Applied Cancer Research

RESEARCH ARTICLE

Open Access

Molecular binary classification of NSCLC: miR-375 is a potential biomarker to differentiate SQCC from ADCC in Indian NSCLC patients

Sangeet Bhaumik, Firoz Ahmad and Bibhu Ranjan Das*

Abstract

Background: Current developments in the targeted therapies of non-small-cell lung carcinoma demand accurate classification to dodge the adverse drug response and to yield maximum therapeutic outcome. Accurate classification depends on the classical hematoxylin and eosin staining and immunohistochemistry techniques. In selected critical cases, inter-observer variability, lack of standardization, tumor heterogeneity, and degree of differentiation makes it difficult to classify NSCLC. During the last decade, microRNAs (miRNAs) have been proven to be a promising biomarker in the field of oncology from diagnosis to therapy. The present study developed a binary classification method based on the expression of three well-known miRNAs: miR-205, miR-196b, miR-375, since it is the most demanding criteria to the clinicians trying to provide better therapy to the patients.

Methods: Quantitative real-time polymerase chain reaction was performed for 90 NSCLC samples. Receiver-Operator Characteristic Curve and Discriminant Function Analysis was done to classify the NSCLC. A discriminant formula was developed to calculate the Z-score of miR-375 ($Z3 = -0.637 + (0.439 \times NCt_{miR\text{-}375}) + (-0.245 \times NCt_{miR\text{-}21})$).

Results: The miR-375 classified NSCLC into SQCC and ADCC with higher accuracy. miR-375 appeared to differentiate SQCC from ADCC accurately in the test and validation set, signifying a sensitivity and specificity of 96.7% and 93.1%, respectively.

Discussion: miR-375 is over-expressed in ADCC and suppressed in SQCC. This evidence accentuated the oncogenic and tumor suppressor nature in ADCC and SQCC respectively.

Conclusion: miR-375 was proven to be the prominent biomarker of accurate NSCLC classification. The current study developed a molecular binary classification method in adjunctive of IHC which will help the clinicians in better classifying NSCLC and designing therapy.

Keywords: Carcinoma, Non-small-cell lung; Squamous cell carcinoma, Adenocarcinoma, MicroRNAs

Background

Lung cancer (LC) is the leading cause of cancer-related death worldwide and contributed to 12.9% of all new cancer cases diagnosed in 2012 [1]. The most common form of lung cancer is the Non-Small-Cell Lung Cancer (NSCLC), which accounts for 85–90% of all lung carcinoma [2]. There are two subtypes of NSCLC: Adenocarcinoma (ADCC), and Squamous Cell Carcinoma (SQCC). Traditional therapy considered NSCLC as a single homogeneous entity. However, targeted therapy changed this idea. The therapeutic response of ADCC and SQCC was different to chemo / targeted therapies. For instance, EGFR (epidermal growth factor receptor) tyrosine kinase inhibitor therapy showed promising outcomes only for those who had mutated EGFR allele, prevalent in ADCC. Anaplastic lymphoma kinase (Alk) translocation positive ADCCs, showed high sensitivity to crizotinib therapy [3]. The clinical outcome of cisplatin/ pemetrexed therapy to ADCC patients was far better

* Correspondence: brdas@srl.in
Research and Development, SRL LTD, Plot No. 1, Prime Square Building, S.V. Road, Goregaon West, Mumbai 400062, India

BioMed Central

© The Author(s). 2017 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

than cisplatin/gemcitabine therapy [4]. Due to angiogenic heterogeneity and proliferative behavior, SQCC poses a threat in the clinic and there is a recognized high need for new therapies for SQCC [5]. Hence, accurate classification of NSCLC became mandatory to facilitate a better outcome.

Till date, NSCLC classification is majorly dependent on hematoxylin-eosin staining (HE) and immunohistochemistry (IHC). In 33% of critical cases due to tumor cell paucity in the absence of tissue architecture, or poor differentiation an adjunctive molecular assay side by side with IHC will help clinicians to classify NSCLC better. 10–30% NSCLC tends to be left as Not Otherwise Specified (NOS) even after a careful examination under the microscope by expert pathologists, due to the difficulties in translating all the morphogenic features available from the cytology/biopsy samples [6]. Hence an adjunctive method is necessary to confirm the NSCLC classification in selected cases.

During last few decades, scientists and clinicians have been hunting for molecular markers which could provide an unbiased classification of NSCLC. It has been long believed and recently proven that miRNAs express differentially in several types of cancers and promising biomarker for NSCLC classification. One study successfully identified ADCC and SQCC by using miR-205 expression level in the western population [7], and another study established a panel of three microRNAs (miR-205, miR-196b, and miR-375) to execute the same function in the Japanese population [8]. Interestingly, the idea of miR-205 as a promising biomarker to identify SQCC and ADCC in any ethnic group was refuted by a study from Italy [9]. It is possible that the expression pattern of a particular miRNA can vary from one ethnicity to another. Hence, it was very important to understand that which microRNAs are the best suitable markers for NSCLC sub-classification in Indian NSCLC patients. In the present study hypothesized that expression of miR-205, miR-196b, and miR-375 might not be the same as other populations, and so the biomarker for NSCLC classification may vary. To the best of our knowledge, this is the first study to evaluate the role of miR-205, miR-196b, and miR-375 to classify NSCLC in Indian lung cancer patients.

Methods

The present study was conducted at the Research and Development Division of SRL Ltd., Mumbai, India. The study included 90 formalin-fixed-paraffin-embedded (FFPE) tumor samples from lung lesions, bronchial

Table 1 Clinicopathological details of 90 NSCLC patients

Clinicopathological features	Total samples, *n* = 90	n (%)	ADCC (%)	SQCC (%)	*P* Value
Total		90(100)	44 (48.9)	46 (51.1)	
Age	Median age (range)	61(34–85)	60 (38–77)	62.5 (34–85)	
	≥60	50 (55.6)	23 (52.3)	27 (58.7)	0.54
	<60	40 (44.4)	21 (47.7)	19 (41.3)	
Gender	Male	71 (78.9)	30 (68.2)	41 (89.1)	**0.023**
	Female	19 (21.1)	14 (31.8)	5 (10.9)	
Differentiation grade[a]	PD	44 (48.8)	21 (47.7)	23 (50)	0.827
	MD	33 (36.7)	17 (38.6)	16 (34.8)	
	WD	14 (15.5)	6 (13.6)	7 (15.2)	
Training Set	Number of cases	30 (100)	15 (50)	15 (50)	0.361
	Male	24 (80)	11 (73.3)	13 (86.7)	
	Female	6 (20)	4 (26.7)	2 (13.3)	
Differentiation grade	PD	13 (43.3)	8 (53.3)	5 (33.3)	0.47
	MD	11 (36.7)	4 (26.7)	7 (46.7)	
	WD	6 (20)	3 (20)	3 (20)	
Validation set	Number of cases	60 (100)	29 (48.3)	31 (51.7)	
	Male	47 (78.3)	19 (65.5)	28 (90.3)	**0.02**
	Female	13 (21.7)	10 (34.5)	3 (9.7)	
Differentiation grade	PD	31 (51.7)	13 (44.8)	18 (58)	
	MD	22 (36.7)	13 (44.8)	9 (29)	0.99
	WD	7 (11.6)	3 (10.4)	4 (13)	

[a]*PD* poorly differentiated, *MD* moderately differentiated, *WD* well differentiated
Bold data signify the $p < 0.05$

biopsies, etc. derived from lung cancer patients. The study is by the Declaration of Helsinki and approved by Institute's ethical committee. Treatment and outcome were not analyzed. Table 1 depicts the details of the clinical characteristics of all patients. Histologic diagnosis of all the cases were made by two independent histopathologists based on the most recent WHO classification lung tumors [10]. Briefly, the HE-stained slides were initially looked for the absence of keratinization or intercellular bridges and/or Thyroid transcription factor 1 (TTF1) positive staining to identify ADCC (Fig. 1). Similarly, the presence of intercellular bridges and/or P63 positive staining were the key criteria for diagnosing SQCC (Fig. 1). Leica auto Steiner was used for HE staining, and IHC was done manually. In the case of any disagreement, the consensus was achieved by reviewing the cases with third independent histopathologists. Schematic diagram of experimental strategy is shown in Fig. 2.

RNA isolation

Total RNA including small RNA was extracted using mirVana™ miRNA Isolation Kit (Ambion) as per the manufacturer's instructions. Before RNA isolation, separate HE slides were reviewed by a pathologist to assure 50% tumor content is suitable for RNA isolation. At least 10 mg tissue was processed for RNA isolation of NSCLC, tumor differentiation grade, as suitable for RNA extraction.

qRT-PCR

The relative levels of miR-205, miR-196b, miR-375, and miR-21 were quantified by quantitative reverse-transcriptase-polymerase-chain-reaction (qRT-PCR) using TaqMan® MicroRNA Assays (Part Number: 4,427,975). U6 snRNA was the normalized and positive control. All the miRNAs were amplified in duplicate. A relative level was calculated for each miRNAs using ΔCt (Cycle threshold) method, concerning U6 snRNA.

Data analysis and statistics

XLSTAT software was used to calculate Mann-Whitney test, Receiver-Operator Characteristic (ROC) curve. The IBM SPSS software (ver. 20.0) was used to perform the discriminant analysis. The average Ct value of the duplicates ($AvgCt_{miR\text{-}205}$, $AvgCt_{miR\text{-}196b}$, $AvgCt_{miR\text{-}375}$) was calculated, and replicates with Ct differing by >1 cycle from each other were excluded. If the $AvgCt_{U6}$ was not within 40 cycles, the assay was repeated. The normalized Ct (NCt) was calculated by subtracting $AvgCt_{U6}$ from the $AvgCt_{miR\text{-}205}$, $AvgCt_{miR\text{-}196b}$, $AvgCt_{miR\text{-}375}$, and $AvgCtmiR_{\text{-}21}$. Sample score for each miRNA was obtained by using the formula = $AvgCt_{miR} - [(AvgCt_{miR\text{-}21} + AvgCt_{U6})/2] = NCt_{miR} - (NCt_{miR\text{-}21}/2)$ [11]. In the Mann-Whitney test, box and whisker plot analysis were done to observe the difference in sample score between SQCC and ADCC for each miRNA (Fig. 3). ROC curves were plotted to determine the cut-off values for each miRNA (Fig. 4). Cut-off values differentiating SQCCs and ADCCs pointed out the youden index (sensitivity + specificity − 1) in ROC plot. The Z score was calculated using Discriminant Function Analysis (DFA), which is more profound and unbiased than ROC analysis [8]. A discriminant score is based on the weighted combination of the independent variables. Discriminant score = a + b1X1 + b2X2 + ……. + bnXn. "X" is the predictor and "b" is the discriminant

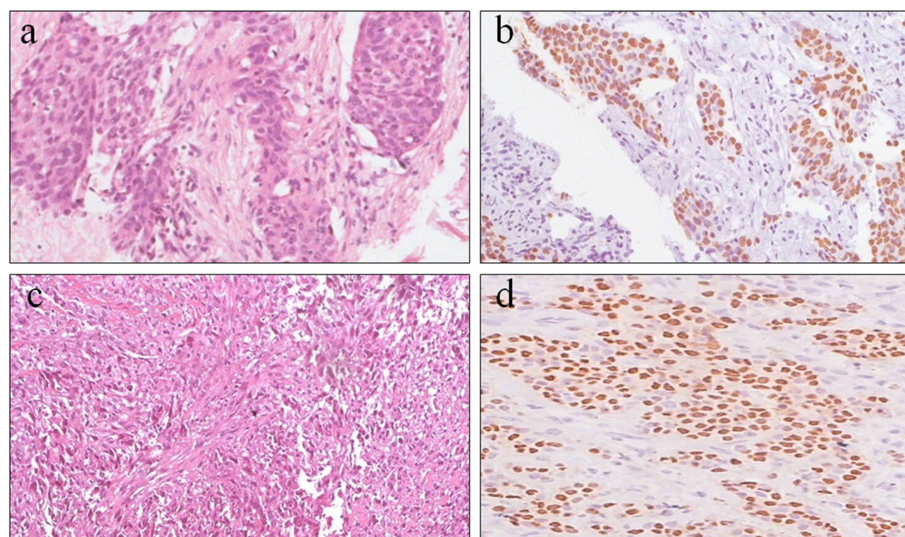

Fig. 1 HE and IHC staining pattern in ADCC and SQCC. **a** HE staining of ADCC, **b** TTF-1 positive staining in ADCC, **c** HE staining of SQCC, **d** P63 positive staining in SQCC

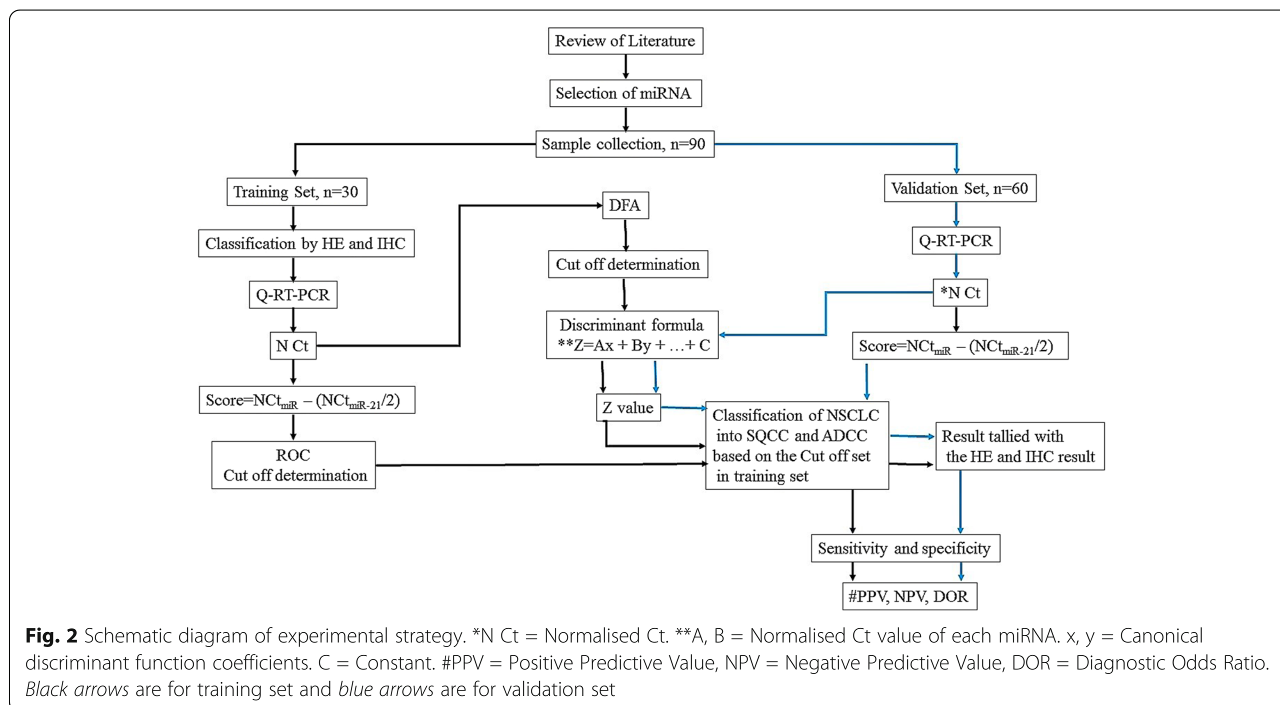

Fig. 2 Schematic diagram of experimental strategy. *N Ct = Normalised Ct. **A, B = Normalised Ct value of each miRNA. x, y = Canonical discriminant function coefficients. C = Constant. #PPV = Positive Predictive Value, NPV = Negative Predictive Value, DOR = Diagnostic Odds Ratio. *Black arrows* are for training set and *blue arrows* are for validation set

coefficient and "a" is constant. In the current study, the predictor is the normalized Ct value of a particular miRNA. The function at group centroids was the identifier of cut-off value. The discriminant formula and generalized cut-off value were generated with the help of SPSS software (ver. 20.0). For example – If the Ct value of a miR-375 is 25 and Ct value of miR-21 and U6-SnRNA is 22 and 27 respectively then the normalized Ct of miR-375 = (25–27) = −2 and normalized Ct of miR-21 = (22–27) = −5. According to the discriminant formula developed for miR-375 is Z3 = (−0.637) + (0.439 x NCtmiR-375) + (− 0.245 x NCtmiR-21) = (−0.637) + {0.439 x (−2)} + {− 0.245 x (−5)} = (−0.637) + (−0.878) + (1.225) = −0.29. This score corresponds to ADCC. Table 2 represents the discriminant formulae for each miRNA. Positive Predictive Value (PPV), Negative Predictive Value (NPV), and Diagnostic Odds Ratio (DOR) were calculated to measure the accuracy of this binary classification of NSCLC using the expression pattern of miR-375.

Results

Clinical characteristics of NSCLC patients

The present study analyzed 90 tumor tissues from NSCLC patients. Table 1 represents the clinical characteristics of all patients, and male NSCLC cases (71, 78.9%) were higher than those of female (19, 21.1%). There was a significant trend of increased number of SQCC in male patients (41, 89.1%) in comparison with female patients (5, 10.9%). The frequency of NSCLC cases was higher in older patients (≥60 years, $n = 50$, 55.6%) than those of younger patients (<60 years, $n = 40$, 44.4%). The median age of the cohort was 61 years, ranging from 34 to 85 years. Distributions across various

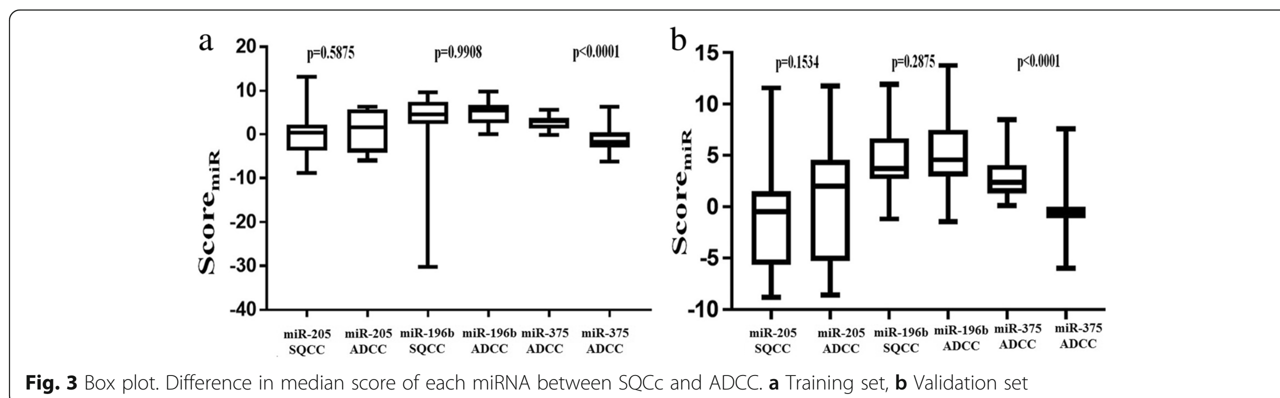

Fig. 3 Box plot. Difference in median score of each miRNA between SQCc and ADCC. **a** Training set, **b** Validation set

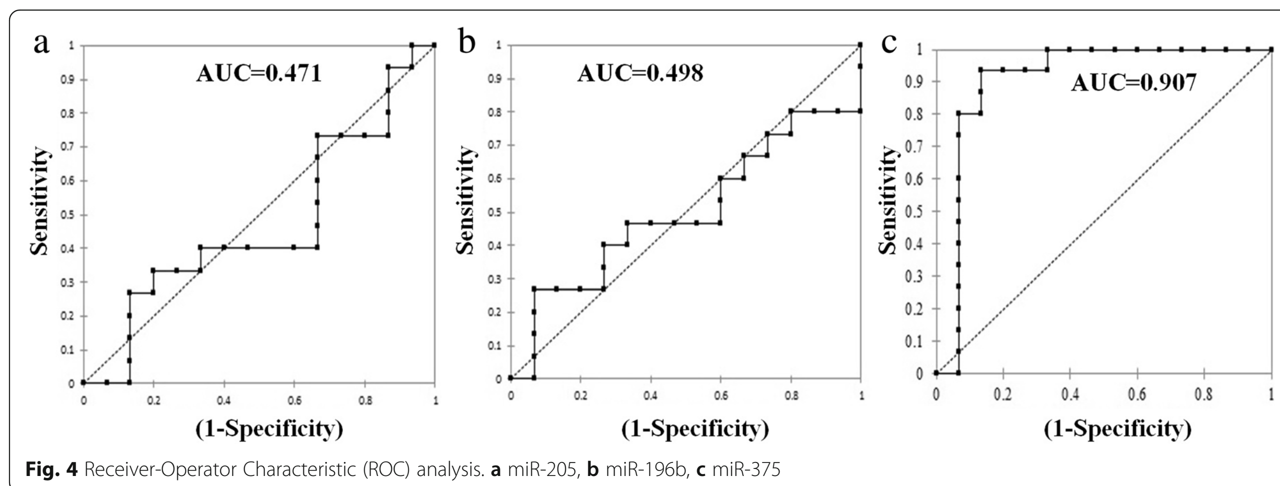

Fig. 4 Receiver-Operator Characteristic (ROC) analysis. **a** miR-205, **b** miR-196b, **c** miR-375

subtypes were as follows: ADCC: 44 (48.9%) with median age 60 years ranging from 38 to 77 years, SQCC: 46 (51.1%) with median age 62.5 years ranging from 34 to 85 years (Table 1). Histological diagnosis was done on all 90 cases, TTF1 and P63 staining was done on 30 adenocarcinomas and 8 squamous cell carcinomas respectively.

Most of the tumor samples were poorly differentiated (PD) (44, 48.9%), and 33 (36.7%) were moderately differentiated (MD), while only 14 (15.5%) cases were well differentiated (WD). The sample cohort was divided into a training set ($n = 30$) and a validation set ($n = 60$) (Table 1).

Optimization of miR-205, miR-196b, miR-375 to classify NSCLC with the ROC and DFA

Initially, the expression levels of miR-205, miR-196b, and miR-375, miR-21 and U6 snRNA of 30 samples (Training set) were measured in duplicate to investigate the appropriate miRNAs for NSCLC subtyping. There was a trend of increased Ct value of miR-375 in SQCC compared to ADCC ($p = 0.0478$, Table 3), while the miR-205 and miR-196b showed no significant difference between Ct value among SQCC and ADCC (Table 3). There was no significant difference between miR-21 Ct value among SQCC and ADCC ($p = 0.965$). Average Ct of miR-21 in ADCC was 20.66 and in SQCC was 20.58. Ct value of each miRNA was normalized by subtracting the Ct value of U6 snRNA. The normalised Ct value was used to calculate the sample score (Sample score = $AvgCt_{miR} - [(AvgCt_{miR\text{-}21} + AvgCt_{U6})/2] = NCt_{miR} - (NCt_{miR\text{-}21}/2)$). As all the sample scores from both groups were independent and ordinal, the Mann-Whitney test revealed that there was a significant difference between the median score of SQCC and ADCC for miR-375 (Median of SQCC = 0.4745, Median of ADCC = 1.682, $p < 0.0001$) (Table 3, Fig. 3a). There were no significant differences observed between the SQCC and ADCC for miR-205 ($p = 0.5875$) and miR-196b ($p = 0.9908$) (Table 3, Fig. 3a). ROC curve was generated for each miRNA, and the Area Under Curve (AUC) of the three miRNAs were as follows: miR-205 – 0.471, miR-196b – 0.498, miR-375 – 0.907 (Fig. 4). As the AUC represented the accuracy of the test, it clearly showed that only miR-375 provided better accuracy, as compared to miR-205 and miR-196b. Also, a cutoff was set at 0.584 for miR-375 with a test sensitivity of 93.33% and specificity of 86.67% (Table 3), and the

Table 2 Discriminant formulae[a]

Sr.No.	miRNA	Discriminant function
1.	miR-205, miR-21	$Z1 = 2.233 - (0.132 \times {}^{b}NCt_{miR\text{-}205}) + (0.261 \times NCt_{miR\text{-}205})$
2.	miR-196b, miR-21	$Z2 = 1.987 - (0.073 \times NCt_{miR\text{-}196b}) + (0.191 \times NCt_{miR\text{-}21})$
3.	miR-375, miR-21	$Z3 = (-0.637) + (0.439 \times NCt_{miR\text{-}375}) + (-\ 0.245 \times NCt_{miR\text{-}21})$
4.	miR-205, miR-196b, miR-21	$Z4 = 2.149 - (0.108 \times NCt_{miR\text{-}205}) - (0.041 \times NCt_{miR\text{-}196b}) + (0.191 \times NCt_{miR\text{-}21})$
5.	miR-205, miR-375, miR-21	$Z5 = (-0.495) - (0.03 \times NCt_{miR\text{-}205}) + (0.432 \times NCt_{miR\text{-}375})\ \text{-}(0.216 \times NCt_{miR\text{-}21})$
6.	miR-375, miR-196b, miR-21	$Z6 = (-0.5982) + (0.438 \times NCt_{miR\text{-}375}) - (0.034 \times NCt_{miR\text{-}196b}) - (0.234 \times NCt_{miR\text{-}21})$
7.	miR-205, miR-196b, miR-375, miR-21	$Z7 = (-0.514) - (0.016 \times NCt_{miR\text{-}205}) - (0.03 \times NCt_{miR\text{-}196b}) + (0.434 \times NCt_{miR\text{-}375}) + (0.22 \times NCt_{miR\text{-}21})$

[a]Discriminant formulae were developed with the help of IBM SPSS software
[b]*NCt* normalised Ct ($Ct_{miR} - Ct_{U6snRNA}$)

Table 3 t Test and Mann-Whitney test

	Training set	miR-205	miR-196b	miR-375
[a]t-test of Ct value	SQCC: Mean ± SEM, $n = 15$	25.72 ± 1.392	28.34 ± 2.181	28.5 ± 1.059
	ADCC: Mean ± SEM, $n = 15$	27.75 ± 1.528	31.59 ± 0.9288	25.47 ± 1.013
	p Value	0.3340	0.1814	0.0478
	Validation set			
	SQCC: Mean ± SEM, $n = 31$	24.72 ± 1.097	30.39 ± 0.6117	28.96 ± 0.717
	ADCC: Mean ± SEM, $n = 29$	26.55 ± 1.056	31.05 ± 0.8598	25.64 ± 0.8151
	p Value	0.2351	0.5293	0.0033
[b]Mann-Whitney test of sample score	Training set			
	SQCC: Median, $n = 15$	0.4745	4.619	3.082
	ADCC: Median, $n = 15$	1.682	5.462	−1.694
	p Value	0.5875	0.9908	<0.0001
	Validation set			
	SQCC: Median, $n = 31$	−0.4515	3.709	2.37
	ADCC: Median, $n = 29$	1.976	4.573	−0.413
	p Value	0.1534	0.2875	<0.0001

[a]t Test was done to check the significant difference between the mean Ct of SQCC and ADCC for the training set and validation set. [b]Mann-Whitney test was done to check the significant difference between the median sample score of SQCC and ADCC for the training set and validation set

sample was considered SQCC if the score ≥ cut off. If the score < cut off, the sample was ADCC. The cut-off value of miR-205 was 4.264. Out of 15 SQCC, only 2 (13.33%) were successfully categorized and 10 (66.67%) out of 15 ADCC were successfully classified by miR-205 by miR-205 (Table 4). The cut-off value of miR-196b was 7.511. Out of 15 SQCC, only 4 (26.67%) were successfully categorized and 14 (93.33%) out of 15 ADCC were successfully classified by miR-196b by miR-205 (Table 4) The sample sorted number versus $\mathrm{Score_{miR}}$ plots corroborated the findings of ROC analysis (Fig. 5). Both the groups (SQCC and ADCC) were evenly distributed on the both sides of the cut-off line in the case of miR-375 (Fig. 5c), but not in the case of miR-205 and miR-196b (Fig. 5a, b). Discriminant analysis calculated the z-value using normalized Ct of single and multiple miRNAs, which revealed 93.3% test sensitivity and 86.7% test specificity for miR-375 (Table 5, Fig. 6a, c). Sensitivity and specificity of miR-205 and miR-196b were very poor compared to miR-375 (Table 5). ROC and DFA both showed similar PPV (87.5%, 95% CI = 61.65%–98.45%), NPV (92.86%, 95% CI = 66.13%–99.82%), and DOR (91, 7.3–1126.9) (Table 6).

t Test of Ct value in validation set ($n = 60$) replicated the result found in training set (miR-205: $p = 0.2351$, miR-196b: $p = 0.5293$, miR-375: $p = 0.0033$). Similar results were obtained for the Mann-Whitney test (miR-205: $p = 0.1534$, miR-196b: $p = 0.2875$, miR-375: $p < 0.0001$) (Table 3, Fig. 3b). The cut-off value, established by the ROC and DFA for each miRNAs, were used to differentiate SQCC from ADCC in the validation set ($n = 60$), which provided a test sensitivity of 96.7% and specificity of 93.1% for miR-375 (Fig. 6b, d). The validation set produced a similar outcome as the training

Table 4 Sensitivity (%) and specificity (%) Calculated from ROC curves

	Variables	miR-205, miR-21	miR-196b, miR-21	miR-375, miR-21
ROC	Cut-off[a]	4.264	7.511	0.584
	AUC	0.471	0.498	0.907
	p Value	0.789	0.983	<0.0001
	Sensitivity (Training set)	13.33%	26.67%	93.3%
	Specificity(Training set)	66.67%	93.33%	86.7%
	Sensitivity (Validation set)	9.67%	12.9%	96.7%
	Specificity (Validation set)	72.41%	75.86%	93.1%

[a]Cut-off value is the score that correspond to the youden index (sensitivity + specificity-1) in ROC curve. Sample was positive if the score ≥ cut off. Sensitivity (%) was determined as [true positive / (true positive + False negative)] and specificity (%) was [True negative / (True negative + False positive)]. True positive: Samples which were diagnosed as SQCC by the histopathologists and with score ≥ cut off. False negative: Samples which were characterised as SQCC but score < cut off. True negative: Samples which were characterised as ADCC by histopathologists and score < cut off. False negative: Samples which were initially classified as ADCC but score ≥ cut off

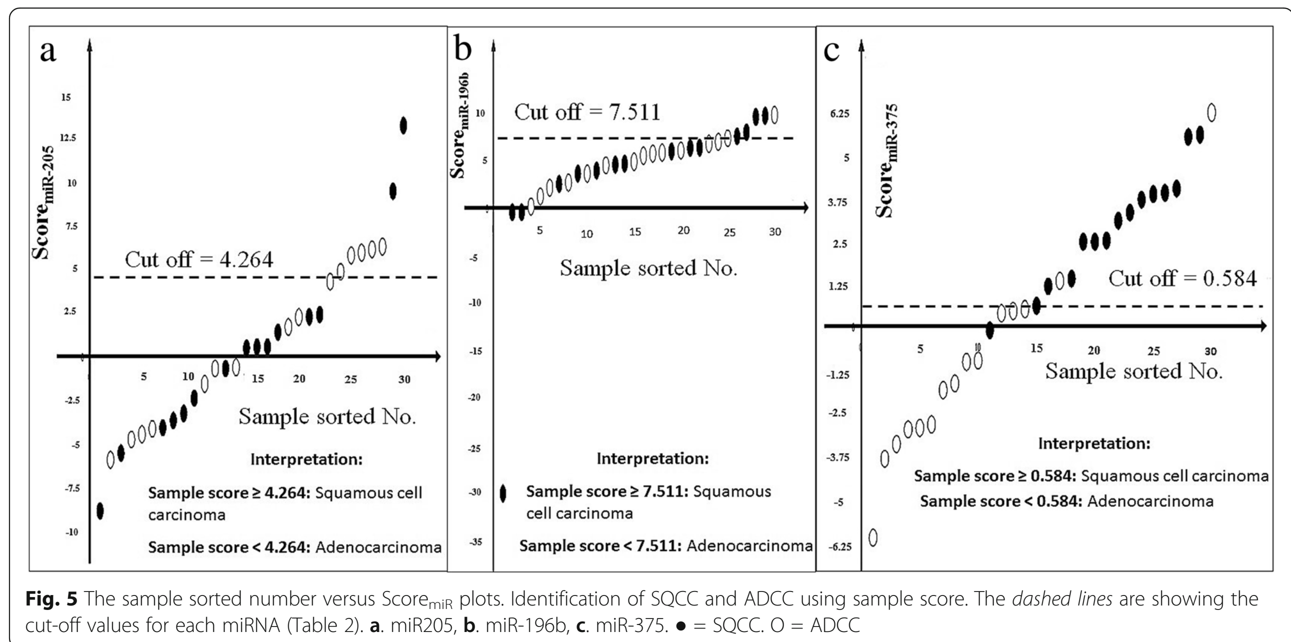

Fig. 5 The sample sorted number versus $Score_{miR}$ plots. Identification of SQCC and ADCC using sample score. The *dashed lines* are showing the cut-off values for each miRNA (Table 2). **a**. miR205, **b**. miR-196b, **c**. miR-375. ● = SQCC. O = ADCC

set. Interestingly, DOR for miR-375 (405, 95% CI = 34.7–4722.3) was increased drastically in the validation set compared to the training set. Also, PPV (93.75% 95% CI = 79.19%–99.23%) and NPV (96.43%, 95% CI = 81.65%–99.91%) (Table 6) showed an increased trend for both ROC and DFA observed in the validation set compared to the training set. 44 (95.65%) SQCC were identified by miR-375 out of 46 SQCC and 40 (90.90%) ADCC were identified by miR-375 out of 44 ADCC. According to sample score method, overall 40 out of 44 (91%) were classified as ADCC and 44 out of 46 (95.7%) were correctly classified as SQCC using miR-375 as a marker (Table 7).

Discussion

Initially, miR-375 was discovered as a key player in glucose metabolism and insulin secretion [12, 13], and was later found to be associated with tumorigenesis and deregulated in several types of cancer. Growing evidence suggested that it was down-regulated in esophageal SQCC, Barrett's carcinoma, glioma, colorectal cancer, hepatocellular carcinoma, lung cancer, and associated with advanced-stage metastasis and unfavorable treatment outcome [14–17]. Also, evidence from the previous study suggested that miR-375 plays a dual role in NSCLC as it was over-expressed in ADCC and suppressed in SQCC, proving the oncogenic nature in ADCC and tumor suppressor in SQCC [18]. This evidence accentuated the principal role of miR-375 tumorigenesis, and the current study found significant up-regulation of miR-375 in ADCC and down-regulation in SQCC, which is consistent with the previous findings. In the present study, a diagnostic method was developed based on the growing evidence which highlighted the significance of miR-205, miR-196b, miR-375 as molecular markers for the binary classification of NSCLC. Recent evidence on miR-375 expression in SQCC and ADCC were consistent with the present study. Lebanony et al. [7] investigated and developed a diagnostic method of NSCLC classification based on

Table 5 Sensitivity[a] (%) and specificity (%) Calculated from discriminant function analysis

	Variables	miR-205, miR-21	miR-196, miR-21	miR-375, miR-21	miR-205, miR-196b, miR-21	miR-205, miR-375, miR-21	miR-375, miR-196, miR-21	miR-205, miR-196b, miR-375, miR-21
Training set	Sensitivity	46.7%	53.3%	93.3%	46.7%	86.7%	86.7%	80%
	Specificity	66.7%	66.7%	86.7%	66.7%	86.7%	86.7%	86.7%
Validation set	Sensitivity	54.8%	48.3%	96.7%	51.6%	96.7%	96.7%	93.5%
	Specificity	65.5%	75.8%	93.1%	68.9%	93.1%	93.1%	93.1%

[a]Sensitivity (%) was determined as [true positive / (true positive + False negative)] and specificity (%) was [True negative / (True negative + False positive)]. Cut-off value = "Zero". If the Z score was positive, sample was designated as SQCC and negative sample was designated as ADCC. Z score was calculated with the help of the formulas mentioned in the Table 2. True positive: Samples which were diagnosed as SQCC by the histopathologists and with positive Z score. False negative: Samples which were characterised as SQCC but with negative Z score. True negative: Samples which were characterised as ADCC by histopathologists and with negative Z score. False negative: Samples which were initially classified as ADCC but with the positive Z score

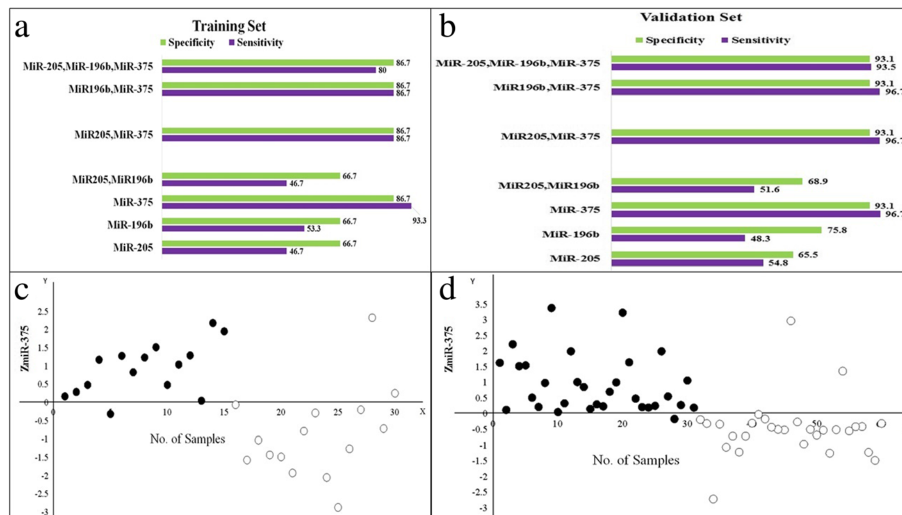

Fig. 6 Comparison of sensitivity and specificity obtained by DFA. **a** Sensitivity and specificity of the training set. **b** Sensitivity and specificity of validation set. **c** Graphical representation of training set DFA for mir-375. Z value of SQCCs is positive and negative for ADCC when Z-score is plotted on the Y axis and sample number on the X axis. **d** Graphical representation of validation set DFA for mir-375. Z value of SQCCs is positive and negative for ADCC when Z-score is plotted on the Y axis and sample number on the X axis. ● = SQCC. O = ADCC

miR-205, and Patnaik et al. [19] proved the diagnostic ability of miR-205 and miR-375. Later, the classifying role of this miR-205 was dismissed by another study [9]. Hamamoto et al. [8] emphasized the diagnostic role of a panel of miRNAs including miR-205, miR-196b, miR-375. The role of miR-205 as an individual biomarker for NSCLC classification was arguable as contradictory evidence were reported earlier [9]. The present study developed a diagnostic method based on differential expression of miR-375 to classify NSCLC into SQCC and ADCC. It not only included ROC and DFA but also calculated PPV, NPV, and DOR to achieve more comprehensive and excellent accuracy, enabling clinicians to better identify SQCC and ADCC with the help of miR-375 and with the additional contribution by miR-21. Interestingly in the present study, Ct value of miR-21 did not vary significantly among SQCC and ADCC (Training set: $p = 0.9658$ Validation set: $p = 0.4367$). miR-375 plays a key role in cancer development and Epithelial Mesenchymal Transition (EMT) pathway by targeting ELAV-like neuron-specific RNA binding protein 4 (ELAVL4), XIAP associated factor 1 (XAF1), monocyte to macrophage differentiation-associated (MMD), and chloride channel accessory 2 (CLCA2) genes. ELAVL4 regulates the radiosensitivity of the cancerous cells [20], and it expresses in neuritis and lung tissue [21]. XAF1 acts as a tumor suppressor in colon cancer [22]. Another target, the CLCA2 protein expression status, could be an excellent way to identify primary SQCC and ADCC of the lung [23]. The previous study reported that the major target gene of miR-375 were the key members of calcium, insulin, Jak-STAT, MAPK, mTOR, PPAR, TGF-beta and Wnt pathways [24].

Table 4 and Table 5 shows the sensitivity and specificity obtained by ROC and DFA for miR-205 and miR-196b. The sensitivity and specificity of these two miRNAs were considerably poor compared to miR-375. Interestingly, there was an increasing pattern of

Table 6 PPV[a], NPV[b] AND DOR[c] of miR-375 and miR-21 by ROC and DFA

	Variables	ROC	95% CI	DFA	95% CI
Training set	PPV	87.5%	61.65% - 98.45%	87.5%	61.65% - 98.45%
	NPV	92.86%	66.13% - 99.82%	92.86%	66.13% - 99.82%
	DOR	**91**	7.3 - 1126.9	**91**	7.3 - 1126.9
Validation set	PPV	93.75%	79.19% - 99.23%	93.75%	79.19% - 99.23%
	NPV	96.43%	81.65% - 99.91%	96.43%	81.65% - 99.91%
	DOR	**405**	34.73 - 4722.3	**405**	34.73 - 4722.3

[a]*PPV* positive predictive value, *PPV* (True positive / True positive + False positive). [b]*NPV* negative predictive value, *NPV* (True negative / True negative + False negative). [c]*DOR* diagnostic odds ratio, *Diagnostic Odds Ratio* [(True positive / False positive) / (False negative / True negative)]
Bold data signify DOR>1 which is the indication of better test performance

Table 7 Comparison of miR-375 and morphoimmunological evaluation according to the sample score

	miR-375	miR-375	Morphoimmunological evaluation
Training set	ADCC	13 (86.67%)	15
	SQCC	14 (93.33%)	15
Validation set	ADCC	27 (93.1%)	29
	SQCC	30 (96.77%)	31
Total	ADCC identified by miR-375		40/44 (91%)
	SQCC identified by miR-375		44/46 (95.7%)

sensitivity and specificity in the DFA when miR-205 or miR196b is associated with miR-375. The higher accuracy of miR-375 dragged the outcome towards higher sensitivity and specificity since miR-205 and miR-196b showed poor outcome when considered individually. In NSCLC, miR-205 is frequently amplified and differentially expressed [25, 26]. Although the biologic role of this miRNA in the development of SQCC was unclear, the previous study documented frequent up-regulation in SQCC [27]. Up-regulation of this miRNA inhibited the expression of PTEN, PHLPP2 and activated the AKT/FOXO3a, AKT/mTOR pathway that may lead to tumor cell proliferations, enhanced blood vessels formation and took part in EMT pathway [28]. However, in the current study, there was no significant difference in the expression level of miR-205 in both histologic subtypes (Training set: $p = 0.3340$, Validation set: $p = 0.2351$, Table 3). miR-196b was proven to be involved in disease in cancer pathogenesis [29]. Deregulation of this miRNA was common in multiple cancers including gastric cancer, glioblastoma, and breast cancer. The role of miR-196b in lung cancer remained an enigma. However, over-expression of miR-196b inhibits HOXA9 and induces the invasive properties in NSCLC cells [30].

Conclusions

In conclusion, the present study trained and validated a binary diagnostic assay to identify SQCC and ADCC. miRNA could be used as an adjunctive diagnostic criterion in difficult cases where immunomorphology alone is inconclusive. A combine approach will certainly improve the overall diagnostic yield and better classification. miR-375 was proven to be the molecular marker of binary classification of NSCLC with 96.7% sensitivity and 93.1% specificity. ROC analysis set the cut-off value of 0.584 and considered SQCC if the Z score was positive. Interestingly, DFA for the training set and the validation set replicated the same result of ROC analysis, which only strengthens the diagnostic ability of miR-375. Further study is required on critical cases to validate the test. The current study is first in India to report the diagnostic ability of three miRNAs (miR-205, miR-196b, miR-375). In conclusion, this is an adjunctive assay which clinicians can use to confirm the classification of NSCLC side by side with HE and IHC in selected cases. However, it is also true that molecular approach may at times misclassify the tumor in a fraction of cases. Hence it will be a good idea to use this approach in conjunction with other lab tests. Similarities and dissimilarities of the present findings with findings of other research work may be attributed to the influence of differences in ethnic origins as well as differential environmental exposure to unknown carcinogenic agents.

Abbreviations

ADCC: Adenocarcinoma; AKT/FOXO3a: AKT/Forkhead box O3a; AKT/mTOR: AKT/mechanistic target of rapamycin; ALK: Anaplastic lymphoma kinase; AUC: Area under curve; Avg: Average; CI: Confidence interval; CLCA2: Chloride channel accessory 2; DFA: Discriminant function analysis; DOR: Diagnostic odds ratio; EGFR: Epidermal growth factor receptor; ELAVL4: ELAV like neuron-specific protein 4; EMT: Epithelial-mesenchymal transition; FFPE: Formalin-fixed paraffin embedded; HE: hematoxylin-eosin; HOXA9: Homeobox protein Hox – A9; IHC: Immunohistochemistry; Jak – STAT: Janus kinase – Signal transducer and activator of transcription; LC: Lung cancer; MAPK: Mitogen-activated protein kinases; MD: moderately differentiated; miR: miRNA; miRNA: microRNA; MMD: Monocyte to macrophage differentiation; NCt: Normalised Ct; NOS: Not otherwise specified; NPV: Negative predictive value; NSCLC: Non – small – cell lung cancer; PD: poorly differentiated; PHLPP2: PH domain leucine-rich-repeat-containing protein phosphatase 2; PPAR: Peroxisome proliferator – activated receptor; PPV: Positive predictive value; qRT-PCR: Quantitative reverse transcriptase – polymerase chain reaction; RNA: Ribonucleic acid; ROC: Receiver operator characteristic; snRNA: Small nuclear RNA; SQCC: Squamous cell carcinoma; TGF-beta: Transforming growth factor beta; TTF1: Thyroid transcription factor 1; WD: Well differentiated; Wnt: Wingless – related integration site, Phosphatase and tensin homolog; XAF1: XIAP associated factor 1

Acknowledgements

Authors are grateful to the management of SRL LTD for providing the necessary infrastructure for the present research.

Funding

There was no funding for this study.

Availability of data and materials

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Authors' contributions

SB participated in concept design, carried out the data acquisition or analysis and interpretation of data and manuscript writing. FA participated in concept design, carried out the data acquisition or analysis and interpretation of data and manuscript writing. BRD participated in concept design, carried out the data acquisition or analysis and interpretation of data and manuscript writing. All authors read and approved the final manuscript.

Ethics approval and consent to participate
The study is in accordance with the declaration of Helsinki, with written consent available from subjects. The "SRL – Ethics Committee" reviewed, discussed and approved the study with reference number – 30/2012.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

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

Received: 22 February 2017 Accepted: 20 June 2017
Published online: 25 August 2017

References
1. Chen W, Zheng R, Zeng H, Zhang S. Epidemiology of lung cancer in China. Thorac Cancer. 2015;6:209–15. doi:10.1111/1759-7714.12169.
2. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clin Proc. 2008;83:584–94. doi:10.4065/83.5.584.
3. Płużański A, Piórek A, Krzakowski M. Crizotinib in the treatment of non-small-cell lung carcinoma. Contemp Oncol (Poznań, Poland). 2012;16:480–4. doi:10.5114/wo.2012.32477.
4. Scagliotti GV, Parikh P, von Pawel J, Biesma B, Vansteenkiste J, Manegold C, et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. J Clin Oncol. 2008;26:3543–51. doi:10.1200/JCO.2007.15.0375.
5. Salgia R. Fibroblast growth factor signaling and inhibition in non-small cell lung cancer and their role in squamous cell tumors. Cancer Med. 2014;3:681–92. doi:10.1002/cam4.238.
6. Rossi G, Pelosi G, Barbareschi M, Graziano P, Cavazza A, Papotti M. Subtyping non-small cell lung cancer: relevant issues and operative recommendations for the best pathology practice. Int J Surg Pathol. 2013;21:326–36. doi:10.1177/1066896913489346.
7. Lebanony D, Benjamin H, Gilad S, Ezagouri M, Dov A, Ashkenazi K, et al. Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from nonsquamous non-small-cell lung carcinoma. J Clin Oncol. 2009;27:2030–7. doi:10.1200/JCO.2008.19.4134.
8. Hamamoto J, Soejima K, Yoda S, Naoki K, Nakayama S, Satomi R, et al. Identification of microRNAs differentially expressed between lung squamous cell carcinoma and lung adenocarcinoma. Mol Med Rep. 2013;8:456–62. doi:10.3892/mmr.2013.1517.
9. Del W, Cantaloni C, Cucino A, Girlando S, Silvestri M, Bragantini E, et al. miR-205 expression levels in Nonsmall cell lung cancer do not always distinguish Adenocarcinomas from Squamous cell carcinomas. Am J Surg Pathol. 2011;35:268–75. doi:10.1097/PAS.0b013e3182068171.
10. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization classification of lung Tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. J Thorac Oncol. 2015;10:1243–60. doi:10.1097/JTO.0000000000000630.
11. Bishop JA, Benjamin H, Cholakh H, Chajut A, Clark DP, Westra WH. Accurate classification of non-small cell lung carcinoma using a novel microRNA-based approach. Clin Cancer Res. 2010;16:610–9. doi:10.1158/1078-0432.CCR-09-2638.
12. Poy MN, Hausser J, Trajkovski M, Braun M, Collins S, Rorsman P, et al. miR-375 maintains normal pancreatic alpha- and beta-cell mass. Proc Natl Acad Sci U S A. 2009;106:5813–8. doi:10.1073/pnas.0810550106.
13. Avnit-Sagi T, Kantorovich L, Kredo-Russo S, Hornstein E, Walker MD. The promoter of the pri-miR-375 gene directs expression selectively to the endocrine pancreas. PLoS One. 4:e5033. https://doi.org/10.1371/journal.pone.0005033.
14. Li J, Li X, Li Y, Yang H, Wang L, Qin Y, et al. Cell-specific detection of miR-375 downregulation for predicting the prognosis of esophageal squamous cell carcinoma by miRNA in situ hybridization. PLoS One. 8:e53582. https://doi.org/10.1371/journal.pone.0053582.
15. Leidner RS, Ravi L, Leahy P, Chen Y, Bednarchik B, Streppel M, et al. The microRNAs, MiR-31 and MiR-375, as candidate markers in Barrett's esophageal carcinogenesis. Genes Chromosomes Cancer. 2012;51:473–9. doi:10.1002/gcc.21934.
16. Chang C, Shi H, Wang C, Wang J, Geng N, Jiang X, et al. Correlation of microRNA-375 downregulation with unfavorable clinical outcome of patients with glioma. Neurosci Lett. 2012;531:204–8. doi:10.1016/j.neulet.2012.10.021.
17. Dai X, Chiang Y, Wang Z, Song Y, Lu C, Gao P, et al. Expression levels of microRNA-375 in colorectal carcinoma. Mol Med Rep. 2012;5:1299–304. doi:10.3892/mmr.2012.815.
18. Huang W, Hu J, Yang D, Fan X, Jin Y, Hou Y, et al. Two microRNA panels to discriminate three subtypes of lung carcinoma in bronchial brushing specimens. Am J Respir Crit Care Med. 2012;186:1160–7. doi:10.1164/rccm.201203-0534OC.
19. Patnaik S, Mallick R, Kannisto E, Sharma R, Bshara W, Yendamuri S, et al. MiR-205 and MiR-375 microRNA assays to distinguish squamous cell carcinoma from adenocarcinoma in lung cancer biopsies. J Thorac Oncol. 2015;10:446–53. doi:10.1097/JTO.0000000000000423.
20. Choi KJ, Lee J-H, Kim KS, Kang S, Lee Y-S, Bae S. Identification of ELAVL4 as a modulator of radiation sensitivity in A549 non-small cell lung cancer cells. Oncol Rep. 2011;26:55–63. doi:10.3892/or.2011.1260.
21. Abdelmohsen K, Hutchison ER, Lee EK, Kuwano Y, Kim MM, Masuda K, et al. miR-375 inhibits differentiation of neurites by lowering HuD levels. Mol Cell Biol. 2010;30:4197–210. doi:10.1128/MCB.00316-10.
22. Tu SP, Sun YW, Cui JT, Zou B, Lin MCM, Gu Q, et al. Tumor suppressor XIAP-Associated factor 1 (XAF1) cooperates with tumor necrosis factor-related apoptosis-inducing ligand to suppress colon cancer growth and trigger tumor regression. Cancer. 2010;116:1252–63. doi:10.1002/cncr.24814.
23. Shinmura K, Igarashi H, Kato H, Kawanishi Y, Inoue Y, Nakamura S, et al. CLCA2 as a novel immunohistochemical marker for differential diagnosis of squamous cell carcinoma from adenocarcinoma of the lung. Dis Markers. 2014;2014:619273. doi:10.1155/2014/619273.
24. Jin Y, Liu Y, Zhang J, Huang W, Jiang H, Hou Y, et al. The expression of miR-375 is associated with carcinogenesis in three subtypes of lung cancer. PLoS One. 10:e0144187. https://doi.org/10.1371/journal.pone.0144187.
25. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell. 2006;9:189–98. doi:10.1016/j.ccr.2006.01.025.
26. Markou A, Tsaroucha EG, Kaklamanis L, Fotinou M, Georgoulias V, Lianidou ES. Prognostic value of mature microRNA-21 and microRNA-205 overexpression in non-small cell lung cancer by quantitative real-time RT-PCR. Clin Chem. 2008;54:1696–704. doi:10.1373/clinchem.2007.101741.
27. Xing L, Todd NW, Yu L, Fang H, Jiang F. Early detection of squamous cell lung cancer in sputum by a panel of microRNA markers. Mod Pathol. 2010;23:1157–64. http://dx.doi.org/10.1038/modpathol.2010.111
28. Cai J, Fang L, Huang Y, Li R, Yuan J, Yang Y, et al. miR-205 targets PTEN and PHLPP2 to augment AKT signaling and drive malignant phenotypes in non-small cell lung cancer. Cancer Res. 2013;73:5402–15. doi:10.1158/0008-5472.CAN-13-0297.
29. Coskun E, von der Heide EK, Schlee C, Kühnl A, Gökbuget N, Hoelzer D, et al. The role of microRNA-196a and microRNA-196b as ERG regulators in acute myeloid leukemia and acute T-lymphoblastic leukemia. Leuk Res. 2011;35:208–13. doi:10.1016/j.leukres.2010.05.007.
30. Son JW, Yu S-L, Lee DC, Lee SY, Park CG, Lee HY, et al. Abstract 4353: microRNA-196b induces epithelial-to-mesenchymal transition through targeting HOXA9 in non-small cell lung cancer cells. Cancer Res. 2014;74(19 Supplement):4353. doi:10.1158/1538-7445.AM2014-4353.