Carcinogenesis vol.33 no.5 pp.1046–1054, 2012
doi:10.1093/carcin/bgs100
Advance Access publication February 13, 2012

MicroRNA profiling and prediction of recurrence/relapse-free survival in stage I lung cancer

Yan Lu[†], Ramaswamy Govindan[1,†], Liang Wang[2,†], Peng-yuan Liu, Boone Goodgame[1], Weidong Wen[3], Ananth Sezhiyan[1], John Pfeifer[1], Ya-fei Li[2], Xing Hua, Yian Wang[3], Ping Yang[2,†] and Ming You[3,4,*]

Department of Physiology and the Cancer Center, Medical College of Wisconsin, Milwaukee, WI 53226, USA, [1]The Alvin J. Siteman Cancer Center, Washington University in St Louis, St Louis, MO 63110, USA, [2]Department of Health Science Research, Mayo Clinic, Rochester, MN 55905, USA, [3]Department of Surgery, Washington University in St Louis, St Louis, MO 63110, USA and [4]Department of Pharmacology and Toxicology and the Cancer Center, Medical College of Wisconsin, Milwaukee, WI 53226, USA

[*]To whom correspondence should be addressed. Tel: 414 805 8228;
Fax: 414 805 8281;
Email: myou@mcw.edu

About 30% stage I non-small cell lung cancer (NSCLC) patients undergoing resection will recur. Robust prognostic markers are required to better manage therapy options. MicroRNAs (miRNAs) are a class of small non-coding RNAs of 19–25 nt and play important roles in gene regulation in human cancers. The purpose of this study is to identify miRNA expression profiles that would better predict prognosis of stage I NSCLC. MiRNAs extracted from 527 stage I NSCLC patients were profiled on the human miRNA expression profiling v2 panel (Illumina). The expression profiles were analyzed for their association with cancer subtypes, lung cancer brain metastasis and recurrence/relapse free survival (RFS). MiRNA expression patterns between lung adenocarcinoma and squamous cell carcinoma differed significantly with 171 miRNAs, including Let-7 family members and miR-205. Ten miRNAs associated with brain metastasis were identified including miR-145[*], which inhibit cell invasion and metastasis. Two miRNA signatures that are highly predictive of RFS were identified. The first contained 34 miRNAs derived from 357 stage I NSCLC patients independent of cancer subtype, whereas the second containing 27 miRNAs was adenocarcinoma specific. Both signatures were validated using formalin-fixed paraffin embedded and/or fresh frozen tissues in independent data set with 170 stage I patients. Our findings have important prognostic or therapeutic implications for the management of stage I lung cancer patients. The identified miRNAs hold great potential as targets for histology-specific treatment or prevention and treatment of recurrent disease.

Introduction

Stage I non-small cell lung cancer (NSCLC) patients are usually treated with surgical resection and ∼30% will eventually have a recurrence (1). It's necessary to develop biomarkers that would reliably identify those at high risk of relapse after surgery for modified adjuvant therapy to potentially improve survival. Global transcriptome analysis has been used extensively in identifying gene expression-based prognostic signatures, but none of them have been proven robust enough for clinical application (2). MicroRNAs (miRNAs) are small non-coding RNA molecules ∼19–30 nt that regulate ∼60% human genes. Mammalian miRNAs are generally encoded in introns in pre-messenger RNA (mRNA) or the 3′untranslated region of mRNA. They reduce gene expression by binding to complementary regions of mRNA and either blocking translation or degrading mRNA through the argonaut complex. The alteration of miRNA regulation has been implicated in carcinogenesis and disease progression (3,4). Generally, one miRNA is predicted to regulate several hundred genes. As a result, miRNA profiling could serve as a better classifier than gene expression profiling.

Several recent studies were published that correlated miRNA expression with outcomes in lung cancer using microarray (5–8). Yanaihara *et al.* indicated that high hsa-miR-155 expression was a significantly unfavorable prognostic factor in lung cancer. Raponi *et al.* (6) showed MiR-146b alone had the strongest prediction accuracy at 78% for stratifying prognostic groups of squamous cell carcinoma (SCC). Landi *et al.* (7) observed markedly different miRNA expression profiles between adenocarcinoma (ADC) and SCC. In their study, no miRNAs were associated with survival in ADC, whereas miR-25, miR-34c-5p, miR-191, let-7e and miR-34a strongly predicted survival in SCC. Patnaik *et al.* (8) demonstrated six miRNA classifiers (miR-200b, miR-30c-1, miR-510, miR-630, miR-124 and miR-585) for recurrence. These studies collectively showed the possibility of using miRNA expression profiles to predict outcomes in lung cancer patients.

Compared with mRNA expression studies, it's easier to study miRNA expression in archived formalin-fixed paraffin embedded (FFPE) specimens using microarray. Small miRNAs are not affected by formalin fixation-induced cross-linking, and this may result in FFPE miRNA expression levels similar to those found in fresh-frozen tissue (9). Archived human tissues in paraffin blocks are a rich research resource for miRNA as tissues found in most pathology departments are available only in FFPE states. If validated in appropriately conducted prospective studies, miRNA expression assays would be readily applicable and useful in clinical laboratories where FFPE tissue blocks are widely available but prospectively collected frozen tissues for mRNA analysis are not. Here, we conducted a miRNA profiling study on a cohort of 357 stage I NSCLC using FFPE specimens from Washington University School of Medicine (WUSM) in St Louis. We identified three miRNA signatures for recurrence/relapse-free survival (RFS) prediction in all subtypes of stage I NSCLC or in ADC or SCC specifically. An independent cohort of 170 stage I patients from Mayo Clinic was used to validate the effectiveness of these signatures. In addition, 10 miRNAs were identified to associate with brain metastasis.

Abbreviations: ADC, adenocarcinoma; AUC, area under the curve; cDNA, complementary DNA; FDR, False discovery rates; FFPE, formalin-fixed paraffin embedded; miRNA, microRNA; mRNA, messenger RNA; NSCLC, non-small cell lung cancer; RFS, recurrence/relapse-free survival; ROC, received operating characteristic; SCC, squamous cell carcinoma; WUSM, Washington University School of Medicine.

[†]These authors contributed equally to this work.

Materials and methods

Patients and tissue samples

The testing group included FFPE tumor tissues from 357 patients, which were diagnosed with stage I NSCLC and underwent surgical resection between 1990 and 2005 in WUSM. The cancer subtypes include ADC, SCC, large cell carcinoma, bronchioloalveolar carcinoma, adenosquamous carcinoma and large cell neuroendocrine carcinoma (Table I). Follow-up ended on December 2009 or at death. FFPE samples are made available from these patients. Tumor tissue blocks were sectioned, stained with routine hematoxylin and eosin and reviewed to identify areas of >70% tumor cellularity. Details regarding treatment and follow-up procedures were published previously (1). The validation group included 85 FFPE and 85 fresh-frozen tumor tissues from 170 stage I NSCLC in Mayo Clinic. Nearly, two-third of patients were ADCs (Table I). All patients had surgery between 1997 and 2008. All tumor blocks were reviewed and confirmed by a Mayo pathologist before RNA extraction. None of patients received any postoperative therapy in this study. The study was approved by the WUSM Human Studies Committee and Mayo Clinic Institute Review Board.

miRNA expression microarray experiments

For each case, hematoxylin and eosin-stained slides were used to identify areas containing at least 70% tumor nuclei with minimal necrosis. For FFPE tissues,

© The Author 2012. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/carcin/article/33/5/1046/2463491 by U.S. Department of Justice user on 16 August 2022

Table I. Clinical characteristics of subjects with stage I NSCLC in the analyzed datasets

	Washington University, dataset 2	Mayo Clinic, dataset 4	
	FFPE	FFPE	Frozen
Total number of samples	357	85	85
Mean age (range)	66 (34–88)	64 (41–85)	64 (27–89)
Sex			
Male	175	16	29
Female	182	69	56
Mean follow-up (years)			
Total RFS	4.36	4.43	4.53
Alive	5.79	4.71	4.84
Dead	1.70	3.30	3.65
Stage			
IA	174	60	46
IB	163	25	39
Histological type			
ADC	189	50	60
SCC	106	5	1
Others	62	30	24

the slides and blocks were aligned and the neoplastic areas were punched with sterile, single-use 1 mm punches (Miltex, Tuttlingen, Germany). RNA was then extracted from the tissue cores using standard techniques detailed in the Supplementary material, available at *Carcinogenesis* Online. For fresh-frozen tissues, samples were pulverized at −80°C and total cellular RNA was collected using TRIzol RNA isolation reagent (Invitrogen, Carlsbad, CA) and purified using the miRNeasy Mini Kit and RNase-free DNase Set (Qiagen, Valencia, CA) according to the manufacturer's protocols. Illumina human miRNA expression profiling v2 panel on the 12-sample beadchip was used for this study. The human v2 miRNA panel contains 1146 assays, for detecting >97% of the miRNAs described in the miRBase database (miRBase Release 12.05). Total RNA was polyadenylated using Poly(A) polymerase. Biotinylated oligo-deoxythymidine primer with a universal PCR sequence at its 5′ end was used to make labeled complementary DNA (cDNA). The biotinylated cDNA was hybridized to miRNA-specific oligos and the mixture was bound to streptavidin-conjugated paramagnetic particles to select the cDNA/oligo complexes. After the oligo annealing, miss-hybridized and non-hybridized oligos were washed away. DNA polymerase was used to extend the specific primer. The extended products were eluted and after PCR amplification, the labeled single-stranded products were hybridized to the beadchip overnight at 45°C. Fluorescence intensity was measured by the Illumina BeadArray Reader. The Gene Expression Omnibus accession number for microarray data of this study is GSE29135.

Quantitative real-time PCR for miRNA expression

Quantitative real-time PCR analysis of HS_142.1, hsa-miR-512-5p, hsa-miR-425, hsa-miR-615-5p, hsa-miR-668, has-miR-888, hsa-miR-34b, hsa-miR-34b*, hsa-miR-34c-3p and hsa-miR-34c-5p was done by using an miRNA-specific TaqMan MicroRNA Assay Kit (Applied Biosystems, Foster City, CA) and an Applied Biosystems 7500 Fast thermocycler system. One hundred nanograms of RNA was converted to cDNA using the ABI miRNA reverse transcription kits and miRNA-specific primers (Applied Biosystems). After reverse transcription, cDNA of patient and 10 μl of Universal PCR Master Mix without AmpErase Uracil N-glycosylase (Applied Biosystems) was added to commercially available PCR primers and 6-carboxyfluorescein-labeled TaqMan probes (Applied Biosystems). RNU44 and RNU6B small nuclear RNAs were used for normalization of input RNA/cDNA levels. Each measurement was done in duplicate and no-template controls were included for each assay. The variation of ΔC_T ($C_{T,target} - C_{T,control}$, where C_T is cycle number at which the fluorescence signal exceeds background) with template dilution was evaluated. The larger ΔC_T, the lower miRNA expression level is. *T*-test was performed to determine differences between the group with recurrence in 2 years and the group without recurrence in >7 years. The association of ΔC_T with RFS was assessed using Cox regression models and log-rank tests. A *P*-value of <0.05 was considered to indicate statistical significance.

Statistical and bioinformatics analysis

Data preprocessing. Principal component analysis was used to diagnose the batch effects and the efficacy of the procession of removing them. All the miRNA data were characterized by data source (WUSM and Mayo), sample beadchips (usually 12 samples per beadchip) and tissue types (FFPE and fresh-frozen tissues) (Plot A–C of Supplementary Figure S1, available at *Carcinogenesis* Online), suggesting large batch effects among microarrays. The source of the data is the major components contributing to batch effects; data points are largely clustered into two groups. There are also large system difference between some beadchips and tissue types. To reduce these potential batch effects, analysis of variance was used to remove the system bias from different institutes, sample beadchips and tissue types. These analyses were performed by the Partek Genomics Suite™ software package (www.partek.com). After data transformation, data points from different institutes, plates and tissue types were mixed uniformly (Plot D–F of Supplementary Figure S1, available at *Carcinogenesis* Online). It implies that the systematic biases were largely removed after adjustment and enhances confidence in the results. MiRNAs with low variance across samples (i.e. coefficient of variation <1% on the log scale) were filtered out to alleviate multiple test problems in the following data analyses.

Comparison of ADC with SCC. The miRNA expression differences between stage I lung ADC and squamous cell lung cancer were tested for each individual miRNAs using two-sided *t*-test.

Identification of miRNAs related to lung cancer brain metastasis. To identify miRNAs related to lung cancer brain metastasis, the expression was compared between lung cancer samples with brain metastasis samples and lung cancer samples with metastasis at other sites. Also, the comparison was performed between recurrence-free lung cancer for >5 years and lung cancer relapsed within 5 years. The differentially expressed miRNAs in the former, but not in the later comparison, were defined as miRNAs related to lung cancer brain metastasis ($P < 0.01$).

Identification of miRNAs associated with survival outcome. The outcome variable is RFS (the time from diagnosis to the first evidence of locoregional or distant recurrence). Patients were censored at the earliest of the following timepoints: death, development of second primary NSCLC, locoregional or distant recurrence or last medical contact. Univariate or multivariate Cox proportional hazards regression analyses adjusted for cancer stages (IA and IB) were performed for each miRNA using all 357 FFPE samples from WUSM. Similar analyses were also performed for each cancer subtype (ADC or SCC) separately. The proportional hazards assumption for stage was investigated by examining the scaled Schoenfeld residuals. Cancer stage displayed significant deviation from the proportional hazards assumption; therefore, it was taken as strata in the Cox proportional hazards model. False discovery rates (FDR) were estimated (10), which was implemented in R library code file (robust-fdr.R, downloaded from www.stjuderesearch.org/depts/biostats). miRNAs with FDR <0.05 for their expression in regression models were defined as RFS-related miRNAs.

Construction of predictor signatures based on miRNA expression. To develop cancer subtype non-specific or cancer subtype-specific miRNA signatures predictive of RFS respectively, survival analyses were performed based on all FFPE samples and only ADC or SCC patients from WUSM, respectively, using the RFS-related miRNAs. Partial Cox regression method was performed to construct predictive components (11). Then, these components were used in the Cox model for building predictive models for cancer patient survival. The principle components were chosen in the model to maximize discrimination ability in the model. The risk scores were calculated by $f(x) = \sum_{j=1}^{G} \beta_j^* (X_j - \bar{x}_j)$, where G represents the number of miRNAs, β_j^* represents the estimated coefficient of the jth miRNA, X_j represents expression levels of the jth miRNA in all the samples and $\bar{x}_j = 1/n \sum_{i=1}^{n} x_{ij}$, where n is sample size and x_{ij} is the expression level of miRNA j from sample i. All the samples were classified into high-and low-risk groups according to the risk scores. Patients with risk scores less than zero potentially have long-term RFS and those larger than zero have short-term RFS after surgical resection. To choose an appropriate subset of genes for signature, we carried out a forward selection and backward elimination procedure. The forward selection procedure includes (i) increase one gene each time and perform the partial Cox regression analysis as described above to obtain the prediction accuracy using the chosen subset of genes; (ii) add the gene which has the maximum prediction accuracy to gene set and (iii) repeat steps 1 and 2 until the prediction accuracy is maximized. These three steps were performed for each gene as the beginning, and then a series of gene sets were obtained. The frequencies of genes in these gene sets were ranked and the top 30% of genes were selected into backward elimination procedure. The backward elimination procedure includes (i) remove one gene each time and perform the partial Cox regression analysis to obtain the prediction accuracy using the chosen subset of genes; (ii) remove the gene which has the maximum prediction accuracy if removed in gene set and (iii) repeat steps 1 and 2 until the prediction accuracy is maximized. The prediction accuracy (discrimination ability) was assessed by Somers' Dxy rank correlation of

Downloaded from https://academic.oup.com/carcin/article/33/5/1046/2463491 by U.S. Department of Justice user on 16 August 2022

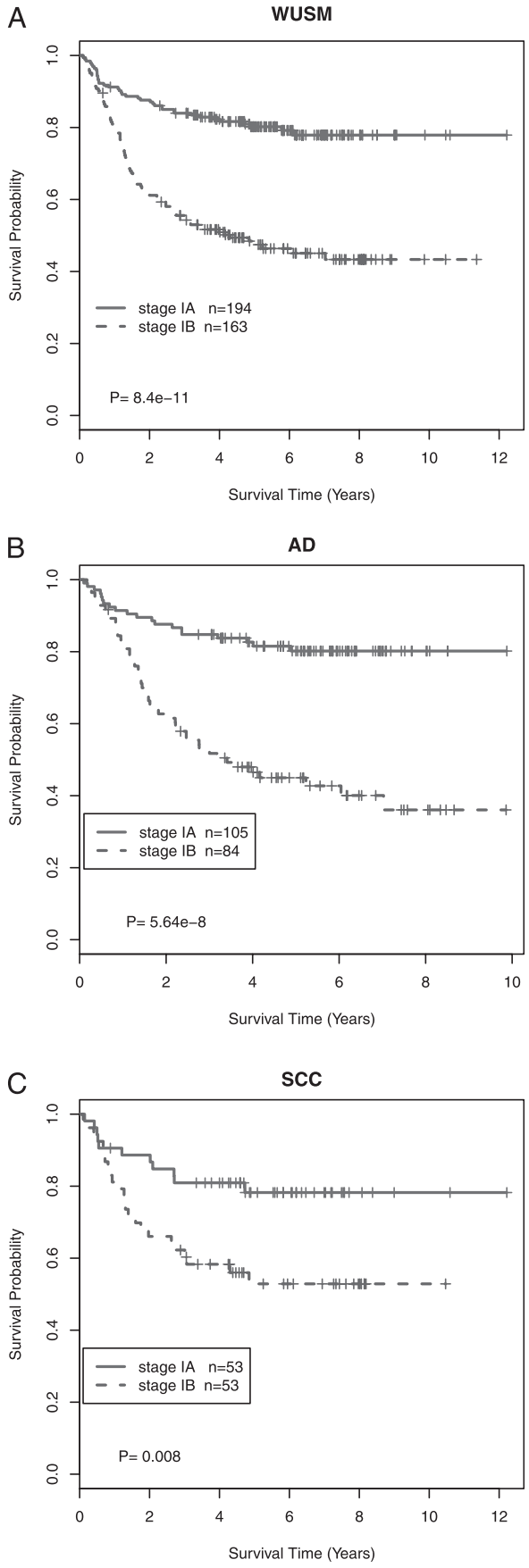

Fig. 1. Kaplan–Meier survival curves of RFS for stage I NSCLC. (**A**) All NSCLC patients; (**B**) ADC patients; (**C**) SCC patients. The *P*-values were calculated by log-rank test.

estimated risk score and real survival time. Somers' Dxy is related to the C-index by Dxy = 2(C−0.5). *C* is the corresponding received operating characteristic (ROC) curve area, which is a graphical representation of the pairs of false-positive test results (specificity) and true-positive test results (sensitivity) for the realizations of a quantitative test.

Comparison of miRNA predictors with stage information. Kaplan–Meier survival analyses were implemented to estimate the survival functions after the samples were classified into two risk groups according to their risk scores. Differences of the survival risk between the two risk groups were assessed using the Mantel–Haenszel log-rank test. The similar analyses were done using stage information (IA and IB). The larger area between the two risk groups and its associated smaller *P* value from the Mantel–Haenszel log-rank test implicate a better classification model.

To evaluate the predictive performance of the proposed survival signatures, time-dependent ROC analysis for censored data was performed and area under the curve (AUC) was employed as our criteria to assess survival predictions. The larger the AUC, the better the prediction model is. AUC = 0.5 indicates no predictive power, whereas AUC = 1 represents perfect predictive performance.

Confirmation of miRNA signatures in independent patient population. The miRNA signatures were tested in FFPE samples and fresh-frozen tissues from Mayo Clinic, separately. Kaplan–Meier survival analysis, Mantel-Haenszel log-rank test and time-dependent ROC analysis were performed using the risk scores which were calculated by a linear combination of the expression values for the miRNAs in the signature weighted by their estimated regression coefficients.

All these statistical analyses were implemented using the R statistical package (http://www.r-project.org/).

Bioinformatics analysis. DAVID bioinformatics resource was used to conduct Gene Ontology term and Kyoto Encyclopedia of Genes and Genomes pathway analysis on predicted targets of the miRNAs from the predictive signature. Most frequently represented pathways are assigned a *P* value calculated with a modified version of Fisher exact test (*P* value cutoff of <0.1), showing significance of the association as compared with a random list using the human genome as a background. TargetScanHuman v5.1 (http://www.targetscan.org/vert_50/) was used for predicting targets of known miRNAs in miRBase. Targets for novel miRNAs were predicted from the microT v3.0 (http://diana.cslab.ece.ntua.gr/microT/).

Results

miRNA differentially expressed between ADC and SCC

Overall, miRNA expression profiles strongly differentiated ADC from SCC in stage I patients. These include 145 miRNAs differentially expressed between the two subtypes ($P < 0.0001$ and fold change >1.2); 53 overexpressed in SCC and 92 ovrexpressed in ADC (Table II). Twenty-three of 92 miRNAs ovrexpressed in ADC were also observed in the study of Landi *et al.* (7), including multiple members of the let-7 family. Hsa-miR-205 was significantly overexpressed in SCC when compared with ADC ($P = 1.35 \times 10^{-26}$, 3.4-fold higher in SCC), which is consistent with that observed in previous studies (12,13).

miRNAs associated with survival outcome

The unadjusted survival analyses and the analyses adjusted by cancer stage (IA and IB) produced 369 and 338 significant miRNAs (FDR <0.01), respectively, for all cancer subtype combined, including 299 overlapped in both analyses (Supplementary Table S1 is available at *Carcinogenesis* Online). Due to the observed significant differences in miRNA expression between ADC and SCC, survival analyses were also conducted separately for ADC and SCC patients. The unadjusted survival analyses identified 263 miRNAs associated with RFS in ADC (FDR <0.05). One hundred and eighty-four miRNAs were associated with RFS in the adjusted analyses with $P < 0.05$ but not significant in global tests (FDR > 0.05). Comparing the sets of 263 and 184 miRNAs, 158 of them found overlapped (Supplementary Table S2 is available at *Carcinogenesis* Online). The unadjusted survival analyses identified 114 miRNAs associated with RFS in SCC patients ($P < 0.01$) with 43 of them significant at the global level (FDR < 0.05). One hundred and eighty-one miRNAs were associated with RFS in SCC patients by the adjusted survival analyses (FDR < 0.05). One hundred and thirteen of them were common in SCC (Supplementary Table S3 is available at *Carcinogenesis* Online).

Downloaded from https://academic.oup.com/carcin/article/33/5/1046/2463491 by U.S. Department of Justice user on 16 August 2022

Table II. miRNAs that significantly differentiate ADC from SCC

miRNAs	*P*-value	FC	miRNAs	*P*-value	FC	miRNAs	*P*-value	FC
hsa-miR-205	1.35×10^{-26}	−3.42	hsa-miR-378	1.04×10^{-7}	−1.22	hsa-miR-101*	4.12×10^{-6}	1.33
hsa-miR-944	7.51×10^{-29}	−3.20	hsa-miR-942	9.64×10^{-6}	−1.22	hsa-miR-32	5.47×10^{-5}	1.33
hsa-miR-149	1.21×10^{-9}	−1.90	hsa-miR-663	5.91×10^{-5}	−1.21	hsa-miR-624*	1.22×10^{-7}	1.33
hsa-miR-584	4.36×10^{-12}	−1.76	hsa-miR-618	1.15×10^{-5}	−1.20	hsa-miR-132	2.39×10^{-5}	1.33
HS_170	1.61×10^{-07}	−1.66	hsa-miR-34a*	9.88×10^{-11}	1.90	hsa-miR-642	4.05×10^{-5}	1.33
HS_74	7.09×10^{-06}	−1.61	hsa-miR-375	2.04×10^{-13}	1.82	***hsa-miR-146b-5p***	1.33×10^{-11}	1.32
HS_116	3.56×10^{-10}	−1.61	***hsa-miR-29b-1******	4.60×10^{-14}	1.81	hsa-miR-652	1.50×10^{-06}	1.32
hsa-miR-1268	1.71×10^{-6}	−1.60	hsa-miR-511	5.96×10^{-9}	1.79	***hsa-miR-29a***	3.60×10^{-13}	1.31
hsa-miR-875-5p	1.68×10^{-24}	−1.59	hsa-miR-503	1.77×10^{-9}	1.75	hsa-miR-193a-3p	2.79×10^{-5}	1.31
hsa-miR-769-3p	7.03×10^{-7}	−1.58	***hsa-miR-29b-2******	8.28×10^{-13}	1.74	hsa-miR-155	1.72×10^{-6}	1.31
HS_303_b	3.11×10^{-10}	−1.57	hsa-miR-497	1.26×10^{-8}	1.70	Has-miR-361-5p	4.86×10^{-5}	1.31
hsa-miR-124	1.66×10^{-6}	−1.57	hsa-miR-542-5p	7.27×10^{-8}	1.66	hsa-miR-744	5.26×10^{-5}	1.31
hsa-miR-518b	1.65×10^{-10}	−1.56	hsa-miR-135a	2.82×10^{-6}	1.64	hsa-miR-424*	3.86×10^{-5}	1.30
solexa-1460-671	4.31×10^{-8}	−1.56	hsa-miR-338-3p	1.61×10^{-7}	1.62	hsa-miR-362-3p	1.61×10^{-7}	1.30
hsa-miR-1290	8.73×10^{-6}	−1.55	hsa-miR-181c	3.02×10^{-10}	1.62	hsa-miR-15a	3.53×10^{-6}	1.30
hsa-miR-519d	1.34×10^{-8}	−1.54	hsa-miR-628-3p	2.43×10^{-9}	1.61	hsa-miR-100*	3.62×10^{-6}	1.29
HS_287	5.38×10^{-6}	−1.51	hsa-miR-1	1.52×10^{-5}	1.60	***hsa-miR-151-5p***	1.07×10^{-6}	1.29
HS_56	1.44×10^{-6}	−1.51	***hsa-let-7e***	2.05×10^{-9}	1.54	***hsa-let-7g****	1.80×10^{-6}	1.29
hsa-miR-1269	1.10×10^{-5}	−1.50	hsa-miR-135b	3.15×10^{-5}	1.53	hsa-miR-29b	1.11×10^{-10}	1.29
hsa-miR-1181	3.44×10^{-8}	−1.50	hsa-miR-326	2.90×10^{-17}	1.52	hsa-miR-450a	6.56×10^{-7}	1.29
hsa-miR-196b	8.44×10^{-9}	−1.48	hsa-miR-92b	1.33×10^{-11}	1.52	hsa-miR-132*	6.21×10^{-6}	1.28
hsa-miR-1202	3.27×10^{-16}	−1.48	hsa-miR-450b-5p	1.05×10^{-7}	1.52	***hsa-let-7c***	9.12×10^{-5}	1.28
hsa-miR-129-3p	$2.57 \times ^{-06}$	−1.46	hsa-miR-140-5p	1.07×10^{-6}	1.51	***hsa-miR-29c***	2.44×10^{-9}	1.28
HS_17	1.63×10^{-5}	−1.46	hsa-miR-148b	4.13×10^{-5}	1.50	hsa-miR-339-3p	4.14×10^{-7}	1.28
HS_155	1.33×10^{-5}	−1.45	hsa-miR-22*	7.34×10^{-6}	1.48	***hsa-let-7g***	6.20×10^{-6}	1.28
hsa-miR-128b:9.1	5.29×10^{-8}	−1.44	***hsa-let-7d***	1.53×10^{-5}	1.47	hsa-miR-126*	2.43×10^{-5}	1.27
HS_284.1	1.67×10^{-8}	−1.44	***hsa-miR-29a****	2.81×10^{-10}	1.44	solexa-3126-285	1.38×10^{-5}	1.27
HS_254	6.00×10^{-5}	−1.43	hsa-miR-139-5p	1.93×10^{-6}	1.43	hsa-miR-502-3p	2.28×10^{-7}	1.27
HS_257	9.78×10^{-5}	−1.39	hsa-miR-660	6.10×10^{-6}	1.43	hsa-miR-181b	3.76×10^{-12}	1.27
hsa-miR-1180	6.78×10^{-6}	−1.39	***hsa-miR-195***	4.18×10^{-6}	1.41	***hsa-miR-181a****	1.00×10^{-8}	1.27
hsa-miR-128a:9.1	5.58×10^{-6}	−1.38	***hsa-let-7f***	3.40×10^{-6}	1.40	***hsa-miR-26b***	1.14×10^{-6}	1.26
HS_149	4.72×10^{-5}	−1.38	hsa-miR-542-3p	9.70×10^{-9}	1.40	hsa-miR-361-3p	4.53×10^{-6}	1.26
hsa-miR-492	2.11×10^{-5}	−1.38	hsa-miR-548b-3p	1.58×10^{-7}	1.40	***hsa-miR-26a***	2.82×10^{-7}	1.26
solexa-9081-91	4.48×10^{-6}	−1.38	hsa-miR-34a	2.57×10^{-13}	1.40	hsa-miR-768-5p:11.0	2.08×10^{-10}	1.25
HS_166.1	4.62×10^{-7}	−1.38	hsa-miR-508-3p	1.88×10^{-5}	1.39	solexa-8211-102	1.24×10^{-5}	1.25
hsa-miR-671:9.1	7.98×10^{-5}	−1.37	hsa-miR-628-5p	5.33×10^{-8}	1.39	solexa-51-13984	7.72×10^{-5}	1.25
hsa-miR-378*	5.55×10^{-8}	−1.37	hsa-miR-625	1.37×10^{-6}	1.39	hsa-miR-532-5p	2.04×10^{-5}	1.24
HS_78	3.99×10^{-5}	−1.37	hsa-miR-140-3p	2.96×10^{-9}	1.38	***hsa-let-7a***	1.31×10^{-5}	1.24
hsa-miR-576-3p	2.48×10^{-5}	−1.37	hsa-miR-340	4.29×10^{-6}	1.37	hsa-miR-130a	7.06×10^{-6}	1.24
HS_120	6.11×10^{-5}	−1.36	hsa-miR-212	1.49×10^{-5}	1.37	***hsa-miR-29c****	1.05×10^{-7}	1.24
HS_55.1	1.20×10^{-5}	−1.35	hsa-miR-101	3.47×10^{-5}	1.36	***hsa-miR-30d***	1.05×10^{-9}	1.21
HS_53	1.42×10^{-5}	−1.34	***hsa-miR-26b****	3.95×10^{-7}	1.36	hsa-miR-30e*	1.03×10^{-5}	1.21
hsa-miR-873	8.32×10^{-5}	−1.33	hsa-miR-489	1.00×10^{-5}	1.35	hsa-miR-664*	2.67×10^{-6}	1.21
hsa-miR-1293	1.15×10^{-6}	−1.32	hsa-miR-424	1.61×10^{-5}	1.35	hsa-miR-30c	5.63×10^{-5}	1.21
solexa-9655-85	1.32×10^{-5}	−1.31	hsa-miR-664	1.29×10^{-7}	1.35	***hsa-miR-30b***	2.90×10^{-7}	1.21
HS_169	3.14×10^{-5}	−1.30	hsa-miR-653:9.1	2.75×10^{-8}	1.34	hsa-miR-24-2*	3.24×10^{-5}	1.20
hsa-miR-1249	1.08×10^{-6}	−1.28	hsa-miR-768-3p:11.0	1.49×10^{-7}	1.34	hsa-miR-189:9.1	3.42×10^{-7}	1.20
HS_188	3.55×10^{-5}	−1.24	hsa-miR-532-3p	5.90×10^{-5}	1.33	***hsa-miR-491-5p***	5.82×10^{-5}	1.20
hsa-miR-1307	8.98×10^{-6}	−1.22						

Bold and italic miRNAs were also identified in the study of Landi *et al.* (2010), *Clin Cancer Res.* FC was fold changes calculated by dividing average expression levels of ADC by those of SCC.

Signatures to predict survival using miRNA expression profile

To determine whether a subset of the RFS-related miRNAs can be used to predict survival of stage I NSCLC, risk scores were derived from survival analyses using the partial Cox regression method. Kaplan–Meier survival analyses were performed after the samples were classified into high-and low-risk groups according to the risk scores. Using all 357 FFPE tissues from WUSM as training samples, we identified a miRNA signature including 34 miRNAs (Signature I) to predict 5 years RFS in stage I NSCLC. We also identified a 27-miRNA signature (Signature II) from stage I lung ADC patients and a 17-miRNA signature (Signature III) from stage I SCC patients, respectively. Signature I was identified from the miRNA list of adjusted survival analyses by forward and backward algorithm; Signature II was identified from the miRNA list of unadjusted survival analyses by forward and backward algorithm and Signature III was identified from miRNA list of adjusted survival analyses by forward algorithm. The associations of expression levels with survival for each miRNAs in the signatures are shown in Table II and Table III. Four miRNAs overlapped in Signature I and II, including HS_142.1, hsa-miR-425, hsa-miR-615-5p and hsa-miR-512-5p and another five miRNAs overlapped in Signature I and III, including HS_215, HS_188, hsa-miR-892b, hsa-34a* and hsa-miR-888.

Kaplan–Meier survival curves could distinguish RFS among stage IB from stage IA NSCLC to some extent (Figure 1A–C). However, the high- and low-risk groups are more significantly different in their RFS when using miRNA expression profiles ($P \leq 1.53 \times 10^{-8}$), which showed larger area between two groups and smaller *P*-value from the Mantel–Haenszel log-rank test (Figure 2A–C). To estimate the predictive power of survival analysis using expression profiles, time-dependent ROC analyses for 5 years RFS were performed for these signatures. Cox model with risk scores estimated by miRNA expression gave the better

Downloaded from https://academic.oup.com/carcin/article/33/5/1046/2463491 by U.S. Department of Justice user on 16 August 2022

Table III. miRNAs in the signature predictive of stage I NSCLC

miRNAs	*P*-value	FDR	Hazard ratio	miRNAs	*P*-value	FDR	Hazard ratio
Signature independent of cancer subtype							
HS_215	**1.54×10^{-9}**	**0.000000672**	**0.61 (0.52-0.72)**	solexa-3044-295	0.0014	0.0029	0.71 (0.58–0.88)
solexa-826-1288	1.35×10^{-7}	0.00000611	0.48 (0.37–0.63)	**HS_188**	**0.0015**	**0.0029**	**0.67 (0.52-0.86)**
hsa-miR-888	**3.36×10^{-7}**	**0.00000946**	**0.36 (0.25-0.54)**	HS_303_a	0.0017	0.0033	1.82 (1.25–2.65)
HS_142.1	1.16×10^{-5}	0.0000783	0.49 (0.36–0.68)	hsa-miR-519e	0.0017	0.0033	0.46 (0.29–0.75)
hsa-miR-1207-3p	4.54×10^{-5}	0.000226	0.55 (0.41–0.73)	hsa-miR-124	0.002	0.0037	0.77 (0.65–0.91)
hsa-miR-603	0.000057	0.00027	0.68 (0.57–0.82)	hsa-miR-941	0.0036	0.0063	1.38 (1.11–1.72)
hsa-miR-769-3p	0.0001	0.0004	0.76 (0.66–0.87)	hsa-miR-34a	0.0038	0.0065	1.63(1.17–2.26)
HS_6	0.0004	0.0012	0.72 (0.6–0.86)	hsa-miR-494	0.0038	0.0066	0.82 (0.72–0.94)
hsa-miR-31*	0.0004	0.0012	1.21 (1.09–1.35)	hsa-miR-222	0.0042	0.0069	1.5 (1.14–1.98)
hsa-miR-892b	**0.0004**	**0.0011**	**0.52 (0.36-0.74)**	**hsa-miR-34a***	**0.0044**	**0.0072**	**1.23 (1.07-1.43)**
HS_15.1	0.0006	0.0016	0.62 (0.47–0.82)	hsa-miR-205	0.0048	0.0077	0.85 (0.77–0.95)
HS_231	0.0006	0.0014	0.55 (0.4–0.78)	hsa-miR-151-5p	0.0049	0.0077	1.53 (1.14–2.05)
hsa-miR-185*	0.0008	0.0018	0.67 (0.53–0.85)	solexa-9124-90	0.0053	0.008	0.6 (0.42–0.86)
hsa-miR-425	***0.0008***	***0.0018***	***2.04 (1.35-3.09)***	hsa-miR-132	0.0054	0.0081	1.38 (1.1–1.72)
hsa-miR-615-5p	***0.0008***	***0.0019***	***0.5 (0.33-0.75)***	hsa-miR-192*	0.0055	0.0083	1.32 (1.08–1.6)
hsa-miR-940	0.001	0.0023	0.56 (0.39–0.79)	hsa-miR-455-3p	0.0059	0.0086	1.29 (1.08–1.54)
hsa-miR-512-5p	***0.0012***	***0.0025***	***0.61 (0.46-0.82)***	HS_122.1	0.0066	0.0094	0.4 (0.2–0.77)
Signature of ADC							
hsa-miR-615-5p	0.000024	0.00596	0.30 (0.17–0.52)	HS_285	0.005	0.028	0.42 (0.23–0.77)
hsa-miR-34b	0.0000699	0.00683	0.65 (0.52–0.8)	hsa-miR-377*	0.005	0.0281	0.53 (0.34–0.83)
has-miR-1248	0.0000875	0.00743	0.65 (0.53–0.81)	hsa-miR-588	0.0058	0.0292	0.53 (0.34–0.83)
HS_59	0.0001	0.0078	0.46 (0.31–0.69)	hsa-miR-34c-5p	0.0082	0.0349	0.78 (0.64–0.94)
HS_283_a	0.0002	0.0084	0.39 (0.24–0.64)	HS_47	0.0092	0.0363	0.41 (0.21–0.8)
hsa-miR-34c-3p	0.0002	0.008	0.66 (0.53–0.82)	HS_43.1	0.0106	0.0381	0.58 (0.39–0.88)
hsa-miR-512-3p	0.0005	0.0118	0.41 (0.25–0.68)	hsa-miR-512-5p	0.0125	0.0401	0.58 (0.38–0.89)
HS_78	0.0014	0.0176	0.63 (0.47–0.83)	HS_74	0.0147	0.0441	0.77 (0.62–0.95)
hsa-miR-34b*	0.0014	0.0177	0.77 (0.66–0.91)	hsa-miR-658	0.0163	0.0454	0.60 (0.40–0.91)
hsa-miR-934	0.0022	0.0213	0.41 (0.23–0.72)	hsa-miR-548b-3p	0.0174	0.0459	0.66 (0.47–0.93)
HS_60	0.0036	0.0263	0.49 (0.31–0.79)	HS_142.1	0.0176	0.046	0.58 (0.36–0.91)
hsa-miR-572	0.0041	0.0267	0.65 (0.48–0.87)	hsa-miR-425	0.019	0.0463	2.26 (1.14–4.46)
hsa-miR-662	0.0044	0.0272	0.61 (0.44–0.86)	hsa-miR-1255a	0.0209	0.0479	0.58 (0.36–0.92)
HS_187	0.0045	0.0272	0.52 (0.33–0.82)				
Signature of SCC							
hsa-miR-888	0.0000178	0.00477	0.15 (0.07–0.36)	HS_159	0.00431	0.02569	0.39 (0.20–0.74)
hsa-miR-801:9.1	0.000019	0.00564	0.37 (0.24–0.59)	HS_188	0.00451	0.0263	0.47 (0.28–0.79)
HS_215	0.00011	0.00756	0.54 (0.39–0.74)	hsa-miR-671:9.1	0.00686	0.03481	0.65 (0.47–0.89)
HS_197	0.00031	0.0088	0.18 (0.07–0.46)	HS_65	0.00741	0.03612	0.56 (0.37–0.86)
HS_263.1	0.00035	0.00884	0.12 (0.04–0.39)	hsa-miR-193a-3p	0.00785	0.03736	1.82 (1.17–2.82)
HS_83.1	0.00036	0.00886	0.23 (0.10–0.52)	solexa-499-2217	0.00795	0.03798	0.20 (0.06–0.66)
hsa-miR-892b	0.00091	0.01183	0.30 (0.15–0.61)	hsa-miR-595	0.01025	0.04503	0.42 (0.21–0.81)
hsa-miR-939	0.00115	0.01254	0.10 (0.02–0.39)	hsa-miR-34a*	0.01032	0.04507	1.52 (1.10–2.11)
hsa-miR-1180	0.00362	0.02325	0.50 (0.32–0.8)				

Italic were overlapped miRNAs in the signature of ADC. Bold were also in the signature of SCC.

predictive performance with AUCs from 86 to 91%, whereas stage information results in lower AUCs from 64 to 69% (Figure 2D–F). This clearly demonstrates that miRNA expression signatures combined with stage information has higher classification power than the staging method to predict survival of stage I NSCLC patients.

To validate the miRNA expression results from microarray experiments, the relative expression of 10 miRNAs associated with RFS was determined by quantitative real-time–PCR analysis on 12 samples with recurrence within 2 years and 12 samples without recurrence for >7 years. In these 10 miRNAs, most of them are overlapped between two signatures and four of them are members in hsa-miR34 family. HS_215 (target mature sequence: ATGTCACTCGGCTCGGCCCAC) was replaced by hsa-miR-668 (target mature sequence: TGTCACTCGGCTCGGCCCAC) because the primer of HS_215 is not available. We confirmed the expression results for all these miRNAs except hsa-miR-425 (Supplementary Figure S2 is available at *Carcinogenesis* Online). The ΔC_T had significant difference between recurrence and no-recurrence groups for eight miRNAs ($P < 0.05$). Hsa-miR-888 is marginally significant ($P = 0.055$). Samples were also divided into high and low expression groups according to the ΔC_T of each miRNA. Kaplan–Meier survival curves showed that the high and low expression groups are significantly different in their RFS (*P* value from 3.99e-5 to 0.048); the high expression groups have low risk of recurrence, which is consistent with the results from microarray data that all these miRNAs had negative hazard ratios.

Confirmation of miRNA expression signatures in an independent dataset

The robustness of these signatures in predicting survival in stage I lung cancer were further tested using miRNA expression data from Mayo Clinic, including 85 FFPE and 85 fresh-frozen tissues, in which 110 samples were ADC and 6 were SCC. Due to very limited SCC, only Signature I from all stage I NSCLC patients and Signature II from ADC were tested for validation. For this dataset, long- and short-term survival patients were not well distinguishable by stage information for all stage I patients ($P = 0.03$) (Figure 3A). Kaplan–Meier analysis using miRNA Signature I demonstrated that the high- and low-risk groups are much more significantly different in their RFS ($P = 5.67 \times 10^{-7}$) (Figure 3B). Cox model with risk scores estimated by miRNA expression gave the better predictive performance with the AUCs of 81%, but Cox model with stage information results in AUCs of 60% only (Figure 3C). Kaplan–Meier analysis using miRNA Signature II, which was generated from FFPE ADC data, also had better performance than using stage information only in 110 ADC samples

Downloaded from https://academic.oup.com/carcin/article/33/5/1046/2463491 by U.S. Department of Justice user on 16 August 2022

(Figure 3D–F). Interestingly, although the signatures were derived from FFPE tissues, both signatures are highly predictive of survival of stage I NSCLC patients using either FFPE or fresh-frozen tissues.

miRNAs associated with brain metastasis

In the patients from WUSM, 147 lived for >5 years without recurrence, 25 had brain metastasis and 100 had metastasis to other sites. Sixteen differentially expressed miRNAs were identified comparing no-recurrence group with brain metastasis group ($P < 0.01$). Excluding the miRNAs differentially expressed between no-recurrence group and other sites metastasis group, 10 miRNAs remained which correlate with brain metastasis, including hsa-miR-450b-3p, hsa-miR-29c*, hsa-miR-145*, hsa-miR-148a*, hsa-miR-1, hsa-miR-30d, hsa-miR-187, hsa-miR-218, hsa-miR-708* and hsa-miR-375 (Supplementary Table S4 is available at *Carcinogenesis* Online).

Discussion

Recently, there were several studies published to correlate miRNA expression with outcomes in NSCLC using microarray (6–8). For example, the mirVana Bioarray (version 2; Ambion) that contains 328 human miRNA probes was used to identify lung SCC miRNA signatures using 61 snap-frozen lung SCC in Raponi *et al.*'s study (6). Landi *et al.* (7) analyzed 440 human miRNAs in 290 FFPE tissues of NSCLC patients including both ADC and SCC but only 40% samples from stage I patients. Patnaik *et al.* (8) applied Exiqon miRCURY locked nucleic acid microarrays 10.0 including 752 human miRNAs to obtain expression profiles for 37 cases with recurrence and 40 cases without recurrence of stage I NSCLC. In our study, we used 357 FFPE samples to construct miRNA signatures for prognosis in stage I NSCLC using Illumina human miRNA expression profiling v2 panel. Comparing with the three published studies, our study is the largest study of miRNA expression focusing on stage I lung cancer to date. The platform we used includes 1146 human miRNAs, which contained large majority of human miRNAs. The larger sample size of ADC and SCC can reduce false positives and increase statistical power in detecting survival-related miRNAs and cancer-subtype specific signatures. The most complete set of human miRNAs presenting on our miRNA platform increased chance to identify a stable signature for predicting the outcome. As a result, two miRNA expression signatures were identified that can accurately predict which stage I lung cancer patients may benefit from more aggressive therapy. The first signature contains 34 miRNAs derived from 357 stage I NSCLC without regarding cancer subtypes, whereas the second one is ADC-specific and contains 27 miRNAs derived from 189 stage I ADCs. This is highly significant because patients diagnosed with stage I NSCLC have variable prognoses. The generalization of our miRNA signatures was demonstrated in an independent dataset from Mayo Clinic, which consisted of half FFPE and half fresh-frozen tissues, and most of the samples are ADCs. Although the signatures were derived from FFPE tissues, they are highly predictive of survival of stage I NSCLC using either FFPE or fresh-frozen tissues. These results demonstrate the possibility of using miRNA signatures generated from FFPE samples to predict patient survival in the clinic. The fact that miRNA expression levels can be determined using FFPE specimens allows immediate and widespread use in the clinic as these results are confirmed in large independent population samples. We believe that this is an important and novel finding that has not been convincingly reported previously and has important implications for the clinical management of lung cancer patients. Several miRNA studies have been performed in lung SCC (6,7); however, we couldn't use their data to validate our signature for SCC because very few miRNAs were overlapped between miRNA array platforms. In general, fewer miRNAs are associated with RFS in ADC than SCC in our survival analysis. The SCC signature tends to have higher AUC than the ADC signature in our training datasets, which is consistent with a recent study (7).

Our miRNA signatures consisted of several miRNAs whose targets are involved in multiple important cancer-related pathways (Supplementary Table S5 is available at *Carcinogenesis* Online). Engineered knockdown of miR-31 substantially represses lung cancer cell growth and tumorigenicity in a dose-dependent manner (14). More strikingly, several laboratories have reported that members of miR-34 family are directly regulated by p53, which induces apoptosis, cell cycle arrest and senescence (15–18). This reinforces the awareness that miR-34 are central mediators of p53 function. MiR-34 family members may be the key players in tumor development by being located centrally within the p53 tumor suppressor network, and there is a balanced state between epigenetic modification and p53 regulating the expression of miR-34. Landi *et al.* (7) reported a 5-miR SCC signature including hsa-miR-34c-5p and hsa-miR-34a which both associate with poor survival. In our data, Signature II predictive of stage I ADC suggests that lower expression of hsa-miR-34b, hsa-miR-34b*, hsa-miR-34c-3p and hsa-miR-34c-5p are associated with poor survival with hazards ratio <1 which were confirmed by real-time–PCR. However, hsa-miR-34a and hsa-miR-34a* did not show the association with poor survival on RFS. This discrepancy may be attributable to different roles of hsa-miR-34 members in stage I and advanced stages (stages II–IV) of lung cancer. Fifty-eight percent of samples in the Landi *et al.*'s study are of advanced stages (19).

MiRNAs play important roles in gene regulations. We thus surveyed the expressions of predicted miRNA targets (Supplementary Table S5 is available at *Carcinogenesis* oOnline) in previous microarray studies of lung cancer. E2F3, histone deacetylase-2, CDT1, RFC2, KIF11 and TBL1X were on the list of signatures of the study of Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma (20), whereas SLIT1 and THBS1 were specifically associated with overall survival and RFS in stage I patients of that study (20). NTRK3, PCDHGA12 and PRKACA were on the list of 64-gene signature for survival of stage I NSCLC in our meta-analysis (21). Insulin receptor were on the list of 37-gene survival signature for lung ADC (22). Both of protein and gene expression level of SOD2 were related to survival of lung ADC (23). Loss of expression of SDC1 is associated with biologic aggressiveness and poor outcome for NSCLC patients (24). PCDHGC3 and PTPRC were on the list of a recurrence signature in lung ADC and SCC, respectively (25,26). CCND2 and FN1 were related to survival of lung SCC (27,28). It is interesting to follow up the miRNAs whose targets are also found to correlate to survival in lung cancer.

Interestingly, a large set of miRNAs was observed differentially expressed between lung ADC and SCC. This may suggest different miRNA-mediated signaling pathways involved in the pathogenesis of these two histologies. Multiple members of the Let-7 family were downregulated in SCC as compared with ADC. Let-7 was shown previously to regulate the expression of the RAS lung cancer oncogenes HRAS, KRAS and NRAS (19). The 3′untranslated regions of the human RAS genes contain multiple let-7 complementary sites, allowing Let-7 to regulate RAS expression. Let-7 expression is lower in lung tumors than in normal lungs, whereas RAS protein is significantly higher in lung tumors, providing a possible mechanism of Let-7 in lung cancer (19). KRAS accounts for 90% of RAS mutations in lung ADCs (29) and KRAS mutations are uncommon in lung SCC (30). This implies that aberrant expression of Let-7 may contribute to lung SCC, whereas KRAS mutations counteract the suppressing effects of Let-7 on lung ADCs.

More than 50% of brain metastases are associated with NSCLC. In our study, 10 miRNAs were identified to associate with brain metastasis. Among these, miR-145 has previously been shown to suppress cell invasion and metastasis by directly targeting the metastasis gene mucin 1 (MUC1) (31). A MUC1-secreting human breast cancer cell line MA11 established with cells isolated from a bone marrow sample using immunomagnetic beads conjugated to the anti-MUC1 antibody BM-2 showed a selective preference for metastasizing to the brain in athymic nude mice (32). MiR-29c is downregulated so as to upregulate mRNAs encoding extracellular matrix proteins that are involved in metastasis in nasopharyngeal carcinomas (33). MiR-218 was found to inhibit invasion and metastasis of gastric cancer by targeting the Robo1 receptor (34).

Downloaded from https://academic.oup.com/carcin/article/33/5/1046/2463491 by U.S. Department of Justice user on 16 August 2022

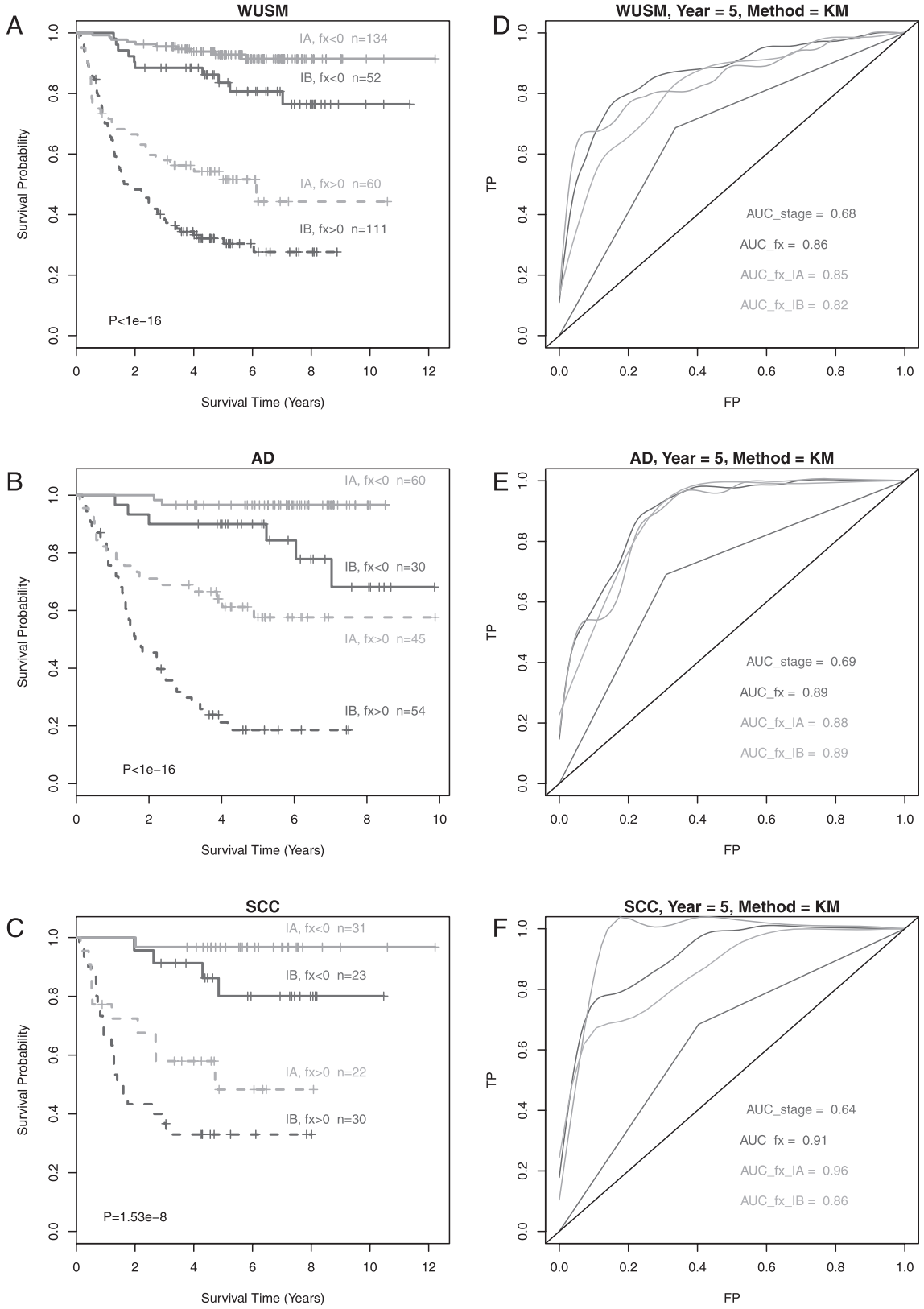

Fig. 2. Survival analyses of stage I NSCLC. (**A**–**C**) Kaplan–Meier survival curves of RFS using miRNA signatures: (A) signature from all stage I NSCLC patients (Signature I); (B) signature from stage I ADC patients (Signature II); (C) signature from stage I SCC patients (Signature III). (**D**–**F**) AUC for estimating 5 years RFS using survival models based on stage information or miRNA expression data respectively: (D) comparison of staging method with risk scores estimated by Signature I in all stage I NSCLC patients; (E) comparison of staging method with risk scores estimated by Signature II in stage I ADC patients; (F) comparison of staging method with risk scores estimated by Signature III in stage I SCC patients. FP, false positive; TP, true positive.

Downloaded from https://academic.oup.com/carcin/article/33/5/1046/2463491 by U.S. Department of Justice user on 16 August 2022

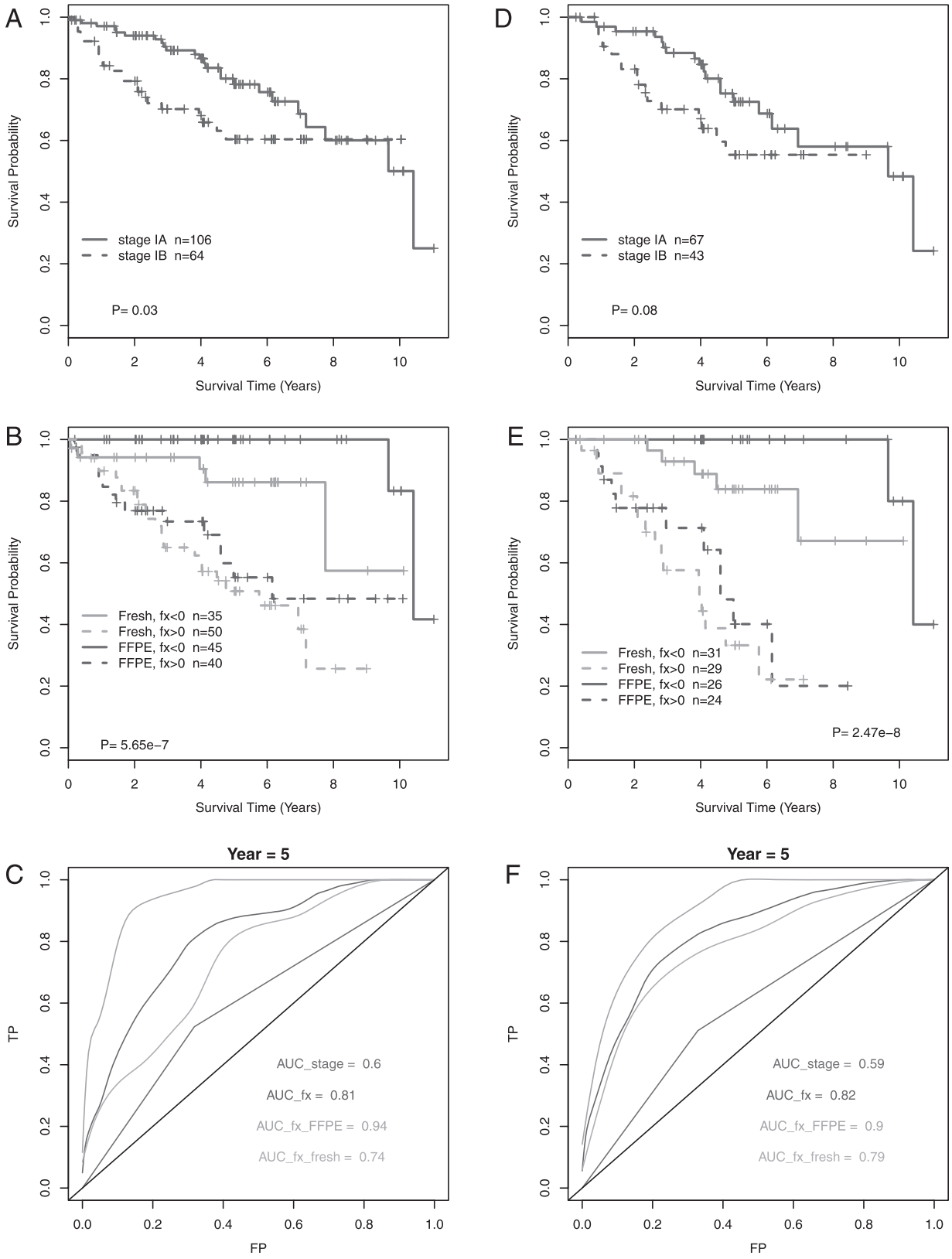

Fig. 3. Validation of the two signatures in an independent testing set from Mayo Clinic. (**A**), (**B**) and (**C**) are validation of the 34-miRNA signature (Signature I) in 170 stage I patients. (**D**), (**E**) and (**F**) are validation of the 27-miRNA ADC-specific signature (Signature II) in 110 stage I ADC patients. A and D are Kaplan–Meier analyses of RFS using staging method. B and E are Kaplan–Meier analyses of RFS using risk scores estimated by miRNA signature. C and F are ROC analyses for estimating 5 years RFS based on stage information or miRNA signature.

Downloaded from https://academic.oup.com/carcin/article/33/5/1046/2463491 by U.S. Department of Justice user on 16 August 2022

In summary, we have identified two miRNA signatures that are highly predictive of survival of stage I NSCLC patients. This has important prognostic or therapeutic implications for the future management of NSCLC patients. We also identified a large set of miRNAs including Let-7 family members and miR-205 whose expression profiles strongly differed between ADC and SCC. Furthermore, 10 miRNAs were identified to associate with brain metastasis. These miRNAs hold great potential as targets for histology-specific treatment or preventing and treating recurrent diseases.

Supplementary material

Supplementary Tables S1–S5 and Figures S1–S2 can be found at http://carcin.oxfordjournals.org/.

Funding

This work was supported by National Institutes of Health grant 1R01CA129533-01A1 (to M.Y.), fund from A Healthier Wisconsin (to Y.L. and P.-Y.L.) and a fund from Department of Laboratory Medicine and Pathology, Mayo Clinic (to L.W.).

Acknowledgements

We are indebted to all the patients and their physicians.

Conflict of Interest Statement: None declared.

References

1. Goodgame,B. *et al*. (2008) A clinical model to estimate recurrence risk in resected stage I non-small cell lung cancer. *Am. J. Clin. Oncol.*, **31**, 22–28.
2. Subramanian,J. *et al*. (2010) Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J. Natl Cancer Inst.*, **102**, 464–474.
3. Calin,G.A. *et al*. (2006) MicroRNA signatures in human cancers. *Nat. Rev. Cancer*, **6**, 857–866.
4. Johnson,C.D. *et al*. (2007) The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Res.*, **67**, 7713–7722.
5. Yanaihara,N. *et al*. (2006) Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*, **9**, 189–198.
6. Raponi,M. *et al*. (2009) MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. *Cancer Res.*, **69**, 5776–5783.
7. Landi,M.T. *et al*. (2010) MicroRNA expression differentiates histology and predicts survival of lung cancer. *Clin. Cancer Res.*, **16**, 430–441.
8. Patnaik,S.K. *et al*. (2010) Evaluation of microRNA expression profiles that may predict recurrence of localized stage I non-small cell lung cancer after surgical resection. *Cancer Res.*, **70**, 36–45.
9. Zhang,X. *et al*. (2008) An array-based analysis of microRNA expression comparing matched frozen and formalin-fixed paraffin-embedded human tissue samples. *J. Mol. Diagn.*, **10**, 513–519.
10. Pounds,S. *et al*. (2006) Robust estimation of the false discovery rate. *Bioinformatics*, **22**, 1979–1987.
11. Li,H. *et al*. (2004) Partial Cox regression analysis for high-dimensional microarray gene expression data. *Bioinformatics*, **20** (suppl. 1), i208–i215.
12. Lebanony,D. *et al*. (2009) Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from nonsquamous non-small-cell lung carcinoma. *J. Clin. Oncol.*, **27**, 2030–2037.
13. Fassina,A. *et al*. (2011) Classification of non-small cell lung carcinoma in trans-thoracic needle specimens using microRNA expression profiling. *Chest*, **140**, 1305–1311.
14. Liu,X. *et al*. (2010) MicroRNA-31 functions as an oncogenic microRNA in mouse and human lung cancer cells by repressing specific tumor suppressors. *J. Clin. Invest.*, **120**, 1298–1309.
15. Bommer,G.T. *et al*. (2007) p53-mediated activation of miRNA34 candidate tumor-suppressor genes. *Curr. Biol.*, **17**, 1298–1307.
16. Chang,T.C. *et al*. (2007) Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol. Cell*, **26**, 745–752.
17. Corney,D.C. *et al*. (2007) MicroRNA-34b and microRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth. *Cancer Res.*, **67**, 8433–8438.
18. He,L. *et al*. (2007) A microRNA component of the p53 tumour suppressor network. *Nature*, **447**, 1130–1134.
19. Johnson,S.M. *et al*. (2005) RAS is regulated by the let-7 microRNA family. *Cell*, **120**, 635–647.
20. Shedden,K. *et al*. (2008) Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat. Med.*, **14**, 822–827.
21. Lu,Y. *et al*. (2006) A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. *PLoS Med.*, **3**, e467.
22. Guo,L. *et al*. (2006) Constructing molecular classifiers for the accurate prognosis of lung adenocarcinoma. *Clin. Cancer Res.*, **12**, 3344–3354.
23. Chen,G. *et al*. (2003) Protein profiles associated with survival in lung adenocarcinoma. *Proc. Natl Acad. Sci. USA*, **100**, 13537–13542.
24. Shah,L. *et al*. (2004) Expression of syndecan-1 and expression of epidermal growth factor receptor are associated with survival in patients with non-small cell lung carcinoma. *Cancer*, **101**, 1632–1638.
25. Larsen,J.E. *et al*. (2007) Expression profiling defines a recurrence signature in lung squamous cell carcinoma. *Carcinogenesis*, **28**, 760–766.
26. Larsen,J.E. *et al*. (2007) Gene expression signature predicts recurrence in lung adenocarcinoma. *Clin. Cancer Res.*, **13**, 2946–2954.
27. Sun,Z. *et al*. (2008) Non-overlapping and non-cell-type-specific gene expression signatures predict lung cancer survival. *J. Clin. Oncol.*, **26**, 877–883.
28. Skrzypski,M. *et al*. (2008) Three-gene expression signature predicts survival in early-stage squamous cell carcinoma of the lung. *Clin. Cancer Res.*, **14**, 4794–4799.
29. Forbes,S. *et al*. (2006) Cosmic 2005. *Br. J. Cancer*, **94**, 318–322.
30. Brose,M.S. *et al*. (2002) BRAF and RAS mutations in human lung cancer and melanoma. *Cancer Res.*, **62**, 6997–7000.
31. Sachdeva,M. *et al*. (2010) MicroRNA-145 suppresses cell invasion and metastasis by directly targeting mucin 1. *Cancer Res.*, **70**, 378–387.
32. Rye,P.D. *et al*. (1996) Brain metastasis model in athymic nude mice using a novel MUC1-secreting human breast-cancer cell line, MA11. *Int. J. Cancer*, **68**, 682–687.
33. Sengupta,S. *et al*. (2008) MicroRNA 29c is down-regulated in nasopharyngeal carcinomas, up-regulating mRNAs encoding extracellular matrix proteins. *Proc. Natl Acad. Sci. USA*, **105**, 5874–5878.
34. Tie,J. *et al*. (2010) MiR-218 inhibits invasion and metastasis of gastric cancer by targeting the Robo1 receptor. *PLoS Genet.*, **6**, e1000879.

Received November 21, 2011; revised January 16, 2012; accepted February 6, 2012

Downloaded from https://academic.oup.com/carcin/article/33/5/1046/2463491 by U.S. Department of Justice user on 16 August 2022