

# microRNA miR-196a-2 and Breast Cancer: A Genetic and Epigenetic Association Study and Functional Analysis

Aaron E. Hoffman,<sup>1</sup> Tongzhang Zheng,<sup>1</sup> Chunhui Yi,<sup>1</sup> Derek Leaderer,<sup>1</sup> Joanne Weidhaas,<sup>2</sup> Frank Slack,<sup>3</sup> Yawei Zhang,<sup>1</sup> Trupti Paranjape,<sup>2</sup> and Yong Zhu<sup>1</sup>

Departments of <sup>1</sup>Epidemiology and Public Health and <sup>2</sup>Therapeutic Radiology, Yale University School of Medicine; <sup>3</sup>Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut

## Abstract

**Increasing evidence has suggested that microRNAs (miRNA) play an important role in tumorigenesis. As transcriptional regulators, altered miRNA expression may affect many cancer-related biological pathways, indicating that miRNAs can function as tumor suppressors and/or oncogenes. We first performed a genetic association analysis by screening genetic variants in 15 miRNA genes and detected that a common sequence variant in hsa-miR-196a-2 (rs11614913, C→T) was significantly associated with decreased breast cancer risk (for homozygous variant: odds ratio, 0.44; 95% confidence interval, 0.28-0.70). Hypermethylation of a CpG island upstream (-700 bp) of the miR-196a-2 precursor was also associated with reduced breast cancer risk (odds ratio, 0.35; 95% confidence interval, 0.15-0.81). By delivering expression vectors containing either wild-type or mutant precursors of miR-196a-2 into breast cancer cells, we showed that this variant led to less efficient processing of the miRNA precursor to its mature form as well as diminished capacity to regulate target genes. A whole-genome expression microarray was done and a pathway-based analysis identified a cancer-relevant network formed by genes significantly altered following enforced expression of miR-196a-2. Mutagenesis analysis further showed that cell cycle response to mutagen challenge was significantly enhanced in cells treated with variant miR-196a-2 compared with cells treated with the wild-type. Taken together, our findings suggest that miR-196a-2 might have a potentially oncogenic role in breast tumorigenesis, and the functional genetic variant in its mature region could serve as a novel biomarker for breast cancer susceptibility. [Cancer Res 2009;69(14):5970-7]**

## Introduction

It was proposed over three decades ago that some non-protein-coding transcripts may regulate gene activity, a missing piece to the central dogma of molecular biology (1). This novel group of small noncoding RNAs with post-transcriptional regulatory function, later termed microRNAs (miRNA), has now been shown to negatively alter protein expression in most organisms, from *Caenorhabditis elegans* to humans (2-5). Interestingly, over 52.5% of the miRNAs elucidated thus far are located within fragile sites on

chromosomes, suggesting that these transcripts may be relevant for carcinogenesis (6, 7).

Research linking miRNAs to cancer is still in its infancy, but the information obtained to date has suggested that miRNA may play an important role in a variety of carcinogenic processes (6, 8-11). miRNAs have been shown to act as tumor suppressors by repressing oncogene expression or may themselves act as oncogenes by repressing other tumor suppressors (11-13). Recent evidence has also shown that a global reduction in miRNA processing promotes carcinogenesis, and miRNA profiling data have been successfully applied to classify tumors and predict prognosis in several cancer types (14-16).

In addition, previous studies have suggested that CpG island methylation in miRNA regions may influence miRNA function, thereby altering carcinogenic processes. Weber and colleagues (17) found that nearly half of all identified miRNAs were associated with CpG sites, and an experimental assay revealed detectable methylation levels at several miRNA loci across normal and malignant cell lines. Lu and colleagues (18) also found that hypermethylation in the let-7a-3 miRNA gene was associated with survival in patients with ovarian cancer, and Lehmann and colleagues (19) suggested that aberrant hypermethylation of miRNAs may be an important early event in breast cancer development.

Given an important role for miRNAs in cancer development, we hypothesize that genetic and epigenetic variations in miRNAs may alter miRNA function and consequently influence susceptibility to cancer. To date, however, few published molecular epidemiologic studies have investigated such associations at the population level to determine whether miRNAs may serve as valuable cancer biomarkers. In the current study, we tested this hypothesis in breast cancer by performing a genetic and epigenetic association study and functional analysis.

## Patients, Materials, and Methods

**Study population.** The study population consisted of subjects enrolled in a previous breast cancer case-control study conducted in Connecticut. Details regarding subject recruitment and participant characteristics have been described in previous publications (20, 21). Briefly, cases were incident, histologically confirmed breast cancer patients (*International Classification of Diseases for Oncology* codes 174.0-174.9) between 30 and 80 years of age with no previous diagnosis of cancer other than nonmelanoma skin cancer, who were alive at the time of the interview. Cases were identified either from computerized patient information at Yale-New Haven Hospital in New Haven County, Connecticut or from nearby Tolland County, Connecticut via hospital records by the Rapid Case Ascertainment Shared Resource at the Yale Cancer Center. Yale-New Haven Hospital controls were patients at Yale-New Haven

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

**Requests for reprints:** Yong Zhu, Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT 06520. Phone: 203-785-4844; Fax: 203-737-6023; E-mail: yong.zhu@yale.edu.

©2009 American Association for Cancer Research.  
doi:10.1158/0008-5472.CAN-09-0236

Hospital who underwent breast-related surgery for histologically confirmed benign breast diseases. Tolland County controls were identified either through random-digit dialing (for subjects ages <65 years) or through the Health Care Finance Administration files (for subjects ages  $\geq$ 65 years). Among Yale-New Haven Hospital subjects, participation rates were 71% for controls and 77% for cases; among Tolland County subjects, participation rates were 61% for controls and 74% for cases. The study was approved by institutional review boards at Yale University, the Connecticut Department of Public Health, and the National Cancer Institute. Participation was voluntary, and written informed consent was obtained. Thus, 441 cases and 479 controls had DNA samples available for genetic association analysis, and of these, 77 cases were untreated. Supplementary Table S4 presents the distribution of selected baseline characteristics for cases and controls.

**Single nucleotide polymorphism identification and genotyping.** We used a recent release of the single nucleotide polymorphism (SNP) database dbSNP b125 on National Center for Biotechnology Information Assembly B35 from HapMap Data Release #20/phase II to search for genetic variation in miRNA genes. The miRNA base target positions on human chromosomes were first determined using the *Search MicroCosm* database produced by the Sanger Institute.<sup>4</sup> These positions were then explored using HapMap to locate SNPs within the region of a given miRNA gene. Genotyping for all SNPs was done at Yale University's W.M. Keck Foundation Biotechnology Research Laboratory using the Sequenom MassARRAY multiplex genotyping platform (Sequenom). Duplicate samples from 100 study subjects and 40 replicate samples from each of two blood donors were interspersed throughout each batch for all genotyping assays. The concordance rates for quality-control samples were >95% for all assays. All genotyping scores, including quality-control data, were rechecked by different laboratory personnel and the accuracy of each assay was confirmed.

**CpG island identification and methylation analysis.** Using the CpG Island Searcher Web tool,<sup>5</sup> a CpG island was identified upstream (-700 bp) of the region encoding pri-miRNA-196a-2. The MethPrimer program<sup>6</sup> was then used to design methylation-specific PCR primers within the identified CpG island region. The two methylated primers were aatttcgatagtagttaatagaacg (left) and ctaaattctcgacaacacga (right), and the two unmethylated primers were ttttaatttgatagtagttaatagaatg (left) and cctaattctcaacaacacaa (right). Genomic DNA samples were bisulfite treated using the EZ DNA Methylation Kit (Zymo Research) according to the manufacturer's protocol. The presence of methylation was determined by quantitative PCR using the Power SYBR Green Kit (Applied Biosystems), and the percentage of methylation was estimated for each sample using the formula: methylation index =  $[1 / (1 + 2^{-(CT_u - CT_{me})})] \times 100\%$ , as described previously (18), where  $CT_u$  = average cycle threshold obtained from duplicate quantitative PCRs using the unmethylated primers and  $CT_{me}$  = average cycle threshold obtained using the methylated primers.

**Cell culture and construction of miR-196a-2 expression vector.** Human breast adenocarcinoma cells (MCF-7) were obtained from the American Type Culture Collection and maintained in DMEM (Invitrogen) supplemented with 10% fetal

bovine serum (Invitrogen), 0.01 mg/mL bovine insulin, and 1% penicillin/streptomycin (Sigma-Aldrich). An expression plasmid vector (pRNAT-U6.1/Neo) containing a neomycin resistance gene and a coral green fluorescent protein marker under cytomegalovirus promoter control was purchased from GeneScript. Individuals known to be homozygous for either allele of pre-miR-196a-2 (C or T) were identified by genotyping (as detailed above), and genomic DNA from these subjects was amplified to generate pRNA-miR-196a-2-C and pRNA-miR-196a-2-T, containing the mature miRNAs, along with a flanking region on each side (130 bp on left and 120 bp on right). PCR products (360 bp) were purified and inserted into the pRNA vector according to the manufacturer's protocol (GeneScript). The ligation mixture was then transformed into competent DH5a cells (Invitrogen) and plated on LB-G418 plates. Positive clones were chosen and grown in liquid medium, and the constructed plasmid vectors were extracted from culture using a MiniPrep kit (Qiagen). Vectors were then sequenced to verify the accuracy of the insert, and each vector was transfected into cells using Lipofectamine 2000 transfection reagent (Invitrogen). Transfection efficiency was verified by fluorescence microscopy, and stable cell lines were established by adding G418 (Sigma-Aldrich) to the culture medium.

**RNA isolation and miR-196 detection.** Total RNA samples were isolated using the RNA Mini Kit (Qiagen), with on-column DNA digestion. The primers used for detecting the miR-196a-2 precursor were acccagcaaccaagtcta (left) and atctggaggagaagggaagg (right), and expression was assessed using the  $2^{-\Delta\Delta Ct}$  method with RNA content normalized to the housekeeping gene *hypoxanthine phosphoribosyltransferase 1*. To determine the levels of mature miRNA for both products of the miR-196a-2 stem loop (miR-196a and miR-196a\*), miRNA-specific quantitative real-time PCR was done using the NCode kit (Invitrogen), which polyadenylates the miRNAs before cDNA synthesis. The cDNA was then amplified using one universal PCR primer targeting on the polyadenylated region of the miRNA (supplied with kit) and a miRNA-specific forward primer (miR-196a sequence: taggtagtttcatgttgggg, miR-196a\*-C: cggcaacaagaactgCctgag, and miR-196a\*-T: cggcaacaagaactgTctgag), which is complementary to the miRNA cDNA. Amplification was measured by SYBR Green and fold change in miRNA abundance relative to empty vector was calculated using the  $2^{-\Delta\Delta Ct}$  method normalized to miR-16. All quantitative real-time PCRs were done in triplicate.

**Genome-wide expression microarray and pathway analysis.** Gene expression differences in pre-miR-196a-C and pre-miRNA-196a-T transfected cells were interrogated by whole-genome microarray (Agilent, 41K chip, done by MoGene). RNA was isolated for array analysis from biological replicates of cells transfected with an empty vector (negative control), pre-miR-196a-C, or pre-miR-196a-T. Fold changes in gene expression relative to the negative control and false discovery rate (FDR) adjusted *P* values were determined for each gene in cells from each treatment. Transcripts were identified as significantly influenced by each miRNA treatment if they fit the criteria of  $FDR < 0.01$  and fold change  $> |2.0|$ . Significantly altered transcripts were investigated for network and functional interrelatedness using the Ingenuity Pathway Analysis software tool (Ingenuity Systems).<sup>7</sup> This software scans the set of input genes to identify networks using information in the

<sup>4</sup> <http://microrna.sanger.ac.uk/sequences/>

<sup>5</sup> <http://www.cpgislands.com/>

<sup>6</sup> <http://www.urogene.org//methprimer>

<sup>7</sup> <http://www.ingenuity.com>

Ingenuity Pathways Knowledge Base, an extensive, manually curated database of functional interactions extracted from peer-reviewed publications (22). A Fisher's exact test, based on the hypergeometric distribution, is then done to determine the likelihood of obtaining at least the same number of molecules by chance (from a random input set), as actually overlap between the input gene set and the genes present in each identified network. A subset of breast cancer relevant genes from the top identified network were validated by quantitative PCR. All microarray data were uploaded to the Gene Expression Omnibus database<sup>8</sup> accession no. GSE15576.

**Cell cycle assay.** Each cell treatment group (empty vector, pre-miR-196a-C, and pre-miR-196a-T) was treated with either 0.015% (v/v) methyl methanesulfonate (DNA damage-inducing agent) or mock treatment (0.015% PBS, pH 7.4) for 1 h. Methyl methanesulfonate is an alkylating agent that interacts directly with DNA and creates lesions that are normally repaired via the base excision or nucleotide excision repair pathways (23). As such, after methyl methanesulfonate treatment, we expected to see a delay in the G<sub>2</sub>-M phase of the cell cycle, as these repair mechanisms operate before cell cycle progression. Cells were then washed two times in PBS and incubated under normal conditions for 24 h. Each cell population was then harvested and stained with propidium iodide to quantify DNA content for determination of cell cycle distributions. Cell fluorescence was determined by flow cytometry (FACSCalibur; Becton Dickinson), and gates were set to include cells with green fluorescent protein activity only (to include only successfully transfected cells). Cell phases were determined with the CellQuest Software (Becton Dickinson) using the Watson-Pragmatic algorithm (24). All data are the result of three biological replicates of each treatment condition.

**Statistical analysis.** All statistical analyses were done using the SAS statistical software (SAS Institute) unless otherwise noted. A  $\chi^2$  test was used to test for departures from Hardy-Weinberg equilibrium for each SNP in the control population. Tests for trend were conducted by assigning the ordinal values 0, 1, and 2 to the genotypes in rank order of wild-type, heterozygous, and homozygous variant genotypes, respectively. Adjusted odds ratios (OR) with 95% confidence intervals (95% CI) were calculated in the overall population, and in premenopausal and postmenopausal women separately, by logistic regression to estimate the relative risk associated with each genotype. Adjusted ORs were calculated by unconditional logistic regression to control for age, ethnicity, family history of breast cancer in first-degree relatives, menopausal status, body mass index, smoking status, family income, and study site. For analysis of methylation in the region upstream of miR-196a-2, the median methylation index in the controls was used to divide the population into two categories of "high" or "low" methylation, and an OR for high methylation was calculated by logistic regression, controlling for the same covariates as the genotyping analysis. Methylation index in cases and controls was also compared using the Wilcoxon rank-sum test (normal approximation). Cell cycle response to mutagen challenge was measured by comparing the proportion of cells in each phase in the mutagen-treated group with the same proportion in the mock-treated group based on triplicate

experiments for each experimental condition. The normality of each of these sets of proportions was assessed using the Shapiro-Wilk test, and no *P* value suggested a significant departure from the normal distribution at  $\alpha = 0.05$ . Similarly, equality of variances was assessed using the Folded *F* method, and all tests suggested no departure from equal variances. As such, the two-sample pooled (homoscedastic) Student's *t* test was used. All *P* values are two-sided.

## Results

**A common SNP (rs11614913) in hsa-miR-196a-2 is associated with breast cancer risk.** There were a total of 462 miRNA genes in the *MicroCosm* database produced by the Sanger Institute at the time of our search. By mapping the chromosomal positions of each miRNA on HapMap, we were able to initially identify 15 genetic variations in 15 different miRNA genes (Supplementary Table S1). These sequence variants were located in various regions of the miRNA gene, including the mature sequence, the stem-loop structure, and the primary miRNA region. Of the 15 original SNPs, 5 were rarely polymorphic (minor allele frequency < 0.1), 1 failed to amplify in the genotyping assay, and 2 had low-quality amplification and allelic distributions that significantly departed from that which would be expected under Hardy-Weinberg equilibrium ( $P < 0.05$ ). These SNPs were thus excluded from further analysis. Each remaining SNP was analyzed in the full population as well as in a secondary analysis with the population stratified by menopausal status (Supplementary Table S2). The homozygous variant genotype of a common SNP (rs11614913; minor allele frequency = 40%) located in miR-196a-2 was significantly associated with reduced breast cancer risk in the full population (OR, 0.44; 95% CI, 0.28-0.70), as were the combined variant genotypes (heterozygous + homozygous variant; OR, 0.75; 95% CI, 0.57-0.98;  $P_{\text{trend}} = 0.002$ ; Table 1). Similar results were obtained after stratifying by menopausal status and when restricting the sample to Caucasians only (data not shown). Note that this SNP would remain significant at  $\alpha = 0.05$  even after employing the conservative Bonferroni adjustment for multiple comparisons ( $P_{0.002} * 15 = 0.030$ ).

**Methylation of a CpG-rich region upstream of hsa-miR-196a-2 is associated with breast cancer risk.** In addition to the genetic variant identified in hsa-miR-196a-2, a CpG island was also identified in the region upstream of the miRNA precursor (Fig. 1). The level of methylation in this region was determined by methylation-specific PCR for all women who had not undergone radiotherapy or chemotherapy due to the potential for these treatments to influence DNA methylation ( $n = 77$  controls and 73 cases). Cases were significantly less likely to have high methylation at this CpG site (OR, 0.36; 95% CI, 0.16-0.82;  $P = 0.02$ ). Methylation indices were also compared using the Wilcoxon rank-sum test; again, methylation was significantly lower among cases ( $P = 0.03$ ; Table 1). A comparison of untreated cases with treated cases was done, which revealed no significant differences in any patient characteristics (data not shown).

**Effect of the variant on miR-196a levels *in vitro*.** hsa-miR-196a-2 is composed of two different mature miRNAs (miR-196a and miR196a\*), which are processed from the same stem-loop (Fig. 1). The SNP (rs11614913) lies in the mature sequence of miR-196a\* but may influence either miRNA by affecting processing of the pre-miRNA to its mature form. To investigate the influence of the SNP on miRNA processing and target binding efficiency, three separate

<sup>8</sup> <http://www.ncbi.nlm.nih.gov/projects/geo/>

**Table 1.** Epidemiologic analyses of a genetic variant in pre-miR-196a-2 and upstream CpG island methylation

Genetic analysis				Methylation analysis				P
Genotype	Cases, n (%)	Controls, n (%)	OR* (95% CI)	Methylation index	Cases, n (%)	Controls, n (%)	OR* (95% CI)	
C/C	181 (42.5)	166 (35.6)	—	Low	50 (68.5)	38 (49.4)	—	0.0168
T/C	209 (49.1)	229 (49.1)	0.84 (0.63-1.12)	High	23 (31.5)	39 (50.6)	0.35 (0.15-0.83)	
T/T	36 (8.4)	71 (15.3)	0.44 (0.28-0.70)	Wilcoxon rank-sum test of methylation level				
T/C or T/T	245 (57.5)	300 (64.4)	0.75 (0.57-0.98)		Cases	Controls	OR*	P
Trend			0.002	Median	1.02	1.35	—	0.0286

\*Adjusted for age, race, menopausal status, family history of breast cancer, body mass index, income, smoking, and study site.

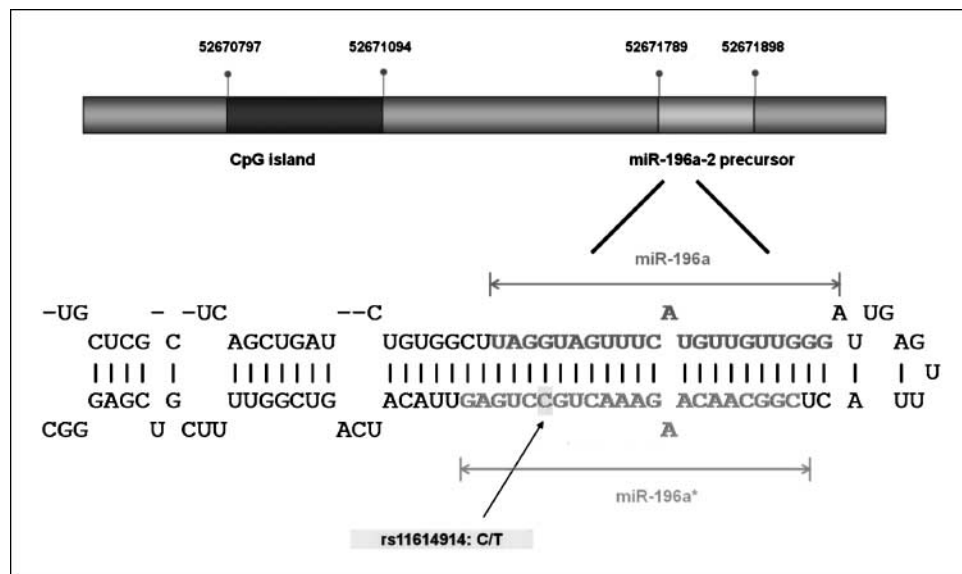
expression vectors were constructed, which contained either pre-miR-196a-C, pre-miR-196a-T, or no miRNA insert (empty vector/negative control). MCF-7 cells were vector transfected and selected for 2 weeks with G418, and RNA was isolated for subsequent analysis by quantitative real-time PCR to determine the levels of precursor and mature miRNA. Relative to the empty vector, both pre-miR-196a-C-treated and pre-miR-196a-T-treated cells had approximately equal levels of the precursor (9.8- and 11.1-fold increase, respectively). Mature miR-196a levels were increased 9.3-fold in cells transfected with pre-miR-196a-C compared with empty vector but only 4.4-fold in cell transfected with pre-miR-196a-T. In cells transfected with pre-miR-196a-C and those transfected with pre-miR-196a-T, mature miR-196a\* was increased 3.1- and 3.4-fold, respectively (Fig. 2). These results indicate that miR-196a is the primary mature product of the pre-miR-196a-2 hairpin, and processing of this mature sequence is affected by the SNP. Note that the results shown for miR-196a\* were obtained using a primer sequence complementary to the C allele. Similar results were obtained when using a primer for the T allele (data not shown), which indicates insufficient specificity for the assay to discriminate between this single-base substitution.

**Effect of the variant on expression of miR-196a-2-mediated genes.** A whole-genome expression microarray was also done using

RNA isolated from cells transfected with each of the vectors (empty vector, pre-miR-196a-C, and pre-miR-196a-T). RNA was isolated separately from two independently transfected populations (biological replicates), and genome-wide expression was determined in miR-196a-C cells compared with empty vector and in miR-196a-T cells compared with empty vector. A total of 684 transcripts fit the criteria for altered expression (FDR < 0.01 and fold change > |2|) following introduction of pre-miR-196a-C, whereas fewer than half that number (263) were significantly altered following introduction of pre-miR-196a-T. Because the epidemiologic evidence suggests that the variant allele is associated with protection from breast cancer risk, and the variant leads to diminished regulatory capacity and decreased mature miR-196a levels, this data indicate that miR-196a may have oncogenic properties in breast cancer.

**Cancer-related network formed by miR-196-mediated genes.** To determine which biological pathways may be influenced by the products of pre-miR-196a-2, the set of genes with altered expression following introduction of the pre-miR-196a-C vector was examined using the Ingenuity Pathway Analysis software package. A list of all networks significantly associated with the input genes is available in Supplementary Table S3. The most significant disease associated with the altered genes was cancer,

**Figure 1.** CpG island location and stem-loop structure of hsa-miR-196a-2, including mature sequences of miR-196a and miR-196a\*, and the location of the genetic variant.



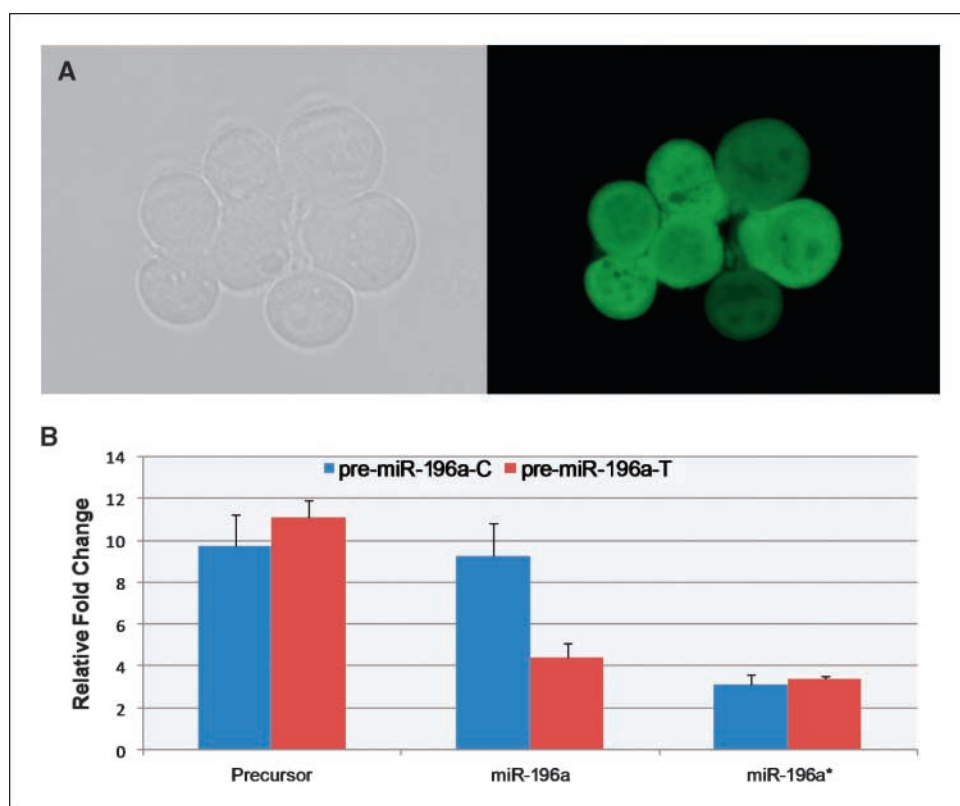
with 137 cancer-related molecules identified. In addition, a significant pathway ( $P = 1.0E^{-48}$ ) was identified with known relevance for "Cancer, Cell Death, and Reproductive System Disease" (Fig. 3). This network consists of 35 molecules, 2 of which (hCG and MAP2K1/2) refer to gene families with multiple subunits, leaving 33 molecules that can be identified uniquely in the microarray dataset (Table 2). To validate the microarray results, a subset of 9 breast cancer relevant genes identified in this network as differentially regulated by pre-miR-196a-C were validated by quantitative PCR. In all cases, the direction of the fold change was the same when analyzed by array or PCR, and there was excellent correlation between the two platforms ( $r = 0.979$ ; Supplementary Table S5).

All genes in the network were then compared in the miR-196a-C and miR-196a-T populations to determine the effect of the SNP on target gene regulation. Twenty-nine of the 33 genes were significantly differentially expressed in the miR-196a-C population, whereas only 3 were significantly altered in the miR-196a-T cells, again suggesting diminished regulatory capacity of the variant form. Cancer-relevant transcripts in this network included tumor protein p63 and the tumor suppressors growth arrest and DNA damage inducible  $\gamma$  and inhibin  $\beta$ B. Also present in this network were two members of the S100 family of genes (S100A8 and S100A9), which are prognostic indicators for poor clinical outcome in some breast tumors. All of these genes were significantly altered following introduction of miR-196a-C but were unchanged in the miR-196a-T cells, suggesting that this SNP has the potential to influence genes with relevance for breast tumor promotion and progression.

**Cell cycle response to mutagen challenge affected by miR-196a-2.** In addition to cancer generally, the network analysis identified 86 differentially expressed molecules with relevance for cellular growth and proliferation. To test whether miR-196a-2 may influence cell cycle phenotypically, we performed an analysis of cell cycle regulation using cell populations from each vector treatment group. After methyl methanesulfonate treatment, we expected to see a delay in the G<sub>2</sub>-M phase of the cell cycle, as repair mechanisms operate before cell cycle progression. As expected, all cells exhibited a substantial G<sub>2</sub> delay following mutagen challenge (empty vector: 4.1 times as many cells in G<sub>2</sub> compared with mock-treated, Student's  $t$  test,  $P = 0.026$ ; pre-miR-196-C: 2.6 times as many G<sub>2</sub> cells in mutagen-treated versus mock-treated,  $P = 0.040$ ; and pre-miR-196-T: 6.5 times as many G<sub>2</sub> cells,  $P = 0.002$ ; Fig. 4). However, when comparing the response of pre-miR-196a-C-treated cells versus those treated with pre-miR-196a-T, the G<sub>2</sub> delay was significantly larger in the variant (T) population ( $P = 0.020$ ), again suggesting that the variant has a protective effect on tumorigenesis. Of note: although we may have expected intermediate G<sub>2</sub> response in the variant population, given the intermediate level of mature miR-196a in this population, this group appears to, in fact, display an enhanced response. However, this difference in G<sub>2</sub> delay in the variant population was not significantly different from control ( $P = 0.218$ ), and it remains unclear whether the sequence variant in mature miR-196a\* influences targets relevant for cell cycle progression.

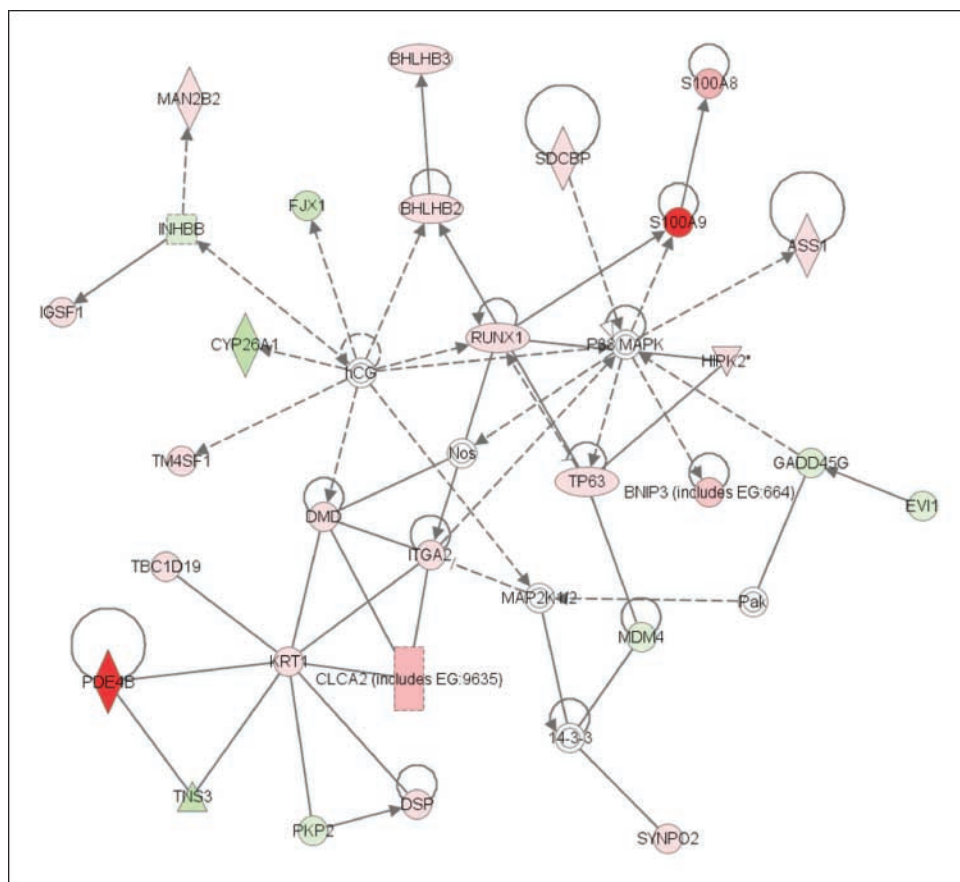
## Discussion

After analyzing 15 SNPs identified from the public SNP database in DNA samples extracted from patients with incident



**Figure 2.** Successful vector transfection of MCF-7 cells and increased expression of the miRNA precursor, mature miR-196a, and mature miR-196a\*. *A*, cluster of cells visualized using both light (*left*) and fluorescence (*right*) microscopy. Expression of green fluorescent protein denotes successful transfection of cells with the expression vector. *B*, levels of precursor and mature miR-196a and miR-196a\* were measured in cells transfected with either pre-miR-196a-C or pre-miR-196a-T. These data indicate that mature miR-196a is the primary product of the miR-196a-2 hairpin, and processing of this mature sequence is affected by the SNP. Mean  $\pm$  SE of triplicate experiments; fold changes are normalized to the negative control.

**Figure 3.** Network of altered genes following miR-196a-C vector transfection. This network was identified by the Ingenuity Pathway Analysis software as significantly associated with the set of genes with altered expression following introduction of pre-miR-196a-C ( $P = 1.0E^{-48}$ ). According to the manually curated Ingenuity Pathway Analysis database, this network has relevance for "Cancer, Cell Death, and Reproductive System Disease."



breast cancer and controls, we report here a significant association between a common sequence variant in miR-196a-2 and predisposition to breast cancer. This association remains significant after correcting for multiple comparisons and is corroborated by *in vitro* evidence, including (a) the variant negatively influences endogenous processing of the miRNA precursor to its mature form; (b) the variant has a phenotypic effect on target gene expression, as fewer genes were altered following introduction of the pre-miR-196a-T vector relative to the pre-miR-196a-C vector; (c) a pathway analysis revealed 137 cancer-related transcripts with significantly altered expression following introduction of pre-miR-196a; and (d) cell cycle response to mutagen challenge in the form of G<sub>2</sub> delay was directly affected in cells transfected with pre-miR-196a-C, and this effect was abrogated in cells transfected with pre-miR-196a-T.

In addition to the *in vitro* findings, a methylation analysis of a CpG island located upstream of the miRNA indicated that increased methylation was associated with decreased breast cancer risk, as cases had significantly higher methylation levels than controls. However, because chemical and radiation therapy may influence methylation patterns, only samples from untreated cases could be used for this analysis, thus significantly reducing the available sample size. In addition, RNA samples were not available for these subjects; thus, expression levels cannot be directly associated with level of methylation. However, a previous study reported a strong correlation ( $r = 0.80$ ) between methylation levels near a miRNA gene and miRNA expression in a subgroup of breast cancer samples (19). If higher degree of methylation correlates with

decreased miRNA expression, these data would be consistent with all other findings, which suggest that miR-196a has oncogenic properties. Further study will be necessary to confirm whether methylation at this site does, in fact, correlate with pre-miR-196a expression and to determine whether this area may indeed serve as a novel epigenetic biomarker for breast cancer susceptibility.

A previous study of 5 common miRNA SNPs conducted in China also identified the same polymorphism (rs11614913) as significantly associated with breast cancer risk (25). Interestingly, although the same allele was associated with risk (C allele), this was the rare allele in the Chinese population, although it was the common allele in our, predominantly Caucasian, population. In another study investigating the same five miRNA SNPs and non-small cell lung cancer survival in China, the same group identified the C allele of rs11614913 as significantly associated with poor survival, indicating the possibility that this miRNA may have implications for prognosis in addition to cancer risk (26). This study also suggests that the pathways influenced by miR-196a are not relevant only for breast cancer but are also important for at least one other unrelated cancer type.

Several members of the HOX gene family are known to be targeted by miR-196a (27). Although many of these genes showed up at poor intensity in our array, four members of the HOX family, HOXB2, HOXB3, HOXC13, and HOXB5, were significantly down-regulated in cells treated with pre-miR-196a-C, with average fold changes of -10.21 (FDR < 0.001), -3.23 (FDR < 0.001), -1.88 (FDR = 0.001), and -2.18 (FDR = 0.038), respectively. Of these, only HOXB2 was significantly down-regulated in pre-miR-196a-T-treated cells,

although the magnitude of the effect was much smaller than that observed in the pre-miR-196a-C population (fold change = -3.69), again suggesting that the SNP results in decreased target gene repression.

The network analysis showed the potential for miR-196a-2 to be operative in several biological pathways, including a high confidence network ( $P = 1.0E^{-48}$ ) with relevance for cancer. Genes in this network included at least two tumor suppressors that were significantly down-regulated following pre-miR-196a-C introduction but were not significantly altered in the pre-miR-196a-T cells. Growth arrest and DNA damage inducible  $\gamma$ , which was down-regulated >2-fold, is epigenetically silenced in several tumors and acts as a tumor suppressor by negatively regulating cell proliferation (28). Inhibin  $\beta$ B, which was also down-regulated >2-fold, is a negative regulator of growth that has been proposed as a tumor suppressor for some cancer types and is expressed in

normal breast tissue (29, 30). Tumor protein p63, which was >6-fold up-regulated following pre-miR-196a-C introduction, may act as an oncogene despite having significant sequence similarity to p53 (31). In addition, two genes encoding calcium-binding proteins (S100A8 and S100A9) were also significantly up-regulated by 15- and 47-fold, respectively. These proteins form a heterodimer, which has been implicated in tumor promotion, possibly mediated by its role as a proinflammatory cytokine (32). In addition, these genes are up-regulated in several tumors, and their overexpression has been associated with poor prognosis in invasive ductal cell carcinoma of the breast (33). Down-regulation of tumor suppressors and up-regulation of oncogenes by introduction of pre-miR-196a-C would be consistent with overall oncogenic activity, and the diminished regulatory activity of pre-miR-196a-T is consistent with a protective role for the T allele of the SNP.

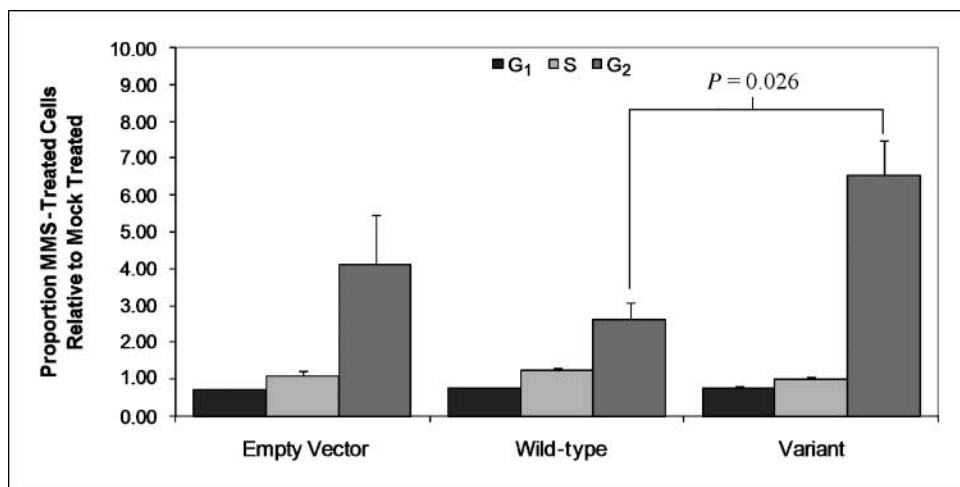
**Table 2.** Network of differentially expressed genes following enforced expression of pre-miR-196a-C and pre-miR-196a-T

Molecule*	Agilent/RefSeq accession no.	pre-miR-196a-C		pre-miR-196a-T	
		Fold change <sup>†</sup>	FDR <sup>†</sup>	Fold change <sup>†</sup>	FDR <sup>†</sup>
14-3-3	NM_006826	-1.45	0.0368	-1.35	0.4054
ASS1	NM_054012	2.75	<0.0001	1.97	0.5898
BHLHB2	NM_003670	2.55	0.0001	2.08	0.1735
BHLHB3	NM_030762	2.06	0.0014	1.77	0.0930
BNIP3	NM_004052	11.18	<0.0001	4.00	0.0140
CLCA2	NM_006536	14.66	0.0004	3.74	0.6009
CYP26A1	NM_057157	-4.46	<0.0001	-2.46	0.0012
DMD	NM_004010	2.71	<0.0001	-1.22	0.8941
DSP	BE693409	6.73	0.0031	5.12	0.0856
EVI1	NM_005241	-2.12	<0.0001	-1.00	0.2974
FJX1	NM_014344	-3.25	<0.0001	-2.70	<0.0001
GADD45G	BI710972	-2.25	0.0022	-1.69	0.1147
HIPK2	BC041926	2.82	<0.0001	1.52	0.0944
IGSF1	NM_001555	4.04	<0.0001	1.34	0.3329
INHBB	NM_002193	-2.41	0.0002	-1.08	0.3479
ITGA2	A_32_P208076	2.65	<0.0001	1.17	0.0228
KRT1	NM_006121	6.01	<0.0001	2.08	0.5909
MAN2B2	NM_015274	2.14	<0.0001	1.67	0.1184
MDM4	BX640923	-2.18	0.0001	-1.25	0.5899
Nos[1]	NM_000620	-1.07	1.0430	1.14	0.9981
P38MAPK	NM_001315	-1.00	1.0104	-1.02	1.1542
Pak[1]	NM_002576	1.09	0.9498	1.38	0.4638
PDE4B	NM_001037341	57.00	<0.0001	1.74	0.8578
PKP2	NM_004572	-2.26	0.0005	-1.64	0.0289
RUNX1	NM_001001890	2.63	<0.0001	1.42	0.1689
S100A8	NM_002964	15.08	<0.0001	4.98	0.1538
S100A9	NM_002965	43.07	<0.0001	8.06	0.6146
SDCBP	BM968705	3.28	<0.0001	2.18	0.0949
SYNPO2	AL833294	5.53	0.0001	-1.19	0.7234
TBC1D19	NM_018317	3.15	<0.0001	2.29	0.0171
TM4SF1	NM_014220	4.30	<0.0001	1.28	0.6111
TNS3	NM_022748	-4.02	<0.0001	-2.94	<0.0001
TP63	NM_003722	6.25	0.0006	-1.06	0.0743

\*hCG and MAP2K1/2 were also included in the network. However, as these molecules are composed of multiple subunits, no unique microarray identifier is available.

<sup>†</sup>Fold change and FDR represent the mean of two biological replicates.

**Figure 4.** Cell cycle response to mutagen challenge affected by miR-196a-2. Cells transfected with empty vector, pre-miR-196a-C, and pre-miR-196a-T were either mock-treated or treated with a DNA damage-inducing agent. The expected G<sub>2</sub> delay was significantly diminished in pre-miR-196a-C relative to pre-miR-196a-T ( $P = 0.020$ ), showing phenotypic effect of the SNP. Mean  $\pm$  SE of triplicate experiments. MMS, methyl methanesulfonate.



Although our findings show that the SNP does have a phenotypic effect in terms of both pre-miRNA processing and target gene regulation, these data are not useful in determining which genes are direct targets of mature miR-196a and miR-196a\* and which are indirectly influenced. As such, an exploration into the nature of the physical interactions between these miRNAs and their potential targets should be the subject of future investigations. Furthermore, given the earlier finding showing a role for this SNP in lung cancer prognosis, similar studies are warranted to determine whether there are prognostic implications for this variant among patients with breast cancer.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Acknowledgments

Received 1/21/09; revised 4/28/09; accepted 5/8/09; published OnlineFirst 6/30/09.

**Grant support:** NIH grants CA122676, CA110937, and CA108369.

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.

We thank Ewa Menet (Cell Sorter Facility) and Irina Tikhonova (Keck Center) for assistance with the cell cycle assay and Sequenom genotyping analysis, respectively.

## References

- Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. *Science* 1969;165:349–57.
- Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 1993;75:843–54.
- Reinhart BJ, Slack FJ, Basson M, et al. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 2000;403:901–6.
- Song L, Tuan RS. microRNAs and cell differentiation in mammalian development. *Birth Defects Res C Embryo Today* 2006;78:140–9.
- Pasquinelli AE. Demystifying small RNA pathways. *Dev Cell* 2006;10:419–24.
- Calin GA, Sevignani C, Dumitru CD, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A* 2004;101:2999–3004. Epub 2004 Feb 29.
- Calin GA, Croce CM. MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res* 2006;66:7390–4.
- Takamizawa J, Konishi H, Yanagisawa K, et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res* 2004;64:3753–6.
- Calin GA, Dumitru CD, Shimizu M, et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 2002;99:15524–9. Epub 12002 Nov 15.
- Gregory RI, Shiekhattar R. MicroRNA biogenesis and cancer. *Cancer Res* 2005;65:3509–12.
- Esquela-Kerscher A, Slack FJ. Oncomirs—microRNAs with a role in cancer. *Nat Rev Cancer* 2006;6:259–69.
- Wijnhoven BP, Michael MZ, Watson DI. MicroRNAs and cancer. *Br J Surg* 2007;94:23–30.
- Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T. New microRNAs from mouse and human. *RNA* 2003;9:175–9.
- Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet* 2007;39:673–7.
- Volinia S, Calin GA, Liu CG, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A* 2006;103:2257–61. Epub 2006 Feb 22.
- Iorio MV, Ferracin M, Liu C-G, et al. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* 2005;65:7065–70.
- Weber B, Stresmann C, Brueckner B, Lyko F. Methylation of human microRNA genes in normal and neoplastic cells. *Cell Cycle* 2007;6:1001–5.
- Lu L, Katsaros D, de la Longrais IA, Sochirca O, Yu H. Hypermethylation of let-7a-3 in epithelial ovarian cancer is associated with low insulin-like growth factor-II expression and favorable prognosis. *Cancer Res* 2007;67:10117–22.
- Lehmann U, Hasemeier B, Christgen M, et al. Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. *J Pathol* 2008;214:17–24.
- Zheng T, Holford TR, Mayne ST, et al. Risk of female breast cancer associated with serum polychlorinated biphenyls and 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene. *Cancer Epidemiol Biomarkers Prev* 2000;9:167–74.
- Zhu Y, Brown HN, Zhang Y, Stevens RG, Zheng T. Period3 structural variation: a circadian biomarker associated with breast cancer in young women. *Cancer Epidemiol Biomarkers Prev* 2005;14:268–70.
- Calvano SE, Xiao W, Richards DR, et al. A network-based analysis of systemic inflammation in humans. *Nature* 2005;437:1032–7.
- Viau CM, Guecheva TN, Sousa FG, et al. SnCl(2)-induced DNA damage and repair inhibition of MMS-caused lesions in V79 Chinese hamster fibroblasts. *Arch Toxicol* 2009 [Epub ahead of print].
- Watson JV, Chambers SH, Smith PJ. A pragmatic approach to the analysis of DNA histograms with a definable G<sub>1</sub> peak. *Cytometry* 1987;8:1–8.
- Hu Z, Liang J, Wang Z, et al. Common genetic variants in pre-microRNAs were associated with increased risk of breast cancer in Chinese women. *Hum Mutat* 2009;30:79–84.
- Hu Z, Chen J, Tian T, et al. Genetic variants of miRNA sequences and non-small cell lung cancer survival. *J Clin Invest* 2008;118:2600–8.
- Yekta S, Shih IH, Bartel DP. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 2004;304:594–6.
- Ying J, Srivastava G, Hsieh WS, et al. The stress-responsive gene GADD45G is a functional tumor suppressor, with its response to environmental stresses frequently disrupted epigenetically in multiple tumors. *Clin Cancer Res* 2005;11:6442–9.
- Reis FM, Luisi S, Carneiro MM, et al. Activin, inhibin and the human breast. *Mol Cell Endocrinol* 2004;225:77–82.
- Sharifi N, Lechleider RJ, Farrar WL. Transforming growth factor- $\beta$  receptor III downregulation in prostate cancer: is inhibin B a tumor suppressor in prostate? *J Mol Endocrinol* 2007;39:329–32.
- Westfall MD, Pietenpol JA. p63: molecular complexity in development and cancer. *Carcinogenesis* 2004;25:857–64.
- Gebhardt C, Nemeth J, Angel P, Hess J. S100A8 and S100A9 in inflammation and cancer. *Biochem Pharmacol* 2006;72:1622–31.
- Arai K, Takano S, Teratani T, et al. S100A8 and S100A9 overexpression is associated with poor pathological parameters in invasive ductal carcinoma of the breast. *Curr Cancer Drug Targets* 2008;8:243–52.