Human Cancer Biology

MicroRNA Expression Differentiates Histology and Predicts Survival of Lung Cancer

Maria Teresa Landi1, Yingdong Zhao2, Melissa Rotunno1, Jill Koshiol1, Hui Liu4, Andrew W. Bergen5, Maurizia Rubagotti6, Alisa M. Goldstein1, Ilona Linnoila3, Francesco M. Marincola4, Margaret A. Tucker1, Pier Alberto Bertazzi6, Angela C. Pesatori6, Neil E. Caporaso1, Lisa M. McShane2, and Ena Wang4

Abstract

Purpose: The molecular drivers that determine histology in lung cancer are largely unknown. We investigated whether microRNA (miR) expression profiles can differentiate histologic subtypes and predict survival for non–small cell lung cancer.

Experimental Design: We analyzed miR expression in 165 adenocarcinoma and 125 squamous cell carcinoma (SQ) tissue samples from the Environment And Genetics in Lung cancer Etiology (EAGLE) study using a custom oligo array with 440 human mature antisense miRs. We compared miR expression profiles using t tests and F tests and accounted for multiple testing using global permutation tests. We assessed the association of miR expression with tobacco smoking using Spearman correlation coefficients and linear regression models, and with clinical outcome using log-rank tests, Cox proportional hazards, and survival risk prediction models, accounting for demographic and tumor characteristics.

Results: MiR expression profiles strongly differed between adenocarcinoma and SQ ($P_{global} < 0.0001$), particularly in the early stages, and included miRs located on chromosome loci most often altered in lung cancer (e.g., 3p21-22). Most miRs, including all members of the let-7 family, were downregulated in SQ. Major findings were confirmed by quantitative real time-polymerase chain reaction (qRT-PCR) in EAGLE samples and in an independent set of lung cancer cases. In SQ, the low expression of miRs that are downregulated in the histology comparison was associated with 1.2- to 3.6-fold increased mortality risk. A five-miR signature significantly predicted survival for SQ.

Conclusions: We identified a miR expression profile that strongly differentiated adenocarcinoma from SQ and had prognostic implications. These findings may lead to histology-based therapeutic approaches.

Clin Cancer Res; 16(2); 430–41. ©2010 AACR.

The two most common histologic subtypes of non–small cell lung cancer (NSCLC) are squamous cell carcinomas (SQ) and adenocarcinomas, mainly derived from epithelial cells that line the larger airways and the peripheral small airways, respectively (1). Demographic and etiologic characteristics of adenocarcinoma and SQ have been described in refs. (2, 3). and inherited predisposition to distinct histologic types has been identified in refs. (4, 5). Mutations in the RAS protooncogene (6) and epidermal growth factor receptor (7) have been mostly found in lung adenocarcinoma, and are typically mutually exclusive (8). Although these and other differences are documented, the precise molecular features that characterize and distinguish the two histologies are largely unknown. Consequently, treatment across NSCLC lesions is similar and shows comparable poor efficacy. A refined understanding of the underlying histologic distinctions may help elucidate histology-specific patterns that can be exploited therapeutically and may improve the selection of molecular markers for early detection. We sought to explore microRNA (miR) expression differences by histology and their significance in terms of prognosis and treatment of lung cancer.

MiRs are a class of short, highly conserved noncoding RNAs involved in numerous developmental processes. MiRs regulate gene expression through incomplete base-pairing to a complementary sequence in the 3′ untranslated region of a target mRNA, resulting in translational repression and, to a lesser extent, accelerated turnover of the target transcript (9). Deregulation of miRs has been
MicroRNA Expression and Lung Cancer Histology

Translational Relevance

Recent clinical trials for non–small cell lung cancer showed that response to therapy may vary by histology. We identified a microRNA signature that strongly differentiated histology subtypes, had prognostic implications, and predicted survival from squamous cell carcinoma. These findings may suggest targets for histology-specific treatments of non–small cell lung cancer in the future.

Materials and Methods

EAGLE samples

The EAGLE study design and the subjects’ characteristics have been previously described in ref. (16). Briefly, EAGLE is a population-based, case-control study of lung cancer conducted in Italy between 2003 and 2005, which includes 2,100 primary lung cancer cases and 2,120 age-, sex-, and residence-matched population controls. The study was approved by the institutional review boards of the enrolling hospitals and the National Cancer Institute (NCI). All participating subjects signed an informed consent form. Lung cancer histology and the presence of malignant cells in the tissue blocks were ascertained by the EAGLE local pathologists and were formalin-fixed, paraffin-embedded (FFPE) tissue blocks.

We conducted a large miR expression study in 290 subjects, including 165 adenocarcinoma and 125 SQ patients, within the Environment And Genetics in Lung cancer Etiology (EAGLE) study (16). Using a custom-made miR microarray, detailed epidemiologic and clinical data, and lung tissue blocks, we investigated whether miR expression profiles can distinguish adenocarcinoma from SQ or predict survival. We also explored the effect of tobacco smoking on miR expression. The large study size allowed the adjustment or stratification for potential confounders (e.g., age, sex, smoking, and tumor stage, grade, size, lymph node involvement, presence of metastasis, and histology) when appropriate. We confirmed our major findings by qRT-PCR in 49 samples from EAGLE and 76 samples from an independent set of lung cancer cases.

Chip array analysis and quality control procedures

We used a custom-made, two-channel oligo array including 713 human, mammalian, and viral mature antisense miRs plus two internal controls with seven serial dilutions, using one EBV cell line as a reference sample (17). The in-house miR array was tested for hybridization efficiency (intensity range log10, 2-16), reproducibility,
Statistical analysis

Class comparisons. Tests for expression differences between histologies were conducted for individual miRs using two-sided t-tests or F-tests, considering P-values of <0.05 as significant. Adjustment or stratification for age, sex, histology, tumor characteristics, and smoking was done when appropriate (details in Supplementary Material S1).

The Benjamini and Hochberg method was used to estimate the false discovery rate (19). To account for multiple testing, global permutation tests with 10,000 permutations were conducted for each class comparison to confirm overall significance of the expression profile differences (details on permutation tests in Supplementary Material S1).

Association between miR expression and tobacco smoking. When relating miR expression to a specified continuous outcome [e.g., for the analysis of cigarettes per day (CPD)], Spearman correlation coefficients were calculated as a measure of correlation and to compute parametric P values. In addition, linear regression models were used to examine the association between CPD and miR expression data using the lm function in the R statistical package.7

Survival analysis. For survival analyses, patients who were alive at the date of last follow-up (n = 149) or who died due to causes other than lung cancer (n = 18) were censored. Univariate Cox proportional hazards models were fit to test individual miR expression levels for association with survival. Regression coefficients from these models were tested using a two-sided Wald test, considering P values of <0.05 as significant. First, we examined the association between expression of individual miRs and survival in unadjusted and adjusted analyses. For SQ, we excluded the three female patients and one never smoker, and adjusted for age and stage. For adenocarcinoma, we adjusted for age, stage, sex, and smoking. We also conducted the same analyses restricted to subjects with resectable tumors (stages I-A, I-B, II-A, II-B, and III-A) to identify miRs with prognostic potential for surgical candidates. Kaplan-Meier survival curves were plotted for individuals with miR expression levels above and below the median using the R statistical package.

To test the combinations of miRs that could predict the risk of dying from lung cancer, we used the supervised principal component method of Bair and Tibshirani (20) in BRB-ArrayTools (21). We calculated the first three principal component linear combinations of expression levels from miRs that were univariately correlated with survival at P < 0.01 (using Cox regression), producing three “super miRs.” A Cox proportional hazards model was fit to relate survival time to these three super miRs. This method provides a regression coefficient (weight) for each super miR for the calculation of a prognostic index, or weighted combination of the super miRs. To avoid overfitting due to the initial supervised selection of miRs to use in defining the prognostic index, we used 10-fold cross-validation. A high prognostic index value corresponded to a high predicted hazard of death due to lung cancer with correspondingly poor predicted survival. Kaplan-Meier survival curves were plotted for cases predicted to have above average risk (prognostic index above the median in the cross-validated model) and for cases predicted to have below average risk (prognostic index below the median in the cross-validated model). Details on the cross-validation procedure and permutation test are in Supplementary Material S1.

Correlation between miRs and mRNA expression. We estimated the predicted targets of the top miRs differentiating the two histology groups using TargetScan (22). We measured the mRNA expression of the putative target genes using Affymetrix U133A microarray chips (23) in 47 snap-frozen tissue samples from adenocarcinoma patients for whom we also had miR expression data. We estimated the correlation (based on a linear model) between miR expression and gene expression for each miR-probe pair and computed a global permutation P value (using 1,000 permutations) based on the number of miR-probe correlations that were significant (P value < 0.01) at the individual level.

Molecular function classification of genes targeted by miRs in the histology comparison. We used Gene Ontology (24) to assign the targeted genes whose expression was significantly correlated with the top miRs differentiating the histology groups to functional categories. The Gene Ontology categories were rank ordered using GoMiner (25).

Confirmation and replication of miR expression

We confirmed array results by qRT-PCR in 49 samples from EAGLE that had sufficient tumor miRNA remaining after the array analysis. We also replicated our analysis in 76 fresh-frozen lung cancer tissue samples from an independent Caucasian population obtained from the Cooperative Human Tissue Network (CHTN). The 49 EAGLE and 76 CHTN samples are described in Supplementary Material S2. We selected five miRs to confirm the array results by qRT-PCR using Taqman miRNA assays (Applied Biosystems; procedure details in Supplementary Material S1).

Results

Analysis of adenocarcinoma versus SQ

In the overall unadjusted analysis, miR expression profiles strongly differentiated adenocarcinoma from SQ (class comparison global permutation test P < 0.0001), with 127 miRs differentially expressed at P < 0.05 and with 86 of the 127 miRs differentially expressed at P < 0.001 (list in Supplementary Material S3). To verify whether these results were affected by gender, age, or smoking, we repeated the same analyses in male smokers only (84 adenocarcinoma, 121 SQ), adjusting for age. In this adjusted analysis, 34 miRs were confirmed at P < 0.0001.
We explored the correlation of the top 34 miRs differentiating the two histology groups with the mRNA expression of the TargetScan-predicted targets of these miRs. As expected, each miR is predicted to target hundreds of genes (Table 3). The expression of 5 of the 34 top miRs (14.7%) was found to be overall significantly associated with the expression of their putative target genes, whereas only about 2 were expected by chance. Results summaries are reported in Table 3, and the corresponding list of probes, genes, correlation coefficients, and FDR values are shown in the Supplementary Material (S4). The Gene Ontology (24) functional categories indicate that these genes

<table>
<thead>
<tr>
<th>miRNAs</th>
<th>P*</th>
<th>FDR†</th>
<th>GM in adenocarcinoma‡</th>
<th>GM in SQ§</th>
<th>Fold-change adenocarcinoma/SQ∥</th>
<th>Unique ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>has-let-7g</td>
<td>&lt;1e-07</td>
<td>&lt;1e-07</td>
<td>0.19</td>
<td>0.09</td>
<td>2.18</td>
<td>MIMAT0000414</td>
</tr>
<tr>
<td>has-let-7b</td>
<td>1e-07</td>
<td>6.6e-06</td>
<td>0.49</td>
<td>0.27</td>
<td>1.83</td>
<td>MIMAT0000063</td>
</tr>
<tr>
<td>has-let-7c</td>
<td>1e-07</td>
<td>6.6e-06</td>
<td>0.35</td>
<td>0.18</td>
<td>1.95</td>
<td>MIMAT0000064</td>
</tr>
<tr>
<td>has-miR-29a</td>
<td>2e-07</td>
<td>7.9e-06</td>
<td>0.35</td>
<td>0.18</td>
<td>1.98</td>
<td>MIMAT0000086</td>
</tr>
<tr>
<td>has-let-7f</td>
<td>2e-07</td>
<td>7.9e-06</td>
<td>0.18</td>
<td>0.09</td>
<td>2.14</td>
<td>MIMAT0000067</td>
</tr>
<tr>
<td>has-miR-453¶</td>
<td>4e-07</td>
<td>1.13e-05</td>
<td>1.25</td>
<td>1.46</td>
<td>0.86</td>
<td>MIMAT0001630</td>
</tr>
<tr>
<td>has-let-7d</td>
<td>4e-07</td>
<td>1.13e-05</td>
<td>0.46</td>
<td>0.30</td>
<td>1.54</td>
<td>MIMAT0000065</td>
</tr>
<tr>
<td>has-miR-98</td>
<td>6e-07</td>
<td>1.32e-05</td>
<td>0.25</td>
<td>0.12</td>
<td>2.06</td>
<td>MIMAT0000096</td>
</tr>
<tr>
<td>has-let-7i</td>
<td>6e-07</td>
<td>1.32e-05</td>
<td>0.25</td>
<td>0.13</td>
<td>1.84</td>
<td>MIMAT0000415</td>
</tr>
<tr>
<td>has-miR-26a</td>
<td>8e-07</td>
<td>1.58e-05</td>
<td>0.63</td>
<td>0.28</td>
<td>2.27</td>
<td>MIMAT0000082</td>
</tr>
<tr>
<td>has-miR-509-3p</td>
<td>1e-06</td>
<td>1.8e-05</td>
<td>0.69</td>
<td>0.89</td>
<td>0.78</td>
<td>MIMAT0002881</td>
</tr>
<tr>
<td>has-miR-30b</td>
<td>1.4e-06</td>
<td>2.31e-05</td>
<td>0.21</td>
<td>0.11</td>
<td>1.85</td>
<td>MIMAT0000420</td>
</tr>
<tr>
<td>has-miR-146b-5p</td>
<td>2.1e-06</td>
<td>2.97e-05</td>
<td>0.05</td>
<td>0.03</td>
<td>1.71</td>
<td>MIMAT0002809</td>
</tr>
<tr>
<td>has-miR-106b</td>
<td>2.1e-06</td>
<td>2.97e-05</td>
<td>0.11</td>
<td>0.07</td>
<td>1.52</td>
<td>MIMAT0000680</td>
</tr>
<tr>
<td>has-let-7a</td>
<td>2.3e-06</td>
<td>3.04e-05</td>
<td>0.27</td>
<td>0.16</td>
<td>1.73</td>
<td>MIMAT0000062</td>
</tr>
<tr>
<td>has-mir-663</td>
<td>3.5e-06</td>
<td>4.31e-05</td>
<td>1.32</td>
<td>0.94</td>
<td>1.40</td>
<td>MIMAT0003326</td>
</tr>
<tr>
<td>has-miR-30d</td>
<td>3.7e-06</td>
<td>4.31e-05</td>
<td>0.63</td>
<td>0.39</td>
<td>1.62</td>
<td>MIMAT0000245</td>
</tr>
<tr>
<td>has-mir-17</td>
<td>4.7e-06</td>
<td>5.17e-05</td>
<td>0.09</td>
<td>0.05</td>
<td>1.60</td>
<td>MIMAT0000070</td>
</tr>
<tr>
<td>has-mir-498¶</td>
<td>5.3e-06</td>
<td>5.52e-05</td>
<td>1.31</td>
<td>0.87</td>
<td>1.51</td>
<td>MIMAT0002824</td>
</tr>
<tr>
<td>has-mir-26b</td>
<td>6.4e-06</td>
<td>6.34e-05</td>
<td>0.40</td>
<td>0.23</td>
<td>1.75</td>
<td>MIMAT0000083</td>
</tr>
<tr>
<td>has-let-7e</td>
<td>1.11e-06</td>
<td>1.e-04</td>
<td>0.49</td>
<td>0.37</td>
<td>1.32</td>
<td>MIMAT0000066</td>
</tr>
<tr>
<td>has-mir-654-5p§</td>
<td>1.28e-06</td>
<td>1.15e-04</td>
<td>0.53</td>
<td>0.43</td>
<td>1.24</td>
<td>MIMAT0003330</td>
</tr>
<tr>
<td>has-mir-181a</td>
<td>1.65e-06</td>
<td>1.42e-04</td>
<td>0.57</td>
<td>0.38</td>
<td>1.52</td>
<td>MIMAT0000256</td>
</tr>
<tr>
<td>has-mir-103</td>
<td>2.21e-05</td>
<td>1.78e-04</td>
<td>0.38</td>
<td>0.25</td>
<td>1.51</td>
<td>MIMAT000101</td>
</tr>
<tr>
<td>has-mir-195</td>
<td>2.25e-05</td>
<td>1.78e-04</td>
<td>0.21</td>
<td>0.12</td>
<td>1.66</td>
<td>MIMAT0000461</td>
</tr>
<tr>
<td>has-mir-191</td>
<td>2.38e-05</td>
<td>1.81e-04</td>
<td>0.40</td>
<td>0.29</td>
<td>1.41</td>
<td>MIMAT0000440</td>
</tr>
<tr>
<td>has-mir-20a</td>
<td>2.8e-05</td>
<td>1.99e-04</td>
<td>0.05</td>
<td>0.03</td>
<td>1.61</td>
<td>MIMAT0000075</td>
</tr>
<tr>
<td>has-mir-106a</td>
<td>2.81e-05</td>
<td>2.32e-04</td>
<td>0.07</td>
<td>0.05</td>
<td>1.56</td>
<td>MIMAT000103</td>
</tr>
<tr>
<td>has-mir-29c</td>
<td>3.46e-05</td>
<td>2.32e-04</td>
<td>0.43</td>
<td>0.26</td>
<td>1.66</td>
<td>MIMAT0000681</td>
</tr>
<tr>
<td>has-mir-29b</td>
<td>3.51e-05</td>
<td>2.32e-04</td>
<td>0.46</td>
<td>0.26</td>
<td>1.79</td>
<td>MIMAT0001000</td>
</tr>
<tr>
<td>has-mir-491-5p</td>
<td>5.51e-05</td>
<td>3.52e-04</td>
<td>1.11</td>
<td>0.84</td>
<td>1.33</td>
<td>MIMAT0002807</td>
</tr>
<tr>
<td>has-mir-19b</td>
<td>5.76e-05</td>
<td>3.56e-04</td>
<td>0.07</td>
<td>0.04</td>
<td>1.59</td>
<td>MIMAT0000074</td>
</tr>
<tr>
<td>has-mir-107</td>
<td>7.63e-05</td>
<td>4.58e-04</td>
<td>0.37</td>
<td>0.25</td>
<td>1.45</td>
<td>MIMAT0001014</td>
</tr>
<tr>
<td>has-mir-16</td>
<td>9.43e-05</td>
<td>5.49e-04</td>
<td>0.16</td>
<td>0.10</td>
<td>1.56</td>
<td>MIMAT000069</td>
</tr>
</tbody>
</table>

NOTE: MiRs are sorted by the parametric P value from the univariate test. The model was adjusted by age. Abbreviation: GM, geometric mean.

*Parametric P value.
†FDR, false discovery rate calculated by the method of Benjamini and Hochberg (19).
‡Geometric mean of miR expression in adenocarcinoma samples compared with the EBV reference sample.
§Geometric mean of miR expression in SQ samples compared with the EBV reference sample.
∥Ratio of geometric mean ratios of miR expression in adenocarcinoma/SQ.
¶MiRs located on chromosome 19q13.42.
are mainly involved in chemical/cellular homeostasis, peroxisomal transport, and G protein signaling (Supplementary Material S5).

Four miRs among the 34 exhibited >2.0-fold difference: miR-26a, Let-7g, let-7f, and miR-98 (P < 0.0001; Table 2). Interestingly, miR-26a and let-7g are located on chromosome 3p21-22 (26), a common area of chromosomal alterations in lung cancer. MiR-98 is localized to chromosome X within the same cluster of let-7f-2 and is considered part of the let-7 family. Notably, all members of the human let-7 family on our chip (let-7a, let-7b, let-7c, let-7d, let-7e, let-7f, let-7g, let-7i, miR-98, and miR-202; ref. 27) differentiated the two histology groups (P < 0.0001) and were upregulated in adenocarcinoma versus SQ.

In comparing histologies (P < 0.05), we also identified two groups of miRs clustered <10 kb apart from each other in two genomic regions that are often gained or lost in lung cancer (28, 29): chromosome 19q13.42 with seven miRs (i.e., miR-498, miR-520b, miR-517*, miR-373*, miR-526b, miR-518c*, miR-520d-3p, plus let-7e, adjacent on chromosome 19q13.41) and chromosome 14q32.2 also with seven miRs (miR-453, miR-654-5p, miR-299-3p, miR-381, miR-432, miR-342-3p, and miR-370).

We verified whether the differences in miR expression between adenocarcinoma and SQ vary by stage. We found that miR expression differed dramatically by histology in stages I and II (Supplementary Material S6-7), but not in the advanced stages, even when samples from stages IIIb and IV were pooled (n = 57 adenocarcinoma and 31 SQ; P\text{global} = 0.195; Supplementary Material S8). Specifically, in comparing 65 adenocarcinoma versus 52 SQ stage I cases and 43 adenocarcinoma versus 42 SQ stage II cases, there were 99 and 76 miRs, respectively, whose expression strongly differentiated the two histology groups (P\text{global} <

### Table 3. Correlation between miRs differentiating adenocarcinoma from SQ and mRNA expression of predicted target genes

<table>
<thead>
<tr>
<th>MiRNAs</th>
<th>No. of genes*</th>
<th>No. of probes*</th>
<th>No. of correlations</th>
<th>P\text{global}</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-181a</td>
<td>617</td>
<td>708</td>
<td>59</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>hsa-miR-191</td>
<td>384</td>
<td>450</td>
<td>40</td>
<td>0.001</td>
</tr>
<tr>
<td>hsa-miR-107</td>
<td>374</td>
<td>434</td>
<td>21</td>
<td>0.008</td>
</tr>
<tr>
<td>hsa-miR-103</td>
<td>394</td>
<td>497</td>
<td>22</td>
<td>0.017</td>
</tr>
<tr>
<td>hsa-let-7b</td>
<td>731</td>
<td>754</td>
<td>25</td>
<td>0.029</td>
</tr>
</tbody>
</table>

*The corresponding genes and probes are listed in Supplementary Material S4.
0.0001 for each of stage I and II). Among the 76 miRs differentiating the histologies in stage II, 55 (71%) also differentiated the histologies in stage I and included all members of the let-7 family. The largest difference in miR expression (3.4-fold overall and 3.5-fold in male smokers only) was observed for miR-21 in stage II samples ($P < 10^{-5}$). Notably, the expression of this miR was not significantly different by histology in stage I cancers. We observed comparable results in the analysis restricted to the male smokers only, with the expression of 89 and 90 miRs differentiating histologies in stages I and II, respectively, and only 2 miRs differentially expressed in stages IIIb-IV (data not shown).

**Effect of cigarette smoking on miR expression**

We assessed whether tobacco smoking affected miR expression by comparing miR expression in smokers versus nonsmokers. Given the differences in miR expression by histology, we focused this analysis on adenocarcinoma only, the histologic group with a wide range of smoking histories, including never smokers (Table 1). No overall significant results were observed by smoking among the adenocarcinoma cases, even when subjects were stratified by sex and adjusted by age (data not shown), although expression of most miRs, including all members of the let-7 family, seemed to be generally downregulated in the smokers. To further explore the role of smoking, we analyzed the correlation between smoking intensity (number of CPD) and the expression of miRs that were downregulated in SQ compared with adenocarcinoma (let-7 family) and one miR that was upregulated in the same comparison (miR21) in smokers. We found an inverse correlation for the let-7 family in female adenocarcinoma patients, independent from age (Fig. 2A). We confirmed this inverse association in CHTN tissue samples from adenocarcinoma female cases (Fig. 2B). We could not analyze the effect of smoking in SQ because we had only three females in this group. In males, there was no association between smoking intensity and miR expression in either SQ or adenocarcinoma in both the EAGLE and CHTN samples (data not shown). No major differences were observed by current versus former smoking status, years of smoking, age at smoking initiation, or exposure to environmental (passive) smoking (data not shown). The miR21 analysis showed a significant increase of miR expression by CPD in males with SQ ($P = 0.02$; Fig. 2C), but no significant association in adenocarcinoma either in males (data not shown) or females, although there was a suggestive down-regulation of the expression in females (Fig. 2C).

**Survival analysis**

Both the unadjusted survival analysis and the analysis adjusted by age, sex, histology, stage, and smoking status ($n = 290$ for all stages, $n = 251$ for early stages) produced no more significant miRs than would have been expected by chance for SQ and adenocarcinoma combined. Due to the observed biological differences in the miR expression comparisons and the disparate covariate distributions (SQ patients were nearly all male ever smokers), survival analyses were conducted separately for adenocarcinoma and SQ patients.

**Adenocarcinoma.** Although 12 miRs (parametric $P$ value < 0.05) were identified in the unadjusted univariate survival analyses of the adenocarcinoma sample and 15 were identified in the analyses restricted to the early stage, the global tests for the associations were not significant (permuation $P$ value of global test = 0.27 and 0.20, respectively). Similarly, 11 miRs were associated with survival in the analyses adjusted by age, sex, smoking, and stage, and 10 were expected by chance (Supplementary Material S9). A risk prediction analysis for combined miR expression did not show significant associations with survival (data not shown).

**Squamous cell carcinoma.** Both the unadjusted and adjusted analyses showed significant results for the association of miR expression with survival from SQ. Because there were only three females and one nonsmoker, we focused on male smokers. Results based on stages I, II, and IIIA ($n = 107$) are discussed below. Results including all stages ($n = 121$, with 25 miRs with $P < 0.05$) are reported in Supplementary Material S10.

Using a cutoff of $P < 0.05$, there were 33 miRs individually associated with mortality risk in the unadjusted analysis (global permutation $P = 0.026$) and 36 in the age- and stage-adjusted analysis (Supplementary Material S11). Many of these miRs also differentiated the histology groups. The expression of most miRs was inversely related to mortality, ranging from 1.2- to 3.6-fold increased risks for subjects with low expression. Given the strong individual miR associations, we fit a risk prediction model for the combined effect of miRs on survival. Because age ($P = 0.96$) and stage ($P = 0.46$, including only stages I, II, and IIIA) were not associated with survival in this group, and subjects only included male smokers, we conducted an unadjusted analysis. Among the 107 SQ patients included in this analysis, 51 were stage I, 39 were stage II, and 17 were stage IIIA, and the number of deaths due to lung cancer were 22, 17, and 9, respectively.

We found five miRs (miR-25, miR-34c-5p, miR-191, let-7e, and miR-34a) whose expression strongly predicted SQ survival for the 107 male smokers with early-stage SQ tumors. Using the cross-validated supervised principal component method, 62 patients (36 lung cancer deaths) were predicted to have high mortality risk (prognostic index above median) and 45 patients (12 lung cancer deaths) were predicted to have low mortality risk (prognostic index below median). Figure 3 displays the Kaplan-Meier curves comparing the two risk groups ($P$ value for the permutation test was 0.017 based on 10,000 permutations); the corresponding results from the individual miR analysis of the five miRs are shown in Table 4.

**Confirmation and replication of array results by qRT-PCR**

We analyzed five miRs by qRT-PCR in 49 samples remaining from the EAGLE study and 76 CHTN samples.
Specifically, let-7g, miR-26a, and let-7f were chosen to confirm the histology comparison (fold change of >2 in the array comparison); let-7g and let-7f were used to confirm the smoking comparison; and miR-638 and miR-107 were chosen to confirm the survival results. The high expression of miR-638 was a risk factor and the high expression of miR-107 was a protective factor for mortality from SQ in the array data. The three miRs chosen for the histology comparison showed a similar pattern in the qRT-PCR-based results in both the EAGLE samples and the CHTN samples (Supplementary Material S12). Similarly, the qRT-PCR results for the two miRs selected for the smoking comparison (Fig. 2B) and the two miRs selected for the association with SQ survival (Supplementary Material S13) were consistent with the array data. Of note, the array results were based on FFPE samples, whereas the CHTN samples were fresh frozen. Our findings confirm (30) that both samples types elicit comparable miR results.

**Discussion**

Recent clinical trials show that different histologic subtypes of NSCLC have different responses to therapy (31–34). Yet, current treatment is similar across histologies because histology-specific molecular targets that could be therapeutically exploited are lacking. In the largest study of miR expression in lung cancer to date, we found a miR profile that strongly and consistently differentiated adenocarcinoma from SQ. We confirmed the largest differences in an independent set of lung cancer cases and identified a set of putative target genes whose expression was correlated with the expression of miRs differentiating the two histology groups, providing preliminary evidence of miR-target correlations for further experimental validation. Expression differences by histology were highly significant in the early-stage tumors but not significant in the advanced stages, suggesting that in advanced, less...
differentiated tumors, miR expression loses histology specificity. This finding implies that efforts to exploit these differences for mechanistic insight or therapeutic benefit should focus on early-stage tumors. Given the different miR profiles by histology, we conducted survival analyses separately in the two histology groups, with a rigorous statistical approach. Many of the miRs downregulated in early-stage SQ versus adenocarcinoma were associated with increased risk of mortality from SQ, suggesting a role for these miRs in the repression of genes involved in lung cancer progression. This finding may also provide insight into why some early-stage, apparently surgically cured, patients recur with a fatal outcome. Among these miRs, we identified a five-miR signature that strongly predicted survival from SQ.

All members of the let-7 family highly differentiated adenocarcinoma from SQ and increased the mortality risk in SQ, suggesting that this group exerts its influence most profoundly within the early-stage SQ histology. Pioneering studies described let-7 miR as a negative regulator of the oncogenic family of RAS and as a repressor of cell proliferation pathways and NSCLC growth (11, 27, 35). As it has been observed for other pathways [e.g., the epidermal growth factor receptor pathway in NSCLC (36)] or other tumors [e.g., the NRAS/BRAF mutations in melanoma (37)], it is possible that let-7 expression alterations and KRAS mutations are mutually exclusive in lung carcinogenesis, with let-7 having a more predominant effect in SQ and KRAS mutations in adenocarcinoma. If further confirmed by mutational analyses, this finding may improve our understanding of the carcinogenetic pathways leading to NSCLC. Among the let-7 miRs, let-7g showed the largest fold change together with miR 26a, an hypoxia-induced miR known to decrease proapoptotic signaling (38). Interestingly, both let-7g and miR-26a are localized to chromosome 3p21-22. Allele loss and genetic alteration at this locus are the most frequent and earliest genomic abnormalities in NSCLC (39) and are more common in SQ than adenocarcinoma (40, 41). We also found histologic differences for miRs clustering at chromosome 14q32 and 19q13, both loci involved in lung cancer development and progression (28, 29). All these loci have been extensively investigated for potential tumor suppressor genes without clear success. However, if the chromosomal alterations involve miRs that target distant genes, the potentially relevant lung cancer genes likely extend beyond these loci.

Other miRs strongly differentiating the two histology groups included miR-29a, which affects apoptosis (42) and epigenetic normalization of NSCLC (43), and miR-21, which acts as an oncogene or “oncomiR” in many tumor types and plays an important role in tumor metastasis (44, 45). Interestingly, in our study, miR-21 strongly differentiated the histology groups with high levels in adenocarcinoma in stage II but not in stage I tumors, suggesting that miR-21 may be a marker of tumor progression in adenocarcinoma, identifying tumors on the verge of acquiring metastatic potential.

The large majority of miRs were downregulated in SQ versus adenocarcinoma in our study. This pattern suggests that miRs that function as tumor suppressors, such as let-7 and miR-29a, may be more relevant for SQ tumorigenesis because they exhibited poor expression in this tumor type. In contrast, miRs with oncogenic potentials in many tissue types, such as miR-21 and -26a, may be crucial for adenocarcinoma development because they were overexpressed in adenocarcinoma compared with SQ. If functionally confirmed, these miRs may identify histology-specific therapeutic targets, especially for surgically resectable lesions.

Table 4. Association of miR expression with survival of SQ for the five miRs included in the risk prediction model

<table>
<thead>
<tr>
<th>MiRNAs</th>
<th>Hazard ratio (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsa-miR-25</td>
<td>0.46 (0.28-0.75)</td>
<td>0.002</td>
</tr>
<tr>
<td>hsa-miR-191</td>
<td>0.52 (0.33-0.80)</td>
<td>0.003</td>
</tr>
<tr>
<td>hsa-let-7c</td>
<td>0.46 (0.27-0.80)</td>
<td>0.006</td>
</tr>
<tr>
<td>hsa-miR-34c-5p</td>
<td>0.31 (0.13-0.74)</td>
<td>0.008</td>
</tr>
<tr>
<td>hsa-miR-34a</td>
<td>0.47 (0.26-0.84)</td>
<td>0.011</td>
</tr>
<tr>
<td>Adjusted analysis*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsa-miR-25</td>
<td>0.44 (0.27-0.72)</td>
<td>0.001</td>
</tr>
<tr>
<td>hsa-miR-191</td>
<td>0.51 (0.33-0.89)</td>
<td>0.003</td>
</tr>
<tr>
<td>hsa-let-7c</td>
<td>0.45 (0.26-0.79)</td>
<td>0.006</td>
</tr>
<tr>
<td>hsa-miR-34c-5p</td>
<td>0.32 (0.13-0.77)</td>
<td>0.011</td>
</tr>
<tr>
<td>hsa-miR-34a</td>
<td>0.48 (0.27-0.85)</td>
<td>0.011</td>
</tr>
</tbody>
</table>

*Analysis adjusted for stage and age.
The only previous study showing miR expression differences among lung cancer histologies found 5 mature miRs and 1 miR precursor differentiating 65 adenocarcinoma from 39 SQ (14) in an unadjusted analysis of Caucasians and African-Americans, with different smoking patterns than Southern Europeans. Dissimilarity in sample sizes, analytic approaches, assays, and ethnic groups may have contributed to the discrepancies between studies. More-
over, differential expression and variable ratio of mature
versus immature miRs in specific subpopulations of mali-
gnant tissues (45, 46) may account for some differences.

We examined the effect of tobacco smoking on miR ex-
pression and obtained interesting preliminary findings. The
comparison between smokers and never smokers did not
elicit significant results, possibly because of the small num-
ber of never smokers (only five in males). In smokers, we
found that the expression of each member of the let-7 family,
which has tumor suppression potential, was inversely associ-
ated with the number of CPD in females but not in males. We
confirmed this finding also in the independent samples from
the CHTN. Downregulation of let-7 by tobacco smoking was
also observed in rats (47). The gender discrepancy might be
related to hormonal effects because estrogen levels may
modify miR expression (48). Sex and smoking status may
have a differential effect on miR expression as has been ob-
erved for mutational patterns of epidermal growth factor re-
ceptor in NSCLC (49). Interestingly, expression of miR21,
which acts as an oncogene, increased with increasing CPD
in SQ but not adenocarcinoma, suggesting a histology-spe-
cific response to tobacco carcinogens. However, small
numbers or other unknown factors may have contributed
to these results, and larger studies and functional tests are
necessary to confirm our findings. If confirmed, they may
provide an important piece in the mechanistic puzzle relat-
ing tobacco smoking to lung cancer development.

In the survival analysis, we found that the low expres-
sion of several miRs was associated with up to 4-fold ex-
cess mortality from SQ overall and in the subgroup
including only male smokers with stage I to IIIA. Interest-
ingly, many of these miRs were downregulated in SQ ver-
sus adenocarcinoma even in stage I, suggesting an early
response to tobacco carcinogens. However, small
numbers or other unknown factors may have contributed
to these results, and larger studies and functional tests are
necessary to confirm our findings. If confirmed, they may
provide an important piece in the mechanistic puzzle relat-
ing tobacco smoking to lung cancer development.

We confirmed this finding also in the independent samples from
the CHTN. Downregulation of let-7 by tobacco smoking was
also observed in rats (47). The gender discrepancy might be
related to hormonal effects because estrogen levels may
modify miR expression (48). Sex and smoking status may
have a differential effect on miR expression as has been ob-
erved for mutational patterns of epidermal growth factor re-
ceptor in NSCLC (49). Interestingly, expression of miR21,
which acts as an oncogene, increased with increasing CPD
in SQ but not adenocarcinoma, suggesting a histology-spe-
cific response to tobacco carcinogens. However, small
numbers or other unknown factors may have contributed
to these results, and larger studies and functional tests are
necessary to confirm our findings. If confirmed, they may
provide an important piece in the mechanistic puzzle relat-
ing tobacco smoking to lung cancer development.

In conclusion, in the largest study of miR expression in
lung cancer to date, miR expression profiles strongly dif-
fered between adenocarcinoma and SQ, suggesting that
different sets of miRs contribute to the pathogenesis of
different NSCLC histologies and may become targets of
histology-specific treatment in the future. Among miRs
whose expression was reduced in SQ from the early stages,
we identified a profile that predicted survival for SQ. These
miRs may have important implications for prognosis and
treatment of this histology subgroup of lung cancer.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank the EAGLE participants and the large number of collaborators (listed in http://dceg.cancer.gov/eagle)
that made the EAGLE study possible, the CHTN that
provided samples and data for the replication study, and
Drs. Giuseppe Giaccone and Adi Gazdar for their
insightful comments.

Grant Support

Intramural Research Program of NIH, NCI, Division of
Cancer Epidemiology and Genetics, and a 2006 NCI Di-
rector’s Innovation Award to M.T.L.

The costs of publication of this article were defrayed in
part by the payment of page charges. This article must
therefore be hereby marked advertisement in accordance
with 18 U.S.C. Section 1734 solely to indicate this fact.
Received 7/4/09; revised 10/19/09; accepted 11/8/09;
published OnlineFirst 1/12/10.

www.aacrjournals.org

Clin Cancer Res; 16(2) January 15, 2010

439
References


Clinical Cancer Research

MicroRNA Expression Differentiates Histology and Predicts Survival of Lung Cancer

Maria Teresa Landi, Yingdong Zhao, Melissa Rotunno, et al.

*Clin Cancer Res* Published OnlineFirst January 12, 2010.

Updated version

Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-09-1736

Supplementary Material

Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2010/01/18/1078-0432.CCR-09-1736.DC1

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.