Clinical Studies

Evaluation of MicroRNA Expression Profiles That May Predict Recurrence of Localized Stage I Non–Small Cell Lung Cancer after Surgical Resection

Santosh K. Patnaik[1], Eric Kannisto[1], Steen Knudsen[2], and Sai Yendamuri[1]

Abstract

Prognostic markers that can predict the relapse of localized non–small cell lung cancer (NSCLC) have yet to be defined. We surveyed expression profiles of microRNA (miRNA) in stage I NSCLC to identify patterns that might predict recurrence after surgical resection of this common deadly cancer. Small RNAs extracted from formalin-fixed and paraffin-embedded tissues were hybridized to locked nucleic acid probes against 752 human miRNAs (representing 82% of the miRNAs in the miRBase 13.0 database) to obtain expression profiles for 37 cases with recurrence and 40 cases without recurrence (with clinical follow-up for at least 32 months). Differential expression between the two case groups was detected for 49% of the miRNAs (Wilcoxon rank sum test; $P < 0.01$). The performance of expression profiles at differentiating the two case groups was assessed by leave-one-out and Monte Carlo cross-validations. In leave-one-out cross-validation using support vector machines- or top-scoring gene pair classifier methods, which looked for six- or two-miRNA-based classifiers, the identified miRNA expression pattern predicted recurrence with an accuracy of 70% and 83%, and hazard ratio of 3.6 [95% confidence interval (95% CI), 1.8–7.1] and 9.0 (95% CI, 4.4–18.2), respectively. Mean accuracy in Monte Carlo cross-validation using 1,000 random 60–17 splits was 69% (95% CI, 68–70) and 72% (95% CI, 71–72), respectively. The specific miRNAs mir-200b*, mir-30c-1*, mir-510, mir-630, mir-657, and mir-146b-3p and mir-124*, mir-585, and mir-708, respectively, represented most commonly among the 1,000 classifiers identified in Monte Carlo cross-validation by the two methods. MiRNAs mir-488, mir-503, and mir-647 were identified as potential reference miRNAs for future studies, based on the stability of their expression patterns across the 77 cases and the two case-groups. Our findings reinforce efforts to profile miRNA expression patterns for better prognostication of stage I NSCLC. *Cancer Res; 70(1); 36–45. ©2010 AACR.*

Introduction

Lung cancer is the leading cause of death among those afflicted with malignancies in the United States. Surgical resection is the primary treatment for the most common (70%; ref. 1), non–small cell variant of lung cancer (NSCLC), but the cure rate remains disappointing. Even for those in the earliest stage of the disease (stage IA; ref. 2), the 5-year survival rate is only 73% (2). Although most studies have not shown any significant benefit from adjuvant therapies for early-stage lung cancer (3), molecular biological investigations are under way to supplement clinical and pathologic observations identify cases with a bad prognosis so as to administer therapy only to patients at a high risk of recurrence (4).

MicroRNAs (miRNA) are small noncoding RNAs that regulate protein expression. The miRBase miRNA registry (13.0; ref. 5) includes 885 human miRNAs. Alterations in miRNA expression have been seen in several forms of cancer, including those of the lung (6). Three previous studies have suggested that miRNA profiling can be used for prognostication in lung cancer (6–8).

The expression of 752 miRNAs in cancer tissue resected from 77 cases of stage I NSCLC was examined in this study. A comparison between the groups of cases that had recurrence after surgery and those that did not shows that the two case groups have distinct miRNA expression profiles. That miRNA expression can be used to predict recurrence is shown using two different classifier methods in leave-one-out and Monte Carlo cross-validation analyses.

Materials and Methods

Patients and tumor samples. This study was approved by an institutional review board. Eighty pathologic stage I NSCLC

Authors' Affiliations: [1]Division of Thoracic Surgery, Roswell Park Cancer Institute, Buffalo, New York and [2]Medical Prognosis Institute, Hørsholm, Denmark

Corresponding Author: Sai Yendamuri, Division of Thoracic Surgery, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263. Phone: 716-845-4870; Fax: 716-845-7692; E-mail: sai.yendamuri@roswellpark.org.

American Association for Cancer Research

Table 1. Demographic and clinical characteristics of study case groups

	Recurrence	No recurrence
Cases	37	40
Mean age, y (range, SD)	67 (46–84, 10)	67 (39–91, 11)
Sex, male (%)	59	50
Race, non-Hispanic white (%)	92	95
History of smoking, positive (%)	92	88
Stage of cancer		
IA	21	25
IB	16	15
Tumor histology		
Adenocarcinoma	17	18
Bronchioloalveolar carcinoma	9	7
Squamous carcinoma	11	15
Other	0	0
Treatment		
Wedge resection (%)	14	10
Lobectomy (%)	81	83
Adjuvant therapy (%)	8	10
Median time to recurrence, mo (range, SD)*,[†]	13 (4–57, 9)	—
Median follow-up duration, mo (range, SD)*	—	57 (32–119, 24)

*Following surgery.
[†]Time to death for four cases for which time of detection of recurrence was not known.

cases resected during 2000 to 2008 at Roswell Park Cancer Institute, with approximately half known to have a recurrence, were chosen for this study. Clinical data was acquired from Roswell Park Cancer Institute's tumor registry and through chart reviews. Tissue cores with >70% tumor content was obtained from formalin-fixed and paraffin-embedded (FFPE) tissue.

Extraction of RNA. Small RNA was extracted from de-paraffinized and proteinase K–treated FFPE core tissue (20–40 mg) using the High Pure miRNA Isolation Kit (Roche). Concentration and quality of RNA was assessed by absorbance spectrometry and electrophoresis on NanoDrop (Thermo Fisher Scientific, Inc.) and Bioanalyzer 2100 (Agilent Technologies) instruments. For a third of the cases, RNA preparations were of poor quality. Consequently, RNA was extracted again from FFPE tissue.

Quantitation of miRNAs by reverse transcription-PCR. Levels of mature miRNAs *miR-21, miR-30a* (3p), *miR-103, miR-106a, miR-200b,* and *miR-623* and *U6* small nuclear 2 (*U6B*) RNA were measured using TaqMan miRNA assays (Applied Biosystems). Twenty nanograms of RNA was reverse-transcribed in 15 μL at 42°C for 30 min using a miRNA- or *U6B*-specific oligonucleotide. MiRNA-specific primers and 1.33 μL of reverse transcribed reactions were used in triplicate 45-cycle PCR of 20 μL volume in a 7900HT thermocycler (Applied Biosystems); the denaturation step at 95°C was done for 15 s, and the annealing and extension step was done at 60°C for 1 min. Fluorescence from the binding of miRNA-specific, carboxy-fluorescein dye–conjugated probes to amplification products was measured during PCR, and

SDS software (2.3; Applied Biosystems) was used to identify cycle threshold (C_t) values. Statistical analysis and graphical plotting was done using Prism 5.0b software (GraphPad).

Microarray hybridization. This work was performed by Exiqon in two experiments using 20 (10 each from the two groups) or 59 of the total 79 cases that yielded RNA of good quality. Reference RNA was generated by pooling equal amounts of RNA from all samples used in an experiment. miRCURY miRNA Power Labeling Kit (Exiqon) was used to 3′- or 5′-end label 0.5-μg samples or reference RNA with Hy3 or Hy5 (Exiqon) dye, respectively, before they were cohybridized to miRCURY locked nucleic acid (9) microarrays (10.0; Exiqon) on an HS4800 hybridization station (Tecan) for 18 h at 56°C with agitation. Slides were scanned on a G2565BA microarray scanner (Agilent Technologies) for analysis using ImaGene 8.0 software (BioDiscovery). Arrays had probes for multiple small RNAs of various species printed in quadruplicate on randomly distributed spots of 90 μm diameter. Human-specific probes covered 722 known and 30 proprietary (miRPlus, Exiqon) miRNAs, and 12 non-miRNA small RNAs. Raw data is available online in the ArrayExpress database (accession E-TABM-727).[3]

Microarray data analysis. Hy3 and Hy5 signal intensities from 78 hybridizations of good quality were processed together as follows using the limma package (2.17; ref. 10) for Bioconductor (2.3) in R (2.8.1). After background correction using the convolution model–based

---

[3] http://www.ebi.ac.uk/microarray-as/ae
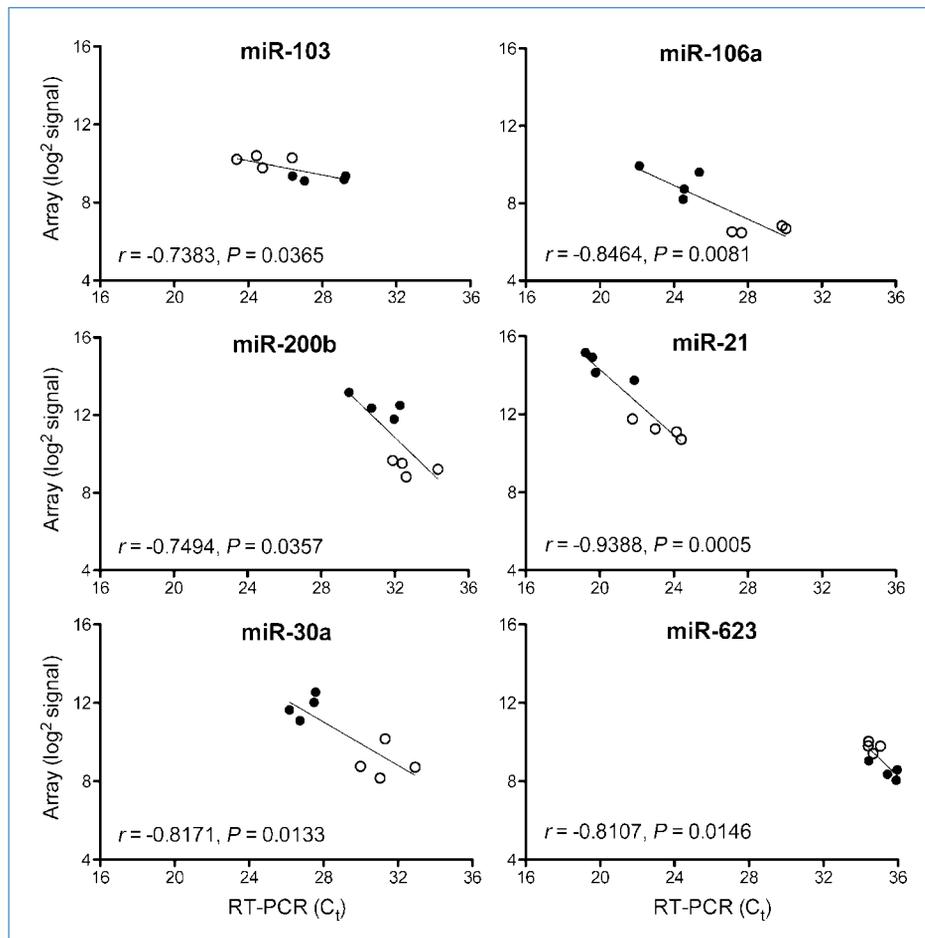
Figure 1. Correlation between RT-PCR C_t and log 2–transformed array Hy3 signal intensity values for miRNAs *miR-21, miR-30a, miR-103, miR-106a, miR-200b,* and *miR-623* ($n = 8$). Pearson correlation coefficients (r), P values, and best-fitting (least squares) lines are shown. Cases with (●) and without (○) recurrence.

normexp method with an offset of 10 (11), values were normalized first by the global Loess regression (12), with a smoothing value of 1/3, and then by limma's "rquantile" method to achieve identical distributions of Hy5 intensities across all arrays. Summarized intensity and intensity ratio (expression relative to reference) values for a probe were the mean of values from its four spots when the maximum was <150% of the minimum; the median was used otherwise. Only data for human RNAs (764) was considered from this point onwards. Two hundred and ninety RNAs, with Hy3 intensities >50% of the within-array interquartile range in >50% of samples, were considered expressed. One sample was excluded as a reassessment of clinical records indicated that the patient did not have stage I NSCLC. Expression values relative to reference were used to identify differentially expressed genes; log 2–transformed values were used for classification and hierarchical clustering using uncentered Pearson correlation and average linkage measures in R or TM4 (13) MultiExperiment Viewer (4.3).

Development and validation of recurrence predictors. The CMA R package (1.3; ref. 14) was used for classification using support vector machines (SVM; linear kernel). For leave-one-out cross-validation, training sets consisting of

all but the single test samples were used to identify the top six differentially expressed miRNAs by the Wilcoxon rank sum test. As is recommended (15), a nested cross-validation loop was used to choose the best value for the C SVM kernel parameter from 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, and 50. Alternately, top-scoring gene pairs (TSP) were used for classification using the tspair R package (1.2; ref. 16). For Monte Carlo cross-validation (17) based on similar methodologies, 1,000 randomly selected 60-sample training sets were used to classify the 17 remaining samples. Model construction in multivariate analysis using Cox proportional hazards regression for clinical variables gender, age, pathologic stage of cancer (IA or IB), tumor histology, smoking history (ever or never), and presence or absence of TSP- or SVM-based classifier signature, with recurrence as the primary outcome, was performed by stepwise elimination of nonsignificant variables ($P > 0.1$) using SPSS software (15.0; SPSS).

Identification of candidate reference miRNAs. A Microsoft Excel add-in (0.953)[4] and the SLqPCR R package[5]

---

[4] http://www.mdl.dk/publicationsnormfinder.htm

[5] http://www.bioconductor.org/packages/2.3/bioc/html/SLqPCR.html

implementing the NormFinder and geNorm algorithms, respectively, were used to analyze expression values relative to the reference range to measure the expression stability of specific miRNAs.

Results

Patient characteristics. Among the 77 surgically treated pathologic stage I NSCLC cases for which miRNA expression profiles were examined, clinical recurrence of the cancer was documented in 37 with a median time of 13 months (range, 4–57) after surgery, with 36 having it within 32 months. The time of detection of recurrence was not known for four, and the time to death was used instead. The 40 recurrence-free cases were followed for at least 32 months, with 17 followed for at least 5 years. The characteristics of the two case groups are summarized in Table 1, and case-specific information is provided in Supplementary Table S1. For univariate analysis, categorical (gender, race, pathologic stage, tumor histology, type of surgery, history of smoking) and continuous (age) clinical variables were compared between the two case groups using χ^2 and two-tailed t tests, respectively, and no statistically significant association with recurrence was found.

MiRNA expression profiling using microarrays. RNA of <100 nucleotides in length was extracted from resected and FFPE-preserved cancer tissue. Median preservation ages for groups for recurrence and no recurrence were 2.5 and 1.5 years, respectively. For RNA samples of good quality, prepared from 79 tissues, the mean of ratio of absorbances at 260 and 280 nm was 1.92 (range, 1.68–2.05; SD, 0.09), of ratio of absorbances at 260 and 230 nm, 1.65 (1.23–2.13; 0.21), and of RNA integrity numbers (18), 2.4 (1.7–2.6; 0.2).

A two-color, locked nucleic acid (9) microarray platform with probes for 752 human miRNAs and 12 human non-miRNA small RNAs was used to quantitate gene expression relative to a reference generated by pooling all RNA samples. Signal intensities were corrected by background subtraction, and global Loess and quantile normalizations. Two hundred and ninety (38%) RNAs (279 miRNAs), with intensities >50% of the within-array interquartile range (corresponding to 1.1–2.5 times the median values) in >50% samples, were deemed as being expressed (Supplementary Table S2).

Validation of microarray data by reverse transcription-PCR. Levels of mature miRNAs *miR-21, miR-30a, miR-103, miR-106a, miR-200b,* and *miR-623,* chosen randomly, were assayed in eight arbitrary RNA samples by reverse transcription-PCR (RT-PCR) to validate expression data obtained by microarray hybridization. Significant correlation was observed between log 2–transformed array Hy3 signal intensities and RT-PCR C_t values for all six miRNAs (Fig. 1). The relative expression for each of the miRNAs with respect to that of the small nuclear RNA *U6B,* quantified as the difference in C_t or log 2–transformed signal intensity values, also correlated significantly in both Pearson and Spearman tests (Supplementary Table S3). For *miR-30a, miR-103* and *miR-106a,* compared with RT-PCR, array-based quantitation reduced intersample differences by 1.5-fold, 4.6-fold, and 2.3-fold, respectively; differences were exaggerated 1.3-fold for *miR-623* and unaffected for *miR-21* and *miR-200b* (Fig. 1).

Tumors that recur and those that do not have different miRNA expressions. Using the nonparametric Wilcoxon rank sum (Mann-Whitney) test at a critical value of $P = 0.01$, 130 (47%) of the 279 expressed miRNAs were found differentially expressed (Supplementary Table S2). In the case group for recurrence, 59 were underexpressed; Table 2 lists 10 each of the most underexpressed and overexpressed miRNAs. Eight of 10 (80%) and 35 of 40 (88%) miRNAs were common to the most differentially expressed miRNAs identified by two other group comparison tests: t test assuming equal group variances (101 miRNAs with $P < 0.01$), and Welch-Satterthwaite t test for unequal group variances with t statistics based on 10,000 grouping permutations (105 miRNAs with $P < 0.01$; Supplementary Table S2).

Unsupervised hierarchical clustering of the 77 cases based on the levels of all 279 expressed miRNAs shows that cases that had recurrence cluster separately from those without recurrence (Fig. 2*A*). Figure 2*B* shows a heat-map illustrating the different miRNA expression profiles of the two groups; 56 miRNAs were differentially expressed between the two groups after multi-test correction for a false discovery rate of 0.5%.

Table 2. Ten each of the differentially expressed microRNAs that are underexpressed or overexpressed the most (recurrence vs. no recurrence case groups)

MicroRNA	P^*	Fold change[†]
Underexpressed		
miR-129-5p	0.001574	0.69
miR-194*	0.000319	0.70
miR-631	0.001873	0.71
miR-200b*	0.000006	0.71
miR-585	0.007110	0.72
miR-623	0.000453	0.73
miR-617	0.000419	0.73
miR-622	0.002722	0.73
miR-638	0.000388	0.73
miRPlus_27560	0.006484	0.75
Overexpressed		
miR-24	0.002632	3.22
miR-141	0.000047	2.86
miR-27b	0.000242	2.67
miR-16	0.000888	2.65
miR-21	0.000373	2.55
miR-30c	0.009052	2.53
miR-106a	0.000373	2.52
miR-15b	0.002223	2.52
miR-23b	0.003432	2.51
miR-130a	0.000111	2.50

*Absolute P value based on the Wilcoxon rank sum test.
[†]Based on the median expression values relative to reference.

Figure 2. *A,* unsupervised clustering of cases based on miRNA expression. Recurrence (*n* = 40; *black*) and no recurrence (*n* = 37; *gray*) groups. *B,* heat-map of log 2–transformed expression values relative to reference of 56 clustered miRNAs identified as differentially expressed between the two groups by the Wilcoxon rank sum test (*P* < 0.01; FDR, 0.5%).

Identification and validation of classifiers for predicting recurrence. Based on the differences between the two groups for the expression of many miRNAs, signatures comprised of multiple miRNAs that could predict recurrence were sought. For leave-one-out cross-validation, a linear SVM classifier based on the six most differentially expressed miRNAs was built using a training set of 76 samples and tested on the one remaining test sample. The linear SVM classifier method has the advantage of having only one adjustable kernel parameter, cost (*C*). To avoid biasing the estimation of the error rate by adjusting this parameter on the test set (15), an internal cross-validation was first performed to select the optimal value for *C*. Four miRNAs, *miR-30c-1*, miR-200b*, miR-510,* and *miR-630,* were present in all 77 six-miRNA classifiers, whereas *miR-657* and *miR-146b-3p* were present in 59 and 28, respectively. Figure 3*A* shows the predictions against a Kaplan-Meier time-to-recurrence plot [hazard ratio (HR), 3.6; 95% confidence interval (95% CI), 1.8–7.1; *P* = 0.0002]. HR was 2.9 (95% CI, 1.4–6.4; *P* = 0.006) for overall survival (Fig. 3*C*). In a Monte Carlo cross-validation analysis using 1,000 random splits into 60-sample training and 17-sample test sets, such six-miRNA classifiers had a mean prediction accuracy of 69% (range, 35–94%; 95% CI, 68–70), and mean (95% CI) sensitivity, specificity, and

positive and negative predictive values, respectively, of 67% (66–68%), 72% (71–73%), 70% (69–71%), and 69% (68–70%). Only 22 (7.9%) of the 279 expressed miRNAs were present in 98% of the 1,000 six-miRNA classifiers, the six most common being *miR-200b** (100%), *miR-30c-1** (77%), *miR-510* (68%), *miR-630* (67%), *miR-657* (29%), and *miR-146b-3p* (27%). The intercase and intergroup variabilities in the expression of miRNAs is shown in Fig. 4*A*.

In leave-one-out cross-validation using the TSP classifier method (19), which considers even nondifferentially expressed genes, prediction accuracy was 83%, with *miR-124**, which is not differentially expressed between the two case groups, and *miR-585* being the paired genes in all 77 classifiers. Figure 3*B* shows the predictions against a Kaplan-Meier time-to-recurrence plot (HR, 9.0; 95% CI, 4.4–18.2; *P* < 0.0001).

HR was 5.2 (95% CI, 2.4–11.4; *P* < 0.0001) for overall survival (Fig. 3*D*). In Monte Carlo cross-validation, the classifiers had a mean prediction accuracy of 72% (range, 24–100%; 95% CI, 71–72), and mean (and 95% CI) sensitivity, specificity, and positive and negative prediction values, respectively, of 68% (67–69%), 75% (74–76%), 73% (72–74%), and 70% (69–71%). Ten (4%) of the 279 expressed miRNAs were present in 98% of the classifiers, the three most common being *miR-124** (60%), *miR-585* (51%), and *miR-708* (17%). The intercase and intergroup variabilities in expression of *miR-124** and *miR-585* is shown as an inset in Fig. 4*A*.

As expected, as the training set size in Monte Carlo cross-validation was increased from 20 to 75, prediction accuracy improved from 60% to 70% and 61% to 81% for the SVM and TSP methods, respectively (Supplementary Fig. S1).
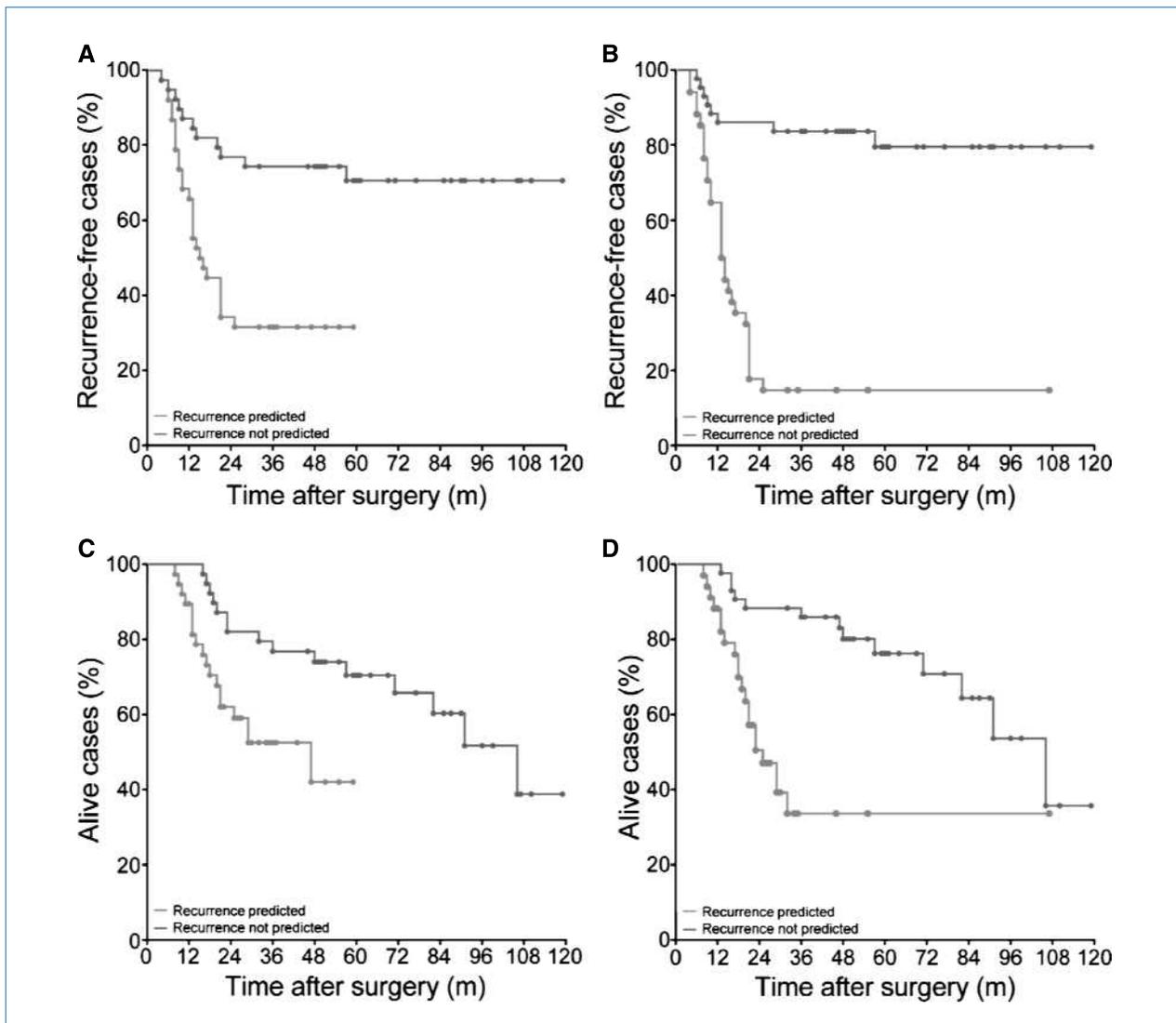
Figure 3. Kaplan-Meier plots of recurrence-free (*A* and *B*) or overall (*C* and *D*) survival predictions (*n* = 77) in a leave-one-out cross-validation analysis using either SVM-based (*A* and *C*) or TSP-based (*B* and *D*) classifiers. Cases predicted to have recurrence are plotted separately (*gray*) from those predicted to be recurrence-free.
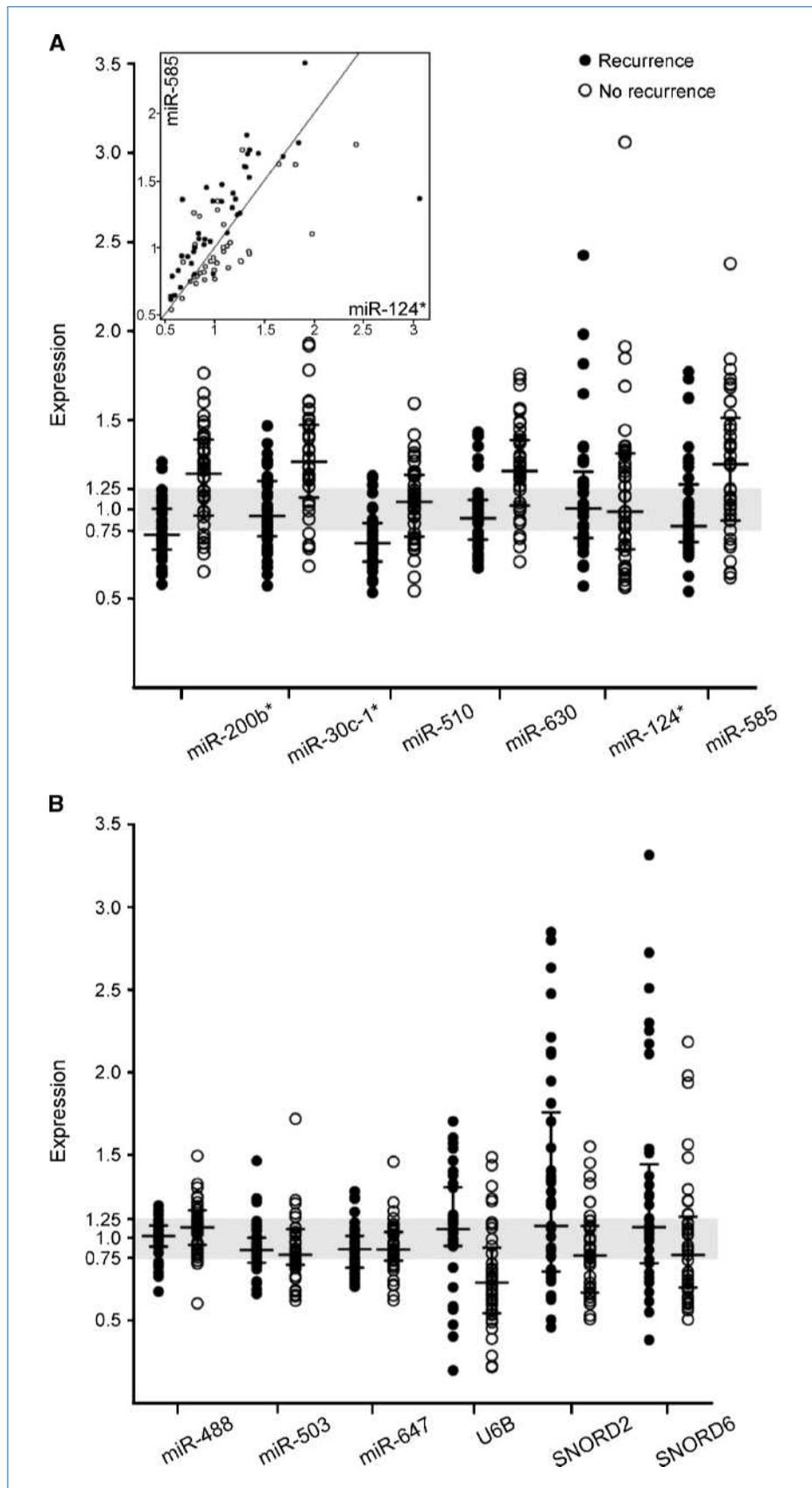
Figure 4. Median and interquartile ranges of expression values relative to reference of *miR-200b*, *miR-30c-1*, *miR-510*, and *miR-630*, and *miR-124** and *miR-585*, which constitute SVM- and TSP-based recurrence-predictive classifiers, respectively (*A*), and of three candidate miRNAs for normalization, *miR-488, miR-503*, and *miR-647*, and of *U6B, SNORD2*, and *SNORD6* non-miRNA RNAs (*B*) are indicated for cases that did (*n* = 37; ●) or did not (*n* = 40; ○) have recurrence. *Inset* in *A,* a scatter plot of *miR-124** and *miR-585* expression levels for all cases.

Multivariate analysis using Cox proportional hazards regression led to a model consisting of only the male gender ($P < 0.05$), and the presence of SVM ($P < 0.05$) and TSP ($P < 0.01$) classifier signatures. Other clinical variables did not add to the prediction of recurrence in a statistically significant fashion.

Identification of miRNAs for normalization. Of 12 non-miRNA small RNAs quantitated by microarray hybridizations, 11 were considered expressed. Except for *SNORD6*, *SNORD13* (*U13*), and *SNORD14B* (*U14B*) small nucleolar RNAs, all others, including *U6* RNAs were differentially expressed between the two case groups (Supplementary Table S2). For use for normalization when quantitating miRNAs in future NSCLC studies, candidate reference miRNAs were therefore sought. Analysis of expression values relative to the reference of all nondifferentially expressed miRNAs across the 77 samples showed that 25 miRNAs had a mean between 0.9 and 1.1, coefficient of variation <25%, and a ratio of intergroup means between 0.9 and 1.1. The 25 miRNAs were ranked for gene expression stability using NormFinder (20), which uses a model-based variance estimation approach and attempts to minimize estimated intragroup and intergroup variations, and geNorm (21), which estimates stability using a pairwise comparison method (Supplementary Table S4). Three miRNAs, *miR-488*, *miR-503*, and *miR-647*, were among the five best reference miRNAs identified using either algorithm. The intercase and intergroup variabilities in expression of the three miRNAs, and of *U6B*, *SNORD2*, and *SNORD6* are shown in Fig. 4*B*.

Pathway enrichment analysis. Pathway enrichment analysis was performed using the NIH DAVID Tools (22) or miR-Gator (23) online software on the gene targets predicted in the miRBase Targets database[6] for the six miRNAs included in the greatest number of SVM classifiers generated in the Monte Carlo cross-validation analysis (*miR-200b**, *miR-30c-1**, *miR-510*, *miR-630*, *miR-657*, and *miR-146b-3p*), and for the three candidate reference miRNAs (*miR-488*, *miR-503*, and *miR-647*). The analysis shows that many of the predicted miRNA targets, including some of those of the candidate reference miRNAs, are involved in pathways affected in cancer (Supplementary Table S5).

Discussion

Prognostication of early stage lung cancer to identify cases likely to recur can help guide the administration of potentially harmful adjuvant therapies to those most likely to benefit from it. With this goal, the expression of 752 miRNAs in 77 stage I NSCLC cases were quantitated using microarrays. This examination of more than four-fifths of known miRNAs (5) using more specific and sensitive locked nucleic acid probes (9) is larger than that described in any published work on lung cancer. Formalin fixation does not adversely affect miRNA quantitation (24), and FFPE tissue, deliberately chosen for easy availability for future applications, was successfully used for extracting RNA

---

[6] http://microrna.sanger.ac.uk/targets

for hybridizations of good quality. Based on a stringent criterion for detection of miRNA expression for more robust analysis, 279 miRNAs were considered expressed (Supplementary Table S2). They included more than half of the 43 miRNAs differentially expressed in lung cancer compared with normal lung as per a stage-oblivious study (6). About half of the miRNAs were differentially expressed between cases with recurrence and those without (Supplementary Table S2), although artifacts (25) may have arisen because of a slight intergroup difference for the ages of preserved tissues (Supplementary Table S2).

Consistent with previously published results (6–8), this study shows that miRNA profiling can be used to predict the recurrence of stage I NSCLC after resection. Yanaihara and colleagues (6), identified the association with prognosis of lung adenocarcinoma for 8 of 147 examined miRNAs; 7 of them were also differentially expressed between this study's case groups. MiRNAs *let-7a-2* and *miR-155* were both associated with poor outcome with a risk ratio of 3. Yu and colleagues (7), evaluated the expression of 157 miRNAs and found a five-miRNA signature to be an independent predictor of overall survival in NSCLC, although for stage I NSCLC (38% of training set), the signature was weakly associated ($P = 0.03$) with relapse-free survival. Two of the five miRNAs (*let-7a* and *miR-221*) were differentially expressed between this study's case groups. By quantitating 328 miRNAs, Raponi and colleagues (8), associated *miR-146b* and *miR-155* with overall survival in squamous cell NSCLC (HR, 5.9 and 2.3, respectively); the former is also identified in this study as a marker for recurrence. Supplementary Figs. S2 and S3 show that the significant miRNAs identified in these studies could predict recurrence among this study's cases as well. This also suggests that miRNA profiling studies on NSCLC, unlike studies that have examined mRNA expression (26, 27), could have more consistent results. Because the previous studies did not quantitate the expression of most of the classifier-constituting miRNAs identified here, it is not possible to validate the SVM- or TSP-based classifiers against the data from those studies.

Although these three studies showed that miRNA profiles can be prognostic, they assessed the expression of <40% of currently known miRNAs using less powerful platforms, required fresh-frozen tissue, included cases with tumors of all stages, or looked for association with overall survival. Patients with lung cancer beyond stage I receive therapy apart from surgery, and this confounds the evaluation of naïve biological aggressiveness as adjuvant therapy affects clinical end points. Patients may also have died from causes other than the cancer. The studies did not detail the extent of surgical resection. This is important because recurrence after limited (wedge) resection is significantly more likely than after lobectomy (28).

Leave-one-out and Monte Carlo cross-validation analyses show that miRNA profiles obtained in this study could predict the recurrence of stage I NSCLC. Misclassification was decreased in the former, suggesting an improvement in classifier performance by increasing training set size, as is shown in Supplementary Fig. S3. Compared with SVM, the simpler

and normalization-dispensable TSP classifier method was more accurate at predicting recurrence (83% versus 70%). Prediction accuracy of a miRNA-based classifier has not been reported thus far. The HRs associated with the classifiers, 3.6 (SVM) and 9.0 (TSP), are better than the only one (2.4) reported for a miRNA-based classifier for recurrence in stage I NSCLC (ref. 7; Fig. 3A and B). This may be because of the higher number of miRNAs examined in this study. The prediction accuracies and HRs for the miRNA-based classifiers identified here are similar to those described using mRNA-based classifiers (4, 26, 29–32). Multivariate analysis supports the conclusion that the miRNA-based classifiers are independent of clinical variables.

As has been observed before (33, 34), expression of small "housekeeping" RNAs varied widely among the cases; seven, including *U6B* which has been used for normalization when quantitating miRNAs in previous studies on NSCLC (6, 7), are differentially expressed between the two case groups (Fig. 4B; Supplementary Table S2). However, this could have resulted from the differential expression of genes responsible for miRNA processing as has been shown to occur as lung cancer evolves (35, 36). Among the miRNAs that were not differentially expressed, levels of *miR-488, miR-503*, and *miR-647* varied the least among cases as well as the two case groups (Fig. 4B). These miRNAs may be the best possible reference

miRNAs for use in future studies on FFPE tissue from early stage NSCLC.

Disclosure of Potential Conflicts of Interest

S. Knudsen focuses on developing and commercializing microarray-based technology for cancer treatment. The other authors disclosed no potential conflicts of interest.

Acknowledgments

We thank Jason North and Wiam Bshara, Department of Pathology, Roswell Park Cancer Institute, for assisting with the selection of tissue cores.

Grant Support

References

1. Owonikoko TK, Ramalingam S. Small cell lung cancer in elderly patients: a review. J Natl Compr Canc Netw 2008;6:333–44.
2. Goldstraw P, Crowley J, Chansky K, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. J Thorac Oncol 2007;2:706–14.
3. Arriagada R, Bergman B, Dunant A, Le Chevalier T, Pignon JP, Vansteenkiste J. Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. N Engl J Med 2004;350:351–60.
4. Potti A, Mukherjee S, Petersen R, et al. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. N Engl J Med 2006;355:570–80.
5. Griffiths-Jones S. miRBase: the microRNA sequence database. Methods Mol Biol 2006;342:129–38.
6. Yanaihara N, Caplen N, Bowman E, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell 2006;9:189–98.
7. Yu SL, Chen HY, Chang GC, et al. MicroRNA signature predicts survival and relapse in lung cancer. Cancer Cell 2008;13:48–57.
8. Raponi M, Dossey L, Jatkoe T, et al. MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. Cancer Res 2009;69:5776–83.
9. Castoldi M, Schmidt S, Benes V, et al. A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA). RNA 2006;12:913–20.
10. Smyth G. Limma: linear models for microarray data. In: Gentleman R, Carey VJ, Huber W, Dudoit S, Irizarry RA, editors. Bioinformatics and computational biology solutions using R and Bioconductor. New York: Springer; 2005. p. 397–420.
11. Ritchie ME, Silver J, Oshlack A, et al. A comparison of background correction methods for two-colour microarrays. Bioinformatics 2007;23:2700–7.
12. Berger JA, Hautaniemi S, Jarvinen AK, Edgren H, Mitra SK, Astola J. Optimized LOWESS normalization parameter selection for DNA microarray data. BMC Bioinformatics 2004;5:194.

13. Saeed AI, Sharov V, White J, et al. TM4: a free, open-source system for microarray data management and analysis. Biotechniques 2003;34:374–8.
14. Slawski M, Daumer M, Boulesteix AL. CMA: a comprehensive Bioconductor package for supervised classification with high dimensional data. BMC Bioinformatics 2008;9:439.
15. Dudoit S, Fridyland J. Classification in microarray experiments. In: Speed T, editor. Statistical analysis of gene expression microarray data. 1st ed. Chapman and Hall/CRC; 2003. p. 93–158.
16. Leek JT. The tspair package for finding top scoring pair classifiers in R. Bioinformatics 2009;25:1203–4.
17. Barrier A, Boelle PY, Roser F, et al. Stage II colon cancer prognosis prediction by tumor gene expression profiling. J Clin Oncol 2006;24:4685–91.
18. Schroeder A, Mueller O, Stocker S, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol 2006;7:3.
19. Geman D, d'Avignon C, Naiman DQ, Winslow RL. Classifying gene expression profiles from pairwise mRNA comparisons. Stat Appl Genet Mol Biol 2004;3:Article19.
20. Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 2004;64:5245–50.
21. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;3:RESEARCH0034.
22. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57.
23. Nam S, Kim B, Shin S, Lee S. miRGator: an integrated system for functional annotation of microRNAs. Nucleic Acids Res 2008;36:D159–64.
24. Xi Y, Nakajima G, Gavin E, et al. Systematic analysis of microRNA

expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. RNA 2007;13:1668–74.

25. Szafranska AE, Davison TS, Shingara J, et al. Accurate molecular characterization of formalin-fixed, paraffin-embedded tissues by microRNA expression profiling. J Mol Diagn 2008;10:415–23.

26. Lu Y, Lemon W, Liu PY, et al. A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. PLoS Med 2006;3:e467.

27. Boutros PC, Lau SK, Pintilie M, et al. Prognostic gene signatures for non-small-cell lung cancer. Proc Natl Acad Sci U S A 2009;106: 2824–8.

28. Ginsberg RJ, Rubinstein L. The comparison of limited resection to lobectomy for T1N0 non-small cell lung cancer. LCSG 821. Chest 1994;106:318–9S.

29. Roepman P, Jassem J, Smit EF, et al. An immune response enriched 72-gene prognostic profile for early-stage non-small-cell lung cancer. Clin Cancer Res 2009;15:284–90.

30. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002;8:816–24.

31. Raponi M, Zhang Y, Yu J, et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. Cancer Res 2006;66:7466–72.

32. Chen HY, Yu SL, Chen CH, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. N Engl J Med 2007; 356:11–20.

33. Peltier HJ, Latham GJ. Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. RNA 2008;14:844–52.

34. Davoren PA, McNeill RE, Lowery AJ, Kerin MJ, Miller N. Identification of suitable endogenous control genes for microRNA gene expression analysis in human breast cancer. BMC Mol Biol 2008; 9:76.

35. Karube Y, Tanaka H, Osada H, et al. Reduced expression of Dicer associated with poor prognosis in lung cancer patients. Cancer Sci 2005;96:111–5.

36. Chiosea S, Jelezcova E, Chandran U, et al. Overexpression of Dicer in precursor lesions of lung adenocarcinoma. Cancer Res 2007;67: 2345–50.

Correction: Online Publication Dates for *Cancer Research* April 15, 2010 Articles

The following articles in the April 15, 2010 issue of *Cancer Research* were published with an online publication date of April 6, 2010 listed, but were actually published online on April 13, 2010:

Garmy-Susini B, Avraamides CJ, Schmid MC, Foubert P, Ellies LG, Barnes L, Feral C, Papayannopoulou T, Lowy A, Blair SL, Cheresh D, Ginsberg M, Varner JA. Integrin $\alpha 4\beta 1$ signaling is required for lymphangiogenesis and tumor metastasis. Cancer Res 2010;70:3042–51. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-3761.

Vincent J, Mignot G, Chalmin F, Ladoire S, Bruchard M, Chevriaux A, Martin F, Apetoh L, Rébé C, Ghiringhelli F. 5-Fluorouracil selectively kills tumor-associated myeloid-derived suppressor cells resulting in enhanced T cell-dependent antitumor immunity. Cancer Res 2010;70:3052–61. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-3690.

Nagasaka T, Rhees J, Kloor M, Gebert J, Naomoto Y, Boland CR, Goel A. Somatic hypermethylation of *MSH2* is a frequent event in Lynch syndrome colorectal cancers. Cancer Res 2010;70:3098–108. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-3290.

He X, Ota T, Liu P, Su C, Chien J, Shridhar V. Downregulation of HtrA1 promotes resistance to anoikis and peritoneal dissemination of ovarian cancer cells. Cancer Res 2010;70:3109–18. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-3557.

Fiorentino M, Judson G, Penney K, Flavin R, Stark J, Fiore C, Fall K, Martin N, Ma J, Sinnott J, Giovannucci E, Stampfer M, Sesso HD, Kantoff PW, Finn S, Loda M, Mucci L. Immunohistochemical expression of BRCA1 and lethal prostate cancer. Cancer Res 2010;70:3136–9. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-4100.

Veronese A, Lupini L, Consiglio J, Visone R, Ferracin M, Fornari F, Zanesi N, Alder H, D'Elia G, Gramantieri L, Bolondi L, Lanza G, Querzoli P, Angioni A, Croce CM, Negrini M. Oncogenic role of *miR-483-3p* at the *IGF2/483* locus. Cancer Res 2010;70:3140–9. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-4456.

Lu W, Zhang G, Zhang R, Flores LG II, Huang Q, Gelovani JG, Li C. Tumor site–specific silencing of *NF-κB p65* by targeted hollow gold nanosphere–mediated photothermal transfection. Cancer Res 2010;70:3177–88. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-3379.

Geng H, Rademacher BL, Pittsenbarger J, Huang C-Y, Harvey CT, Lafortune MC, Myrthue A, Garzotto M, Nelson PS, Beer TM, Qian DZ. ID1 enhances docetaxel cyto-toxicity in prostate cancer cells through inhibition of p21. Cancer Res 2010;70:3239–48. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-3186.

Yoo BK, Chen D, Su Z-z, Gredler R, Yoo J, Shah K, Fisher PB, Sarkar D. Molecular mechanism of chemoresistance by astrocyte elevated gene-1. Cancer Res 2010;70:3249–58. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-4009.

Lu ZH, Shvartsman MB, Lee AY, Shao JM, Murray MM, Kladney RD, Fan D, Krajewski S, Chiang GG, Mills GB, Arbeit JM. Mammalian target of rapamycin activator RHEB is frequently overexpressed in human carcinomas and is critical and sufficient for skin epithelial carcinogenesis. Cancer Res 2010;70:3287–98. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-3467.

Hattermann K, Held-Feindt J, Lucius R, Müerköster SS, Penfold MET, Schall TJ, Mentlein R. The chemokine receptor CXCR7 is highly expressed in human glioma cells and mediates antiapoptotic effects. Cancer Res 2010;70:3299–308. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-3642.

Nadiminty N, Lou W, Sun M, Chen J, Yue J, Kung H-J, Evans CP, Zhou Q, Gao AC. Aberrant activation of the androgen receptor by NF-κB2/p52 in prostate cancer cells. Cancer Res 2010;70:3309–19. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-3703.

Acu ID, Liu T, Suino-Powell K, Mooney SM, D'Assoro AB, Rowland N, Muotri AR, Correa RG, Niu Y, Kumar R, Salisbury JL. Coordination of centrosome homeostasis and DNA repair is intact in MCF-7 and disrupted in MDA-MB 231 breast cancer cells. Cancer Res 2010;70:3320–8. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-3800.

McFarlane C, Kelvin AA, de la Vega M, Govender U, Scott CJ, Burrows JF, Johnston JA. The deubiquitinating enzyme USP17 is highly expressed in tumor biopsies, is cell cycle regulated, and is required for G_1-S progression. Cancer Res 2010;70:3329–39. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-4152.

Dudka AA, Sweet SMM, Heath JK. Signal transducers and activators of transcription-3 binding to the fibroblast growth factor receptor is activated by receptor amplification. Cancer Res 2010;70:3391–401. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-3033.

Cho SY, Xu M, Roboz J, Lu M, Mascarenhas J, Hoffman R. The effect of CXCL12 processing on CD34[+] cell migration in myeloproliferative neoplasms. Cancer Res 2010;70:3402–10. Published OnlineFirst April 13, 2010. doi:10.1158/0008-5472.CAN-09-3977.

Cancer Research

AACR American Association
for Cancer Research

Evaluation of MicroRNA Expression Profiles That May Predict Recurrence of Localized Stage I Non−Small Cell Lung Cancer after Surgical Resection

Santosh K. Patnaik, Eric Kannisto, Steen Knudsen, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-09-3153
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2009/12/22/0008-5472.CAN-09-3153.DC1.html

Cited articles	This article cites 34 articles, 14 of which you can access for free at: http://cancerres.aacrjournals.org/content/70/1/36.full.html#ref-list-1
Citing articles	This article has been cited by 20 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/70/1/36.full.html#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.