Unique microRNA molecular profiles in lung cancer diagnosis and prognosis

Nozomu Yanaihara,[1] Natasha Caplen,[2] Elise Bowman,[1] Masahiro Seike,[1] Kensuke Kumamoto,[1] Ming Yi,[3] Robert M. Stephens,[3] Aikou Okamoto,[4] Jun Yokota,[5] Tadao Tanaka,[4] George Adrian Calin,[6] Chang-Gong Liu,[6] Carlo M. Croce,[6] and Curtis C. Harris[1,*]

[1] Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892
[2] Gene Silencing Section, Office of Science and Technology Partnership, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892
[3] Advanced Biomedical Computing Center, National Cancer Institute-Frederick/SAIC-Frederick Inc., Frederick, Maryland 21702
[4] Department of Obstetrics and Gynecology, The Jikei University School of Medicine, Tokyo 105-8461, Japan
[5] Biology Division, National Cancer Center Research Institute, Tokyo 104-0045, Japan
[6] Molecular Virology, Immunology and Medical Genetics, The Ohio State University Comprehensive Cancer Center, Columbus, Ohio 43210
*Correspondence: curtis_harris@nih.gov

Summary

MicroRNA (miRNA) expression profiles for lung cancers were examined to investigate miRNA's involvement in lung carcinogenesis. miRNA microarray analysis identified statistical unique profiles, which could discriminate lung cancers from noncancerous lung tissues as well as molecular signatures that differ in tumor histology. miRNA expression profiles correlated with survival of lung adenocarcinomas, including those classified as disease stage I. High *hsa-mir-155* and low *hsa-let-7a-2* expression correlated with poor survival by univariate analysis as well as multivariate analysis for *hsa-mir-155*. The miRNA expression signature on outcome was confirmed by real-time RT-PCR analysis of precursor miRNAs and crossvalidated with an independent set of adenocarcinomas. These results indicate that miRNA expression profiles are diagnostic and prognostic markers of lung cancer.

Introduction

Lung cancer is the leading cause of cancer deaths in the world, and its etiology is primarily genetic and epigenetic damage caused by tobacco smoke (Travis et al., 2004). Systematic analysis of mRNA and protein expression levels among thousands of genes has also contributed to defining the molecular network of lung carcinogenesis (Meyerson and Carbone, 2005; Granville and Dennis, 2005). Defects in both the *p53* and *RB*/*p16* pathways are common in lung cancer. Several other genes, such as *K-ras*, *PTEN*, *FHIT*, and *MYO18B*, are genetically altered, though less frequently (Minna et al., 2002; Sekido et al., 2003; Yokota and Kohno, 2004). Although focusing on known genes and proteins has already yielded new information, previously unknown markers such as noncoding RNA gene products may also lend insight into the biology of lung cancer.

MicroRNAs (miRNAs) are small noncoding RNA gene products about 22 nt long that are found in diverse organisms and play key roles in regulating the translation and degradation of mRNAs through base pairing to partially complementary sites, predominately in the untranslated region of the message (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). miRNAs are expressed as long precursor RNAs that are processed by a cellular nuclease, Drosha, before being transported by an Exportin-5-dependent mechanism into the cytoplasm (Yi et al., 2003; Gregory and Shiekhattar, 2005). Once in the cytoplasm, miRNAs are cleaved further by the enzyme DICER (Lee et al., 2002, 2003), and this results in 17–24 nt miRNAs that are associated with a cellular complex that is similar to the RNA-induced silencing complex that participates in RNA interference (Hutvagner and Zamore, 2002). Recently, it was reported that *DICER* expression levels were reduced in a fraction of lung

SIGNIFICANCE

miRNAs are a class of small noncoding RNA genes found to be abnormally expressed in several types of cancer, suggesting that miRNAs play a substantial role in the pathogenesis of human cancers. Lung cancer is the leading cause of cancer deaths in the world, reflecting the need for a better understanding of the mechanisms that underlie lung carcinogenesis. Although focusing on known genes and proteins has already yielded new information, unknown markers may also lend insight into the biology of lung cancer. We showed that lung cancer has extensive alterations of miRNA expression that may deregulate cancer-related genes. Furthermore, the miRNA molecular signature of lung adenocarcinomas, including those without evidence of metastasis, also correlates with patient survival.

cancers with a significant prognostic impact on patient survival (Karube et al., 2005). The biological functions of most miRNAs are not yet fully understood. It has been suggested that the miRNAs are involved in various biological processes, including cell proliferation, cell death, stress resistance, and fat metabolism, through the regulation of gene expression (Ambros, 2003).

Our understanding of miRNA expression patterns and function in normal or neoplastic human cells is just starting to emerge. miRNA genes are frequently located at fragile sites (FRAs), as well as in minimal regions of loss of heterozygosity, minimal regions of amplification, or common breakpoint regions, suggesting that miRNAs might be a new class of genes involved in human tumorigenesis (Calin et al., 2004b). For example, *mir-15-a* and *mir-16-1* are frequently deleted and/or downregulated in patients with B cell chronic lymphocytic leukemia (Calin et al., 2002). Other links between cancer and miRNA have been reported, including reduced expression of *mir-143* and *mir-145* in colorectal cancers (Michael et al., 2003) and *let-7* in lung cancers (Takamizawa et al., 2004), high expression of the precursor *mir-155* in Burkitt's lymphomas (Metzler et al., 2004), and oncogenic function of *mir-17-92* cluster in human B cell lymphomas as well as in lung cancers (He et al., 2005; Hayashita et al., 2005). The precise mechanisms regulating miRNA expression are unknown. However, several mechanisms, including genetic and epigenetic alteration, might affect miRNA expression, and they may lead to alterations in the pattern of target genes expression in cancers. It was shown that miRNA expression patterns have relevance to the biological and clinical behavior of human B cell chronic lymphocytic leukemia and solid tumors, including breast cancers (Calin et al., 2004a; Iorio et al., 2005; Volinia et al., 2006). One or more members of the *let-7* family regulate RAS expression in human cells, and thus, *let-7* may play a major role in human lung carcinogenesis as a tumor suppressor gene (Johnson et al., 2005). Recently, miRNA expression profiles have been shown to be potential tools for cancer diagnosis (Lu et al., 2005). These and other data are consistent with the hypothesis that miRNAs play a substantial role in the pathogenesis of human cancers.

In this study, we investigated the miRNA expression profiles in human lung cancer and miRNA regulation by epigenetic mechanisms and found that the miRNA molecular profile of lung adenocarcinoma correlates with patient survival.

Results

Altered miRNA expression in primary lung cancers and identification of miRNAs associated with clinicopathological features of lung cancer

We analyzed the miRNA expression in 104 pairs of primary lung cancers and corresponding noncancerous lung tissues. We compared miRNA expression of several group pairs as listed in Table 1. Expression profiles were generated by comparing lung cancers, except when comparing lung cancer tissues with corresponding noncancerous lung tissues. We identified miRNAs, which were expressed differently in phenotypical and histological classifications (Table 1). When we compared miRNA expression among lung cancer tissues versus corresponding noncancerous lung tissues, 43 miRNAs had statistical differences in expression between groups (Table 2). In class comparison analysis using our microarray analysis tool, the multivariate permutation test was performed to control multiple comparisons. It provides a specific confidence level for ensuring that the number of false discoveries does not exceed a target level or for ensuring that the proportion of the gene list that are false discoveries does not exceed a target level. Thus, the probability of identifying at least 43 miRNAs by chance at the <0.001 level, if there are no real differences between the classes, was 0 as estimated by the multivariate permutation test. Furthermore, 91% of 104 lung cancers were correctly classified using the leave-one-out crossvalidated class prediction method based on the compound covariate predictor. Based on 2000 random permutations, the p value, which is defined as the proportion of the random permutations that gave a crossvalidated error rate no greater than the crossvalidated error rate with the real data, was < 0.0005.

Several of these miRNAs were associated with FRAs (Table 2). In particular, three miRNAs are located inside FRAs (*hsa-mir-21* at FRA17B, *hsa-mir-27b* at FRA9D, and *hsa-mir-32* at FRA9E). Furthermore, many of these miRNAs are located at frequently deleted or amplified regions in several malignancies (Table 2). For example, *hsa-mir-21* and *hsa-mir-205* are located at the region amplified in lung cancer, whereas *hsa-mir-126** and *hsa-mir-126* are at 9q34.3, a region deleted in lung cancer. Reduced expression of precursor *let-7a-2* and *let-7f-1* was also found in adenocarcinoma and squamous cell carcinoma at a p value cutoff of 0.05, respectively. In the same way, comparison analyses between lung adenocarcinoma versus noncancerous tissues and squamous cell carcinoma versus noncancerous tissues revealed 17 and 16 miRNAs with statistically different expression, respectively (Table S2 in the Supplemental Data available with this article online). Six miRNAs (*hsa-mir-21*, *hsa-mir-191*, *hsa-mir-155*, *hsa-mir-210*, *has-mir-126**, and *hsa-mir-224*) were shared in both histological types of non-small cell lung carcinoma (NSCLC).

Next, we asked whether the microarray data revealed specific molecular signatures for subsets of lung cancer that differ in clinical behavior. For this analysis, we examined the relationship of five types of clinical and pathological information (Table 1). Among them, we identified six miRNAs (*hsa-mir-205*, *hsa-mir-99b*, *hsa-mir-203*, *hsa-mir-202*, *hsa-mir-102*, and *hsa-mir-204-prec*) that were expressed differently in the two most common histological types of NSCLC, adenocarcinoma and squamous cell carcinoma. The expression levels of *hsa-mir-99b* and *hsa-mir-102* were higher in adenocarcinoma. No miRNAs were identified as differently expressed when classified by age, gender, or race in our data set.

Validation of the microarray data by the solution hybridization detection method and real-time RT-PCR analysis

We used the solution hybridization detection method for mature miRNAs and real-time RT-PCR analysis for precursor miRNAs to validate the results from microarray analysis. First, the microarray data of three miRNAs (*hsa-mir-21*, *hsa-mir-126**, and *hsa-mir-205*) were analyzed by the solution hybridization detection method. Seven pairs of primary lung cancers and corresponding noncancerous lung tissues in which sufficient RNA was available were analyzed. The mature forms of *hsa-mir-21* and *hsa-mir-205* were clearly upregulated in lung cancer tissues when compared with the corresponding noncancerous lung tissues (Figure 1A). In contrast, *hsa-mir-126** was downregulated in most of the lung cancer tissues examined. Therefore, the analyses confirmed the microarray data for these three miRNAs.