Carcinogenesis, 2019, Vol. 40, No. 5, 643–650

doi:10.1093/carcin/bgy132
Advance Access publication November 14, 2018
Original Article

ORIGINAL ARTICLE

A 5-microRNA signature identified from serum microRNA profiling predicts survival in patients with advanced stage non-small cell lung cancer

Yajie Zhang[1,2], Jack A.Roth[3], Hao Yu[1], Yuanqing Ye[1], Kunlin Xie[1], Hua Zhao[1], David W.Chang[1], Maosheng Huang[1], Hecheng Li[2], Jieming Qu[4] and Xifeng Wu[1,*]

[1]Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA [2]Department of Thoracic Surgery, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China [3]Department of Thoracic and Cardiovascular Surgery, University of Texas MD Anderson Cancer Center, Houston, TX, USA [4]Department of Pulmonary and Critical Care Medicine, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

*To whom correspondence should be addressed. Tel: +1 713 745 2485; FAX: +1 713 792 4657; Email: xwu@mdanderson.org

Abstract

Circulating microRNAs (miRNAs) are potential biomarkers for cancer diagnosis, screening and prognosis. This study aimed to identify serum miRNAs as predictors of survival in patients with advanced non-small cell lung cancer (NSCLC). We profiled serum miRNAs in a pilot set of four patients with good survival (>24 months) and four patients with poor survival (<6 months). We selected 140 stably detectable miRNAs and 42 miRNAs reported in literature for further analysis. Expression of these 182 miRNAs was measured using high-throughput polymerase chain reaction assay, and their association with 3-year survival in the discovery (n = 345) and validation (n = 177) cohorts was assessed. Five serum miRNAs (miR-191, miR-28-3p, miR-145, miR-328 and miR-18a) were significantly associated with 3-year overall survival in both cohorts. A combined 5-miRNA risk score was created to assess the cumulative impact of these miRNAs on risk of death. Quartile analysis of the risk score showed significant association with 3-year death risk, with a 4.6-, 6.8- and 9.3-month reduction in median survival time for the second, third and fourth quartiles, respectively. Survival tree analysis also identified distinct risk groups with different 3-year survival durations. Data from The Cancer Genome Atlas revealed all five miRNAs were differentially expressed ($P < 0.0001$) in paired tumor and normal tissues. Pathway analysis indicated that target genes of these five miRNAs were mainly enriched in inflammatory/immune response pathways and pathways implicated in resistance to chemoradiotherapy and/or targeted therapy. Our results suggested that the 5-miRNA signature could serve as a prognostic predictor in patients with advanced NSCLC.

Introduction

Lung cancer is the leading cause of cancer-related death worldwide, and non-small cell lung cancer (NSCLC) accounts for 85% of all lung cancer cases (1,2). Despite efforts to improve early disease detection and the development of advanced chemotherapeutic and targeted treatments, the overall survival rate of NSCLC patients remains poor (2). More than 70% of patients with NSCLC have locally advanced (stage III) or metastatic (stage IV) disease at diagnosis, and 5-year survival rates range from 4.8% to 26.4% (3). These rather short survival times, along with the fact that late-stage patients with similar clinical features often have diverse outcomes, pose clinical challenges. The identification of new, specific biomarkers that can be used to monitor tumor progression and response to therapy and predict patient survival would help to overcome these challenges and improve outcomes for patients with NSCLC (4,5).

MicroRNAs (miRNAs) are small non-coding RNAs that post-transcriptionally regulate gene expression by degrading or repressing translation of targeted transcripts, thereby affecting

Received: August 9, 2018; **Revised:** August 9, 2018; Accepted: November 11, 2018

Abbreviations

CI	confidence interval
HR	hazard ratio
miRNA	microRNAs
MST	median survival time
NSCLC	non-small cell lung cancer
TCGA	The Cancer Genome Atlas

processes such as cell proliferation, differentiation and apoptosis (6). The existence of stable cell-free miRNAs in plasma/serum has been repeatedly demonstrated in several studies (7,8), and circulating miRNAs from liquid biopsy could serve as useful biomarkers to correlate with individual patients' distinctive tumor characteristics and response to therapy (9,10). Changes in miRNA expression have shown potential as biomarkers for lung cancer risk and prognosis (11); however, most such studies used targeted approaches or had limited sample sizes (12–14).

In this multiphase study, we first performed pilot global serum miRNA profiling in samples from eight patients with advanced NSCLC who had either good or poor survival. Next, expression levels of the identified stably detectable miRNAs and miRNAs reported in the literature to be important in lung cancer were quantified in discovery and validation cohorts to identify serum miRNA biomarkers that could predict 3-year overall survival in patients with advanced NSCLC. The candidate miRNAs were further analyzed for their relevance to NSCLC using NSCLC miRNA expression data from The Cancer Genome Atlas (TCGA). In addition, potential target genes of the candidate miRNAs were analyzed *in silico* to identify enriched signaling pathways, yielding clues to the underlying biological mechanisms for future experimental verification and functional studies.

Materials and methods

Study population

This study included patients with newly diagnosed, histologically confirmed NSCLC who were recruited at the University of Texas MD Anderson Cancer Center between January 2002 and January 2009. All study participants signed an informed consent document and underwent a 45-min interview conducted by trained MD Anderson staff. A comprehensive epidemiological questionnaire was used to elicit information on demographic characteristics, medical history and history of tobacco use. Immediately after each interview, peripheral blood sample was drawn for isolation of serum and other biological materials. Clinical and follow-up data were abstracted from participants' medical charts. Patients were selected according to the following criteria: (i) the patient had been diagnosed with stage III or IV NSCLC; (ii) serum was available in sufficient volume for RNA isolation and (iii) demographic, clinical and follow-up data for the patient were available. Our analysis for this study was restricted to non-Hispanic whites because the numbers of patients from other racial/ethnic groups in our study population were small. A total of 530 patients were included in our final analysis. The study was approved by MD Anderson Cancer Center's institutional review board.

RNA isolation

Total RNA was isolated from 750-μL serum samples using a miRNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. Synthetic cel-miR-39 was added to each sample as a spike-in control for evaluation of successful extraction (15). Collected RNAs were reconstituted in 30 μL of ultrapure water. The total concentration of small RNA molecules was quantified using a NanoDrop ND-1000 spectrometer (Thermo Fisher Scientific, Wilmington, DE).

MiRNA array profiling

Initial global miRNA screening was performed on serum samples from four patients with advanced NSCLC who had good survival (>24 months) and four sex-, age- and smoking status-matched patients with advanced NSCLC who had poor survival (<6 months). We used a TaqMan Array Human MicroRNA Card Set v3.0 (Applied Biosystems, Foster City, CA), which contained 754 human miRNAs, as previously described (16). miRNAs with a threshold cycle (C_t) value of less than 35 and a missing rate of less than 25% in both groups were considered stably detectable candidates for further analysis. Initial screening yielded 140 stably detected miRNAs for inclusion in the subsequent analyses. Another 42 serum miRNAs reported in the literature to be important in lung cancer were also included in the subsequent analyses (17–20).

Quantitative real-time polymerase chain reaction assay

Expression levels of the 182 selected miRNAs were measured in serum samples using a high-throughput BioMark HD Real-Time Polymerase Chain Reaction system (Fluidigm, San Francisco, CA) and a TaqMan miRNA assay (Thermo Fisher Scientific, Waltham, MA) as previously described (16). Each assay was tested in duplicate. For quality control, data that met any of the following criteria were excluded from further analysis: (i) samples with a C_t value higher than 25 or less than 16 in spike-in miRNAs; (ii) miRNAs with a detection rate lower than 80% or (iii) outliers with data points outside 5 SDs. Among the 182 examined miRNAs, 119 passed these quality control tests (Supplementary Table 1). After data cleaning, the C_t value obtained for each miRNA was normalized to the average expression level of spike-in cel-miR-39 and analyzed with the $2^{-\Delta\Delta Ct}$ method (21).

Analysis of candidate miRNAs using TCGA data

To determine the potential relevance of the candidate miRNAs to NSCLC, we used an NSCLC miRNA expression data set from TCGA. miRNA expression data for paired tumor and normal tissues from 84 patients were downloaded from the cBioPortal containing TCGA level 3 miRNAseq data for NSCLC (www.cbioportal.org; accessed June 2018). The miRNA expression profiling had been performed using the Illumina HiSeq 2000 miRNA sequencing platform (Illumina, Inc., San Diego, CA). The miRNA expression levels for paired tumor and normal tissues were calculated as reads per million miRNAs mapped and were log2-transformed for further analysis.

Target gene and pathway enrichment analyses

Analyses of potential target genes of the identified miRNAs and enriched pathways were conducted using the Web-based tool miRSystem (http://mirsystem.cgm.ntu.edu.tw/), which includes seven algorithms (DIANA, miRanda, miRBridge, PicTar, PITA, rna22 and TargetScan) and two experimentally validated databases (TarBase and miRecords) for predicting miRNA targets. miRSystem also includes five pathway databases (Kyoto Encyclopedia of Genes and Genomes, BioCarta, Pathway Interaction Database, Reactome and Gene Ontology) for annotating the biological functions and canonical pathways of target genes (22).

Statistical analysis

Statistical analyses were performed using Stata software (version 14.0; Stata Corp., LLC, College Station, TX). To assess the association of serum miRNA expression levels with 3-year survival rates, serum miRNA expression levels were dichotomized into 'low' and 'high' groups using the median level as a cutoff. Estimated hazard ratios (HRs) and 95% confidence intervals (CIs) were calculated using Cox proportional hazards models and adjusted for age, sex, smoking status, clinical stage and treatment regimen. Patients with survival times longer than 36 months were censored at 36 months in the Cox regression analysis. The combined miRNA risk score for each patient was a linear combination of the product of the reference-normalized expression level of each miRNA and its Cox regression-corresponding coefficient. The association of miRNA risk scores with 3-year survival was evaluated categorically using a quartile distribution and continuously for trend analysis. Survival tree analysis using STREE software (http://c2s2.yale.edu/software/stree/) was

conducted in the combined data set to identify higher-order miRNA-miRNA interactions affecting survival. STREE uses a log-rank statistical method to select optimal and subsequent splits of data sets. For analysis of TCGA data, the mean miRNA expression levels in tumor and normal tissues were compared using a paired *t*-test. All statistical analyses were two sided. *P* values of less than 0.05 were considered statistically significant.

Results

Patient characteristics

Figure 1 is a schematic flowchart of the study design. The study's discovery cohort included 345 patients, and the validation cohort included 177 patients. The patients' characteristics are summarized in Supplementary Table 2. The median follow-up durations for the discovery, validation and overall cohorts were 32.3, 36.0 and 35.5 months, respectively. The discovery and validation cohorts did not significantly differ with respect to age, sex, smoking status or histology ($P > 0.05$). The two cohorts did differ significantly in terms of TNM stage, treatment modality and vital status, which were adjusted in subsequent analyses.

Association of individual serum miRNA expression with NSCLC survival

To identify miRNA markers that predict patient survival, we assessed the association of expression levels of 119 miRNAs with 3-year NSCLC survival using a two-stage design. In the discovery cohort, expression levels of 47 serum miRNAs were significantly associated with 3-year survival (using median expression levels as the cutoff) (Supplementary Table 3). In the validation cohort, using the same cutoff point, expression of 5 of 47 serum miRNAs (miR-191, miR-28-3p, miR-145, miR-328 and miR-18a) was significantly associated with 3-year survival in the same direction as in the discovery cohort (Supplementary Table 4). In the combined cohort, high expression levels of miR-191 (HR, 1.57; 95% CI, 1.29–1.92; $P = 9.77E{-}06$), miR-28-3p (HR, 1.51; 95% CI, 1.24–1.85; $P = 4.42E{-}05$), miR-145 (HR, 1.49; 95% CI, 1.22–1.81; $P = 8.17E{-}05$), miR-328 (HR, 1.48; 95% CI, 1.22–1.81; $P = 9.50E{-}05$) and miR-18a (HR, 1.34; 95% CI, 1.10–1.63; $P = 3.45E{-}03$) were significantly associated with an increased risk of death (Table 1). The median survival times (MSTs) for patients with high expression of these five serum miRNAs were shorter than those of patients with low expression by 3.6–5.7 months.

Prediction of survival outcomes by the 5-miRNA signature

We constructed a miRNA signature based on expression of the five candidate miRNAs and calculated a risk score for each patient. We then examined the risk scores' association with 3-year survival (Table 1). The patients in the discovery cohort were categorized into four groups according to the quartile distribution of their risk scores. The HRs from the first to fourth risk-score quartiles were 1.00 (reference), 1.46 (95% CI, 1.03–2.08, $P = 0.036$), 1.59 (95% CI, 1.14–2.23, $P = 6.31E{-}03$) and 1.97 (95% CI, 1.41–2.75, $P = 7.07E{-}05$). The same cutoff point obtained from the discovery cohort was applied to the validation cohort. The HRs for the validation cohort were 1.00 (reference), 1.19 (95% CI, 0.71–1.99, $P = 0.506$), 1.84 (95% CI, 1.11–3.07, $P = 0.019$) and 2.18 (95% CI, 1.25–3.82, $P = 6.35E{-}03$) across the risk-score quartiles. In the combined cohort, the HRs were 1.36 (95% CI, 1.02–1.81, $P = 0.038$) for the second quartile, 1.70 (95% CI, 1.29–2.25, $P = 1.66E{-}04$) for the third quartile and 2.01 (95% CI, 1.51–2.66, $P = 1.26E{-}06$) for the highest quartile compared with the lowest quartile, which

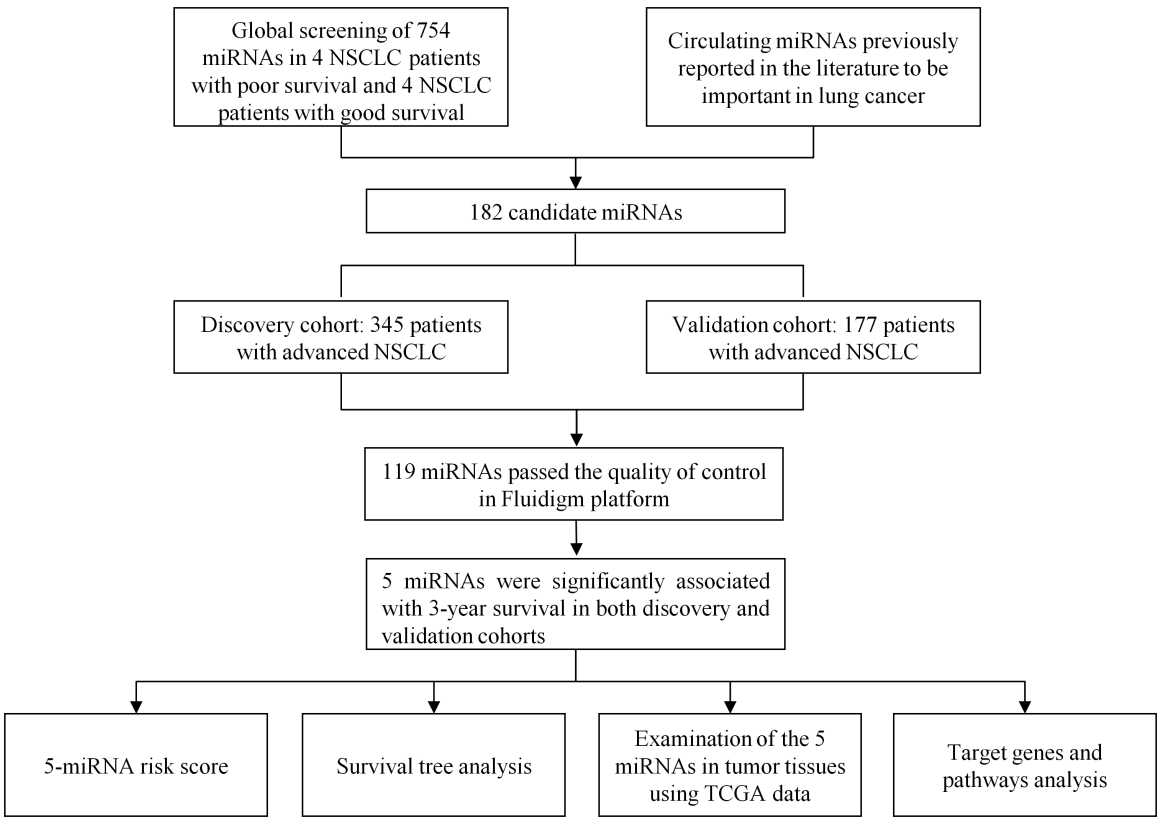

Figure 1. Schematic flowchart of the study design.

Table 1. Association of serum levels of the five candidate miRNAs and the 5-miRNA risk score with 3-year survival in discovery, validation and combined cohorts of patients with advanced non-small cell lung cancer

Cohort	miRNA expression		3-year death, *N* (%)	3-year survival, *N* (%)	Adjusted HR[a](95% CI)	*P* value	MST[b]
Discovery	miR-191	Low	128 (74.0)	45 (26.0)	1 (reference)		10.7
(*n* = 345)		High	161 (93.6)	11 (6.4)	1.54 (1.21–1.96)	5.26E–04	9.5
	miR-28-3p	Low	130 (75.1)	43 (24.9)	1 (reference)		11.2
		High	159 (92.4)	13 (7.6)	1.53 (1.20–1.94)	5.30E–04	8.9
	miR-145	Low	137 (79.2)	36 (20.8)	1 (reference)		10.8
		High	152 (88.4)	20 (11.6)	1.40 (1.11–1.77)	5.21E–03	9.0
	miR-328	Low	139 (80.3)	34 (19.7)	1 (reference)		10.8
		High	150 (87.2)	22 (12.8)	1.36 (1.08–1.73)	9.55E–03	9.2
	miR-18a	Low	136 (78.6)	37 (21.4)	1 (reference)		10.3
		High	153 (89.0)	19 (11.0)	1.30 (1.03–1.64)	0.030	9.5
	Risk score	Q1 (low)	68 (72.3)	26 (27.7)	1 (reference)		14.1
		Q2	64 (79.0)	17 (21.0)	1.46 (1.03–2.08)	0.036	10.8
		Q3	77 (93.9)	5 (6.1)	1.59 (1.14–2.23)	6.31E–03	9.6
		Q4 (high)	80 (90.9)	8 (9.1)	1.97 (1.41–2.75)	7.07E–05	8.8
		Trend				6.95E–05	
Validation	miR-191	Low	66 (68.8)	30 (31.2)	1 (reference)		24.4
(*n* = 177)		High	55 (67.9)	26 (32.1)	1.53 (1.05–2.24)	0.028	16.2
	miR-28-3p	Low	67 (68.4)	31 (31.6)	1 (reference)		24.1
		High	54(68.4)	25 (31.6)	1.50 (1.02–2.20)	0.038	14.5
	miR-145	Low	61 (65.6)	32 (34.4)	1 (reference)		24.6
		High	60 (71.4)	24 (28.6)	1.78 (1.22–2.60)	2.83E–03	14.6
	miR-328	Low	57 (65.5)	30 (34.5)	1 (reference)		25.1
		High	64 (71.1)	26 (28.9)	1.68 (1.15–2.46)	7.50E–03	14.6
	miR-18a	Low	57 (65.5)	30 (34.5)	1 (reference)		24.5
		High	64 (71.1)	26 (28.9)	1.48 (1.03–2.13)	0.033	16.3
	Risk score	Q1 (low)	30 (65.2)	16 (34.8)	1 (reference)		26.6
		Q2	31 (67.4)	15 (32.6)	1.19 (0.71–1.99)	0.506	24.1
		Q3	35 (72.9)	13 (27.1)	1.84 (1.11–3.07)	0.019	16.0
		Q4 (high)	25 (67.6)	12 (32.4)	2.18 (1.25–3.82)	6.35E–03	14.0
		Trend				1.99E–03	
Combined	miR-191	Low	194 (72.1)	75 (27.9)	1 (reference)		15.6
(*n* = 522)		High	216 (85.4)	37 (14.6)	1.57 (1.29–1.92)	9.77E–06	10.8
	miR-28-3p	Low	197 (72.7)	74 (27.3)	1 (reference)		16.0
		High	213 (84.9)	38 (15.1)	1.51 (1.24–1.85)	4.42E–05	10.3
	miR-145	Low	198 (74.4)	68 (25.6)	1 (reference)		14.7
		High	212 (82.8)	44 (17.2)	1.49 (1.22–1.81)	8.17E–05	11.0
	miR-328	Low	196 (75.4)	64 (24.6)	1 (reference)		15.7
		High	214 (81.7)	48 (18.3)	1.48 (1.22–1.81)	9.50E–05	10.5
	miR-18a	Low	193 (74.2)	67 (25.8)	1 (reference)		14.7
		High	217 (82.8)	45 (17.2)	1.34 (1.10–1.63)	3.45E–03	11.1
	Risk score	Q1 (low)	98 (70.0)	42 (30.0)	1 (reference)		18.9
		Q2	95 (74.8)	32 (25.2)	1.36 (1.02–1.81)	0.038	14.3
		Q3	112 (86.2)	18 (13.8)	1.70 (1.29–2.25)	1.66E–04	12.1
		Q4 (high)	105 (84.0)	20 (16.0)	2.01 (1.51–2.66)	1.26E–06	9.6
		Trend				1.60E–07	

Q, quartile.
[a]Adjusted by age, sex, smoking status, clinical stage and treatment regimen.
[b]In months.

corresponded to 4.6-, 6.8- and 9.3-month reductions in MST, respectively (Figure 2).

We conducted further analysis of the association between the 5-miRNA signature and 3-year survival in the combined cohort stratified by clinicopathological risk factors known to affect NSCLC outcomes, including sex, age, smoking status, TNM stage, histological grade, histological type and treatment regimen. Similar and significant associations between the 5-miRNA signature and survival were observed for most of the different strata, except for patients who were never-smokers, which might be due to the small sample size of this subset of the patient cohort (Supplementary Table 5).

Survival tree analysis

To explore potential higher-order interactions between miRNAs and to define subgroups that have distinct survival prospects, we performed a survival tree analysis using the five identified miRNAs in the combined cohort. Four miRNAs—miRNA-191, miRNA-28-3p, miRNA-18a and miRNA-328—demonstrated interactions that led to five terminal nodes (Figure 3A). The initial split on the survival tree was miRNA-191 expression, indicating that this miRNA is the primary factor contributing to the observed variation in overall survival. With terminal node 1 (patients with low expression of both miRNA-191 and miRNA-28-3p) as the reference group, the HRs for the other four terminal nodes ranged from 1.17 to 1.90.

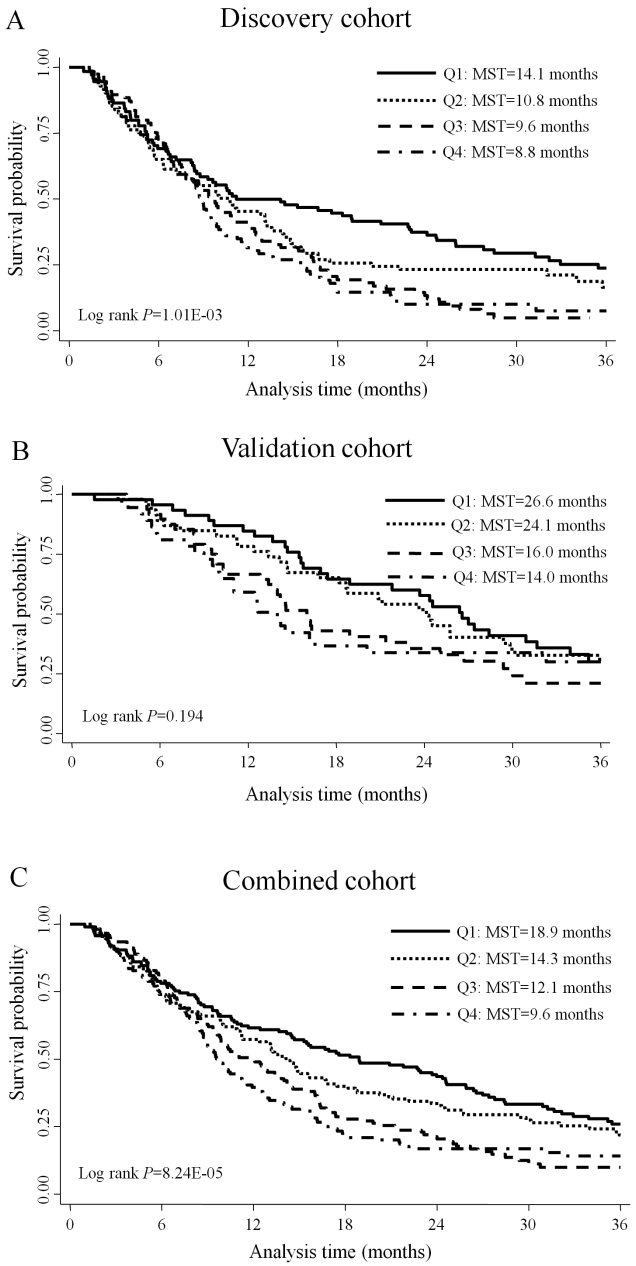

Figure 2. Kaplan–Meier plots comparing estimates of 3-year survival probability for patients with advanced NSCLC grouped by quartile rankings of the 5-miRNA risk score. Plots show the discovery (A), validation (B) and combined (C) sets. Q, quartile.

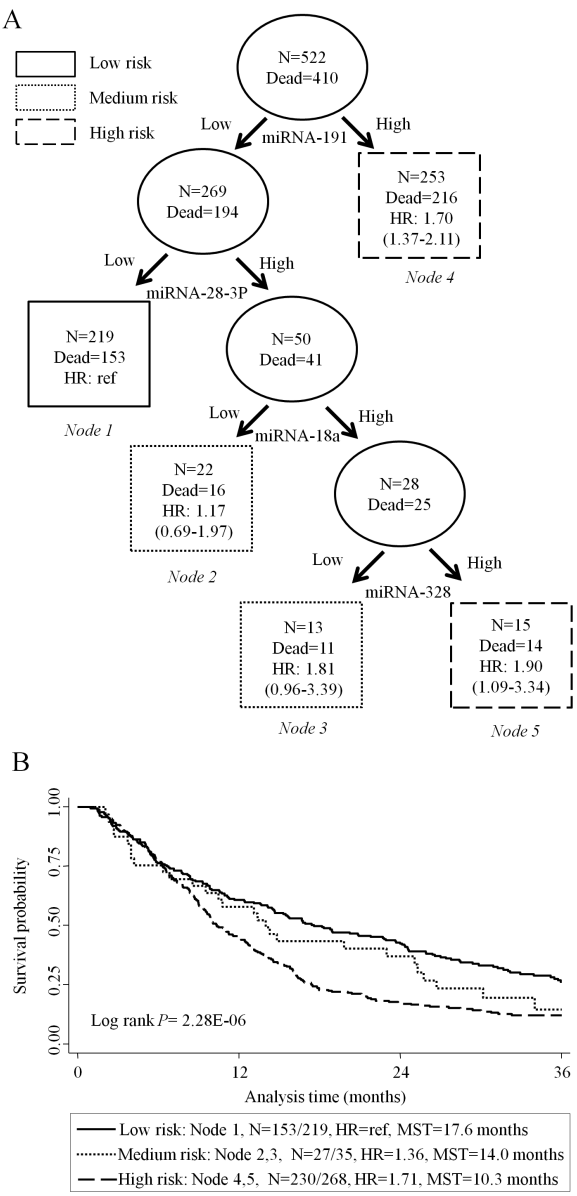

Figure 3. Potential miRNA–miRNA interactions identified by survival tree analysis in the combined set. (A) Survival tree analysis of miR-191, miR-28-3p, miR-18a and miR-328 based on low-expression and high-expression groups. Each node contains the number of patients (*N*), number of patients who died (Dead), hazard ratio (HR), and 95% confidence interval (in parentheses). (B) Kaplan–Meier plots comparing estimates of 3-year survival probability in low-risk (node 1), medium-risk (nodes 2 and 3) and high-risk (nodes 4 and 5) groups as determined by the survival tree analysis. *N* represents number of patients with events (deaths) in the 3-year period/total number of patients in the data set. MST, median survival time.

The five terminal nodes were further differentiated into three distinct groups, with MSTs of 17.6-, 14.0- and 10.3- months for low-risk (node 1), medium-risk (nodes 2 and 3), and high-risk (nodes 4 and 5) groups, respectively (log-rank *P* = 2.28E–06; Figure 3B).

Analysis of the candidate miRNAs in TCGA NSCLC tumor tissue data

Analysis of TCGA NSCLC miRNA sequencing data revealed that all five candidate miRNAs were differentially expressed ($P < 0.0001$) in tumor and normal tissues (Figure 4). Four of the five miRNAs—miR-191 (*P* = 2.68E–12), miR-28-3p (*P* = 2.35E–08), miR-328 (*P* = 7.99E–09) and miR-18a (*P* = 1.58E–07)—were significantly overexpressed in NSCLC tumors compared with paired normal tissues. In contrast, miR-145 was more highly expressed (*P* = 8.53E–14) in normal tissues than in paired tumor samples.

Target gene prediction coupled with pathway analysis

To explore the biological mechanisms underlying the involvement of these five miRNAs in advanced NSCLC survival, we performed target gene prediction coupled with pathway analysis.

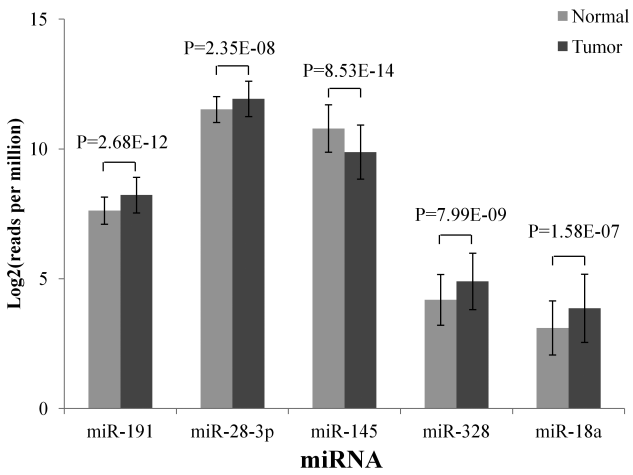

Figure 4. Analysis of the expression of five candidate mRNAs in paired NSCLC tumor and normal tissues based on TCGA data. Illumina HiSeq microRNA sequencing data from TCGA were downloaded and analyzed for 84 paired NSCLC tumors and normal tissues. All five miRNAs showed significant differences in expression in tumors compared with paired normal tissues. The bar graph indicates the means in log2 scale with the standard deviations shown as error bars. *P* values were determined from paired *t*-test.

A total of 1183 potential target genes (data not shown) regulated by the five miRNAs were identified by both prediction algorithms and experiment-supported databases. Pathway analysis demonstrated that the predicted targets of these five miRNAs were mainly enriched ($P < 0.05$) in 19 pathways (Table 2), 9 of which were associated with inflammatory and immune responses, such as the tumor necrosis factor , CD40/CD40L, signal transducer and activator of transcription 5/interleukin 2 (STAT5/IL2) and platelet aggregation signaling pathways. Four pathways, including Ras, mitogen-activated protein kinase, ATM and estrogen receptor signaling pathways, were closely associated with resistance to chemoradiotherapy and/or targeted therapy in advanced NSCLC. The predicted target genes were also involved in several other critical pathways involved in NSCLC, such as those regulating apoptosis, the cell cycle and DNA repair.

Discussion

NSCLC is a heterogeneous disease with varied outcomes. To date, tumor stage and clinical factors including non-squamous histology and blood and lymphatic vascular invasion have been established as prognostic factors (23,24). Nevertheless, the prognosis of patients with the same disease stage and clinical features may vary substantially (25). Moreover, assessment of these factors often requires invasive procedures or repeat biopsies, which are frequently challenging or infeasible and may lead to complications in patients with advanced NSCLC (26). Therefore, it is of paramount importance to develop noninvasive methods that accurately predict survival in patients with advanced NSCLC and can identify individuals at higher risk of mortality for more intensive monitoring (27).

To the best of our knowledge, our current study is one of the largest to evaluate serum miRNAs and prognosis in patients with advanced NSCLC. From global serum miRNA profiling and selected profiling of serum miRNAs previously found to be important in lung cancer, we found that a 5-miRNA signature comprising miR-191, miR-28-3p, miR-328, miR-145 and miR-18a was significantly associated with overall survival in patients with advanced NSCLC. Among these five miRNAs, the expression of miR-191 was the most strongly associated with 3-year survival. A recent study indicated that miR-191 expression levels were higher in cancerous tissues than in adjacent noncancerous tissues from NSCLC patients and that miR-191 may promote the proliferation and migration of lung cancer cell lines under chronic hypoxic conditions by downregulating nuclear factor 1α (28). Moreover, miR-191 is a known oncomiR that is involved in the neoplastic and metastatic properties of malignantly transformed human bronchial epithelial cells by promoting epithelial–mesenchymal transition and conferring cancer stem cell-like properties, highlighting that miR-191 may be a therapeutic target (29). Our analyses demonstrated that patients with high serum levels of miR-328 had an MST 5.2 months shorter than that of patients with low levels of miR-328. Du *et al.* (30) reported that miR-328 was overexpressed in A549 lung cancer cells and that its downregulation inhibited the invasion and migration capacity of NSCLC cells. Furthermore, miR-328 was reported to be associated with NSCLC brain metastasis, possibly through its promotion of tumor cell migration (31).

The relationship of miR-145 with NSCLC is still controversial. Shen *et al.* (32) reported that low expression of miR-145 in NSCLC tissue was associated with poor histological differentiation and predicted poor prognosis. Another study also indicated that miR-145 was downregulated in lung cancer and that miR-145 inhibited the migration and invasion of lung cancer cells by targeting the gene *FSCN1* (33). Our analysis of miRNA expression in TCGA data also indicated that miR-145 was significantly downregulated in tumor tissues compared with paired normal tissues. Paradoxically, however, we also found that high levels of serum miR-145 were associated with poor survival among patients with advanced NSCLC. The reasons for this discrepancy are unclear. It is possible that miR-145 plays different roles in tumor cells and in the circulation. Supporting this possibility, Wang *et al.* (34) found that serum miR-145 was overexpressed in NSCLC patients compared with healthy control subjects. As for miR-18a, consistent with our results, a study by Xu *et al.* (35) reported that high expression of circulating miR-18a was an independent risk factor for overall survival in NSCLC patients. Another study revealed that miR-18a was significantly upregulated in NSCLC tissues and that it promoted carcinogenesis by targeting interferon regulatory factor 2 (36). We also observed that elevated serum levels of miR-28-3p were associated with poor survival in patients with advanced NSCLC. Although no associations of miR-28-3p expression with lung cancer prognosis have been reported in the literature, miR-28 has been linked to the development and progression of other cancers (37,38).

In addition to our analysis of individual miRNAs, we also evaluated the collective prognostic value of the five miRNAs by creating a risk score that correlated with 3-year survival in both the discovery and validation cohorts. We identified novel associations among these five miRNAs that affected patient survival. Furthermore, stratified analyses by clinicopathological factors that may influence late-stage NSCLC survival showed similar associations, suggesting these associations were mostly not confounded by patient characteristics. Survival tree analysis also revealed potential higher-order interactions between miRNA-191, miRNA-28-3p, miRNA-18a and miRNA-328.

We identified the target genes regulated by the five miRNAs in the signature and determined the implicated biological pathways. Most of these pathways, including tumor necrosis factor, CD40/CD40L, STAT5/IL2 and platelet aggregation were related to the inflammatory and immune responses signaling. It is widely

Table 2. miRSystem enriched pathway analysis of the 5-microRNA signature

Database	Pathway[a]	Identification	Genes	Targets	miRNAs	Empirical *P* value[b]
BioCarta	BioCarta stress pathway	—	25	3	2	2.93E–03
Pathway interaction database	**TNF receptor signaling pathway**	200102	46	3	2	4.83E–03
KEGG	MAPK signaling pathway	4010	272	8	4	5.46E–03
Reactome	**Platelet aggregation (plug formation)**	REACT 278	37	3	3	5.85E–03
Pathway interaction database	**CD40/CD40L signaling**	200037	30	3	2	6.88E–03
Pathway interaction database	**IL2 signaling events mediated by STAT5**	200185	30	3	2	7.47E–03
Pathway interaction database	ATM pathway	200072	34	3	3	8.45E–03
Pathway interaction database	**TCR signaling in naive CD8+ T cells**	200075	51	3	2	0.011
Pathway interaction database	Validated nuclear estrogen receptor α network	200159	63	3	2	0.025
Reactome	**Integrin αIIbβ3 signaling**	REACT 15523	27	2	3	0.028
Pathway interaction database	Caspase cascade in apoptosis	200174	56	3	2	0.039
KEGG	Small cell lung cancer	5222	84	3	2	0.043
KEGG	**Systemic lupus erythematosus**	5322	136	3	2	0.043
Reactome	**Signaling by interleukins**	REACT 22232	106	3	4	0.045
Pathway interaction database	Regulation of Ras family activation	200213	33	2	2	0.045
Pathway interaction database	**Signaling events mediated by PTP1B**	200042	52	3	4	0.045
Reactome	Meiotic recombination	REACT 27271	54	2	3	0.045
Pathway interaction database	Urokinase-type plasminogen activator (uPA) and uPA receptor-mediated signaling	200140	42	2	4	0.046
Pathway interaction database	EphA forward signaling	200143	34	2	4	0.047

ATM, ATM serine/threonine kinase; CD40L, CD40 ligand; EphA, ephrin receptor A; IL2, interleukin 2; KEGG, Kyoto Encyclopedia of Genes and Genomes; MAPK, mitogen-activated protein kinase; PTP1B, protein tyrosine phosphatase, non-receptor type 1; STAT5, signal transducer and activator of transcription 5; TCR, T-cell receptor; TNF, tumor necrosis factor.

[a]Inflammatory and immune response signaling pathways are shown in bold font.

[b]Empirical *P* values were compared with 1000 random selections.

believed that inflammatory and immune responses are important in the tumor microenvironment and play decisive roles in the initiation, proliferation, invasion and metastasis of NSCLC (39,40). Inflammatory molecules and effectors produced during chronic inflammation may directly affect lung cancer development and progression through the activation of transcriptional factors or effectors (e.g. NF-κB, AP-1 and STAT) (41). Moreover, the acute inflammatory and immune responses triggered during chemoradiotherapy can attenuate treatment effectiveness and lead to the development of chemoresistance or toxicities, both of which worsen prognosis (42,43). Some of the five identified miRNAs including miR-145 and miR-18a have been found to be involved in the regulation of the inflammatory and immune responses and cancer progression (44,45). In addition, the identified miRNAs participate in several other critical cancer-related pathways involved in the resistance to chemoradiotherapy and/or targeted therapy, apoptosis, the cell cycle and DNA repair, all of which play important roles in NSCLC progression and metastasis.

Our study has several strengths. First, its large overall sample size—530 patients with advanced-stage NSCLC—enabled us to perform a multistage study using discovery and validation cohorts. Second, we obtained comprehensive clinical, demographic and other data that enabled us to account for the effects of confounders and identify interactions between circulating miRNA levels and patient characteristics that affect NSCLC survival. Third, the large number of serum miRNAs screened using both the global and targeted approaches enabled us to identify several potential noninvasive biomarkers for predicting survival in patients with advanced NSCLC. However, our study population was restricted to non-Hispanic whites, so our findings may not be extrapolated to other racial/ethnic groups. The use of blood samples to obtain miRNA data also has inherent limitations, as blood analyses can be affected by confounding factors such as sample storage time, the subject's health conditions and the extraction method. To lessen these effects, we imposed stringent quality controls and used a consistent processing method for all blood samples, so any variations were likely minimal.

In conclusion, we identified a 5-miRNA signature that was significantly associated with 3-year survival in patients with advanced NSCLC. Larger multicenter and prospective studies are necessary to further validate our findings. Future functional

investigations may clarify the biological mechanisms by which the candidate miRNAs contribute to NSCLC progression and outcomes.

Supplementary material

Supplementary data are available at *Carcinogenesis* online

Funding

Cancer Prevention and Research Institute of Texas (RP1300502); National Cancer Institute (P50 CA070907 and R01 CA176568) and MD Anderson's Center for Translational and Public Health Genomics supported by Duncan Family Institute for Cancer Prevention and Risk Assessment.
Conflict of Interest Statement: None declared.

References

1. Blandin Knight, S. *et al.* (2017) Progress and prospects of early detection in lung cancer. Open Biol., 7, pii: 170070.
2. Siegel, R.L. *et al.* (2017) Cancer statistics, 2017. CA. Cancer J. Clin., 67, 7–30.
3. Richards, T.B. *et al.* (2017) Lung cancer survival in the United States by race and stage (20012009): findings from the CONCORD-2 study. Cancer, 123 (suppl. 24), 5079–5099.
4. Vargas, A.J. *et al.* (2016) Biomarker development in the precision medicine era: lung cancer as a case study. Nat. Rev. Cancer, 16, 525–537.
5. Neal, J.W. *et al.* (2015) Developing biomarker-specific end points in lung cancer clinical trials. Nat. Rev. Clin. Oncol., 12, 135–146.
6. Pritchard, C.C. *et al.* (2012) Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. Cancer Prev. Res. (Phila)., 5, 492–497.
7. Fortunato, O. *et al.* (2014) Assessment of circulating microRNAs in plasma of lung cancer patients. Molecules, 19, 3038–3054.
8. Allegra, A. *et al.* (2012) Circulating microRNAs: new biomarkers in diagnosis, prognosis and treatment of cancer (review). Int. J. Oncol., 41, 1897–1912.
9. Mitchell, P.S. *et al.* (2008) Circulating microRNAs as stable blood-based markers for cancer detection. Proc. Natl. Acad. Sci. US A, 105, 10513–10518.
10. Schwarzenbach, H. *et al.* (2014) Clinical relevance of circulating cell-free microRNAs in cancer. Nat. Rev. Clin. Oncol., 11, 145–156.
11. Świtlik, W.Z. *et al.* (2017) Circulating miRNAs as non-invasive biomarkers for non-small cell lung cancer diagnosis, prognosis and prediction of treatment response. Postepy Hig. Med. Dosw. (Online), 71, 649–662.
12. Su, K. *et al.* (2016) Diagnostic and prognostic value of plasma microRNA-195 in patients with non-small cell lung cancer. World J. Surg. Oncol., 14, 224.
13. Powrózek, T. *et al.* (2016) Application of plasma circulating microRNA-448, 506, 4316, and 4478 analysis for non-invasive diagnosis of lung cancer. Tumour Biol., 37, 2049–2055.
14. Zhao, Q. *et al.* (2015) Circulating miRNAs is a potential marker for gefitinib sensitivity and correlation with EGFR mutational status in human lung cancers. Am. J. Cancer Res., 5, 1692–1705.
15. Li, Y. *et al.* (2012) Method for microRNA isolation from clinical serum samples. Anal. Biochem., 431, 69–75.
16. Wang, Y. *et al.* (2013) Pathway-based serum microRNA profiling and survival in patients with advanced stage non-small cell lung cancer. Cancer Res., 73, 4801–4809.
17. Boeri, M. *et al.* (2011) MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. Proc. Natl. Acad. Sci. USA, 108, 3713–3718.
18. Lei, Z. *et al.* (2014) MiR-142-3p represses TGF-β-induced growth inhibition through repression of TGFβR1 in non-small cell lung cancer. FASEB J., 28, 2696–2704.
19. Li, J. *et al.* (2014) MicroRNA-146 up-regulation predicts the prognosis of non-small cell lung cancer by miRNA in situ hybridization. Exp. Mol. Pathol., 96, 195–199.
20. Garofalo, M. *et al.* (2011) EGFR and MET receptor tyrosine kinase-altered microRNA expression induces tumorigenesis and gefitinib resistance in lung cancers. Nat. Med., 18, 74–82.
21. Livak, K.J. *et al.* (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta C(T)) method. Methods, 25, 402–408.
22. Lu, T.P. *et al.* (2012) miRSystem: an integrated system for characterizing enriched functions and pathways of microRNA targets. PLoS One, 7, e42390.
23. Ramnefjell, M. *et al.* (2017) Vascular invasion is an adverse prognostic factor in resected non-small-cell lung cancer. APMIS, 125, 197–206.
24. Shao, Q. *et al.* (2017) Clinical investigation into the initial diagnosis and treatment of 539 patients with stage IV lung cancer. Onco. Targets. Ther., 10, 535–541.
25. Zhu, C.Q. *et al.* (2014) Prognostic markers in lung cancer: is it ready for prime time? Transl. Lung Cancer Res., 3, 149–158.
26. Hegde, P.V. *et al.* (2016) Mediastinal staging: endosonographic ultrasound lymph node biopsy or mediastinoscopy. Thorac. Surg. Clin., 26, 243–249.
27. Tang, Y. *et al.* (2017) Biomarkers for early diagnosis, prognosis, prediction, and recurrence monitoring of non-small cell lung cancer. Onco. Targets. Ther., 10, 4527–4534.
28. Zhao, J. *et al.* (2017) A novel pathway in NSCLC cells: miR-191, targeting NFIA, is induced by chronic hypoxia, and promotes cell proliferation and migration. Mol. Med. Rep., 15, 1319–1325.
29. Xu, W. *et al.* (2015) MicroRNA-191, by promoting the EMT and increasing CSC-like properties, is involved in neoplastic and metastatic properties of transformed human bronchial epithelial cells. Mol. Carcinog., 54 (suppl 1), E148–E161.
30. DU, C. *et al.* (2016) [Downregulation of miR-18a or miR-328 inhibits the invasion and migration of lung adenocarcinoma A549 cells]. Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi, 32, 1051–1054.
31. Arora, S. *et al.* (2011) MicroRNA-328 is associated with (non-small) cell lung cancer (NSCLC) brain metastasis and mediates NSCLC migration. Int. J. Cancer, 129, 2621–2631.
32. Shen, H. *et al.* (2015) Low miR-145 expression level is associated with poor pathological differentiation and poor prognosis in non-small cell lung cancer. Biomed. Pharmacother., 69, 301–305.
33. Zhang, Y. *et al.* (2015) MicroRNA-145 inhibits migration and invasion by down-regulating FSCN1 in lung cancer. Int. J. Clin. Exp. Med., 8, 8794–8802.
34. Wang, R.J. *et al.* (2015) Serum miR-125a-5p, miR-145 and miR-146a as diagnostic biomarkers in non-small cell lung cancer. Int. J. Clin. Exp. Pathol., 8, 765–771.
35. Xu, X. *et al.* (2018) High circulating miR-18a, miR-20a, and miR-92a expression correlates with poor prognosis in patients with non-small cell lung cancer. Cancer Med., 7, 21–31.
36. Liang, C. *et al.* (2017) MicroRNA-18a-5p functions as an oncogene by directly targeting IRF2 in lung cancer. Cell Death Dis., 8, e2764.
37. Pashaei, E. *et al.* (2017) Meta-analysis of miRNA expression profiles for prostate cancer recurrence following radical prostatectomy. PLoS One, 12, e0179543.
38. Huang, Z. *et al.* (2017) A novel serum microRNA signature to screen esophageal squamous cell carcinoma. Cancer Med., 6, 109–119.
39. Guo, D. *et al.* (2018) Prognostic value of systemic immune-inflammation index in patients with advanced non-small-cell lung cancer. Future Oncol., 14, 2643–2650.
40. Shi, L. *et al.* (2015) Targeting roles of inflammatory microenvironment in lung cancer and metastasis. Cancer Metastasis Rev., 34, 319–331.
41. Lin, W.W. *et al.* (2007) A cytokine-mediated link between innate immunity, inflammation, and cancer. J. Clin. Invest., 117, 1175–1183.
42. Nakano, S. *et al.* (1995) 5-Fluorouracil and cisplatin sequential chemotherapy; mechanism of action and clinical application in advanced, unresectable non-small cell lung cancer. Intern. Med., 34, 288–291.
43. Pu, X. *et al.* (2014) Inflammation-related genetic variations and survival in patients with advanced non-small cell lung cancer receiving first-line chemotherapy. Clin. Pharmacol. Ther., 96, 360–369.
44. Shinohara, H. *et al.* (2017) Regulated polarization of tumor-associated macrophages by mir-145 via colorectal cancer-derived extracellular vesicles. J. Immunol., 199, 1505–1515.
45. Dong, P. *et al.* (2018) Control of PD-L1 expression by miR-140/142/340/383 and oncogenic activation of the OCT4-miR-18a pathway in cervical cancer. Oncogene, 37, 5257–5268.