

## Spotlight on Molecular Profiling

# MicroRNA expression profiles for the NCI-60 cancer cell panel

Paul E. Blower,<sup>1</sup> Joseph S. Verducci,<sup>2,3</sup> Shili Lin,<sup>2,3</sup> Jin Zhou,<sup>3</sup> Ji-Hyun Chung,<sup>1</sup> Zunyan Dai,<sup>1</sup> Chang-Gong Liu,<sup>4</sup> William Reinhold,<sup>5</sup> Philip L. Lorenzi,<sup>5</sup> Eric P. Kaldjian,<sup>6</sup> Carlo M. Croce,<sup>4</sup> John N. Weinstein,<sup>5</sup> and Wolfgang Sadee<sup>1</sup>

<sup>1</sup>Program of Pharmacogenomics, Department of Pharmacology and the Comprehensive Cancer Center, College of Medicine, Departments of <sup>2</sup>Statistics, <sup>3</sup>Mathematical Biosciences Institute, and <sup>4</sup>Department of Molecular Virology, Immunology and Medical Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio; <sup>5</sup>Genomics and Bioinformatics Group, Laboratory of Molecular Pharmacology, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland; <sup>6</sup>GeneLogic, Inc., Gaithersburg, Maryland

### Abstract

Advances in the understanding of cancer cell biology and response to drug treatment have benefited from new molecular technologies and methods for integrating information from multiple sources. The NCI-60, a panel of 60 diverse human cancer cell lines, has been used by the National Cancer Institute to screen >100,000 chemical compounds and natural product extracts for anticancer activity. The NCI-60 has also been profiled for mRNA and protein expression, mutational status, chromosomal aberrations, and DNA copy number, generating an unparalleled public resource for integrated chemogenomic studies. Recently, microRNAs have been shown to target particular sets of mRNAs, thereby preventing translation or accelerating mRNA turnover. To complement the existing NCI-60 data sets, we have measured expression levels of microRNAs in the NCI-60 and incorporated the resulting data into the CellMiner program package for integrative analysis.

Received 1/4/07; revised 3/13/07; accepted 3/23/07.

**Grant support:** National Institute of General Medical Sciences grant GM61390; NSF Agreement no. 0112050; and the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

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.

**Requests for reprints:** Paul Blower, Department of Pharmacology, The Ohio State University, 5072 Graves Hall, 333 West 10th Avenue, Columbus, OH 43210. Phone: 614-688-5565; Fax: 614-292-7232. E-mail: blower.7@osu.edu

Copyright © 2007 American Association for Cancer Research.

doi:10.1158/1535-7163.MCT-07-0009

Cell line groupings based on microRNA expression were generally consistent with tissue type and with cell line clustering based on mRNA expression. However, mRNA expression seemed to be somewhat more informative for discriminating among tissue types than was microRNA expression. In addition, we found that there does not seem to be a significant correlation between microRNA expression patterns and those of known target transcripts. Comparison of microRNA expression patterns and compound potency patterns showed significant correlations, suggesting that microRNAs may play a role in chemoresistance. Combined with gene expression and other biological data using multivariate analysis, microRNA expression profiles may provide a critical link for understanding mechanisms involved in chemosensitivity and chemoresistance. [Mol Cancer Ther 2007;6(5):1483–91]

### Introduction

Genomic and proteomic studies have yielded a wealth of novel insights into molecular targets and mechanisms of cancer chemosensitivity and resistance (1–10). Nonetheless, progress in translating those insights into effective therapies has been relatively slow. One way to accelerate the advance toward molecularly based cancer therapy is to integrate various types of molecular information on the same set of cancer samples to develop as comprehensive as possible a molecular portrait of the cells and their pharmacology. An attractive model system for that type of “integromic” enterprise (11) is the panel of 60 diverse human cancer cell lines (the NCI-60) used by the National Cancer Institute (NCI) to screen >100,000 chemical compounds and natural product extracts for anticancer activity since 1990 (12). Included in the panel are nine broad categories of cancer cells: leukemias, melanomas, and cancers of breast, central nervous system (CNS), colon, lung, ovarian, prostate, and renal origin. Dose-response curves generated by the screening process provide 50% growth inhibitory (GI<sub>50</sub>) values for each compound-cell line pair. Screening data for ~43,000 nonproprietary compounds are publicly available.<sup>7</sup>

To take advantage of the pharmacologic profiling, the NCI-60 have also been the subject of numerous genomic, proteomic, and other “-omic” profiling studies (13). Profiles determined using cell materials obtained by methods of cell culture and harvesting plug-compatible

<sup>7</sup> [http://dtp.nci.nih.gov/docs/cancer/cancer\\_data.html](http://dtp.nci.nih.gov/docs/cancer/cancer_data.html)

with those used for the study to be described here have included the following: transcript profiling on multiple platforms (2, 3, 7, 14, 15); proteomic profiling using two-dimensional gels (16) and lysate arrays (15, 17, 18); analysis of single nucleotide polymorphism (19); DNA resequencing for mutational status (20); comparative genomic hybridization for DNA copy number changes (21); spectral karyotyping for chromosomal aberrations (22–24); and promoter region DNA methylation studies for cancer-related genes (19). The collection of data sets related to the NCI-60 provides an unparalleled public resource for integrated chemogenomic studies aimed at elucidating molecular targets, identifying biomarkers for personalization of therapy, and understanding mechanisms of chemosensitivity and chemoresistance. To highlight the possibilities, *Molecular Cancer Therapeutics* launched a new series in November 2006 under the rubric “Spotlight on Molecular Profiling” with three articles on molecular characterization of the NCI-60 (20, 25, 26). Such studies have led to hypotheses tested in well-controlled experimental systems, with promising candidates progressing toward the clinic. A well-integrated set of such molecular profile databases obtained under strictly controlled, standardized protocols has been incorporated into the CellMiner program package,<sup>8</sup> which also includes tools for querying and combining the data sets. Various molecular profile data sets on the NCI-60 are also available online.<sup>9</sup> Because cell lines have been removed from their *in vivo* context and selected for growth in culture, they cannot be considered accurate surrogates for clinical tumors. However, they offer a number of advantages for chemogenomic studies (19). The cell lines are reasonably stable and reproducible over extended time periods; they are available in large quantities; and they are manipulable experimentally (e.g., by transfection or by selection for drug resistance).

MicroRNAs are small noncoding RNAs of 21 to 25 nucleotides that negatively modulate protein expression (27–29). One strand of the mature double-stranded microRNA is incorporated into the RNA-induced silencing complex, which down-regulates target mRNAs either by degradation or by translational inhibition (28). MicroRNAs play important roles in normal regulation of gene expression for developmental timing, cell proliferation, and apoptosis. Moreover, altered microRNA expression is implicated in cancers. MicroRNAs have also been shown to play critical roles in cancer biology (30–35). For example, Volinia et al. (35) identified microRNAs that are differentially expressed in six solid tumors compared with normal tissues. However, their effect on chemotherapy has yet to be studied systematically. MicroRNAs could provide a critical link for understanding chemosensitivity/resistance patterns in the NCI-60. In particular, they could, in principle, help to explain discrepancies between mRNA and protein

levels. Those discrepancies seriously complicate the use of mRNA profiles to study chemoresistance (7–10).

We have used custom pin-spotted microarrays to measure all known microRNAs (as of January 2006) in the NCI-60. The data have been deposited in the NCI Genomics and Bioinformatics Group’s CellMiner database,<sup>8</sup> which provides a variety of NCI-60 databases at the DNA, RNA, protein, and pharmacologic levels. CellMiner is a user-friendly, queryable, web-based relational database (under MySQL) that facilitates navigation and integrative analysis of the molecular profiles. The data have also been deposited in ArrayExpress (ref. 36; accession no. E-MEXP-1029) and in the NCI Developmental Therapeutics Program databases.<sup>9</sup> This report provides an analysis of the distribution of microRNA expression across the NCI-60 (37). We assess data quality, compare cell line clusters based on microRNA and mRNA expression patterns, and analyze microRNAs that differentiate among tissue types or that distinguish among other cell line groupings.

## Materials and Methods

### The NCI-60 Cancer Cell Lines

Cell stocks were obtained from the NCI Developmental Therapeutics Program. The Genomics and Bioinformatics Group then cultured them, harvested RNA, and purified the RNA by a method (see below) that preserves the small microRNA species. To permit broad, integrative use of the data, the cells were cultured under essentially the same conditions as were used for the drug screen, for mRNA expression profiling, and for other molecular studies of the NCI-60 by the Genomics and Bioinformatics Group. In particular, the cells were grown in tissue culture flasks at 37°C in 5% CO<sub>2</sub> in RPMI 1640 with L-glutamine and 10% fetal bovine serum, without antibiotics. Total RNA was extracted at ~80% confluence using Trizol (Invitrogen) according to the manufacturer’s instructions. The time from incubator to stabilization of the preparation was kept to <1 min. For details, see Shankavaram et al. (15).

### Microarray Hybridization

MicroRNA labeling and hybridization were done as previously described (38) using 5 µg of total RNA. Our pin-spotted microRNA microarray (ref. 37; Ohio State University Comprehensive Cancer Center, version 3.0) contains probes for 321 mature human microRNAs, spotted in duplicate. Included are 627 human microRNA probes, typically with more than one probe for a given mature microRNA as well as probes for most precursor microRNAs. Hybridization signals were detected with Streptavidin-Alexa647 conjugate, and scanned images (Axon 4000B) were quantified using the Genepix 6.0 software (Axon Instruments).

### Experimental Design

Because it was logistically infeasible to run microarrays for all of the samples simultaneously, data for microRNA expression in the NCI-60 cell lines were collected in three batches, 20 cell lines + 4 controls (24 samples) per batch for

<sup>8</sup> <http://discover.nci.nih.gov>

<sup>9</sup> <http://dtp.nci.nih.gov>

a total of 72 samples. To select cell lines for assignment to each batch, the lines were cluster-ordered based on gene expression (2) and divided into three groups of 20 cell lines with reasonably similar expression profiles. For each batch, a balanced random selection was taken from each of the three groups, and one cell line was chosen as a control from each batch. The controls, selected for tissue diversity and moderate doubling time, were (a) prostate cell line PC-3, (b) leukemia K-562, and (c) lung cancer A549. To enable us to assess within- and between-batch variability, the controls were run in each of the three batches, and, for each batch, one of the controls was run in triplicate. That design is illustrated schematically as follows, with letters representing the three controls and periods representing other cell lines:

Batch A: aabc.....a.....  
 Batch B: abbc.....b.....  
 Batch C: abcc.....c.....

#### Quality Assessment and Normalization of Microarray Data

The data were analyzed using the R software package (39). The signal intensity of each spot was calculated by subtracting local background (based on the median intensity of the area surrounding each spot) from the median signal. After converting any negative values to 1, signal intensities were log<sub>2</sub> transformed and duplicate spots were averaged. In addition, for cluster analyses, we did quantile normalization (40) across the 3 microarray batches, replaced all values <5 with the median of such values, and set the value for each control cell line to the mean of its five replicates.

Using the unnormalized data, we then generated estimated bias curves from the 15 control values (i.e., 5 replicates of each of the 3 control cell lines, PC-3, K-562, and A549). For each microRNA probe and cell line, we averaged expression scores over the five replicates to give a "true" expression level for each of the 1,881 (i.e., 627 × 3) microRNA probe-cell line combinations. For each batch, there were 627 microRNA probes × 5 samples = 3,135 expression values that we paired with the 1,881 "true values" + 1,254 (i.e., 2 × 627) replicated true values. For each batch, we subtracted the paired true expression levels from the observed expression levels to get 3,135 bias values.

#### Affymetrix U133A Expression Data for the NCI-60

The methods used have been described in detail elsewhere (15). Briefly, the protocols for cell culture, cell harvest, and purification of RNA were as described above for microRNA except that the Genomics and Bioinformatics Group purified total RNA using the Qiagen RNeasy Midi Kit according to the manufacturer's instructions. Labeling, hybridization, scanning, and primary data analysis were done at GeneLogic, Inc. as per the previously described protocols (15). The MAS5 algorithm was used to generate signal intensities for the HG\_U133A arrays, and expression values were normalized to a mean target level of 100.

#### Tar Base

TarBase (41) gives experimentally supported targets for microRNAs, with literature references and links to other databases. The version used here was downloaded Sept. 20, 2006.

#### Compound Database

To assess potential associations between microRNA expression and drug potencies, we used the September 2003 release of the NCI antitumor drug screening database. The data were obtained from the NCI DTP website,<sup>7</sup> which contains nonconfidential screening results and chemical structural data. For each compound and cell line, growth inhibition after 48 h of drug treatment had been assessed from changes in total cellular protein using a sulforhodamine B assay (12). The assay provides GI<sub>50</sub> values for all of the compound-cell line pairs. From those data, we obtained a curated subset of 7,794 compounds as follows: For drugs that had GI<sub>50</sub> values recorded using multiple concentration ranges, we chose only the value obtained for the best concentration range. The best concentration range was chosen as the one containing the greatest number of values in the range (low concentration, +0.2; high concentration, -0.2). In other words, values close to either end point were discounted. Compounds with fewer than 30 accepted values were excluded. Finally, to focus on compounds showing good discrimination between cell lines, we selected compounds with  $[\max(\text{GI}_{50}) - \min(\text{GI}_{50})] \geq 2.0$ . That filtering process resulted in a database of 3,089 compounds.

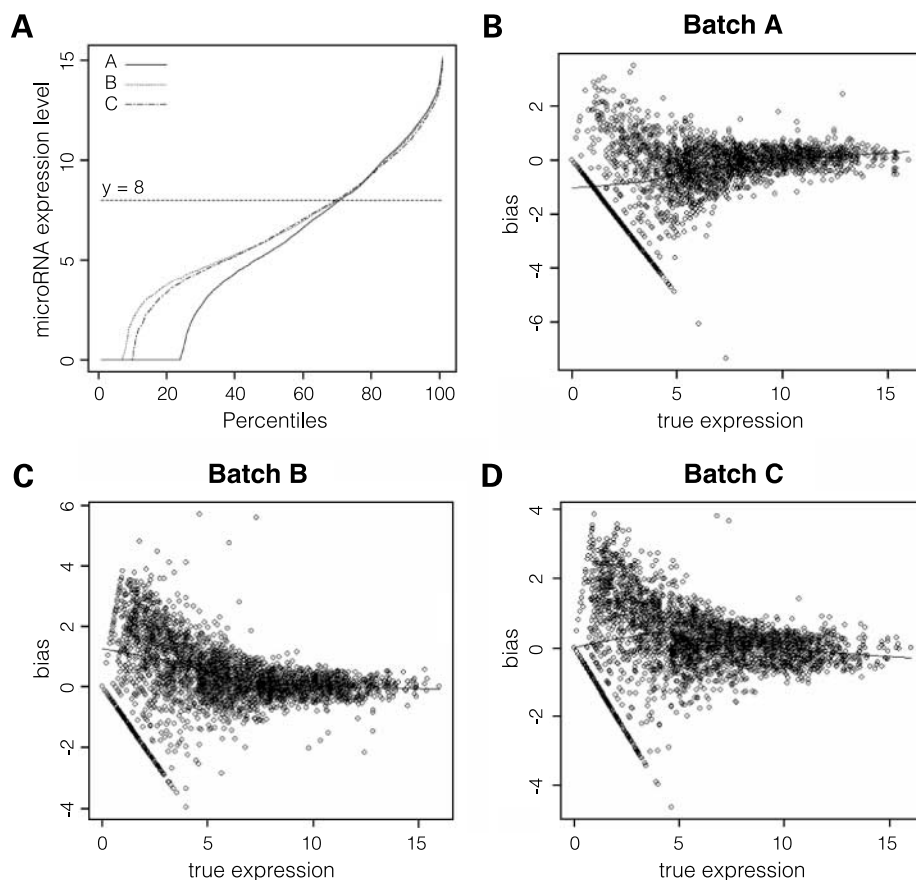
#### Results and Discussion

For microRNA profiles, we compared the distributions of data values across the three batches, 15,048 (i.e., 627 × 24) microRNA probe-cell line measurements per batch. Figure 1A shows quantile plots for the three batches. There was clearly a batch effect at low expression levels, that is for levels of the log<sub>2</sub> value of relative spot fluorescence intensity less than ~8. Batches A, B, and C gave ~24%, ~7%, and ~9% zero values, respectively. The expression levels at the 25th, 50th, and 75th percentiles were as follows:

Batch	25th	50th	75th
A	1.3	5.5	8.7
B	4.3	6.1	8.6
C	4.1	6.2	8.8

Above the 70th percentile (i.e., expression level ~8), the quantile curves for all three batches became nearly identical, with the curve for batch C remaining slightly lower than the other two.

To assess batch effects at the microRNA probe level, we plotted estimated bias curves calculated from the 15 control samples, representing 5 replicates each of the PC-3, K-562, and A549 cell lines. The estimated bias curves in Fig. 1B to



**Figure 1.** Quality assessment of micro-RNA microarray batches. **A**, quantile plots for batches A, B, and C. **B**, estimated bias curve for batch A. Y axis, difference between true and observed expression levels; X axis, true expression level. **C**, estimated bias curve for batch B. **D**, estimated bias curve for batch C.

D show how the difference between true and observed expression levels varies with expression level for each batch. The diagonal line of data points at the lower left of each plot resulted from the constraint that the estimated bias is never less than the true value. All three batches showed greater variability at expression levels less than  $\sim 8$ . Batches A and C were negatively biased at the low end, where observed values tended to be lower than true values. Batch B was positively biased at the low end. That observation is consistent with the quantile plot in Fig. 1A. At the high end, all three batches showed low variability and low bias.

To focus our analysis on microRNAs likely to be most informative and discriminating, we filtered the full set of 627 probes to select those with  $\log_2(\text{exp}) \geq 8$  in at least 10% of the cell lines. That selection process gave 279 microRNAs,  $\sim 46\%$  of the total probes. For further assessment of between-sample variability, we clustered the 72 samples based on the expression patterns of the 279 selected microRNAs using complete linkage clustering and a correlation metric. For each of the three control sets, the five replicates clustered together with no intervening cell lines. The mean pairwise correlation coefficient within replicate sets was  $r = 0.96$  for PC-3 and  $r = 0.94$  for both K-562 and A549.

#### Clustering of NCI-60 Cell Lines Based on MicroRNA Expression

Figure 2A shows the clustering of cell lines based on expression patterns of the 279 selected microRNAs. The mean cell-cell Pearson correlation coefficient was 0.74, indicating an essential biological similarity across the various cell types. Setting a cutoff of 0.7 for intercluster correlation gave nine clusters and one singleton (K-562). Supplementary Fig. S1<sup>10</sup> shows a clustered image map (i.e., clustered heat map; ref. 1) of the 279 selected microRNAs, with color coding for the expression levels.

For comparison, we also clustered the cell lines on the basis of U133A gene expression data obtained previously (15). For that analysis, we used the 1,000 genes with the largest interquartile ranges, again by complete linkage clustering and a correlation metric (Fig. 2B). Cell line groupings based on microRNA expression were generally consistent with tissue type and with cell line clustering based on mRNA expression. The leukemias clustered together, well separated from all other cell types. The melanomas formed a tight cluster and included two cell

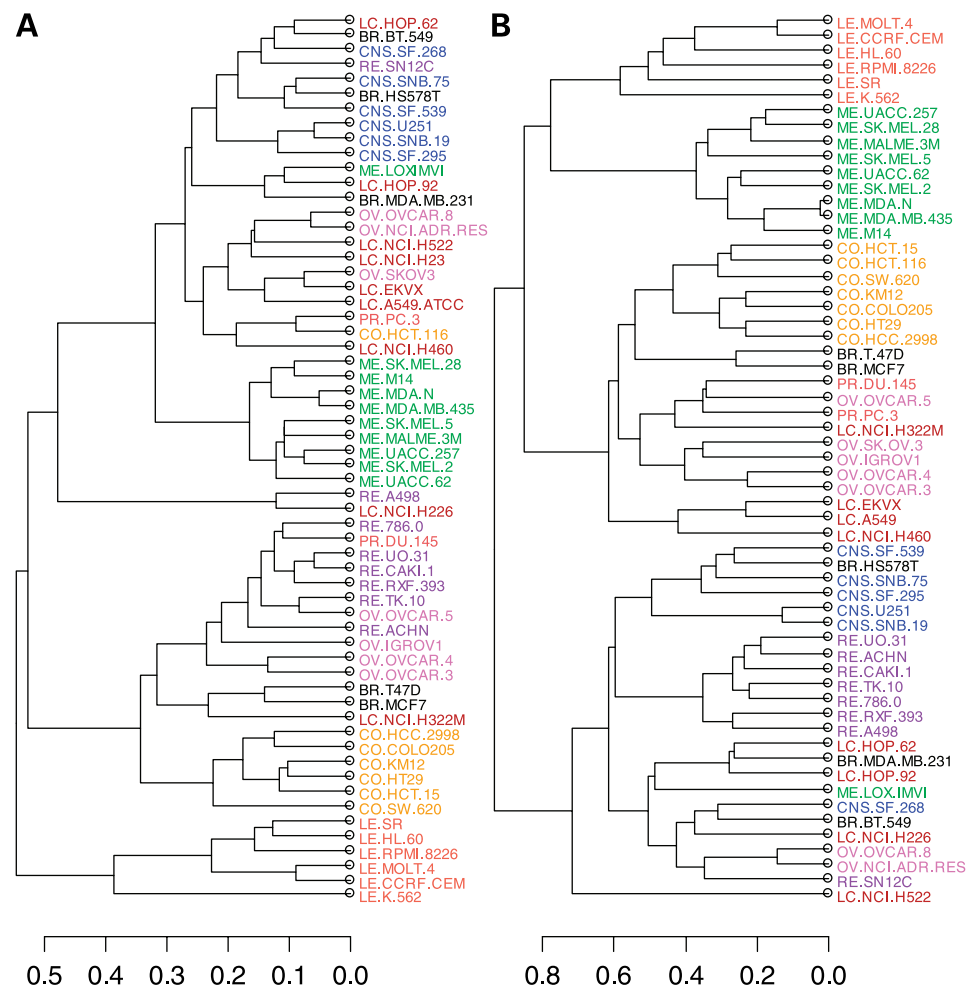
<sup>10</sup> Supplementary data for this article are available at Molecular Cancer Therapeutics Online (<http://mct.aacrjournals.org/>).

lines originally thought to be breast cancers (MDA-MB-435 and MDA-N). A majority of the colon, renal, and CNS cancers also clustered by tissue type. Furthermore, cell pairs known to be very similar clustered together. Included were cell line pairs from the melanoma panel [MDA-MB-435 and MDA-N ( $r = 0.95$ ), MDA-MB-435 and M14 ( $r = 0.90$ )], the CNS panel [SNB-19 and U251 ( $r = 0.92$ )], and the ovarian panel [OVCAR-8 and OVCAR-8/ADR-RES ( $r = 0.93$ )]. MDA-MB-435 was originally thought to be derived from the pleural effusion of a patient with breast cancer, but it is absolutely characteristic of melanotic melanoma in all molecular profiling studies (2, 14, 15). MDA-N is an ErbB2 transfectant of MDA-MB-435. On the basis of comparative genomic hybridization and spectral karyotyping, we identified what was previously called NCI/ADR-RES as a drug-resistant derivative of OVCAR-8 ovarian cancer (i.e., OVCAR-8/ADR-RES; refs. 21, 23). In both cluster dendrograms, the breast and lung cancer cell lines were widely distributed. As with virtually all of our molecular profiles of the NCI-60, the two hormone-dependent ER+ breast cancer lines, MCF7 and T47D, clustered together, but the other breast lines did not. Similarly, the lung lines are widely distributed in the cluster tree no matter what type of

molecular profile at the DNA, RNA, protein, or pharmacologic level is being analyzed. On the basis of the multiple types of molecular data now in hand, we are investigating those relationships further, but that analysis is beyond the scope of the present study.

The fact that cell line groupings based on correlation of microRNA expression patterns largely paralleled the groupings found for transcript expression provides additional evidence for the quality of data from the microarray experiments. It also attests to the hypothesis that microRNA patterns are characteristic of tissue type. To assess that finding more formally for both of the dendrograms in Fig. 2, we set a cutoff to give nine subgroups in each dendrogram. We then determined the minimum number of between-group moves necessary to rearrange the cells into tissue groups. For the microRNA clustering of Fig. 2A, 21 moves were required, whereas only 16 moves were needed for the transcript clustering of Fig. 2B.

Although the leukemias were all more similar to each other than to other cell types, they formed a relatively loose group with mean pairwise correlation of 0.78. K-562 had the lowest mean correlation with the other 59 cell lines (mean  $r = 0.58$ ), much lower than the next lowest, and its



**Figure 2.** Cell line clustering. Complete linkage clustering of cell lines using a correlation metric based on expression levels of 279 microRNAs with log expression level  $\geq 8$  in at least 10% of the cell lines (A) and 1,000 U133A genes with the largest interquartile range (B). Lung cancer cell line NCI-H23 was omitted from (B) because the U133 data are being rerun for this cell line.

**Table 1. MicroRNAs selected using partition around medoids (PAM<sub>c</sub>)**

let-7d	mir-30e-5p	mir-141	mir-301
mir-7	mir-34a	mir-147	mir-320
mir-7-prec	mir-99	mir-181b	mir-321
mir-9	mir-99b-prec	mir-183	mir-326
mir-10a	mir-103	mir-192	mir-335
mir-16	mir-106	mir-196	mir-365
mir-21	mir-124a	mir-196a	mir-371
mir-24	mir-126	mir-219	mir-377
mir-26a	mir-128a	mir-221	mir-382
mir-29c	mir-130a	mir-223	mir-497

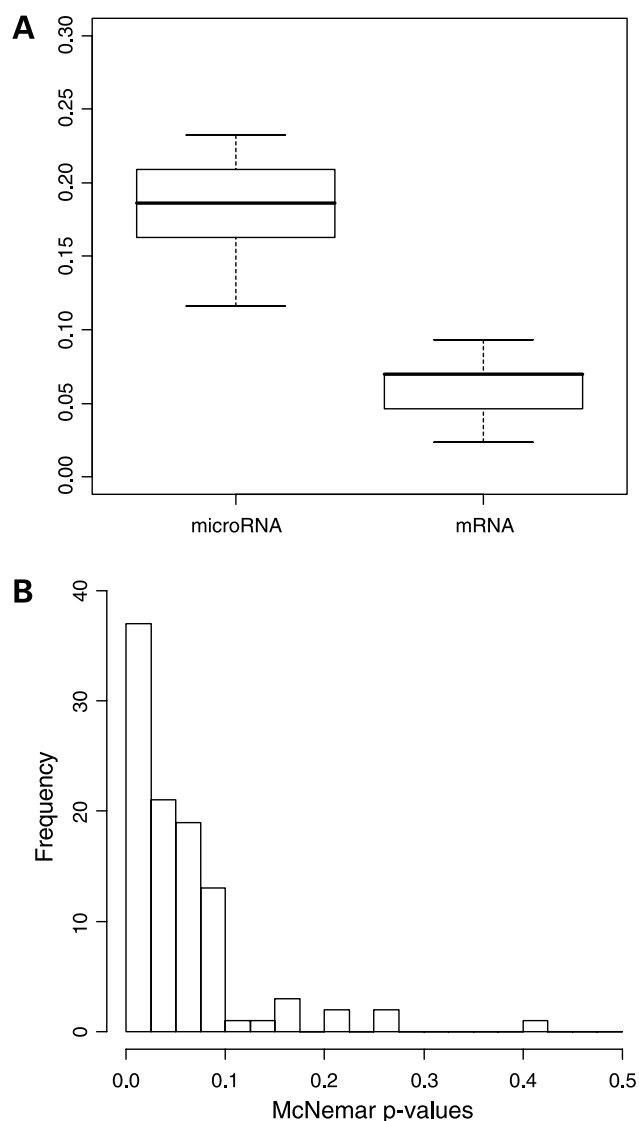
mean correlation with the other leukemias was only 0.66. By chance, K-562 was one of the three controls tested in five replicates, and its microRNA profile showed good reproducibility across replicates. Hence, the outlier status of K-562 in terms of microRNA expression pattern seemed to reflect real differences in biology, not experimental error. K-562 is the only NCI-60 cell type derived from chronic myelogenous leukemia and the only one with the BCR-ABL translocation. We do not know whether either of those characteristics is causally associated with the outlier status of the line.

#### Comparison of MicroRNAs and mRNAs for Prediction of Tissue of Origin

Initially, we used Predictive Analysis for Microarrays (PAM<sub>p</sub>), which performs sample classification based on the nearest shrunken centroid method (42), to investigate the ability of microRNA expression levels to predict tissue of origin. PAM<sub>p</sub> uses Significance Analysis of Microarrays (SAM; ref. 43) to select features (microRNAs in this case) for its classifier. Hereafter, we will refer to the algorithm as SAM-PAM. Because some of the microRNAs selected by SAM might be highly correlated with each other, thereby reducing the independent predictive value of each individual microRNA in multivariate analysis, we sought an alternative way to select more “distinctive” microRNAs. Toward that end, we used Partition Around Medoids [ref. 44; denoted PAM<sub>c</sub> as implemented by the *pam()* function in the *cluster* library in R] to cluster the microRNAs. The PAM<sub>c</sub> algorithm first searches for *k* representative microRNAs (called medoids) among all of the microRNAs to be clustered. After finding the medoids, *k* clusters are constructed by assigning each microRNA to the nearest medoid. One particularly attractive feature of PAM<sub>c</sub>, as implemented in *pam()* is the graphical display, the *silhouette plot*, which can be used to guide selection of the number of medoids (and hence clusters). The silhouette, which takes on values between  $-1$  and  $+1$ , is a measure of how distinct the clusters are; the larger the silhouette, the more distinctive the clusters. Indeed, based on the silhouette plot, we decided to use 40 clusters because the silhouette started to level off at that point, although it continued on a generally upward trajectory with increasing numbers of clusters. Table 1 lists the 40 microRNAs selected by PAM<sub>c</sub>.

The medoids from the clusters comprised the set of microRNA probes that we used in PAM<sub>p</sub> for predictive analysis. The overall approach is termed PAM-PAM. Six-fold cross-validation with 100 replicates was used in both SAM-PAM and PAM-PAM. We used the same analytic scheme to analyze the U133A transcript expression data. However, because the number of mRNAs was large even after preprocessing, we preselected 1,000 genes with the largest interquartile range before running PAM<sub>c</sub>. The medoids of 40 gene clusters were selected for comparative PAM-PAM analysis.

We applied the PAM-PAM analysis to a 43-cell line subset of the NCI-60 comprising the CNS, colon, leukemia, melanoma, ovarian, and renal tissue types.



**Figure 3.** Misclassification rates. Misclassification rates over 100 replications of PAM-PAM: **A**, comparison of rates from the microRNA expression microarray and the U133A mRNA expression array; **B**, histogram of *P* values from the McNemar test.

**Table 2. MicroRNAs differentially expressed between tissue groups**

Panel	Discriminating microRNAs
CNS	mir-21, mir-99, mir-382
Colon	mir-7, mir-10a, mir-29c, mir-103, mir-106, mir-130a, mir-141, mir-183, mir-192, mir-196, mir-196a, mir-335
Leukemia	mir-10a, mir-21, mir-99, mir-106, mir-221, mir-365
Melanoma	mir-141, mir-147, mir-192, mir-335
Ovarian	N/A
Renal	mir-365

NOTE: For each tissue group, we applied SAM to identify microRNAs differentially expressed between the given group and its complement.

The breast and lung cancer cell lines were omitted because their intragroup correlations are low. The prostate cell lines were omitted because there are only two members, and the melanoma cell line LOX IMVI was omitted because it seems to be nonmelanotic and highly undifferentiated (15). The left-hand box plot in Fig. 3A shows the misclassification rates over 100 replications of PAM-PAM based on the microRNA data. The median error rate was 0.18. The right-hand box plot in Fig. 3A shows the analogous results obtained for the 40 U133A probes (i.e., the medoids selected by PAM<sub>c</sub>). The median error rate, 0.07, was smaller. Those results suggest that the mRNA expression data are more informative in discriminating among tissue types, an observation consistent with the relative numbers of moves needed to rearrange the subclusters in Fig. 2 into tissue groups. To assess that observation more formally, we applied the McNemar test (45) in R to the results from each 6-fold cross-validation. The *P* values over the 100 replications are plotted in Fig. 3B, which shows a skewed distribution rather than a U(0, 1), suggesting that, indeed, the U133A data are more informative in tissue type discrimination. However, that observation cannot be attributed unambiguously to the biology; it might, instead, reflect differences in experimental platform and/or methods of analysis between the microRNA and mRNA expression data. As we also found when comparing transcript and protein expression (15), it can be almost impossible to design experiments and analyses that rigorously eliminate all potential sources of technological and selection bias to focus exclusively on the underlying biological bias.

We used SAM itself (43) to identify microRNAs that are differentially expressed among the six tissue groups (CNS, colon, leukemia, melanoma, ovarian, and renal). The analysis was limited to the 40 microRNA medoids previously selected by PAM<sub>c</sub> (Table 1). For each tissue group, we applied SAM for identification of microRNAs differentially expressed between the given group and its complement (comprising the cell lines of the five remaining groups). The  $\delta$  value was chosen to control the false discovery rate at <0.01. Table 2 lists the microRNAs called significant by SAM for each group.

### Correlations with Known mRNA Targets

We next asked whether there is an association between the expression pattern of a microRNA and that of a known target. TarBase (41) provides information on experimentally supported targets for microRNAs. The version downloaded Sept. 20, 2006 listed translationally repressed targets for 47 microRNAs and a total of 72 microRNA-gene pairs. We mapped gene names to HUGO symbols and then to Affymetrix probes on the U133A set. There are typically multiple probes per HUGO symbol, and there are also multiple probes on our microarray for each microRNA (38). All microRNA and mRNA probes were included in the analysis, for a total of 86 microRNA probes and 84 U133A probes.

We calculated the Pearson correlation coefficients for all 7,224 microRNA-mRNA probe pairs. Two-hundred fifteen of the 7,224 correlations corresponded to microRNA-target pairs listed in TarBase, and the remaining 7,009 correlations to non-TarBase pairs. A Wilcoxon rank sum test between the two sets of microRNA-mRNA correlations gave a *P* value of 0.28, suggesting that there is not a significant difference between the correlations of microRNA-target and nontarget pairs. Because microRNAs are predicted to target multiple unrelated genes (46, 47) that are not coexpressed, it is not surprising that microRNA expression levels do not tend to be strongly correlated with the expression of particular target transcripts.

### Correlations with Drug Activity

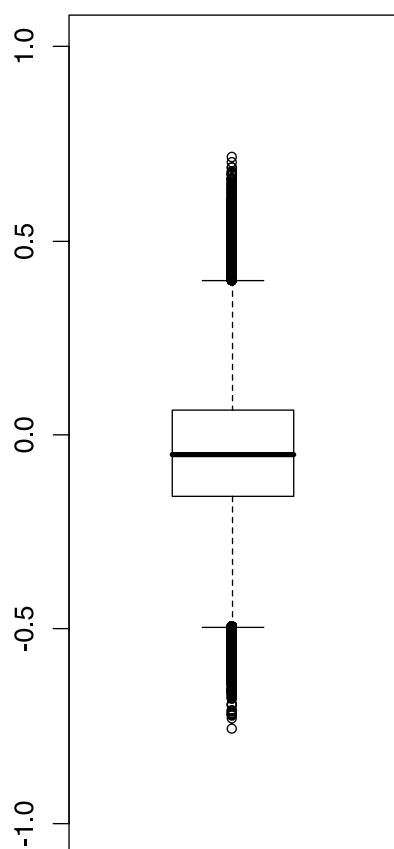
MicroRNAs have already been shown to play critical roles in the cancer biology underlying sensitivity or resistance to specific drugs. For example, Si et al. (48) found that inhibition of mir-21 using antisense oligonucleotides enhanced growth inhibition of MCF7 cells by topotecan (a clinical camptothecin analog) by ~40%. The authors also found that inhibition of mir-21 inhibition resulted in lower levels of the Bcl-2 protein, which is not a direct target but may be partly responsible for increased apoptosis and increased drug sensitivity. In addition, Meng et al. (49) found that mir-21 was highly overexpressed in cholangiocarcinoma (a tumor of the biliary tract), and that inhibition of mir-21 increased sensitivity to gemcitabine in malignant cholangiocytes cell lines.

To assess the potential roles of microRNAs in cancer chemotherapy, we calculated Pearson correlation coefficients across the NCI-60 between expression patterns of the 279 microRNAs and potency patterns of the 3,089 compounds, each selected through the filtering processes described above. In general, the six leukemias were more sensitive to cytotoxic compounds than were the other cell lines. The mean  $-\log(GI_{50})$  value for the leukemia cell lines was 5.9, in contrast to a mean of 5.0 for the nonleukemic cell lines. Because a number of microRNAs were overexpressed or underexpressed in the leukemias (in comparison with the other cell lines), an extreme correlation coefficient might have resulted from the intrinsic sensitivity of the leukemias, not from chemoresistance or chemosensitivity.

Therefore, we calculated compound-microRNA correlations using only the 54 nonleukemia cell lines. There were ~861,000 compound-microRNA correlations, and a qqplot showed that the distribution was nearly normal with only small deviations at very extreme values.

Compound-microRNA correlation coefficients ranged from  $-0.75$  to  $+0.71$ . Statistics from box and whisker plots (Fig. 4) were as follows: For left whisker, 1st quartile, median, 3rd quartile, and right whisker, the values were  $(-0.49, -0.16, -0.04, +0.06, +0.40)$ . Between the left and right whiskers were 99.48% of the correlations, with 1,606 values  $< -0.49$  and 2,896 values  $> 0.40$ . To assess extreme values from the correlation matrix, we compared the 0.005 and 0.995 quantiles (which have values  $-0.447$  and  $0.377$ ) with the corresponding values from a null distribution. The two values were well within the linear region of the qqplot. That assessment showed that the low-end correlation ( $-0.447$ ) was very significant,  $P < 0.001$ , but the high-end correlation ( $0.377$ ) was not significant,  $P = 0.26$ .

Table 3 lists compound-microRNA pairs and 54-cell correlation coefficients for the 20 microRNAs with the



**Figure 4.** Compound-microRNA Correlations. Box plot of Pearson correlation coefficients between expression patterns of the 279 microRNAs and potency patterns of the 3,089 compounds across the 54 nonleukemia cell lines. Values ranged from  $-0.75$  to  $+0.71$ , and critical values were left whisker  $-0.49$ , 1st quartile  $-0.16$ , median  $-0.04$ , 3rd quartile  $0.06$ , and right whisker  $0.40$ .

**Table 3.** Compound-microRNA correlations

NSC	microRNA	<i>r</i>	NSC	microRNA	<i>r</i>
711670	mir-220	$-0.75$	650711	mir-181c	$-0.67$
703783	mir-221	$-0.73$	633274	mir-193a-prec	$-0.67$
703783	mir-222	$-0.72$	713070	mir-196a	$0.67$
721038	mir-140	$-0.71$	713070	mir-326	$0.67$
697472	mir-24	$-0.70$	22842	mir-423	$0.67$
609394	mir-30a-3p	$-0.69$	713070	mir-30b-prec	$0.68$
637992	mir-146a	$-0.69$	22842	mir-494	$0.69$
622597	mir-212-prec	$-0.69$	622926	mir-342	$0.69$
637992	mir-146b	$-0.69$	713070	mir-363	$0.70$
625863	mir-122a	$-0.68$	655765	mir-214	$0.71$

NOTE: The table lists compound-microRNA correlations for the 20 microRNAs with the most extreme values. Pearson correlation coefficients (*r*) were calculated using data for the 54 nonleukemia cell lines, expression values for microRNAs, and  $-\log(GI_{50})$  values for compounds.

most extreme values. A strong correlation between the expression pattern of a microRNA and the growth inhibitory pattern of a drug may indicate a role in drug response. For example, a strong negative correlation may mean that cells expressing higher levels of the microRNA are less sensitive to the compound. Such correlations suggest a potential chemoresistance mechanism and provide a hypothesis that can be tested experimentally. Preliminary studies to assess the potential roles of microRNAs in chemosensitivity and chemoresistance are in progress.

## Conclusion

The collection of pharmacologic and molecular profile data sets on the NCI-60 provides a resource for integrated chemogenomic studies aimed at elucidating molecular targets, identifying biomarkers for personalization of therapy, and understanding mechanisms of chemosensitivity and chemoresistance. However, information on expression patterns of microRNAs had been missing from the NCI-60 repertoire. Given the pervasive regulatory roles of microRNAs in cancer biology, we undertook to measure and analyze the expression of all known microRNAs (as of January 2006) in the NCI-60 cancer cell line panel. The data are available in CellMiner<sup>8</sup> for direct integration with other molecular profiling data sets on the NCI-60. They have also been deposited in ArrayExpress and at the DTP web site. Cell line clustering based on microRNA expression patterns shows that cell groupings are generally consistent with tissue type and with clustering based on mRNA expression. The new microRNA microarray data reported here will make it possible for researchers to use the entire array of integrated NCI-60 molecular databases to study the role of microRNAs in the cellular response to chemotherapy. When combined with gene expression and other biological data in multivariate analyses, microRNAs may provide critical information for an understanding of cancer chemosensitivity and chemoresistance.



## Acknowledgments

We thank the many members of the DTP staff whose efforts make the Molecular Targets Program possible.

## References

- Weinstein JN, Myers TG, O'Connor PM, et al. An information-intensive approach to the molecular pharmacology of cancer. *Science* 1997;275:343–9.
- Scherf U, Ross DT, Waltham M, et al. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000;24:236–44.
- Staunton JE, Slonim DK, Collier HA, et al. Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci U S A* 2001;98:10787–92.
- Blower PE, Yang C, Fligner MA, et al. Pharmacogenomic analysis: correlating molecular substructure classes with microarray gene expression data. *Pharmacogenomics J* 2002;2:259–71.
- Wallqvist A, Rabow AA, Shoemaker RH, Sausville EA, Covell DG. Linking the growth inhibition response from the National Cancer Institute's anticancer screen to gene expression levels and other molecular target data. *Bioinformatics* 2003;19:2212–24.
- Covell DG, Wallqvist A, Huang R, Thanki N, Rabow AA, Lu XJ. Linking tumor cell cytotoxicity to mechanism of drug action: an integrated analysis of gene expression, small-molecule screening and structural databases. *Proteins* 2005;59:403–33.
- Huang Y, Anderle P, Bussey KJ, et al. Membrane transporters and channels: role of the transportome in cancer chemosensitivity and chemoresistance. *Cancer Res* 2004;64:4294–301.
- Huang Y, Blower PE, Yang C, et al. Correlating gene expression with chemical scaffolds of cytotoxic agents: ellipticines as substrates and inhibitors of MDR1. *Pharmacogenomics J* 2005;5:112–25.
- Huang Y, Dai Z, Barbacioru C, Sadee W. Cystine-glutamate transporter SLC7A11 in cancer chemosensitivity and chemoresistance. *Cancer Res* 2005;65:7446–54.
- Dai Z, Barbacioru C, Huang Y, Sadee W. Prediction of anticancer drug potency from expression of genes involved in growth factor signaling. *Pharm Res* 2006;23:336–49.
- Weinstein JN. Integromic analysis of the NCI-60 cancer cell lines. *Breast Dis* 2004;19:11–22.
- Boyd MR, Paull KD. Some practical consideration and applications of the National Cancer Institute In Vitro Anticancer Drug Discovery Screen. *Drug Dev Des* 1995;34:91–109.
- Weinstein JN. "Omic" and hypothesis-driven research in the molecular pharmacology of cancer. *Curr Opin Pharmacol* 2002;2:361–5.
- Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000;24:227–35.
- Shankavaram U, Reinhold WC, Nishizuka S, et al. Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. *Mol Cancer Ther* 2007;6:820–32.
- Myers TG, Anderson NL, Waltham M, et al. A protein expression database for the molecular pharmacology of cancer. *Electrophoresis* 1997;18:647–53.
- Nishizuka S, Charboneau L, Young L, et al. Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc Natl Acad Sci U S A* 2003;100:14229–34.
- Nishizuka S, Chen ST, Gwadrly FG, et al. Diagnostic markers that distinguish colon and ovarian adenocarcinomas: identification by genomic, proteomic, and tissue array profiling. *Cancer Res* 2003;63:5243–50.
- Weinstein JN, Pommier Y. Transcriptomic analysis of the NCI-60 cancer cell lines. *C R Biol* 2003;326:909–20.
- Ikediobi ON, Davies H, Bignell G, et al. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. *Mol Cancer Ther* 2006;5:2606–12.
- Bussey KJ, Chin K, Lababidi S, et al. Integrating data on DNA copy number with gene expression levels and drug sensitivities in the NCI-60 cell line panel. *Mol Cancer Ther* 2006;5:853–67.
- Roschke AV, Lababidi S, Tonon G, et al. Karyotypic "state" as a potential determinant for anticancer drug discovery. *Proc Natl Acad Sci U S A* 2005;102:2964–9.
- Roschke AV, Tonon G, Gehlhaus KS, et al. Karyotypic complexity of the NCI-60 drug-screening panel. *Cancer Res* 2003;63:8634–47.
- Wallqvist A, Huang R, Covell DG, Roschke AV, Gelhaus KS, Kirsch IR. Drugs aimed at targeting characteristic karyotypic phenotypes of cancer cells. *Mol Cancer Ther* 2005;4:1559–68.
- Weinstein JN. Spotlight on molecular profiling: "integromic" analysis of the NCI-60 cancer cell lines. *Mol Cancer Ther* 2006;5:2601–5.
- Lorenzi PL, Reinhold WC, Rudelius M, et al. Asparagine synthetase as a causal, predictive biomarker for L-asparaginase activity in ovarian cancer cells. *Mol Cancer Ther* 2006;5:2613–23.
- Ambros V. The functions of animal microRNAs. *Nature* 2004;431:350–5.
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116:281–97.
- He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 2004;5:522–31.
- Calin GA, Croce CM. MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res* 2006;66:7390–4.
- Calin GA, Croce CM. MicroRNAs and chromosomal abnormalities in cancer cells. *Oncogene* 2006;25:6202–10.
- Jovanovic M, Hengartner MO. miRNAs and apoptosis: RNAs to die for. *Oncogene* 2006;25:6176–87.
- Kent OA, Mendell JT. A small piece in the cancer puzzle: microRNAs as tumor suppressors and oncogenes. *Oncogene* 2006;25:6188–96.
- Engels BM, Hutvagner G. Principles and effects of microRNA-mediated post-transcriptional gene regulation. *Oncogene* 2006;25:6163–9.
- 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.
- Sarkans U, Parkinson H, Lara GG, et al. The ArrayExpress gene expression database: a software engineering and implementation perspective. *Bioinformatics* 2005;21:1495–501.
- Gaur A, Jewell DA, Liang Y, et al. Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. *Cancer Res* 2007;67:2456–68.
- Liu CG, Calin GA, Meloon B, et al. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc Natl Acad Sci U S A* 2004;101:9740–4.
- R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2005. ISBN 3-900051-07-0. Available from: <http://www.R-project.org>.
- Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003;19:185–93.
- Sethupathy P, Corda B, Hatzigeorgiou AG. TarBase: a comprehensive database of experimentally supported animal microRNA targets. *RNA* 2006;12:192–7.
- Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 2002;99:6567–72.
- Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98:5116–21.
- Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. New York: Wiley; 1990.
- Conover WJ. Practical nonparametric statistics. 3rd edition. New York: Wiley; 1999.
- Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005;120:15–20.
- Lim LP, Lau NC, Garrett-Engle P, et al. Microarray analysis shows that some microRNAs down-regulate large numbers of target mRNAs. *Nature* 2005;433:769–73.
- Si ML, Zhu S, Wu H, Lu Z, Wu F, Mo YY. miR-21-mediated tumor growth. *Oncogene* Epub 30 Oct 2006.
- Meng F, Henson R, Lang M, et al. Involvement of human micro-RNA in growth and response to chemotherapy in human cholangiocarcinoma cell lines. *Gastroenterology* 2006;130:2113–29.