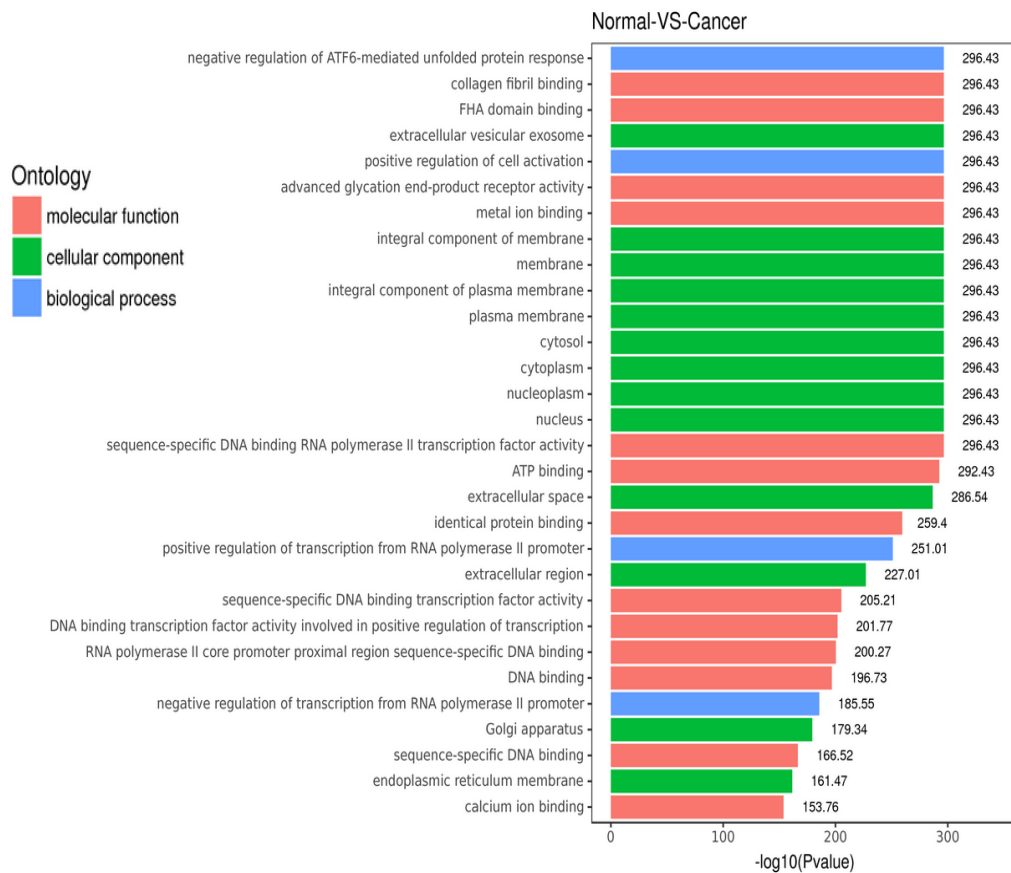


Figure 2 Gene Ontology (GO) analysis of the differential miRNA target genes. GO enrichment histogram (A). The GO enrichment analysis for the top 30 expressed miRNA targets (B). The x-axis denotes the $-\log_{10}(P\text{-value})$ and the P-value is generated by hypergeometric distribution. The y-axis denotes the GO term.

Table 2. Top 10 Upregulated and Downregulated Circulating miRNAs in Patients with NSCLC Compared with Non-Cancer Volunteers

Increased miRNAs	Fold- change (tumor/normal)	P value (FDR-adjusted)	Decreased miRNAs	Fold- change (tumor/normal)	P value (FDR- adjusted)
1. hsa-miR-483-5p	13.7	3.18E-18	1.hsa-miR-200b-3p	-9.5	0.000001
2.NovelmiRNA-78 (ggcuggagaacugagggggg)	13.3	1.76E-17	2.hsa-let-7i-5p	-9.5	0.000001
3.hsa-miR-4488	11.8	3.29E-13	3.hsa-miR-103a-3p	-8.5	0.000078
4.hsa-miR-619-5p	11.5	1.49E-12	4.hsa-miR-200a-3p	-8.1	0.000337
5.NovelmiRNA-133 (cacugcaaccucugccucc)	11.3	5.97E-12	5.hsa-miR-148b-3p	-7.7	0.000203
6.NovelmiRNA-130 (aagucugggauuacagg)	11.3	5.97E-12	6.NovelmiRNA-145 (uauguguguuuguguugugugu aug)	-7.4	0.000601
7.hsa-miR-500a-3p	11.2	9.12E-12	7.hsa-miR-205-5p	-7.4	0.000601
8.hsa-miR-502-3p	11.2	9.12E-12	8.NovelmiRNA-30 (ggcgggcgcgugcgcgggcg)	-7.2	0.001186
9.hsa-miR-486-5p	10.2	4.73E-09	9.NovelmiRNA-46	-7.2	0.001186
10.Novel-52 (gagacucagugaaauuca)	10.1	6.40E-09	10.NovelmiRNA-56 (ggcgggcgcgugcgcgggcg)	-7.2	0.001186

FDR, False Recovery Rate

(Table 5). An increasing AUC was achieved when the number of miRNAs in the panel was increased but was plateaued at 4-6 miRNAs (Figure 6). The optimal miRNAs combination that yields the best AUC of 0.910 is the four miRNAs combination of hsa-miR 23a, hsa-miR 26b, novel-130 and has-miR 4488. Almost all miRNAs including CYRFA 21-1 showed lower sensitivity for the detection of early-stage tumors compared to later stage

cancers, apart from miR4488 (Figure 7).

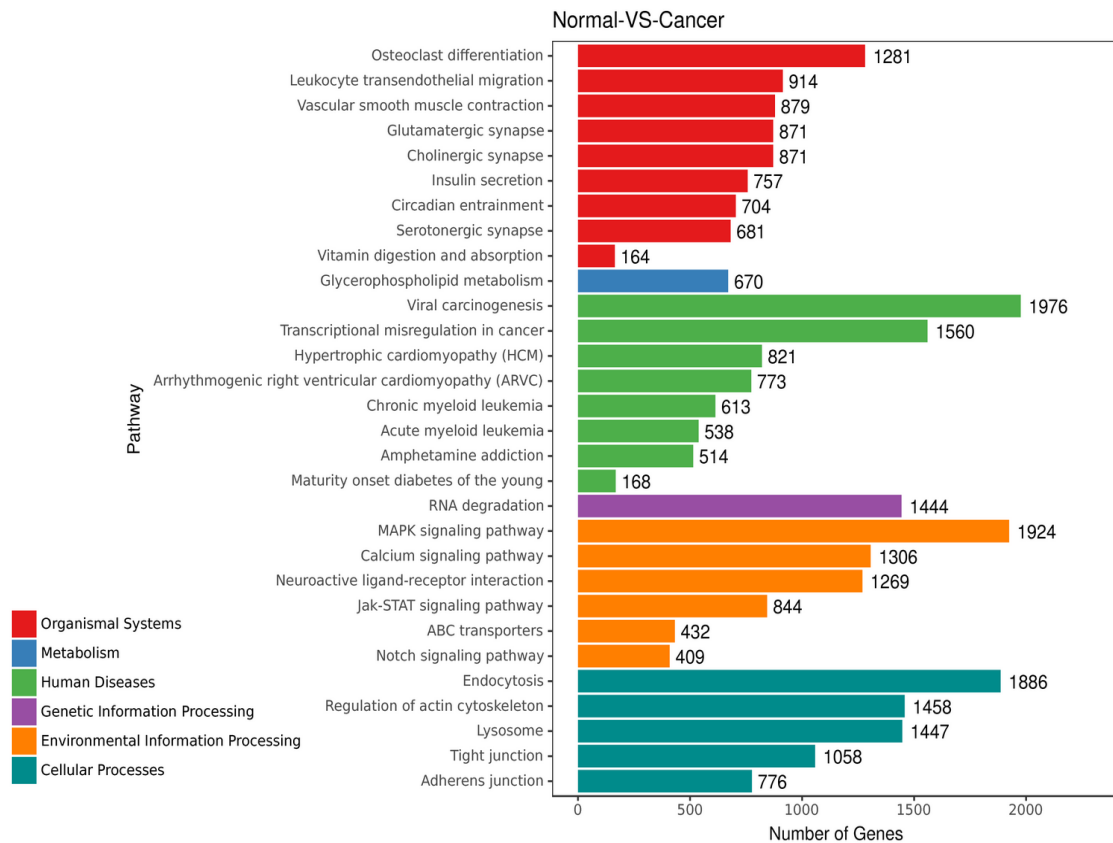
To assess the performance of the 4-miRNA panel in discrimination of clinical samples, we applied the 4-miRNA signature to samples in the validation set, which consisted of 24 control participants, 32 NSCLC patients. The results suggested that the expression level of hsa-miR 23a, hsa-miR 26b, novel-130 and has-miR 4488 was predominately upregulated in NSCLC patients (Figure 8).

Table 3. Characteristic of Circulating miRNAs Expressed in Patients with NSCLC and Non-Cancer Volunteers Validated by Realtime RT-PCR

miRNAs	Mean 2 ^{-(ΔCt)} ± SE in NSCLC	Mean 2 ^{-(ΔCt)} ± SE in control	Mann Whitney U (p value)	Fold change (CA/normal)	Cut-off (2 ^{-ΔCt})	Chi-square (p value)	Diagnostic Odds Ratio (DOR) value	95% CI ^a lower upper
I. miRNAs chosen according to previous publications								
1. hsa-miR let 7a-5p	0.6±0.135	0.3±0.076	0.228	2	0.419	0.362	1.667	0.553 5.022
2. hsa-miR 1-3p	0.5±0.114	0.2±0.064	0.269	2.5	0.369	0.454	1.498	0.519 4.322
3. hsa-miR 21	8.1±3.839	44.6±28.517	0.153	0.2	103.390	0.270	0.273	0.024 3.165
4. hsa-miR 23a	0.3±0.094	0.1±0.062	0.021*	3	0.199	0.007*	11.290*	1.388 91.835
5. hsa-miR 26b	0.7±0.147	0.3±0.088	0.130	2.3	0.433	0.028*	4.230*	16.302
6. hsa-miR 126	1.5±0.362	0.8±0.283	0.155	1.9	1.351	0.710	1.226	0.420 3.581
7. hsa-miR 146b	7.3±3.172	3.9±0.722	0.496	1.9	5.430	0.207	0.486	0.157 1.506
8. hsa-miR 205	0.7±0.199	0.6±0.294	0.443	1.2	1.211	0.239	2.595	0.507 13.275
9. has-miR 375-5p	36.2±6.411	39.1±19.270	0.417	0.9	72.800	0.469	1.846	0.344 9.897
10. hsa-miR 1268b	0.4±0.104	0.2±0.049	0.116	2.0	0.318	0.362	1.667	0.553 5.022
11. hsa-miR 6075	36.7±24.567	16.1±8.108	0.797	2.3	32.812	0.289	3.125	0.345 28.328
II. miRNAs chosen according to small RNA sequencing data								
1. Novel-78	0.8±0.209	0.2±0.073	0.038*	4.0	0.354	0.072	2.800	0.893 8.783
2. Novel-130	0.5±0.116	0.2±0.075	0.042*	2.5	0.332	0.018	4.655	1.212 17.877
3. Novel miR 133	2.9±2.086	0.7±0.161	0.634	4.1	1.002	0.612	1.359	0.414 4.462
4. Novel miR 145	0.4±0.110	0.2±0.094	0.135	2	0.427	0.199	2.234	0.643 7.761
5. hsa-miR 200b-3p	1.6±1.040	1.1±0.473	0.411	1.5	2.057	0.871	0.857	0.134 5.496
6. hsa-miR 483-5p	0.6±0.136	0.2±0.075	0.036*	3	0.351	0.046*	3.339	0.982 11.356
7. hsa-miR 4488	1.0±0.347	0.1±0.044	0.011*	10	0.216	0.008*	4.813*	1.427 16.231

* Statistically significant at p< 0.05, ^a, Confidential Interval (CI); CA, cancer; SE, Standard Error

A



B

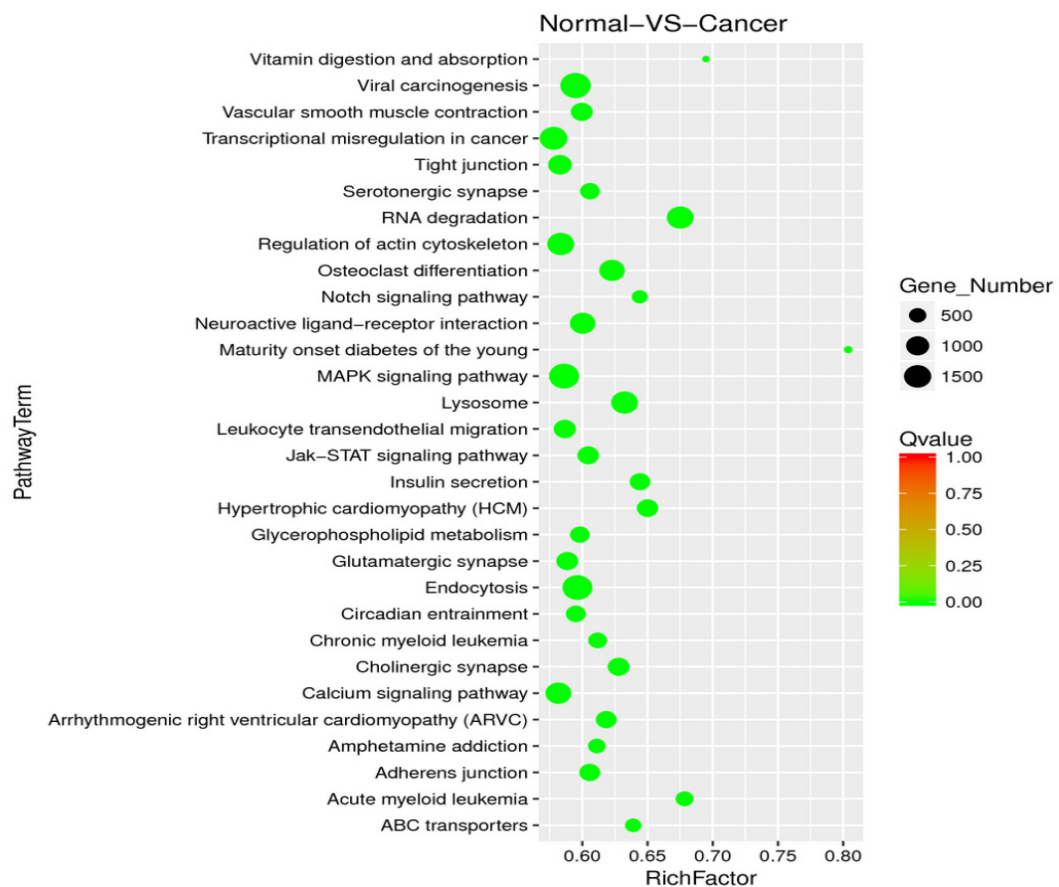


Figure 3. KEGG Pathway Annotations of Differentially Expressed microRNA Target Genes. KEGG enrichment histogram (A). X axis: the number of differentially expressed genes in the KEGG pathway. Y axis: the name of the KEGG pathway. Scatter plot of differential gene KEGG enrichment (B). X axis: rich factor. Y axis specify KEGG pathways. The size of the dot is positively correlated with the number of differential miRNA target genes in the pathway. Color code is to indicate different Q value ranges. The larger the rich factor, the greater the degree of enrichment. The smaller the Q value, the more significant the enrichment

Table 4. Diagnostic Performance of CYFRA21-1 and miRNAs Significantly Associated with the Development of NSCLC Compared between Early Tumour Stage and All Tumour Stage of Cancers

miRNAs	AUC± SE	All Tumor stage			AUC± SE	Early Tumor stage		
		95% CI (lower/upper)		Sensitivity at 100% specificity		95% CI (lower/upper)		Sensitivity at 100% specificity
1. hsa-miR 23a	0.635±0.066	0.506	0.764	0.044	0.423±0.098	0.230	0.616	0.000
2. hsa-miR 26b	0.606±0.067	0.475	0.736	0.156	0.682±0.092	0.501	0.863	0.000
3. Novel miR130	0.638±0.066	0.508	0.767	0.022	0.670±0.105	0.463	0.876	0.000
4. miR 483-5p	0.642±0.065	0.515	0.770	0.222	0.699±0.098	0.506	0.892	0.091
5. miR 4488	0.669±0.063	0.545	0.792	0.267	0.715±0.104	0.511	0.919	0.364
6. CYFRA 21-1	0.774±0.076	0.625	0.923	0.357	0.798±0.095	0.612	0.983	0.167

Discussion

One of the reasons for the extremely poor survival of patients with lung cancer is because of the late diagnosis due to the lack of feasible and reliable screening tests. The aim of this study was to conduct a comprehensive screening of miRNA differentially expressed in serum

of Thai lung cancer patients with NSCLCs compared to non-cancer volunteers and tested for their feasibility as potential non-invasive biomarkers of NSCLCs. Deep sequencing analysis identified a unique set of miRNAs and the top 10 most differentially expressed were chosen for further verification by realtime RT-PCR in comparison to the miRNA candidates chosen according to the previous

Table 5. Diagnostic Performance of Various Combinations of miRNAs Significantly Associated with the Development of NSCLC Compared between Early Tumor Stage and All Tumor Stage of Lung Cancer

Types of combinations	Total tumor stage		Early tumor stage	
	AUC± SE	Sensitivity at 100% specificity	AUC± SE	Sensitivity at 100% specificity
Two miRNAs combinations				
1. Novel-130 & miR 483-5p	0.773±0.054	0.467	0.909±0.052	0.545
2. Novel-130 & miR 4488	0.796±0.051	0.444	0.911±0.054	0.727
3. miR 483-5p & miR 4488	0.778±0.054	0.489	0.925±0.047	0.636
4. miR 23a & miR 26b	0.796±0.051	0.444	0.823±0.074	0.182
5. miR 23a & Novel-130	0.783±0.049	0.489	0.864±0.068	0.545
6. miR 23a & miR 483-5p	0.780±0.049	0.400	0.871±0.066	0.273
7. miR 23a & miR 4488	0.784±0.047	0.511	0.857±0.071	0.545
8. miR 26b & Novel-130	0.841±0.044	0.467	0.953±0.034	0.636
9. miR 26b & miR 483-5p	0.831±0.046	0.422	0.895±0.059	0.455
10. miR 26b & miR 4488	0.865±0.040	0.489	0.935±0.043	0.636
Three miRNAs combinations				
1. Novel-130 & miR 483-5p & miR 4488	0.800±0.051	0.467	0.948±0.037	0.727
1. Novel-130 & miR 483-5p & miR 4489	0.838±0.046	0.511	0.965±0.029	0.727
3. miR 23a & miR 26b & miR 483-5p	0.826±0.048	0.444	0.911±0.055	0.545
4. miR 23a & miR 26b & miR 4488	0.882±0.038	0.556	0.935±0.043	0.636
5. miR 23a & miR 483-5p & miR 4488	0.823±0.048	0.578	0.937±0.043	0.727
6. miR 26b & miR 483-5p & miR 4488	0.882±0.038	0.578	0.972±0.026	0.818
7. miR 23a & Novel-130 & miR 483-5p	0.790±0.052	0.511	0.925±0.047	0.636
8. miR 23a & Novel-130 & miR 4488	0.841±0.045	0.556	0.911±0.054	0.727
9. miR 26b & Novel-130 & miR 483-5p	0.859±0.043	0.489	0.963±0.030	0.727
10. miR 26b & Novel-130 & miR 4488	0.890±0.037	0.533	0.979±0.021	0.818
Four miRNAs combinations				
1. miR 23a & miR 26b & Novel-130 & miR 483-5p	0.857±0.043	0.511	0.972±0.026	0.818
2. miR 23a & miR 26b & Novel-130 & miR 4488	0.901±0.034	0.578	0.979±0.021	0.818
3. miR 23a & miR 26b & miR 483-5p & miR 4488	0.885±0.037	0.600	0.972±0.026	0.818
4. Novel-130 & miR 483-5p & miR 4488 & miR 23a	0.841±0.045	0.556	0.948±0.037	0.727
5. Novel-130 & miR 483-5p & miR 4488 & miR 26b	0.791±0.052	0.511	0.979±0.021	0.818
Five miRNAs combinations				
miR 23a & miR 26b & Novel-130 & miR 483-5p & miR 4488	0.895±0.036	0.600	0.979±0.021	0.818

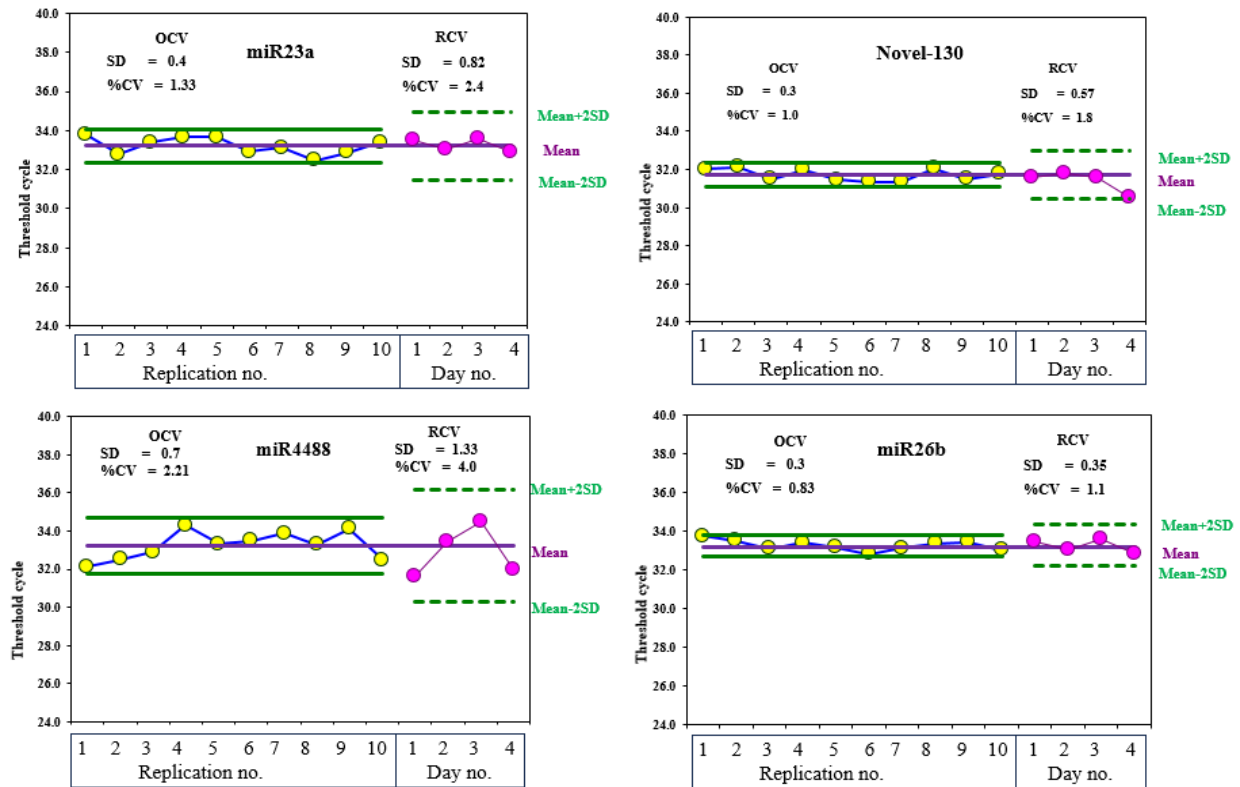


Figure 4. Quality Control (QC) Chart of Detection of miRNAs by Realtime RT-PCR Showing Coefficients of Variation (%CV) of Optimum Condition Variance (OCV) and Routine Condition Variance (RCV)

publications.

miRNAs are involved in the response of cells to both endogenous and environmental stress; thus, representing ideal markers to reflect on etiology fingerprint of the disease. It is evident that circulating miRNA is a feasible

screening biomarker for lung cancer. Nevertheless, in this study the optimal screening miRNAs panel was primarily identified through sequencing data from the local patient population indicating that the etiology of NSCLCs in Northern Thailand may differ from other

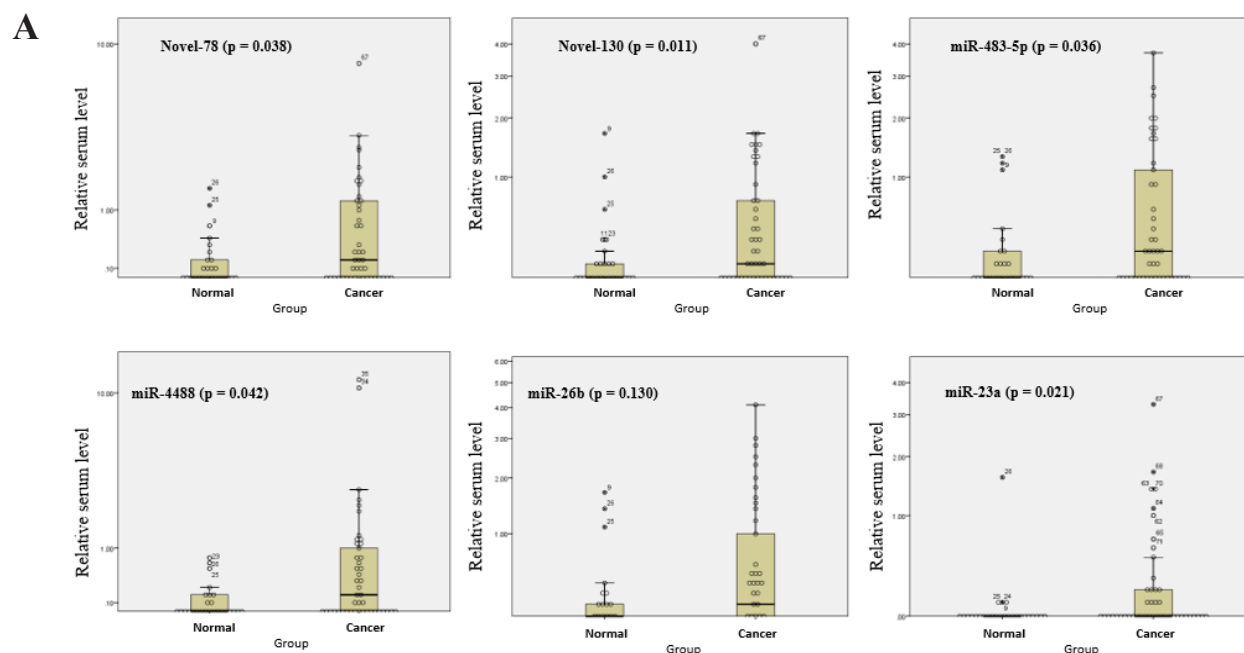


Figure 5. Boxplots Showing Relative Serum miRNAs (in relation to miR484 (house keeping miRNA)) Significantly Associated or Differently Expressed in Patients with NSCLC Compared to Controls (n=26), all NSCLCs (n=45) (A) and different NSCLC sub-type (B); SCC, squamous cell carcinoma (n=7) ; LCC; large cell carcinoma (n=3); ADC; adenocarcinoma (n=35)

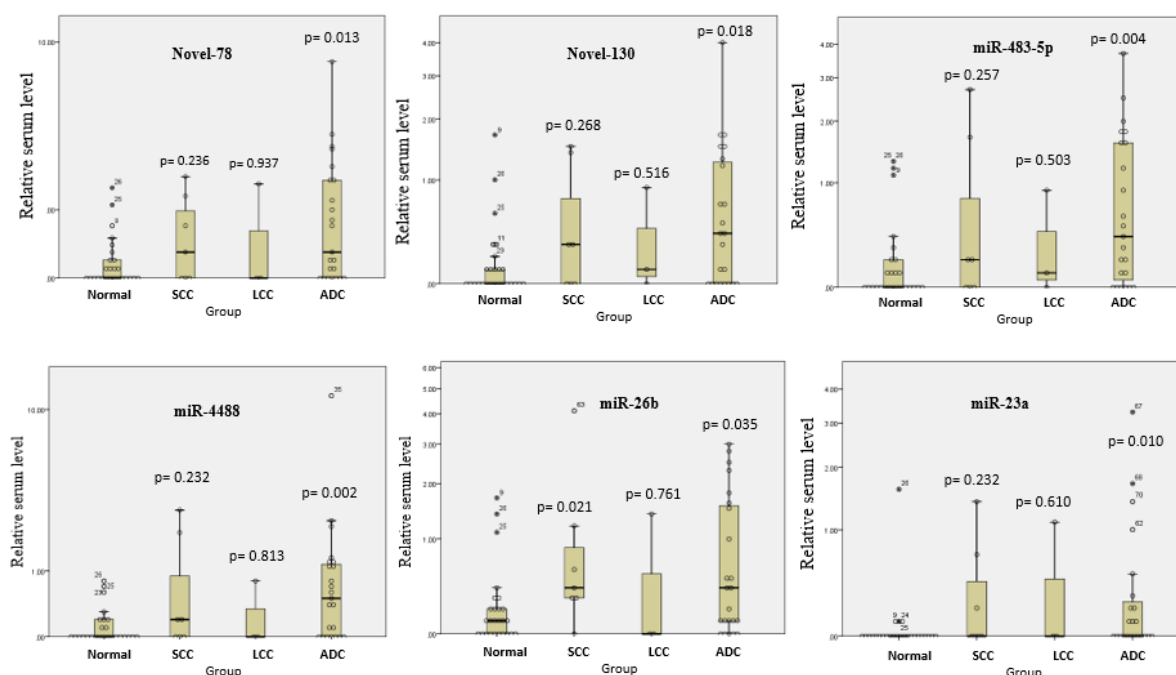
B

Figure 5. Boxplots Showing Relative Serum miRNAs (in relation to miR484 (house keeping miRNA)) Significantly Associated or Differently Expressed in Patients with NSCLC Compared to Controls (n=26), all NSCLCs (n=45) (A) and different NSCLC sub-type (B); SCC, squamous cell carcinoma (n=7) ; LCC; large cell carcinoma (n=3); ADC; adenocarcinoma (n=35); $p < 0.05$ is considered statistically significant (Mann Whitney U test compared to normal controls)

populations and thus require unique panel of miRNAs for their identification. Lung cancer is one of the main cause of cancer death in Thailand (2020). The most problematic area is northern Thailand where lung cancer incidence and mortality were twice as high compared to other parts of the country (Imsamran, 2015). There are many known risk factors causing lung cancer, of which smoking is the main course. In northern Thailand, the prevalence of tobacco smoking started to decrease since 2009 (Pongnikorn et al., 2018), which simultaneously corresponded to a decline in the incidence of squamous and small cell lung carcinomas known to be linked to smoking. However, there has

been an increase of lung adenocarcinomas incidence, strongly linked to environmental factors (Virani et al., 2017) (Khuder, 2001). In our study, 73.3% of lung cancer patients had a history of smoking, environmental factors like radon gas (Somsunun et al., 2022), air pollution/PM2.5 (Nakharutai et al., 2022) may play a crucial role contributing to lung cancer development in the remaining 26.7% of the cases as well as facilitating the effect of cigarette smoke on tumorigenesis.

The panel of miRNA with best AUC was achieved with the combination of 4 miRNAs composed of hsa-miR23a, hsa-miR26b, hsa-miR4488 and novel miR130. Of all 4

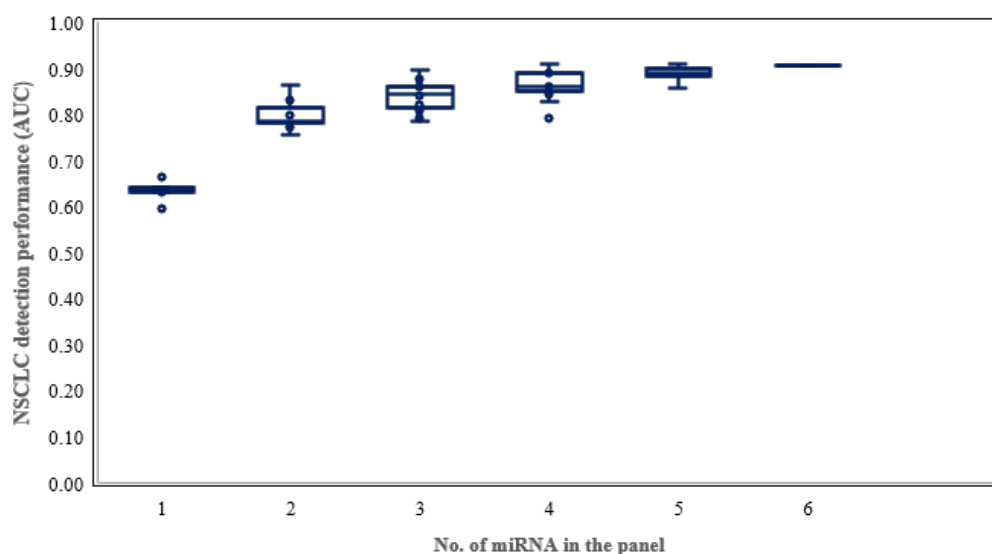


Figure 6. Boxplots Showing AUC Value of Receiver Operating Characteristics (ROC) Curve of Each miRNA and Their Various Combination of 2, 3, 4, 5 and 6 miRNAs.

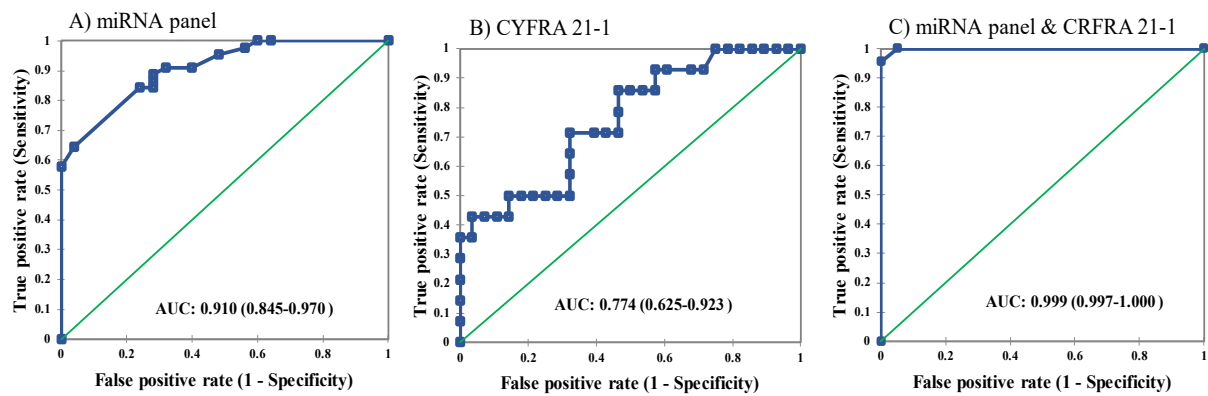


Figure 7 Receiver operating characteristics (ROC) curve of optimal miRNA panel (A) compared to CYFRA-21-1 (B) and optimal miRNA panel combined with CYFRA-21-1 (C)

miRNAs, only hsa-miR4488 showed higher sensitivity with 100 specificities toward identifying NSCLCs at early stage, while others were more sensitive to later stage of lung cancer. Extracellular vesicles miR-4488 secreted by mesenchymal stem cells (MSCs) has been reported to trigger signal transduction in the epithelial-mesenchymal transition (EMT) process and the stimulation of mesenchymal marker expression in epithelial cells, therefore promoting the metastasis of lung cancer (Fang et al., 2022). Accordingly, our small RNA sequencing results showed that most of miRNA differentially expressed in lung cancer patients were those involved in the regulation of cellular component, which is the major cellular changes involved with the EMT. Although hsa-miR4488 showed the ability to suppress angiogenesis, increased its level in serum of breast cancer patients enhanced angiogenesis and

promoted breast cancer metastatic colonization (Zheng et al., 2020). In addition, exosomal hsa-miR-4488 has been shown to involve with initiating NF- κ B signaling and its level was significantly increased in dermatomyositis-associated interstitial lung disease with anti-melanoma differentiation-associated protein 5 antibody-positive subset (DM-ILD-MDA5 Ab(+)) compared to DM-nonILD (Zhong et al., 2021), thus indicated its role in the induction of inflammation.

We have also identified an unknown miRNA, which referred to as “novel-130”, to be significantly increased in cancer patients with NSCLC and inclusion of this miRNA into the screening panel help increasing the ROC, however, its role toward lung cancer development remains to be elucidated. The third miRNA included in the panel is hsa-miR-23a, which has been widely reported to be

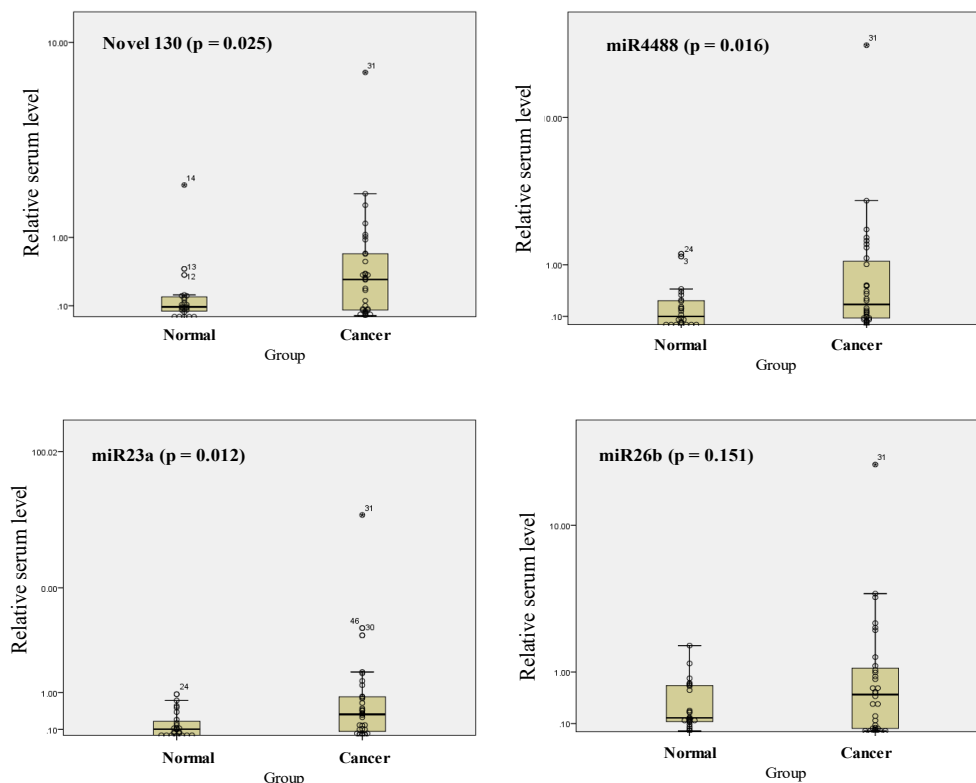


Figure 8. Boxplots Showing Differentially Expressed Serum miRNAs in the Validation Cohort

dysregulated during malignant transformation of normal human tissues (Ramassone et al., 2018). Overexpression of hsa-miR-23a has been reported to correlate with the areca nut extract-induced DNA double-strand break and DNA damage in oral cell line, suggesting the role of hsa-miR-23a in early oral carcinogenesis induced by areca nut (Tsai et al., 2011). In non-small cell lung cancer (NSCLC), expression of hsa-miR-23a was correlated with smoking habit, tumor size, TNM stage and lymph node metastasis (Qu et al., 2015), which consistent with our results that serum has-miR23a was significant increase in patients with late stage of cancer. Serum hsa-miR23a was identified to be increased in various type of human cancers including lung (Qu et al., 2015), breast (Wu et al., 2012), gastric (Yin et al., 2019) and pancreatic carcinoma (Frampton et al., 2015), thus represent a good non-invasive cancer biomarker.

Reports related to the involvement of hsa-miR26b in cancer development is controversy. On one hand, over-expression of hsa-miR26b has been reported to significantly inhibit the proliferation, migration, and invasion of A549 lung cancer cells in vitro and suppressed the growth of established tumors in vivo (Xia et al., 2015), thus indicating its tumor suppressive effect. On the other hand, upregulation of miR-26b in serum of cancer patients occurred after irradiation has been reported to enhance expansion of myeloid-derived suppressor cells (MDSCs), which constitute pre-metastatic niche (PMN) that support metastatic microenvironment in distant organs (Yin et al., 2022) thus indicating its oncogenic effects. It was hypothesized that dying tumor may secrete some factors to help with the survival of the remaining tumors. In addition, overexpression of miR-26b-5p has been reported to significantly decrease the percentage of G0/G1 phase cells and slightly increase the percentage of S phase cells in response to extremely low frequency electromagnetic fields (Liu et al., 2016), which indicated its role in driving proliferation in response to cellular stress. Nevertheless, the results from our study indicated that increased serum hsa-miR26b level was a useful biomarker for lung cancer screening.

Circulating miRNAs have the potential to be non-invasive biomarkers for many diseases, including cancer. Extracellular miRNAs are highly stable, resisting degradation at room temperature for up to 4 days and in deleterious conditions such as boiling, multiple freeze-thaw cycles, and high or low pH (Chen et al., 2008). miRNAs are believed to have 2 major populations of extracellular miRNAs in biological fluids. One is found in vesicles such as exosomes, microvesicles, and apoptotic bodies (Gallo et al., 2012) while the other is associated with proteins (Turchinovich et al., 2011). Although the release of extracellular miRNAs has been demonstrated to be a regulated process (Kosaka et al., 2010), some considered them as by-products of cellular activities, such as cell injury or death (Turchinovich et al., 2012). It is also possible that the level of miRNA in the serum may represent different meaning to its intracellular roles. Accordingly, both hsa-miR26b and hsa-miR4488 have been demonstrated by some studies to have tumour suppressive effect (Zheng et al., 2020) (Xia et al., 2015),

but their serum level was significantly increased in lung cancer patients. Nevertheless, detection of miRNA panel by realtime RT-PCR represents a robust, feasible, and good reproducibility approach for lung cancer detection.

In conclusion, this study has demonstrated that serum miRNA profile is a feasible screening tool for detection of NSCLC and population with different cancer etiology may need to identify their own miRNA panel for optimal screening.

Author Contribution Statement

Concept and design: R Cressey, MTT Han; Sample collection: S Saeteng, A Tantraworasin, S Siwachat, B Chewaskulyong, MTT Han; Acquisition of data: B Chewaskulyong, MTT Han; R Cressey; Analysis and interpretation of data: R Cressey, MTT Han, S Pornprasert, B Chewaskulyong; Writing, reviewing and/or revision of manuscript: R Cressey, MTT Han

Acknowledgements

Funding statement

This research project was supported by the Fundamental Fund 2022, Chiang Mai University (FF65/022) and Moe Thi Thi Han's post-graduation study was supported by the Presidential Scholarship of Chiang Mai University. This paper is a part of an approved student thesis.

Ethical approval

The research protocol was approved by the human ethics committee at the Faculty of Medicine, Chiang Mai University, Thailand (Study code: NONE-2564-08207)) and all patients signed an informed consent.

Data availability statement

The data that support the findings of this study are available from the corresponding author, [RC], upon reasonable request.

Conflict of interest

The author(s) declared no potential conflict of interest with respect to the research, authorship, and/or publication of this article.

References

- Alberg AJ, Ford JG, Samet JM (2007). Epidemiology of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest*, **132**, 29s-55s.
- Asakura K, Kadota T, Matsuzaki J, et al (2020). A miRNA-based diagnostic model predicts resectable lung cancer in humans with high accuracy. *Commun Biol*, **3**, 134.
- Bianchi F, Nicassio F, Marzi M, et al (2011). A serum circulating miRNA diagnostic test to identify asymptomatic high-risk individuals with early stage lung cancer. *EMBO Mol Med*, **3**, 495-503.
- Bica-Pop C, Cojocneanu-Petric R, Magdo L, et al (2018). Overview upon miR-21 in lung cancer: focus on NSCLC. *Cell Mol Life Sci*, **75**, 3539-51.
- Boeri M, Verri C, Conte D, et al (2011). MicroRNA signatures in tissues and plasma predict development and prognosis of

- computed tomography detected lung cancer. *Proc Natl Acad Sci U S A*, **108**, 3713-8.
- Chen X, Ba Y, Ma L, 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.
- de Koning HJ, van der Aalst CM, de Jong PA, et al (2020). Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *N Engl J Med*, **382**, 503-13.
- Duan B, Guo T, Sun H, et al (2017). miR-205 as a biological marker in non-small cell lung cancer. *Biomed Pharmacother*, **91**, 823-30.
- Fang K, Chen X, Qiu F, et al (2022). Serum-Derived Exosomes-Mediated Circular RNA ARHGAP10 Modulates the Progression of Non-Small Cell Lung Cancer Through the miR-638/FAM83F Axis. *Cancer Biother Radiopharm*, **37**, 96-110.
- Force USPST, Krist AH, Davidson KW, et al (2021). Screening for Lung Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*, **325**, 962-70.
- Frampton AE, Castellano L, Colombo T, et al (2015). Integrated molecular analysis to investigate the role of microRNAs in pancreatic tumour growth and progression. *Lancet*, **385** Suppl 1, S37.
- Friedlander MR, Chen W, Adamidi C, et al (2008). Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol*, **26**, 407-15.
- Gallo A, Tandon M, Alevizos I, et al (2012). The majority of microRNAs detectable in serum and saliva is concentrated in exosomes. *PLoS One*, **7**, e30679.
- Gill N, Dhillon B (2022). RNA-seq Data Analysis for Differential Expression. *Methods Mol Biol*, **2391**, 45-54.
- Grimolizzi F, Monaco F, Leoni F, et al (2017). Exosomal miR-126 as a circulating biomarker in non-small-cell lung cancer regulating cancer progression. *Sci Rep*, **7**, 15277.
- Hsu YL, Hung JY, Chang WA, et al (2017). Hypoxic lung cancer-secreted exosomal miR-23a increased angiogenesis and vascular permeability by targeting prolyl hydroxylase and tight junction protein ZO-1. *Oncogene*, **36**, 4929-42.
- Inagaki M, Uchiyama M, Yoshikawa-Kawabe K, et al (2023). Comprehensive circulating microRNA profile as a supersensitive biomarker for early-stage lung cancer screening. *J Cancer Res Clin Oncol*, **149**, 8297-305.
- Jemal A, Siegel R, Ward E, et al (2009). Cancer statistics, 2009. *CA Cancer J Clin*, **59**, 225-49.
- Kadkhoda S, Eslami S, Mahmud Hussien B, et al (2022a). A review on the importance of miRNA-135 in human diseases. *Front Genet*, **13**, 973585.
- Kadkhoda S, Hussien BM, Eslami S, et al (2022b). A review on the role of miRNA-324 in various diseases. *Front Genet*, **13**, 950162.
- Khuder SA (2001). Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. *Lung Cancer*, **31**, 139-48.
- Kosaka N, Iguchi H, Yoshioka Y, et al (2010). Secretory mechanisms and intercellular transfer of microRNAs in living cells. *J Biol Chem*, **285**, 17442-52.
- Liu M, Si Q, Ouyang S, et al (2020). Serum MiR-4687-3p Has Potential for Diagnosis and Carcinogenesis in Non-small Cell Lung Cancer. *Front Genet*, **11**, 597508.
- Liu Y, Liu WB, Liu KJ, et al (2016). Overexpression of miR-26b-5p regulates the cell cycle by targeting CCND2 in GC-2 cells under exposure to extremely low frequency electromagnetic fields. *Cell Cycle*, **15**, 357-67.
- Lu S, Kong H, Hou Y, et al (2018). Two plasma microRNA panels for diagnosis and subtype discrimination of lung cancer. *Lung Cancer*, **123**, 44-51.
- Martin M (2011). Cutadapt removes adapter sequences from high throughput sequencing reads. *EMBnet. journal*, **17**, 1, 10.
- Melkamu T, Zhang X, Tan J, et al (2010). Alteration of microRNA expression in vinyl carbamate-induced mouse lung tumors and modulation by the chemopreventive agent indole-3-carbinol. *Carcinogenesis*, **31**, 252-8.
- Mitchell PS, Parkin RK, Kroh EM, et al (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A*, **105**, 10513-8.
- Nakharutai N, Traisathit P, Thongsak N, et al (2022). Impact of Residential Concentration of PM2.5 Analyzed as Time-Varying Covariate on the Survival Rate of Lung Cancer Patients: A 15-Year Hospital-Based Study in Upper Northern Thailand. *Int J Environ Res Public Health*, **19**.
- National Lung Screening Trial Research T, Aberle DR, Adams AM, et al (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*, **365**, 395-409.
- Pongnikorn D, Daoprasert K, Waisri N, et al (2018). Cancer incidence in northern Thailand: Results from six population-based cancer registries 1993-2012. *Int J Cancer*, **142**, 1767-75.
- Qu WQ, Liu L, Yu Z (2015). Clinical value of microRNA-23a upregulation in non-small cell lung cancer. *Int J Clin Exp Med*, **8**, 13598-603.
- Ramassone A, Pagotto S, Veronese A, et al (2018). Epigenetics and MicroRNAs in Cancer. *Int J Mol Sci*, **19**.
- Shi R, Chiang VL (2005). Facile means for quantifying microRNA expression by real-time PCR. *Biotechniques*, **39**, 519-25.
- Somsunun K, Prapamontol T, Pothirat C, et al (2022). Estimation of lung cancer deaths attributable to indoor radon exposure in upper northern Thailand. *Sci Rep*, **12**, 5169.
- Tsai YS, Lin CS, Chiang SL, et al (2011). Areca nut induces miR-23a and inhibits repair of DNA double-strand breaks by targeting FANCG. *Toxicol Sci*, **123**, 480-90.
- Turchinovich A, Weiz L, Burwinkel B (2012). Extracellular miRNAs: the mystery of their origin and function. *Trends Biochem Sci*, **37**, 460-5.
- Turchinovich A, Weiz L, Langheinz A, et al (2011). Characterization of extracellular circulating microRNA. *Nucleic Acids Res*, **39**, 7223-33.
- Virani S, Bilheem S, Chansaard W, et al (2017). National and Subnational Population-Based Incidence of Cancer in Thailand: Assessing Cancers with the Highest Burdens. *Cancers (Basel)*, **9**.
- Vrijens K, Bollati V, Nawrot TS (2015). MicroRNAs as potential signatures of environmental exposure or effect: a systematic review. *Environ Health Perspect*, **123**, 399-411.
- Wu Q, Wang C, Lu Z, et al (2012). Analysis of serum genome-wide microRNAs for breast cancer detection. *Clin Chim Acta*, **413**, 1058-65.
- Xia M, Duan ML, Tong JH, et al (2015). MiR-26b suppresses tumor cell proliferation, migration and invasion by directly targeting COX-2 in lung cancer. *Eur Rev Med Pharmacol Sci*, **19**, 4728-37.
- Yang X, Zhang Q, Zhang M, et al (2019). Serum microRNA Signature Is Capable of Early Diagnosis for Non-Small Cell Lung Cancer. *Int J Biol Sci*, **15**, 1712-22.
- Yin L, Xu G, Zhu Y, et al (2019). Expression of miR-23a and miR-135 and tumor markers in gastric cancer patients and the significance in diagnosis. *Oncol Lett*, **18**, 5853-8.
- Yin X, Tian M, Zhang J, et al (2022). MiR-26b-5p in small extracellular vesicles derived from dying tumor cells after irradiation enhances the metastasis promoting microenvironment in esophageal squamous cell carcinoma. *Cancer Lett*, **541**, 215746.
- Ying L, Du L, Zou R, et al (2020). Development of a serum

- miRNA panel for detection of early stage non-small cell lung cancer. *Proc Natl Acad Sci U S A*, **117**, 25036-42.
- Yu H, Guan Z, Cuk K, et al (2019). Circulating MicroRNA Biomarkers for Lung Cancer Detection in East Asian Populations. *Cancers (Basel)*, 11.
- Zhang YH, Jin M, Li J, et al (2020). Identifying circulating miRNA biomarkers for early diagnosis and monitoring of lung cancer. *Biochim Biophys Acta Mol Basis Dis*, **1866**, 165847.
- Zheng X, Lu S, He Z, et al (2020). MCU-dependent negative sorting of miR-4488 to extracellular vesicles enhances angiogenesis and promotes breast cancer metastatic colonization. *Oncogene*, **39**, 6975-89.
- Zhong D, Wu C, Xu D, et al (2021). Plasma-Derived Exosomal hsa-miR-4488 and hsa-miR-1228-5p: Novel Biomarkers for Dermatomyositis-Associated Interstitial Lung Disease with Anti-Melanoma Differentiation-Associated Protein 5 Antibody-Positive Subset. *Biomed Res Int*, 2021, 6676107.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.