

A unique set of 6 circulating microRNAs for early detection of non-small cell lung cancer

Ann Rita Halvorsen¹, Maria Bjaanæs^{1,2}, Marissa LeBlanc³, Are M. Holm⁴, Nils Bolstad⁵, Luis Rubio⁶, Juan Carlos Peñalver⁷, José Cervera⁸, Julia Cruz Mojarrieta⁹, Jose Antonio López-Guerrero⁶, Odd Terje Brustugun^{1,2}, Åslaug Helland^{1,2}

¹Department of Cancer Genetics, Institute for Cancer Research, OUS Radiumhospitalet, Oslo, Norway

²Department of Oncology, OUS Radiumhospitalet, Oslo, Norway

³Oslo Centre for Biostatistics and Epidemiology, Research Support Services, Oslo University Hospital, Oslo, Norway

⁴Department of Respiratory Medicine, OUS Rikshospitalet, Oslo, Norway

⁵Department of Medical Biochemistry, OUS Radiumhospitalet, Oslo, Norway

⁶Laboratory of Molecular Biology, Fundación Instituto Valenciano de Oncología, Valencia, Spain

⁷Department of Thoracic Surgery, Fundación Instituto Valenciano de Oncología, Valencia, Spain

⁸Department of Radiology, Fundación Instituto Valenciano de Oncología, Valencia, Spain

⁹Department of Pathology, Fundación Instituto Valenciano de Oncología, Valencia, Spain

Correspondence to: Åslaug Helland, **email:** ahelland@medisin.uio.no

Keywords: circulating microRNAs, biomarker, lung cancer, early detection, COPD

Received: February 15, 2016

Accepted: April 16, 2016

Published: May 14, 2016

ABSTRACT

Introduction: Circulating microRNAs are promising biomarkers for diagnosis, predication and prognostication of diseases. Lung cancer is the cancer disease accountable for most cancer deaths, largely due to being diagnosed at late stages. Therefore, diagnosing lung cancer patients at an early stage is crucial for improving the outcome. The purpose of this study was to identify circulating microRNAs for detection of early stage lung cancer, capable of discriminating lung cancer patients from those with chronic obstructive pulmonary disease (COPD) and healthy volunteers.

Results: We identified 7 microRNAs separating lung cancer patients from controls. By using RT-qPCR, we validated 6 microRNAs (miR-429, miR-205, miR-200b, miR-203, miR-125b and miR-34b) with a significantly higher abundance in serum from NSCLC patients. Furthermore, the 6 miRNAs were validated in a different dataset, revealing an area under the receiver operating characteristic curve of 0.89 for stage I-IV and 0.88 for stage I/II.

Materials and Methods: We profiled the expression of 754 unique microRNAs by TaqMan Low Density Arrays, and analyzed serum from 38 patients with NSCLC, 16 patients suffering from COPD and 16 healthy volunteers from Norway, to explore their potential as diagnostic biomarkers. For validation, we analyzed serum collected from high-risk individuals enrolled in the Valencia branch of the International Early Lung Cancer Action Program screening trial (n=107) in addition to 51 lung cancer patients.

Conclusion: Considering the accessibility and stability of circulating miRNAs, these 6 microRNAs are promising biomarkers as a supplement in future screening studies.

INTRODUCTION

Worldwide, lung cancer is the primary cause of cancer death. Non-small cell lung cancer (NSCLC) represents approximately 85% of all diagnosed lung cancer cases, and adenocarcinomas and squamous cell carcinomas are the main histological subtypes. Despite improvements in targeted therapies, the overall survival rates remain poor, which is partly explained by delayed diagnosis. Only 25% of the patients are diagnosed at an early stage, when the tumor is accessible for curative surgery [1]. Today, several low-dose computed tomography (LDCT) screening studies have been initiated to evaluate the effect of such programs on lung cancer mortality. One concern of using annual CT scanning is that it may lead to harm due to radiation exposure [2]. Another concern of using LDCT as a screening tool is the high risk for false positives. A recently published meta-analysis of LDCT screening reported 235 false positive nodules detected for every 1000 individuals screened [3]. In the National Lung Screening Trial (NLST), 24.2% of enrolled persons at risk screened with LDCT had a positive result. Out of these, 96.4% were false positive. This, in turn, can lead to unnecessary and potentially harmful further diagnostics. However, a reduction in mortality of 20% was observed when screened with LDCT, which indicates that screening is beneficial [4]. Since a blood test is readily accessible, circulating biomarkers for early detection of lung cancer may be an integrated part of future lung cancer screening programs.

MicroRNAs are small non-coding oligonucleotides with capacity to negatively regulate expression of mRNAs by inhibition of protein translation [5]. MicroRNAs are involved in several, essential biological processes, and based on microRNA profiles, tumor cells can be distinguished from normal cells [6]. Interestingly, tumor cells can release microRNAs into circulation which can be detected in fluids like blood [7]. These microRNAs are shown to be remarkably stable and are able to avoid degradation from RNases. In blood, microRNAs can be incorporated into exosomes or bound to proteins like Argonaut [8] or HDL and transported via serum to a target cell with possible gene expression alteration [9]. Much effort has been invested in order to find circulating microRNAs for cancer identification, treatment monitoring and prognostication. Several studies have reported evidence of an altered circulating microRNA expression profile in cancer patients, including patients with breast cancer and lung cancer [10, 11]. However, due to small cohorts, lack of validation and use of different types of microRNA screens, it has proved challenging to reproduce initially promising discoveries within this field.

The purpose of this study was to find circulating biomarkers specific for lung cancer, capable of discriminating lung cancer patients from those with chronic obstructive pulmonary disease (COPD) and healthy normal volunteers.

RESULTS

By using TLDA cards, a total number of 754 microRNAs were screened in serum from lung cancer patients, COPD patients, and healthy volunteers. Due to low abundance of microRNAs in serum, we did a manual inspection of the expression curve of each microRNA. Those with irregular shape of the curve, or microRNAs expressed in less than 30% of the samples were removed. 272 microRNAs passed the abovementioned control-steps, and were used in further analyses. Many of the microRNAs found in blood showed similar expression profile in both groups. In order to capture microRNAs differentially expressed, we performed a PAM analysis, resulting in a list of 37 significantly differentially expressed microRNAs (FDR<0.01) (Supplementary Table 1).

For visualization, a supervised hierarchical clustering based on the 37 microRNAs identified two main microRNA clusters, as displayed in Supplementary Figure 1. In order to find a signature able to distinguish lung cancer patients from controls, microRNAs with a higher abundance in serum from lung cancer patients compared with the controls were further examined. Out of the 37 microRNAs identified by PAM analysis, 10 were upregulated (miR-205, miR-200b, miR-125b, miR-34b, miR-429, miR-203, miR-186, miR-1180, miR-601 and miR-378). Due to low expression of circulating microRNAs, a technical validation was performed on the 10 selected microRNAs. Seven of these (miR-205, miR-200b, miR-125b, miR-34b, miR-429, miR-203, and miR-1180) showed a difference in expression between the two groups and were further validated in the external validation cohort. We did not find any correlation between tumor stage or tumor size and the level of microRNA's in the discovery set or in the validation set.

Validation of selected microRNAs

The seven selected microRNAs were analyzed in the external validation set, using RT-qPCR and single assays. All microRNAs except miR-1180 showed a significantly higher abundance in serum from lung cancer patients compared with those with no detected tumor, and miR-1180 was left out of further analysis. The differences in expression are shown in Table 1.

In order to visualize the sensitivity and specificity for the six microRNAs combined, a ROC curve was calculated (Figure 1). To see how each microRNA performed individually, a ROC curve and the area under the curve (AUC) was calculated for the 6 validated microRNAs (Figure 2). With an AUC on 0.83, miR-200b revealed the highest sensitivity and specificity of the 6 microRNAs.

For the validation set, the cross-validation based area under ROC curve was 0.89 for the 6-microRNAs combined (logistic regression, the coefficients are shown

Table 1: An independent samples t-test was conducted to compare expression of the 6 microRNAs for lung cancer patients and controls

Sample Set	miRNA	p-value (t-test)	Mean difference	95% CI mean difference
Discovery Set	miR125bCt	0.025	-1.80	-3.36 -0.24
	miR200bCt	<0.001	-4.65	-6.35 -2.96
	miR203Ct	<0.001	-1.60	-2.35 -0.86
	miR205Ct	<0.001	-4.45	-6.12 -2.79
	miR34bCt	0.082	-1.48	-3.16 0.19
	miR429Ct	<0.001	-3.25	-4.89 -1.61
Validation Set	miR125bCt	<0.001	-2.38	-3.20 -1.57
	miR200bCt	<0.001	-1.46	-1.83 -1.09
	miR203Ct	<0.001	-2.08	-3.14 -1.02
	miR205Ct	<0.001	-2.56	-3.26 -1.87
	miR34bCt	0.002	-0.74	-1.20 -0.29
	miR429Ct	<0.001	-1.65	-2.13 -1.17

Data are shown for both the discovery set and validation set.

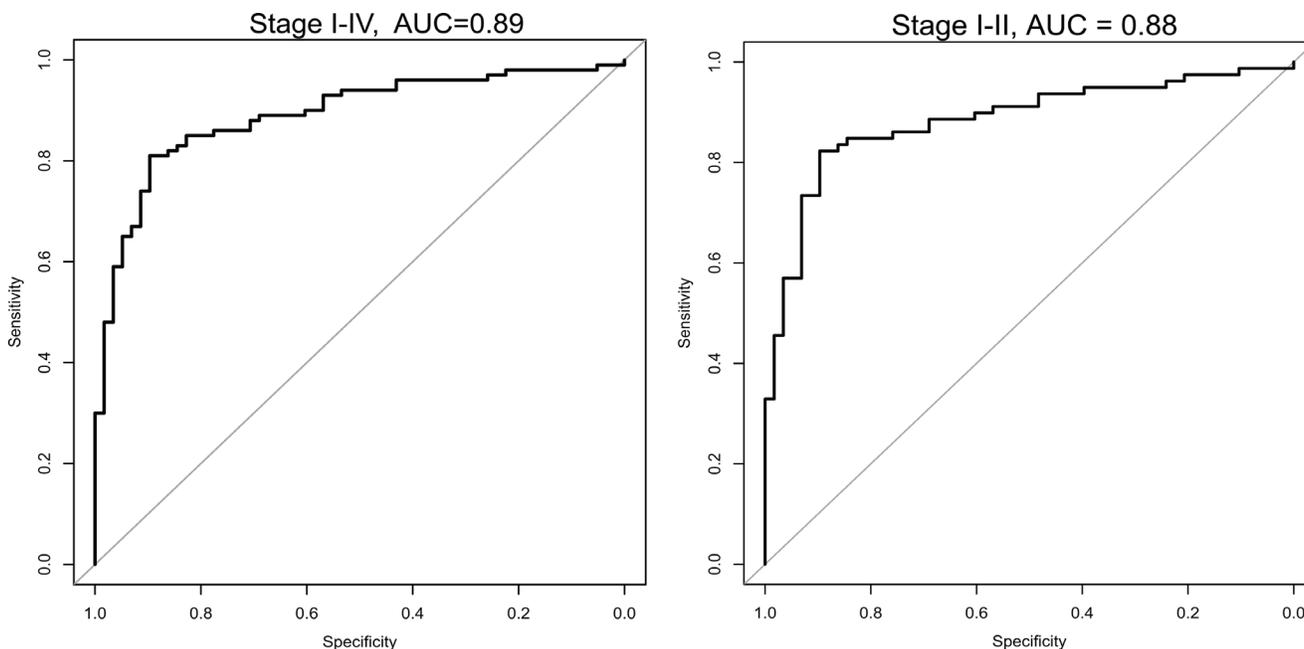


Figure 1: ROC Curve combined for the selected 6 microRNA analyzed in the validation set, for all stages (I-IV) and for only the early stages (I/II). The plot shows the sensitivity versus specificity. The area under the curve (AUC) gives a summary of the sensitivity and specificity across a range of different cut off points.

in Supplementary Table 2). The sensitivity and specificity for the 6-miRNAs were obtained by a threshold value. With a threshold value of 0.37, a sensitivity of 88 % and a specificity of 71 % were found. Depending on clinical priorities, a higher sensitivity can be achieved by lowering the threshold. The pair-wise correlations for the six microRNAs ranged from 0.01 – 0.76, supporting their usefulness in the prediction model (Supplementary Table 3).

In order to emphasize the predictive potential of the six microRNAs in only the early stage patients, stage IIIa and IV were removed, and stage Ia/Ib/IIa/IIb were analyzed separately (logistic regression, Supplementary Table 2). For the 6 microRNAs combined an area under the ROC curve of 0.88 was achieved (Figure 1), and a sensitivity of 85% and a specificity of 74% were found.

Does the microRNA profile in serum reflect the microRNA profile in tumor?

A study based on microRNA expression in lung adenocarcinoma samples and normal lung samples has been performed on the same patients by Bjaanæs and colleagues [12]. They identified 129 significantly differentially expressed microRNAs in lung adenocarcinoma tissue compared with adjacent normal lung tissue. When we compared the microRNA profile in

tumor tissue with the serum profile, only 8 overlapping microRNAs were found. However, 3 of these were in the opposite direction (Supplementary Table 4).

DISCUSSION

By comparing expression levels of serum microRNAs in lung cancer patients with serum from COPD patients and healthy controls, we identified 37 differentially expressed microRNAs. It is important to bear in mind that most of the measured circulating microRNAs in blood are released by other cells than tumor cells [13]. Therefore, we have focused on the circulating microRNAs upregulated in lung cancer patients. With high sensitivity and specificity, six microRNAs distinguished individuals with lung cancer from those without. Focusing on only the early stages, the sensitivity and specificity remained high indicating that these microRNAs are able to capture lung cancer patients very early in the cancer development.

The 6 identified microRNAs are known to be involved in tumor progression. MiR-200b and miR-429 are both members of the miR-200 family, reported to be elevated in serum from many cancer types, including breast, ovarian and pancreatic cancer [14–16]. Independent studies have revealed that the miR-200 family and miR-205 play critical roles in regulating EMT by targeting the

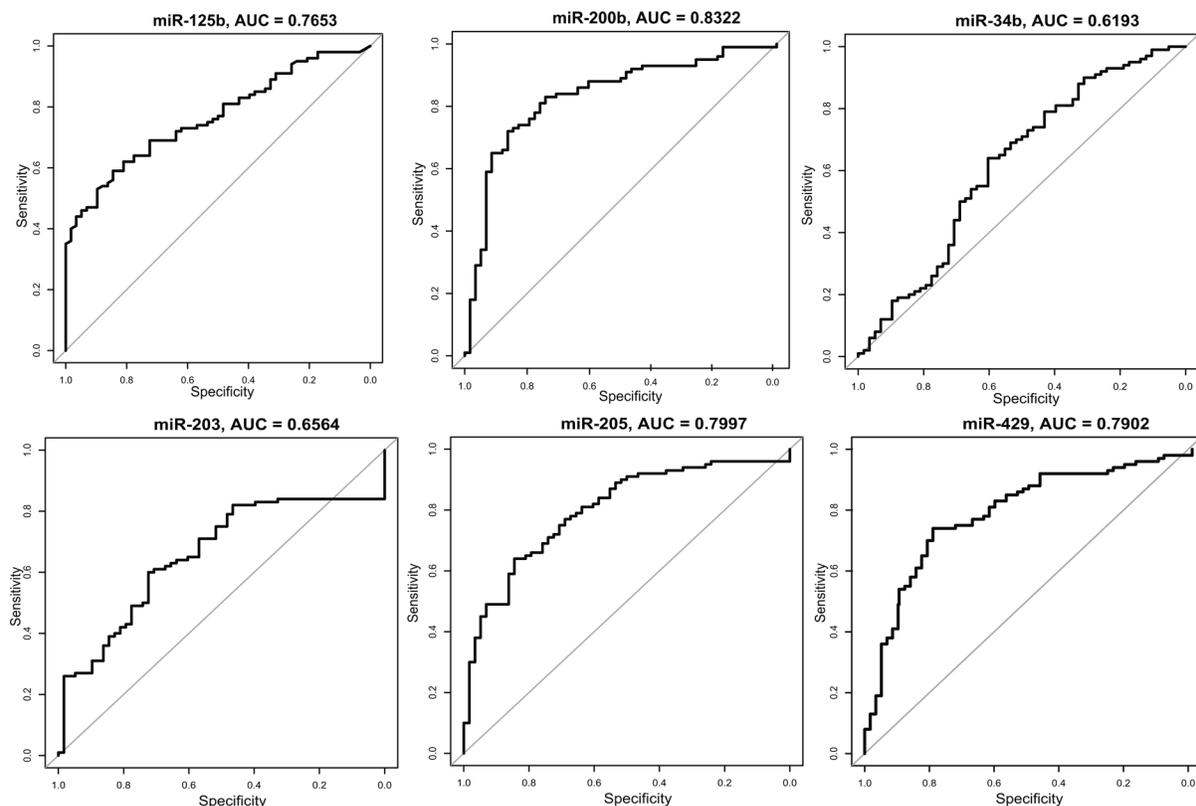


Figure 2: ROC curves for the 6 validated microRNAs analyzed in the validation set. The plots show the sensitivity versus specificity. The area under the curves (AUC) is given for each microRNA.

E-cadherin repressors ZEB1 and ZEB2 [17]. Furthermore, ZEB1 has been reported to repress miR-203 in the ZEB/miR-200 feedback loop [18]. In a recent study, the metastatic capacities of the miR-200s were investigated in breast cancer. The miR-200s were secreted in extracellular vesicles (EV) in human breast cancer cell lines, engulfed in non-metastatic cells, altering the gene expression and facilitating mesenchymal-to-epithelial transition (MET) [19]. Both miR-205 and miR-203 have been reported to be elevated in blood from NSCLC patients [20]. In some studies, miR-205 is described as a tumor suppressor, but others link this microRNA to processes of tumor initiation and progression of cancer [21]. However, several studies have demonstrated that the release of miR-205 is selectively packed into exosomes and secreted by lung cancer cells, and the level of circulating miR-205 decreases after lung tumor removal [22].

In this study, miR-125b was elevated in the serum samples of lung cancer patients. Published reports on miR-125b are somewhat contradictory, and it seems like this microRNA can act both as a tumor suppressor and have an oncogenic role in tumorigenesis [23, 24]. Nevertheless, miR-125b can directly target the tumor suppressor gene *TP53* [25] and the downstream modulator *TP53INP1* [26], contributing to malignancy. Interestingly, miR-34b, which is a transcriptional target of TP53 often found downregulated in tumor [27], showed a higher abundance in sera from lung cancer patients than in the control sera in this study. In another study of circulating microRNA, miR-34b was also found to be upregulated in sera of prostate cancer patients with respect to normal individuals [28].

MicroRNAs for detection of lung cancer

Blood-based screening tests may be a crucial diagnostic tool to detect lung cancer at an early stage. Today, LDCT is widely used for tumor detection, and although it has been shown to reduce lung cancer deaths, the specificity is not optimal. A blood-based test with high sensitivity would be of great value as an addition to LDCT. This could increase both specificity and sensitivity. Since most screening projects have focused on the high-risk population of heavy smokers, we wanted to develop a blood-based screening test to be offered to all at risk. In order to capture lung cancer also among never smokers with a blood based test, 47% never-smokers were included in lung cancer group in the discovery cohort. However, since LDCT screening cohorts don't include never-smokers, and only 2 never-smokers were included in the Valencia-cohort, we were not able to confirm this potential in the external validation, and future studies need to address this further. On the other hand, six microRNAs were validated in a cohort slightly different from the discovery set, indicating the robustness of the identified microRNAs. Many researchers report on biomarkers with high potential for lung cancer detection

[29, 30], but a commercial blood-based screening test is yet not available. Recently, a 4-miR serum based signature able to discriminate lung cancer patients from non-cancer controls with high AUC was identified. Two of the reported microRNAs (miR-141 and miR-200b) belong to the miR-200 family [31]. Interestingly, circulating miR-141, miR-203 and miR-200b have recently also been reported to discriminate between colorectal patients and healthy controls [32]. This is in line with our study, where miR-200b, miR-203 and miR-429 could discriminate between lung cancer and controls. However, most of the identified signatures reported are not consistent with other reported microRNA signatures. This may be due to many factors such as use of limited number of samples or pooled samples in the discovery set [30, 33, 34], or studies based on candidate markers [35, 36]. A promising plasma based microRNA ratio signature was tested in a screening trial revealing a high sensitivity and specificity when combined with LDCT. In that study, 1000 blood samples were collected, identifying 60 of 69 lung cancers and reducing the false positives down to 3.7% [37]. In the present study, we chose another strategy, analyzing serum samples using global normalization and focusing on only microRNAs with higher abundance in serum from lung cancer patients relatively to the controls. A reduction in microRNA abundance is most likely an unspecific reaction, and cannot be explained by the tumor itself [38]. In addition, we have a thorough handling of missing values, a substantial underestimated bias when analyzing low abundant molecules [39]. A lot of microRNAs cannot be detected in the blood due to biological features and probably not due to technical issues, and should be handled with a cut-off value and not being excluded from the analysis [39]. We replaced the missing values with the lowest value in the dataset on linear scale, before log₂ transformation of the data. Here, 70 serum samples were used in the discovery set, 158 samples in the validation set, and 754 unique microRNAs were screened in the discovery set. However, this was explored mainly in adenocarcinomas and should be further investigated in all types of lung cancer histology. Due to tumor heterogeneity and differences between patients, tumors can act very differently and the regulation of tumor cells may vary radically between individuals [29]. We considered that using serum from COPD patients as part of the controls as extremely important in order to avoid a smoking based signature. It's also crucial to validate the findings in an independent cohort, and by confirming our results in the IELCAP study proves the robustness of our study. Two prerequisites for clinically effective biomarkers are high specificity and sensitivity. In our study, the combination of the six microRNAs revealed a higher AUC than any of the single microRNAs. One marker may detect only a portion of the patients, but by combining several biomarkers, the possibility of capturing a high number of the patients with increased diagnostic accuracy might be feasible.

Can the microRNA profile in serum reflect the microRNA profile in tumor?

When we compared the microRNA profile found in tumor with the profile in serum, only 5 microRNAs were significantly differentially expressed in the same direction. Other studies have demonstrated that microRNA profile in serum does not mirror the microRNA profile found in tumor [28, 30]. This may indicate that microRNAs are not randomly shed from tumor, but selected for their ability to regulate important mRNAs. Therefore, microRNA candidates from tumor profiling may be inappropriate as biomarkers in serum.

MATERIALS AND METHODS

Serum samples

In the discovery set, the patients were diagnosed with operable NSCLC from 2006 to 2011, and underwent curatively intended surgical resection at Rikshospitalet, Oslo University Hospital, Norway. The procedure for blood samples handling are outlined in Supplementary Table 5. The study was approved by the Regional Ethics Committee (S-05307), and written informed consent was

obtained from all patients. Clinical data were collected from questionnaires, the medical journal, histology reports and follow up information from the patients' local hospitals. All patients selected for the discovery cohort were diagnosed with adenocarcinoma. A predominance of stage I and II were included due to inoperable tumors in later stages. In order to capture the profile of the never smokers, 47.4% of the lung cancer patients were never smokers. For all selected patients, microRNA profiling in tumor has previously been performed by Bjaanæs and colleagues [12]. Sera from 16 patients diagnosed with COPD and 16 healthy volunteers were used as controls [40]. The overview of the study design is displayed in Figure 3 and the clinical data are shown in Table 2.

Samples from 107 persons enrolled into the International Early Lung Cancer Action Program (IELCAP, Valencia branch), were obtained from the Biobank of the Fundación Instituto Valenciano de Oncología (FIVO). 58 of these had no cancer detected with CT, denoted controls, and 49 were diagnosed with lung cancer. These series corresponded to a high-risk population for lung cancer >50 years old, heavy smokers for more than 15 years but without any history of cancer. All participants were examined with CT and whole blood was collected before diagnostic scanning. In addition,

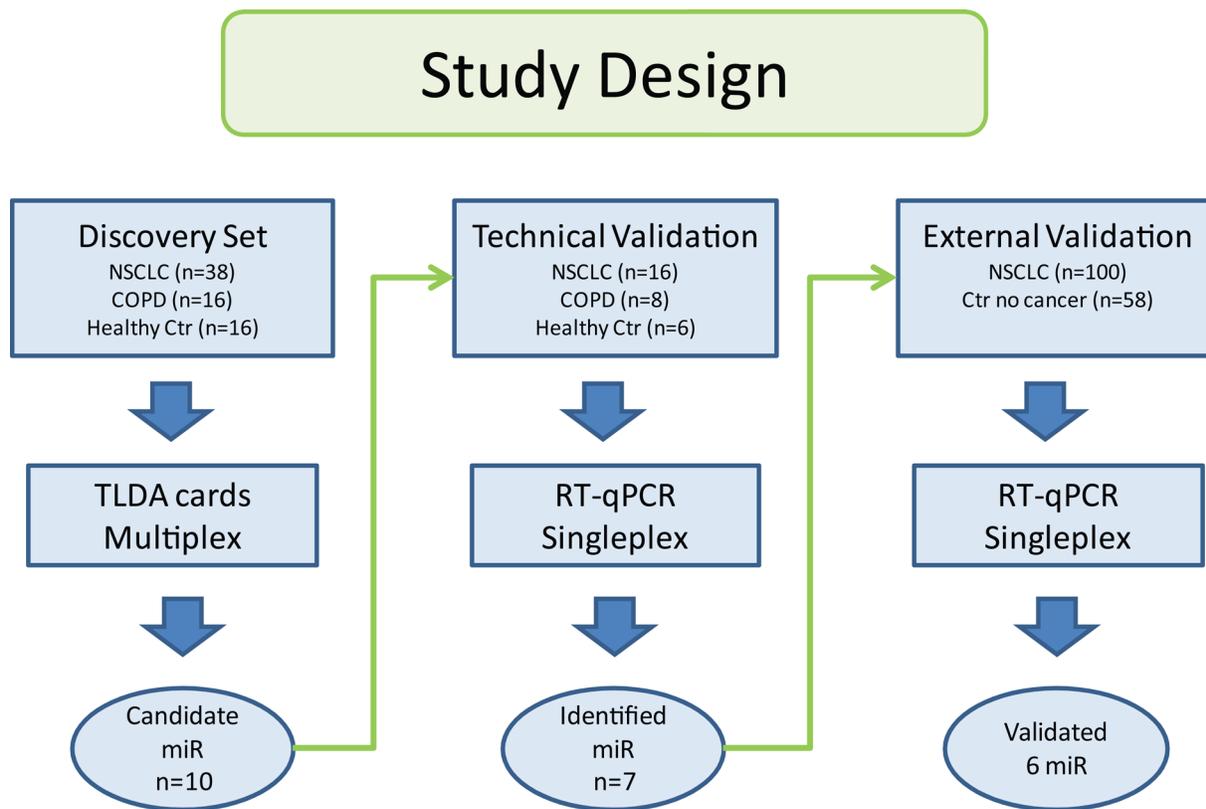


Figure 3: The study was divided into three phases. Samples in the discovery set were used to find candidate microRNAs. Technical validation of the candidate microRNAs was performed on a subset of the same samples. Identified microRNAs from the discovery set were analyzed in the external validation.

Table 2: Information about gender, stage, age and smoking status for all the patients in this study are shown in the table

	Discovery Set			P value	Validation Set		
	NSCLC	COPD	Healthy Control		NSCLC	Healthy Control	P value
	n = 38 No.(percent)	n = 16 No.(percent)	n = 16 No.(percent)		n = 100 No.(percent)	n = 58 No.(percent)	
Gender							
Male	8 (21)	8 (50)	7 (43.8)	p = 0.022†	72 (72)	34 (59)	
Female	30 (79)	8 (50)	9 (56.2)		28 (28)	24 (41)	p = 0.085†
Age(mean)	66.8	66.5	52.9	p = 0.017††	62.6	57.6	p < 0.001†††
Smoking status							
Never smoker	18 (47.4)	0 (0)	3 (18.8)		2 (2)	0 (0)	
Former/current smoker	20 (52.6)	16 (100)	13 (81.2)		87 (87)	58 (100)	
Unknown				p = 0.001†	11 (11)		p = 0.519†††
Packyears mean	6.6*	ND	ND		43.8	30.8	p < 0.001†††
Histology							
Adenocarcinomas	38 (100)				98 (98)		
Squamous cell carcinomas	0 (0)			NA	2 (2)		NA
Tumor stage							
Ia/Ib	24 (63.2)				65 (65)		
IIa/IIb	5 (13.2)				14 (14)		
IIIa	8 (21)				18 (18)		
IV	1 (2.6)			NA	2 (2)		NA

* 47% never-smokers included

†Pearson's Chi-square test

†† Students T-test

††† Fisher's Exact Test

NA: not assessed

ND: no data

51 serum samples from lung cancer patients from the Thoracic surgery department in Valencia were analyzed.

MicroRNA screening with TLDA cards

From 500µl serum, totalRNA was extracted using miRCURY™ RNA isolation kit for biofluids (cat#300113, Exiqon) according to manufactures protocol. TotalRNA (totRNA) quantity and quality (yield, 260/280 ratio and 260/230 ratio) were determined with the NanoDrop ND-1000 spectrophotometer (Thermo Scientific).

First strand synthesis was performed using megaplex™ RT primer pool (pool A and B, cat#4444745, Life Technologies) and TaqMan® MicroRNA reverse

transcription kit (cat#4366596 Life Technologies). For each sample, 30 ng totRNA was added to the reaction mix. In order to increase the sensitivity, preamplification of the cDNA was implemented using megaplex™ preamp primer pool (pool A and B, cat#4444748, Life Technologies) and TaqMan® preamp master mix (cat#4391128, Life Technologies). The samples were diluted with 75µl of 0.1XTE pH 8.0 and then added with TaqMan® Universal PCR Master Mix, NoAmpErase® UNG, 2X. The reaction mix was applied to the TaqMan® Low Density Arrays (TLDA; Life Technologies, cat# 4444913) and loaded onto the 7900HT Fast thermocycler system (Life Technologies) for analysis. The protocol from supplier (PN 4399721) was followed for all the procedures. The

microRNA microarray data have been deposited in the Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo>) with the accession number GSE70080.

Validation with reverse transcription-quantitative PCR (RT-qPCR)

Due to low expression levels of microRNAs in sera, a technical validation was performed with RT-qPCR for the ten candidate microRNAs upregulated in sera from lung cancer patients. Seven of these were further evaluated in the independent screening cohort of serum samples, using the same technique. The microRNAs were first reverse transcribed (MicroRNA reverse transcription kit, Life Technologies) and then quantified using qPCR. The procedures were performed according to standard protocols. The primers, reaction mix components and temperature conditions used for RT and qPCR are listed in Supplementary Table 6. All samples were run in triplicates. The comparative threshold (Ct) was used to evaluate the relative detection level of each microRNA.

Normalization

The raw C_T data were exported to the ExpressionSuite software (v.1.0.3, Life Technologies) for normalization and quality control. Threshold and baseline was automatically calculated for each assay, and a global normalization was performed, as is recommended for large scale microRNA expression profiling. Global normalization first finds the common microRNAs among all samples. The median CT of those microRNAs is used as the normalizer, on a per sample basis [41]. Normalized ΔC_T data were used for calculation of relative gene expression in fold change ($2^{-\Delta\Delta C_T}$) [42]. In the ExpressionSuite software a calculation score for the best choice of endogenous controls was performed. The three most uniformly expressed microRNAs (miR-220, miR-19b, sU6) from the TLDA cards were analyzed and chosen as reference genes for normalization in the validation set. To obtain the ΔC_T , the mean expression level of the three microRNAs ($C_{T_{ref}}$) was subtracted from the mean Ct level ($C_{T_{target}}$) of each sample. Due to low expression of microRNAs in serum, missing values were most likely the results of too few RNA-copies. Not detected measurements were replaced by zero in linear scale. For the logarithmic scale the lowest expressed value for each microRNA replaced the missing values. Pre-amplification was not utilized in the validation and Ct values of <40 indicated the presence of microRNA.

Statistical analysis

The dataset was divided into two groups consisting of 1) lung cancer patients 2) COPD/healthy

controls, and a Prediction Analysis of Microarrays (PAM) was performed in R, ver. 2.12.2 [43]. The “nearest shrunken method” was used to identify those genes best characterizing each class [44]. Unsupervised hierarchical clustering was utilized in J-express [45]. A logistic regression model was built in the external validation set using 6 microRNAs as predictors and case/control status as the outcome. Leave-one-out-cross-validation was performed in R to assess how well the given model will generalize to an independent data set and a receiver operating characteristic curve (ROC) was generated from the cross-validation predictions. Chi-square test (χ^2) and Student T-test were used to evaluate the frequency of clinical variables among lung cancer patients and controls (IBM SPSS Statistics, Version 21.0. Armonk, NY: IBM Corp).

CONCLUSION

A global profiling of microRNAs in serum of lung cancer patients and controls was conducted and a panel of 6-microRNAs was identified in a Norwegian cohort. The microRNAs found in this study showed a significantly higher abundance in sera from lung cancer patients, compared with sera from COPD patients and healthy controls. Validation in a larger independent cohort from Spain confirmed a high sensitivity and specificity, which indicate that these six circulating microRNAs have a great potential to detect early stage lung cancer. Further validation of the six microRNAs identified in this study should be a part of a prospective study for early diagnosis of lung cancer.

ACKNOWLEDGMENTS

Special thanks to Department of Clinical Molecular Biology Division of Medicine, Akershus University Hospital, Norway for the use of their instruments for this project. The authors thank the Biobank of the Fundación Instituto Valenciano de Oncología for providing the biological samples for the validation set of these analyses. The authors thank the Nordic Society of Clinical Chemistry (NFKK) for making samples and data from the Nordic Reference Interval Project Bio-bank and Database (NOBIDA) available to this project. We want to thank Ingerd Solvoll for help with collecting tissue samples and clinical information. The authors gratefully acknowledge the contribution from the patients participating in this study.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

GRANT SUPPORT

This project was supported with grant from The Eurocan platform, Helse Sør-Øst and the Norwegian Cancer Society.

REFERENCES

1. Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, Galen B, Gareen IF, Gatsonis C, Goldin J, Gohagan JK, Hillman B, Jaffe C, Kramer BS, et al. The National Lung Screening Trial: overview and study design. *Radiology*. 2011; 258:243-253.
2. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, Byers T, Colditz GA, Gould MK, Jett JR, Sabichi AL, Smith-Bindman R, Wood DE, et al. Benefits and harms of CT screening for lung cancer: a systematic review. *Jama*. 2012; 307:2418-2429.
3. Gopal M, Abdullah SE, Grady JJ, Goodwin JS. Screening for lung cancer with low-dose computed tomography: a systematic review and meta-analysis of the baseline findings of randomized controlled trials. *Journal of thoracic oncology*. 2010; 5:1233-1239.
4. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD. Reduced lung-cancer mortality with low-dose computed tomographic screening. *The New England journal of medicine*. 2011; 365:395-409.
5. Ma L, Weinberg RA. MicroRNAs in malignant progression. *Cell cycle*. 2008; 7:570-572.
6. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer*. 2006; 6:857-866.
7. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanian EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, Drescher CW, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:10513-10518.
8. Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, Mitchell PS, Bennett CF, Pogosova-Agadjanian EL, Stirewalt DL, Tait JF, Tewari M. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:5003-5008.
9. Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat Cell Biol*. 2011; 13:423-433.
10. Chan M, Liaw CS, Ji SM, Tan HH, Wong CY, Thike AA, Tan PH, Ho GH, Lee AS. Identification of circulating microRNA signatures for breast cancer detection. *Clin Cancer Res*. 2013; 19:4477-4487.
11. Rani S, Gately K, Crown J, O'Byrne K, O'Driscoll L. Global analysis of serum microRNAs as potential biomarkers for lung adenocarcinoma. *Cancer biology & therapy*. 2013; 14:1104-1112.
12. Bjaanaes MM, Halvorsen AR, Solberg S, Jorgensen L, Dragani TA, Galvan A, Colombo F, Anderlini M, Pastorino U, Kure E, Borresen-Dale AL, Brustugun OT, Helland A. Unique microRNA-profiles in EGFR-mutated lung adenocarcinomas. *Int J Cancer*. 2014; 135:1812-1821.
13. Wang J, Zhang KY, Liu SM, Sen S. Tumor-associated circulating microRNAs as biomarkers of cancer. *Molecules (Basel, Switzerland)*. 2014; 19:1912-1938.
14. Madhavan D, Zucknick M, Wallwiener M, Cuk K, Modugno C, Scharpf M, Schott S, Heil J, Turchinovich A, Yang R, Benner A, Riethdorf S, Trumpp A, et al. Circulating miRNAs as surrogate markers for circulating tumor cells and prognostic markers in metastatic breast cancer. *Clin Cancer Res*. 2012; 18:5972-5982.
15. Kan CW, Hahn MA, Gard GB, Maidens J, Huh JY, Marsh DJ, Howell VM. Elevated levels of circulating microRNA-200 family members correlate with serous epithelial ovarian cancer. *BMC Cancer*. 2012; 12:627.
16. Li A, Omura N, Hong SM, Vincent A, Walter K, Griffith M, Borges M, Goggins M. Pancreatic cancers epigenetically silence SIP1 and hypomethylate and overexpress miR-200a/200b in association with elevated circulating miR-200a and miR-200b levels. *Cancer Res*. 2010; 70:5226-5237.
17. Gregory PA, Bracken CP, Bert AG, Goodall GJ. MicroRNAs as regulators of epithelial-mesenchymal transition. *Cell cycle*. 2008; 7:3112-3118.
18. Brabletz S, Brabletz T. The ZEB/miR-200 feedback loop--a motor of cellular plasticity in development and cancer? *EMBO reports*. 2010; 11:670-677.
19. Le MT, Hamar P, Guo C, Basar E, Perdigo-Henriques R, Balaj L, Lieberman J. miR-200-containing extracellular vesicles promote breast cancer cell metastasis. *J Clin Invest*. 2014; 124:5109-5128.
20. Rabinowits G, Gercel-Taylor C, Day JM, Taylor DD, Kloecker GH. Exosomal microRNA: a diagnostic marker for lung cancer. *Clinical lung cancer*. 2009; 10:42-46.
21. Qin AY, Zhang XW, Liu L, Yu JP, Li H, Wang SZ, Ren XB, Cao S. MiR-205 in cancer: an angel or a devil? *Eur J Cell Biol*. 2013; 92:54-60.
22. Aushev VN, Zborovskaya IB, Laktionov KK, Girard N, Cros MP, Herceg Z, Krutovskikh V. Comparisons of microRNA patterns in plasma before and after tumor removal reveal new biomarkers of lung squamous cell carcinoma. *PloS one*. 2013; 8:e78649.
23. Feliciano A, Castellvi J, Artero-Castro A, Leal JA, Romagosa C, Hernandez-Losa J, Peg V, Fabra A, Vidal F, Kondoh H, Ramon YCS, Lleona ME. miR-125b acts as a tumor suppressor in breast tumorigenesis via its novel direct

- targets ENPEP, CK2-alpha, CCNJ, and MEGF9. *PLoS one*. 2013; 8:e76247.
24. Wang X, Zhang Y, Fu Y, Zhang J, Yin L, Pu Y, Liang G. MicroRNA-125b may function as an oncogene in lung cancer cells. *Molecular medicine reports*. 2015; 11:3880-3887.
 25. Jiang F, Liu T, He Y, Yan Q, Chen X, Wang H, Wan X. MiR-125b promotes proliferation and migration of type II endometrial carcinoma cells through targeting TP53INP1 tumor suppressor in vitro and in vivo. *BMC Cancer*. 2011; 11:425.
 26. Le MT, Teh C, Shyh-Chang N, Xie H, Zhou B, Korzh V, Lodish HF, Lim B. MicroRNA-125b is a novel negative regulator of p53. *Genes Dev*. 2009; 23:862-876.
 27. Corney DC, Flesken-Nikitin A, Godwin AK, Wang W, Nikitin AY. MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth. *Cancer Res*. 2007; 67:8433-8438.
 28. Lodes MJ, Caraballo M, Suci D, Munro S, Kumar A, Anderson B. Detection of cancer with serum miRNAs on an oligonucleotide microarray. *PLoS one*. 2009; 4:e6229.
 29. Chen X, Hu Z, Wang W, Ba Y, Ma L, Zhang C, Wang C, Ren Z, Zhao Y, Wu S, Zhuang R, Zhang Y, Hu H, et al. Identification of ten serum microRNAs from a genome-wide serum microRNA expression profile as novel noninvasive biomarkers for nonsmall cell lung cancer diagnosis. *Int J Cancer*. 2012; 130:1620-1628.
 30. Boeri M, Verri C, Conte D, Roz L, Modena P, Facchinetti F, Calabro E, Croce CM, Pastorino U, Sozzi G. MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:3713-3718.
 31. Nadal E, Truini A, Nakata A, Lin J, Reddy RM, Chang AC, Ramnath N, Gotoh N, Beer DG, Chen G. A Novel Serum 4-microRNA Signature for Lung Cancer Detection. *Scientific reports*. 2015; 5:12464.
 32. Sun Y, Liu Y, Cogdell D, Calin GA, Sun B, Kopetz S, Hamilton SR, Zhang W. Examining plasma microRNA markers for colorectal cancer at different stages. *Oncotarget*. 2016; 7:11434-11449. doi: 10.18632/oncotarget.7196.
 33. Keller A, Leidinger P, Gislefoss R, Haugen A, Langseth H, Staehler P, Lenhof HP, Meese E. Stable serum miRNA profiles as potential tool for non-invasive lung cancer diagnosis. *RNA biology*. 2011; 8:506-516.
 34. Foss KM, Sima C, Ugolini D, Neri M, Allen KE, Weiss GJ. miR-1254 and miR-574-5p: serum-based microRNA biomarkers for early-stage non-small cell lung cancer. *Journal of thoracic oncology*. 2011; 6:482-488.
 35. Zheng D, Haddadin S, Wang Y, Gu LQ, Perry MC, Freter CE, Wang MX. Plasma microRNAs as novel biomarkers for early detection of lung cancer. *International journal of clinical and experimental pathology*. 2011; 4:575-586.
 36. Shen J, Liu Z, Todd NW, Zhang H, Liao J, Yu L, Guarnera MA, Li R, Cai L, Zhan M, Jiang F. Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microRNA biomarkers. *BMC Cancer*. 2011; 11:374.
 37. Sozzi G, Boeri M, Rossi M, Verri C, Suatoni P, Bravi F, Roz L, Conte D, Grassi M, Sverzellati N, Marchiano A, Negri E, La Vecchia C, et al. Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: a correlative MILD trial study. *Journal of clinical oncology*. 2014; 32:768-773.
 38. Witwer KW. Circulating microRNA biomarker studies: pitfalls and potential solutions. *Clin Chem*. 2015; 61:56-63.
 39. Stahlberg A, Rusnakova V, Kubista M. The added value of single-cell gene expression profiling. *Briefings in functional genomics*. 2013; 12:81-89.
 40. Rustad P, Simonsson P, Felding P, Pedersen M. Nordic Reference Interval Project Bio-bank and Database (NOBIDA): a source for future estimation and retrospective evaluation of reference intervals. *Scand J Clin Lab Invest*. 2004; 64:431-438.
 41. Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, Vandesompele J. A novel and universal method for microRNA RT-qPCR data normalization. *Genome biology*. 2009; 10:R64.
 42. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif)*. 2001; 25:402-408.
 43. Team RC. (2013). R: language and environment for statistical computing. (URL <http://www.R-project.org/>). R Foundation for Statistical Computing, Vienna, Austria.
 44. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99:6567-6572.
 45. Dysvik B, Jonassen I. J-Express: exploring gene expression data using Java. *Bioinformatics*. 2001; 17:369-370.