# mRNA and microRNA expression profiles of the NCI-60 integrated with drug activities

**Hongfang Liu**[1,4], **Petula D'Andrade**[2], **Stephanie Fulmer-Smentek**[2], **Philip Lorenzi**[1,3], **Kurt W. Kohn**[1], **John N. Weinstein**[1,3], **Yves Pommier**[1,*], and **William C. Reinhold**[1,*]

[1] Laboratory of Molecular Pharmacology, CCR, NCI, NIH, Bethesda, MD 20892

[2] Agilent Technologies, Santa Clara, CA 95052; Frederick, MD

[3] Department of Bioinformatics and Computational Biology, University of Texas M. D. Anderson Cancer Center, Houston, TX 77030

[4] Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC

## Abstract

As part of the Spotlight on Molecular Profiling series, we present here new profiling studies of mRNA and microRNA expression for the 60 cell lines of the NCI DTP drug screen (NCI-60) using the 41,000-probe Agilent Whole Human Genome Oligo Microarray and the 15,000-feature Agilent Human microRNA Microarray V2. The expression levels of ~21,000 genes and 723 human microRNAs were measured. These profiling studies include quadruplicate technical replicates for six and eight cell lines for mRNA and microRNA, respectively, and duplicates for the remaining cell lines. The resulting data sets are freely available and searchable online in our CellMiner database. The result indicates high reproducibility for both platforms and an essential biological similarity across the various cell types. The mRNA and microRNA expression levels were integrated with our previously published 1,429-compound database of anticancer activity obtained from the NCI DTP drug screen. Large blocks of both mRNAs and microRNAs were identified with predominately unidirectional correlations to ~1,300 drugs including 121 drugs with known mechanisms of action. The data sets presented here will facilitate the identification of groups of mRNAs, microRNAs and drugs that potentially affect and interact with one another.

## Keywords

mRNA expression; microRNA expression; NCI-60; drug activity

*Corresponding authors and request of reprints: Yves Pommier, NIH, 9000 Rockville Pike, Bethesda, MD 20892. Phone: 301-496-5944; FAX: 301-402-0752. pommier@mail.nih.gov. William C. Reinhold, NIH, 9000 Rockville Pike, Bethesda, MD 20892. Phone: 301-496-9571; FAX: 301-402-0752. wcr@mail.nih.gov.

## Introduction

MicroRNAs (microRNAs) are small non-coding RNAs, approximately 18–24 nucleotides in length that regulate the levels of their target mRNAs in a highly multiplexed way (1–5). More than 700 human microRNAs have been sequenced according to the latest miRBase release (Release 14). They are abundantly present in all human cells and haven been estimated to target at least ~60% of all genes (2,6,7). It is also well-established that microRNAs regulate cell proliferation and apoptosis, both of which are critical processes in cancer (8,9). microRNAs have been reported to play roles in a number of human malignancies, including leukemias and breast, lung, liver, brain, and colon cancers (1,10–13). An example of this is mir-15a and mir-16-1, which are frequently deleted or down-regulated in chronic lymphocytic leukemia (CLL) (14). More than half of the known microRNA sequences are located in chromosomal regions that are often genetically altered in human cancer, for example at fragile sites or regions of deletion or amplification (11). Over-expression of microRNAs (such as mir-155) in cancers implies their possible function as oncogenes through the negative regulation of tumor suppressor genes and/or genes that inhibit cell differentiation or apoptosis (8,15,16). Conversely, some microRNAs (such as let-7d and mir-127) that are under-expressed in cancers may function as tumor suppressors by inhibiting oncogenes and/or genes that control cell differentiation or apoptosis (15,16).

To study the relationships among expression levels of the various mRNAs and microRNAs, as well as their correlations with drug activity, the NCI-60 panel of human tumor cancer cell lines was used (17,18). The panel, which consists of 60 cell lines from nine tissues of origin, includes melanomas (ME), leukemias (LE), and cancers of breast (BR), kidney (RE), ovary (OV), prostate (PR), lung (LC), central nervous systems (CNS), and colon (CO) origin. Several mRNA expression profiling studies of the NCI-60 have been reported previously (17,19–21), and microRNA profiling studies of the NCI-60 have also been done on 241 human microRNAs using stem-loop real-time PCR (22) and on 321 human microRNAs using microarray (23). Clustering analyses of mRNA and microRNA expression in those studies revealed that some cell types (leukemia, colon, CNS, and melanoma) are grouped in a manner reflecting their tissues of origin. Expression levels of mRNAs and microRNAs were also significantly associated with drug sensitivity or resistance (20,22,23).

To expand and improve our mRNA and microRNA expression data (24), we have profiled the NCI-60 using the Agilent Whole Human Genome Oligo Microarray (mRNA_Agilent) and Agilent Human microRNA Microarray (V2) (microRNA_Agilent), which interrogate ~21,000 genes and 723 microRNAs, respectively. Both profiles include technical replicates. This report describes the generation, quality-control, and integrative analyses of the profile data sets obtained from the two platforms, focusing on the correlations of mRNA and microRNA expression with drug activity (leaving detailed examination of the linkages between mRNAs and microRNAs for a future report). Previous molecular profiles of the NCI-60 by our research group have appeared in the Spotlight on Molecular Profiling series (17,24–29) and elsewhere (20,21,30,31). The profiling data, metadata, and SQL capabilities for querying them are freely available in our CellMiner relational database program[1], where a listing of our Spotlight series publications are kept current.

---

[1]http://discover.nci.nih.gov

## Materials and Methods

### Cell Lines

Frozen stocks of the NCI-60 were obtained from the NCI Developmental Therapeutics program (NCI DTP). The cells were cultured as described previously (20,24), and then thawed, placed in RPMI 1640 (Lonza Walkersville, Inc., Walkersville MD) containing 5% fetal calf serum (Atlantic Biologicals, Norcross, GA) and 2 mM glutamine (Invitrogen Corporation, Carlsbad, CA). For compatibility with our other profiling studies, we used the same batch of serum used by DTP, and the procedures were performed or overseen by the same researcher (WCR).

### mRNA and microRNA Purification and Quality Assessment

For profiling of mRNA, total RNA was extracted using the RNeasy purification kit (Qiagen, Inc., Valencia, CA) according to the manufacturer's instructions. For profiling of microRNA, total RNA was extracted using Trizol (Invitrogen Corporation, Carlsbad, CA) following the manufacturer's recommended procedures. The samples were assayed using the 2100 Expert_Eukaryotic Total RNA Pico assay, and the 2100 Expert_Small RNA Bioanalyzer assay (Agilent Technologies, Santa Clara). Most of the samples showed an RNA Integrity Number (RIN) greater than 9 according to the Total RNA Pico assay. Percentages of microRNA, as determined by the Small RNA assay varied, among samples.

### RNA Labeling, Microarray Hybridization, Profiling, and Quality Control

The mRNA samples were labeled and processed following the Agilent Technologies One-Color Microarray-Based Gene Expression Analysis Protocol v 5.5 with a total RNA input of 200ng. Six cell samples from the NCI-60 panel (BR:MCF7, CO:HCT116, CO:HT29, LE:K562, ME:SK-MEL-2 and RE:CAKl-1) were labeled in quadruplicate, and the remaining samples were labeled in duplicate. All 132 labeled samples were hybridized to the Agilent Whole Human Genome Oligo Microarray (G4112F, design ID 014850, Agilent Technologies, Santa Clara, CA) which contains 41,000 probes measuring around 21,000 unique genes.

For the microRNA profiling, 100 ng of total RNA was labeled as recommended by Agilent Technologies (microRNA Microarray System Protocol v.1.5), using T4 RNA ligase from USB (Bedford Heights, Ohio). Most samples were labeled in duplicate, but eight samples were labeled in quadruplicate (the six mentioned above for mRNA plus LC:A549-ATCC and PR:PC-3). The 136 labeled samples were hybridized to the version 2 Human microRNA microarray (G4470B, design ID 019118, Agilent Technologies) which assesses 723 Human microRNAs and 23 Human viral microRNAs.

The microarrays for profiling both mRNA and microRNA were scanned and data extracted as recommended in the respective protocols. After data extraction with Feature Extraction (version 9.5, Agilent Technologies), the data quality was confirmed using the Agilent QC metrics. All mRNA microarrays passed the QC step, while two microRNA microarrays with borderline quality (one for LE:SR, the other for OV:NCI/ADR-RES) were excluded from further data analysis.

### The Drug Database

A note about nomenclature: we use the term "drug" to indicate chemical compounds tested in the NCI-60 DTP human tumor cell line screen, which was designed to screen up to 3,000 compounds per year for potential anticancer activity[2]. The screen utilizes the NCI-60 to prioritize drugs showing selective growth inhibition or cell killing of particular tumor cell lines.

---

[2]DTP website: http://dtp.nci.nih.gov/branches/btb/ivclsp.html.

To assess potential associations between drug activity and the expression level of either mRNAs or microRNAs, we used the 50% growth inhibitory concentrations (GI50's) determined by DTP. Those data were further curated by our research group available through CellMiner[3]. In our analysis, we used our A1429 data set (32), and also our data set of activity for 121 drugs with putatively known mechanism of actions (33) extracted from the A4463 data set (34). A1429 and A4463 are available through CellMiner. The activity levels are expressed as the negative log of the 50% growth inhibitory concentration (-log10(GI50)), measured using a 48-hour sulphorhodamine B assay.

## Data Analysis

The mRNA and microRNA microarray output text files were loaded into GeneSpring GX and R for analysis. All clustered image maps (CIMs) were obtained using our CIMMiner program package[4] (31).

Note that there is a significant difference between the mRNA and microRNA microarrays: from prior reports the majority of the microRNAs are either not expressed or are expressed at low levels while approximately 60–70% of mRNAs in general are expressed in any individual tissue (35,36). Consequently, we included in our analysis all mRNA probes, but only a subset of the microRNAs. We chose the subset of 365 (of the 723 total) microRNAs that had detectable expression in at least 10% of the cell lines for correlation and clustering analysis. We considered a microRNA to be expressed at a detectable level in a cell line if the corresponding value of the variable "Isdetected" from the Agilent Scanner software Feature Extractor was true in more than half of the corresponding technical replicates where "Isdetected" is a Boolean variable to indicate if the signal is three times higher than the background noise. Data on all mRNA probes and the 365 microRNAs were normalized using GeneSpring GX by i) setting any gProcessedSignal (mRNA_Agilent) or gTotalGeneSignal (microRNA_Agilent) value less than 5 to 5, ii) transforming the gProcessedSignal or gTotalGeneSignal to Logbase 2, and iii) normalizing per array to the 75th percentile of mRNA probes and microRNAs, respectively. After normalization, an intensity value greater than or equal to 0 indicates that the probe was in the top quartile. Average intensities for all 41,000 mRNA probes and 365 microRNAs were obtained for each cell line by averaging the normalized data across replicates.

For the comparison of probe intensities (Figure 1A–D), we used all 41,000 mRNA probes and the 365 microRNAs. For both the cell-cell (Figure 1A and C) and tissue-tissue (Figure 1B and D) comparisons, all possible pair-wise correlations for the available replicates were calculated, and then averaged.

We limited our mRNA clustering (Figure 2A) to probes that showed relatively high and diverse expression across the NCI-60. For each mRNA probe (p), we calculated two values, maximum probe intensity, max(p) and probe inter-quartile range, IQR(p), across the NCI-60. There were 3,032 probes (out of 41,000) that were contained in both the top quartile of max(p) and the top quartile of IQR(p). Those probes were used in our mRNA clustering analysis (Figure 2A). Given the error structures routinely seen in microarray data (in the log frame), restriction of the clustering to only this high-expression and high-diversity data generally yields higher quality clustering. The general expression distribution analysis of mRNA probes was based on the number of cell lines having top quartile expression of the corresponding mRNA probe (Figure 3A).

The microRNA clustering analysis (Figure 2B) was carried out on the 365 microRNAs that had detectable expression in at least 10% of the cell lines. To provide the complete range of

---

[3]CellMiner: http://discover.nci.nih.gov/cellminer.
[4]CIMMiner: http://discover.nci.nih.gov/cimminer.

data for the general expression distribution analysis, Figure 3B was obtained to show the distribution of microRNAs based on the number of cell lines with detectable expression in at least one cell line for the corresponding microRNA (Figure 3B).

To assess relationships between drug activity and expression of mRNA (Figure 4A), we generated CIMs (37) by: i) forming a matrix of the Pearson correlation coefficient for each drug's 60-element vector of activities across the cell lines with each probe's 60-element vector of mRNA expression levels across the cell lines, ii) clustering rows and columns of the resulting matrix, and iii) quantile-color coding of the resulting matrix.

The clustering analysis shown in Figure 4A was performed on the A1429 drug set, and 2,566 probes selected as follows: i) for each probe, obtain the maximum absolute correlation value across 1429 drugs, and ii) select the top 6.25% probes ranked according to their maximum absolute correlation values. Figure 4B shows the corresponding visualizations for the 1429 drugs and the 365 microRNAs.

For each of the 41,000 mRNA probes or 365 microRNAs, we calculated the number of drugs with significant positive correlations at $p < 0.05$ (correlation, R greater than 0.254) and denoted it as Rpos. Similarly, we obtained Rneg, the number of drugs with significant negative correlations with either the mRNA or microRNA.. Then the ratio of the maximum of Rpos and Rneg (Rmax) to the sum of Rpos and Rneg (Rsum) was calculated, that is Rmax/Rsum. This yields the ratio of drugs having significant correlations along the dominant direction, either positive or negative. This ratio has a range between 0.5 and 1. A ratio close to 1 indicates the majority of the significant correlations are in one direction. The two histograms in Supplement Figure 1 show the distribution of the number of mRNA probes and microRNAs with respect to the ratios.

All clustering analyses were average-linkage agglomerative hierarchical. The metric used for clustering based on expression levels or drug sensitivities across the NCI-60 was the Pearson correlation coefficient. For clustering based on correlations (i.e., between drug sensitivities and either mRNA or microRNA expression levels), the metric used was the Euclidean distance.

## Results

### Cell and Tissue-of-Origin Correlations of mRNA and microRNA Expression Levels

Microarray correlation comparisons were made for the designated cell-cell combinations for mRNAs and microRNAs in Figures 1A and 1C, respectively. For the mRNA microarrays, all 41,000 probes were included. For the microRNA microarrays, 365 microRNAs with detectable expression in at least 10% of the cell lines were included. These visualizations provide both assessments of the technical reproducibility and variability among disparate cell lines.

For the mRNA technical replicates, seen as the gray diagonal in Figure 1A (going from top left to bottom right), the mean correlation was 0.992, with a standard deviation, of 0.004 and a range of 0.977–0.997. For the mRNA correlations from different cell lines, the mean correlation decreased to 0.892, with a standard deviation of 0.025 and a range of 0.824–0.987. For the microRNA technical replicates, seen as the gray diagonal in Figure 1C (going from top left to bottom right), the mean correlation was 0.994 with a standard deviation of 0.007, and a range of 0.952–0.999. For the microRNA correlations from different cell lines, the mean decreased to 0.775, with a standard deviation of 0.082 and a range of 0.469–0.965.

The tissue-tissue mRNA and microRNA expression correlations in Figures 1B and 1D, respectively, indicate the levels of variation both within and between tissues of origin. The same probe sets are used as in Figure 1A and C. For mRNA expression comparisons of cells

within a single tissue of origin, the mean correlation was 0.928, with a standard deviation of 0.017, and a range of 0.899–0.946. For mRNA expression for cells from different tissues of origin, the mean correlation dropped to 0.890 with a standard deviation of 0.020 and a range of 0.854–0.923. For microRNA expression comparisons for cells within a single tissue of origin, the mean correlation was 0.864 with a standard deviation of 0.043 and a range of 0.792–0.921. For microRNA expression for cells from different tissues of origin, the mean correlation dropped to 0.768, with a standard deviation of 0.068 and a range of 0.636–0.851.

These correlation results indicate high technical reproducibility of the Agilent mRNA and microRNA platforms. They also demonstrate the generally coherent mRNA and microRNA expressions within groups of cell lines. The most coherent groups being the colon (CO) and CNS and the least coherent being the leukemia (LE) and breast (BR). Greater variation was generally observed for cell lines from different tissues of origin, and greater variability of expression levels among cells for microRNAs than for mRNAs.

### Clustering and Distribution Analyses of mRNA and microRNA Expression in the NCI-60

The expression patterns and clustering of 3,032 mRNA probes selected for both high level and diverse expression across the NCI-60 are shown in Figure 2A. Figure 3A depicts the distribution of the number of cell lines expressing each of 20,146 probes with top-quartile expression levels in at least one cell line. That distribution was bimodal. Probes with generally low expression appear as the predominately blue vertical strips in Figure 2A and on the left side of the histogram in Figure 3A. Probes with generally high expression appear as the predominately red vertical strips in Figure 2A and on the right side of the histogram in Figure 3A. About 10% of the probes had expression levels higher than the 75th percentile in all cell lines. Clustering of the cell lines based on patterns of mRNA expression (vertical axis in Figure 2A) showed, for the most part, clear separation by tissue of origin.

The clustering result for the 365 microRNA set expressed in at least 10% of the cell lines is shown in Figure 2B. Figure 3B shows the histogram of the more inclusive number of microRNA with detectable expression in at least one cell line. The distribution was bimodal. Thirty percent (217 out of 723) of the microRNAs were not detectable in any of the cell lines. Probes expressed in only a small number of cell lines appear as the predominately blue vertical strips in Figure 2B and on the left side in Figure 3B. Probes expressed in most of the cell lines appear as the predominately red vertical strips in Figure 2B and on the right side in Figure 3B. Clustering of the cell lines based on patterns of microRNA expression (vertical axis in Figure 2B) showed, for the most part, clear separation by tissue of origin.

Overall, the distribution of the numbers of cell lines in which either mRNA or microRNA were expressed was bimodal, i.e., most of mRNAs and microRNAs were expressed in either low or high levels in most cell lines (Figure 3A and B, respectively). Noticeably, cell lines from the different tissues of origin tend to cluster in separate groups in terms of both mRNA and microRNA (see annotations on the right of Figure 2A and B, respectively). The entire set of leukemia (6/6 LE) and 9/10 melanoma (ME) cell lines cluster together in the mRNA clustering analysis. For the microRNA, 6/6 LE and 10/10 ME cluster together again, indicating their relatively high degree of coherence with respect to patterns of expression. In contrast, the lung cancer (LC) and CNS cell lines tended to cluster the least both for mRNA and microRNA, indicating their relatively genetic heterogeneity. The ovarian (OV) group of cells is somewhat intermediate with 3 out of 7 cell lines (OGROV1, OVCAR-4 and OVCAR-3) clustering together for both mRNA and microRNA.

### Correlation of Drug Activity with mRNA and microRNA Expression

The clustered image map (CIM) in Figure 4A relates activity patterns for the previously described 1,429-drug set (32) to the expression patterns of 2,564 mRNA probes selected based on high correlation to drug activities in the NCI-60. Supplemental Figure 2 shows the same approach for the set of 121 drugs of known mechanism of action (many of them clinically used). The CIMs in Figure 4B and Supplemental Figure 3 relate the drug activity patterns of the same 1,429-drug and 121-drug sets to the expression patterns of the 365 microRNA set in the NCI-60. A positive correlation (red color in Figures 4A or B, Supplemental 2, and Supplemental 3) indicates that when mRNA or microRNA expression level increases, the activity level of the corresponding drug will increase and therefore the cell are more sensitive. Similarly, a negative correlation (blue color in Figures 4A or B, Supplemental 2, and Supplemental 3) indicates that when mRNA or microRNA expression levels increase, the activity level of the corresponding drug will decrease and therefore the cell are more resistant. The broad patterns of positive and negative correlation between mRNA or microRNA expression and drug activity reflect the existence of coherent blocks within which mRNA or microRNA levels are consistently correlated either positively or negatively to drug activities.

The histograms of the distribution of the ratios (see the Data Analysis section) for all 41,000 mRNA probes and the 365 microRNAs, in Supplement Figure 1A and B, respectively, support the visual observation of a predominant uni-directionality of those correlations found in Figures 4. The mean ratio value for mRNA (in Supplemental Figure 1A) was 0.83 indicating that 83% of the drugs have significant correlation with a mRNA, on average, in a unidirectional fashion. The mean ratio value for microRNA (in Supplemental Figure 1B) was 0.86. Note that the selection of mRNA probes on the basis of high correlation in Figure 4A is biasing toward the generation of patterns that might not be seen, or be as strong, if the probes were not selected in that way. However, the consistency and robustness of the dichotomous patterns seen were also supported by the histogram shown in Supplemental Figure 1A, which was based on all mRNA probes.

The correlation analysis of mRNA and microRNA expression with drug activity shows that most drugs have a coherent relationship to substantial blocks of both transcript and microRNA expression levels, indicated by the coherent positive or negative mRNA and microRNA correlations across drugs.

Figure 5 compares the clustering results for the set of 121 drugs with defined mechanism of action based purely on drug activities (A), based on the cross-correlations of drug activities and mRNA expression levels (B), and on the cross-correlations of the drug activities with microRNA expression levels (C). In all three cases, drugs were generally clustered according to their mechanism of action classes.

## Discussion

The NCI-60 cell lines have been profiled more comprehensively at the DNA, RNA, protein, and pharmacological levels than any other set of cells in existence. The resulting molecular databases therefore constitute a uniquely valuable information resource for understanding cancer biology, assessing molecular pharmacology, and developing new approaches to analyze and interpret high-throughput molecular data. The profiling studies reported here for mRNA and microRNA expression using the recent Agilent microarrays add two important new platforms of high quality data for mRNA and microRNA expression.[5]

---

[5]Those data are electronically available at: http://discover.nci.nih.gov

The present data analyses demonstrate a high degree of reproducibility for both the mRNA and microRNA microarrays, as depicted by the close to 1 Pearson correlation coefficient in diagonal cells in the Figure 1A and 1C CIMs. As indicated in Figures 1A and 1C, and in tabular form in Figures 1B and 1D, the within-tissue-of-origin correlations tends to be higher than those between tissues of different origin for both the mRNAs and the microRNAs. This is true for all of the mRNA comparisons, and all but two of the microRNA comparisons (the exceptions being the LC-PR, and OV-PR, as compared with LC-LC, and OV-OV, respectively). Those results indicate that the tissue-of-origin signature remains for both the mRNA and microRNA transcriptomes, despite adaptation of the cell lines to *in vitro* culture. That observation is consistent with prior results (20, 22, 23). An apparent example of disease sub-typing within a tissue of origin is provided by the high correlation (r = 0.955) for mRNA between CCRF-CEM and MOLT-4 (Figure 1A), both of which are acute lymphocytic leukemias. LOXIMVI, which is unlike the other melanoma lines by both mRNA and microRNA expression, has previously been noted to be highly differentiated and amelanotic, i.e., it lacks melanin and other melanoma signature markers (38).

CIMs of the NCI-60 based on mRNA and microRNA expression levels (Figure 2) also indicate residual tissue-of-origin signatures that are generally consistent with those seen in prior profiling studies (20,22,23). Overall, there were four tissue types (i.e., leukemia, colon, renal, and melanoma) in which cell lines tend to be well clustered by both mRNA and microRNA according to their tissues of origin. The leukemias were the most clearly separated from other tissue-of-origin lines (as indicated by the height of the cluster branches in both figures). Otherwise, notable exceptions were HCC-2998 (CO) and LOXIMVI (ME) for the mRNA-based clustering, and HCT116 (CO), SN12C (RE), ACHN (RE), and LOXIMVI (ME) for the microRNA-based clustering (Figure 2B). Those cell lines have also tended to group with different tissue-of-origin lines in prior profiling studies (20,22,23).

The other five tissue types were more variable. The CNS cell lines clustered well for mRNAs but poorly for microRNAs. Breast, lung, ovarian, and prostate lines formed small and/or widely distributed groups. Some cell pairs known to be similar, or to have been reported to cluster together in previous profiling studies (20,22,23), also clustered together here. Examples include two hormone-dependent estrogen receptor-positive breast cancer cell lines, MCF7 and T47D, OVCAR8 and its drug resistant derivative NCI/ADR-RES, and the CNS cancer cell lines U251 and SNB-19, which probably originated from the same patient (39). We identify two additional robust pairs, LC:HOP92/BR:MDA-MB-231, and OV:OVCAR3/OV:OVCAR4, that have clustered together in both the current and prior mRNA and microRNA profiling studies (20, 22,23).

The bimodal distribution seen in Figure 3A indicates that there are relatively large groups of mRNAs that are either expressed at high levels in a small number of cancer lines (left side of Figure 3A and also the predominant blue strips in Figure 2A), or expressed at high levels ubiquitously across all cell lines (right side of Figure 3A and also the red strips in Figure 2A). Similarly, the bimodal distribution seen in Figure 3B indicates that there are relatively large groups of microRNAs that either have detectable expression in a small number of cell lines or have detectable expression across all cell lines. The bimodal distributions may reflect some mRNA and microRNAs have critical housekeeping functionalities in cells while some have tissue- or cell line-specific functionalities.

CIM visualizations for the correlations of drug activities with expression levels of mRNA in Figure 4A and Supplemental Figure 2 as well as the histogram in Supplemental Figure 1A indicate the presence of large unidirectional blocks (i.e. with either positive or negative correlation). The 1,429 drugs formed two well-separated clusters: Cluster I, which consists of 146 drugs, and Cluster II, which consists of the remaining 1,283 drugs. There are also two

well-organized mRNA clusters. The 871 mRNAs in Cluster A are predominately negatively correlated with the drugs from Cluster I, and positively correlated with the drugs from Cluster II. The 1,251 mRNAs in Cluster B are predominately positively correlated with the drugs in Cluster I, and negatively correlated with the drugs in Cluster II. Functional analysis using KEGG (40) indicates that mRNAs in Cluster A are enriched significantly with ribosome proteins, and mRNAs in Cluster B are enriched significantly with focal adhesion pathway proteins. Based on UniProtKB, the remaining mRNAs (i.e., those in Cluster C) are enriched with glycoproteins. The histogram in Supplemental Figure 1A reinforces the uni-directional preponderance of the correlations, as depicted by the increase in the number of mRNAs on the right side of the histogram.

CIMs for the correlation of drug activity with the expression level of microRNA in Figure 4B and Supplemental Figure 3 also indicate the presence of large unidirectional blocks. As in Figure 4A, the 1,429 drugs formed two well-separated clusters in Figure 4A, with 115 drugs in Cluster I, and the remaining 1,314 drugs in Cluster II. There were 93 drugs that are common to Cluster I in both mRNA and microRNA drug correlations (Figure 4). The chemical structure and corresponding NSC numbers of some of those drugs are listed in Supplemental Figure 4. Well-organized blocks were formed between the drugs in Cluster II and the microRNAs in clusters A and B. The histogram in Supplemental Figure 1B reinforces the unidirectional preponderance of the correlations, as depicted by the increase in the number of microRNAs on the right side of the graph.

The prior chemosensitivity study conducted by Blower et. al. (41) provides information on three microRNAs, let-7i, mir-16, and mir-21 and their relationship to activity patterns of eight out of the 1,429 drugs [Table 1 in (41)]. We found that the previously described change of drug potency following either silencing or forced expression of microRNAs in the A549 cell line (41) is consistent with our correlation results (Table 1). That is, if the chemosensitivity study demonstrated that a change in microRNA expression leads to a change in drug activity, then that same association (in the same direction) was found in our correlation analysis. For example, the correlation (in Table 1) between mir-21 expression and the $-\log10(GI50)$ value of drug NSC622700 is $-0.445$, which indicates that when the expression of mir-21 decreases, the drug activity increases, consistent with prior results (41).

Clustering of 121 drugs with known mechanisms of action based on: i) the correlation between drug activity and mRNA expression (Figures 5B and Supplemental 2) or (ii) the correlation between drug activity and microRNA expression (Figures 5C and Supplemental 3) was largely successful in classifying the drugs by their mechanisms of action. The clustering between drug activity and mRNA was consistent with prior results (20). We find that the Top1 inhibitors and antimitotic agents generally form coherent clusters. Conversely, Top2 inhibitors form several subgroups. One consistent (Top2 inhibitor) subgroup consists of bisantrene, an anthrapyrazole derivative and deoxydoxorubicin, and another consistent subgroup contains m-AMSA, mitoxantrone, and oxanthrazole. The majority of the N-7 alkylating agents (A7) cluster into one large group, but also appeared in several smaller groups. For DNA antimetabolites (DA), all of the purine analogues (thioguanines and thiopurines) appeared together on small branches in all clusters. There are several drug pairs consistently clustered together, including cytarabine (ara-C) and its congener cyclocytidine, mitomycin and its N-methyl derivative porfiromycin, and the two O-6 position alkylating agents CCNU and BCNU.

In summary, we have characterized the NCI-60 mRNA and microRNA data using two novel Agilent platforms, the Agilent Whole Human Genome Oligo Microarray, and the Agilent microRNA microarray Human version 2. Our analysis indicates high reproducibility for both platforms and an essential biological similarity across various cell types at molecular level. Analyses of drug activity reveal substantial blocks of both mRNA and microRNAs whose

expression levels are correlated with anticancer drug response. Future work will focus on systematic analysis of the three-way relationships of mRNA expression, microRNA expression, and drug activity by taking advantage of the high-quality profiles reported here or previously. We will also extend the approach by including new drugs recently approved by the FDA to elucidate their mRNA and microRNA profiles across the NCI-60 and compare them with the "classical" drugs in the present database. Such analyses are now facilitated by free web access to our data and by the possibility of searching directly the CIM shown in Figures 2 and 4.[6]

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## Abbreviations

| DTP | National Cancer Institute Developmental Therapeutics Program |
|---|---|
| NCI-60 | National Cancer Institute of 60 cancer cell lines |
| BR | breast |
| CO | colon |
| CNS | central nervous system |
| LC | non-small cell lung cancer |
| LE | leukemia |
| ME | melanoma |
| OV | ovarian |
| PR | prostate |
| RE | renal |
| CIM | clustered image map |

## References

1. Jay C, Nemunaitis J, Chen P, Fulgham P, Tong AW. mi RNA Profiling for Diagnosis and Prognosis of Human Cancer. DNA and Cell Biology 2007;26:293–300. [PubMed: 17504025]

2. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. Cell 2009;136:642–55. [PubMed: 19239886]

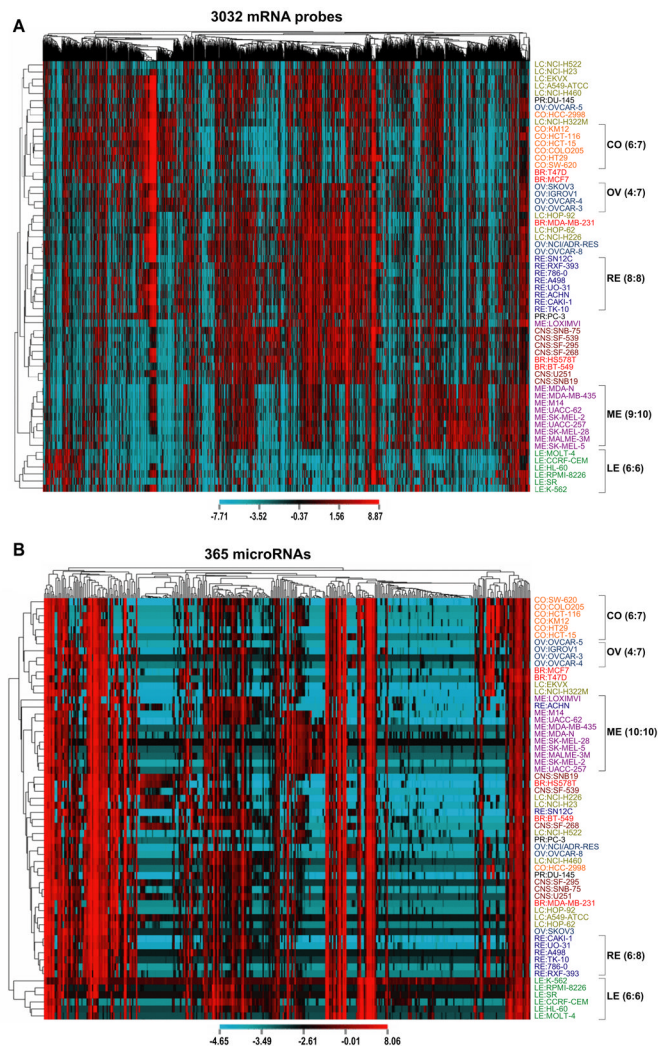3. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004;116:281–97. [PubMed: 14744438]

---

[6]http://discover.nci.nih.gov/cellminerdata/NCI60_MCT/index.html

4. Aravin A, Tuschl T. Identification and characterization of small RNAs involved in RNA silencing. FEBS Lett 2005;579:5830–40. [PubMed: 16153643]

5. Barad O, Meiri E, Avniel A, et al. MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. Genome Res 2004;14:2486–94. [PubMed: 15574827]

6. Friedman JM, Jones PA. MicroRNAs: critical mediators of differentiation, development and disease. Swiss Med Wkly 2009;139:466–72. [PubMed: 19705306]

7. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. PLoS Biol 2005;3:e85. [PubMed: 15723116]

8. Zhang B, Pan X, Cobb GP, Anderson TA. microRNAs as oncogenes and tumor suppressors. Developmental Biology 2007;302:1–12. [PubMed: 16989803]

9. Hwang HW, Mendell JT. MicroRNAs in cell proliferation, cell death, and tumorigenesis. British journal of cancer 2006;94:776–80. [PubMed: 16495913]

10. Akao Y, Nakagawa Y, Naoe T. let-7 microRNA functions as a potential growth suppressor in human colon cancer cells. Biological & pharmaceutical bulletin 2006;29:903–6. [PubMed: 16651716]

11. Calin GA, Croce CM. MicroRNA signatures in human cancers. NATURE REVIEWS CANCER 2006;6:857.

12. Calin GA, Croce CM. MicroRNA-cancer connection: the beginning of a new tale. Cancer Research 2006;66:7390. [PubMed: 16885332]

13. Croce CM, Calin GA. miRNAs, cancer, and stem cell division. Cell 2005;122:6–7. [PubMed: 16009126]

14. Calin GA, Ferracin M, Cimmino A, et al. A MicroRNA Signature Associated with Prognosis and Progression in Chronic Lymphocytic Leukemia. 2005:1793–801.

15. Lowery AJ, Miller N, McNeill RE, Kerin MJ. MicroRNAs as Prognostic Indicators and Therapeutic Targets: Potential Effect on Breast Cancer Management. Clinical Cancer Research 2008;14:360. [PubMed: 18223209]

16. Saumet A, Vetter G, Cougot N, et al. microRNA-Associated Therapies. Current Perspectives in MicroRNAs (miRNA) 2008:395.

17. Weinstein JN. Spotlight on molecular profiling: "Integromic" analysis of the NCI-60 cancer cell lines. Molecular Cancer Therapeutics 2006;5:2601. [PubMed: 17088435]

18. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer 2006;6:813–23. [PubMed: 16990858]

19. 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. [PubMed: 16648555]

20. Scherf U, Ross DT, Waltham M, et al. A gene expression database for the molecular pharmacology of cancer. Nat Genet 2000;24:236–44. [PubMed: 10700175]

21. Weinstein JN. Integromic Analysis of the NCI-60 Cancer Cell Lines. Breast Disease 2004;19:11–22. [PubMed: 15687693]

22. Gaur A, Jewell DA, Liang Y, et al. Characterization of MicroRNA Expression Levels and Their Biological Correlates in Human Cancer Cell Lines. Cancer Research 2007;67:2456. [PubMed: 17363563]

23. Blower PE, Verducci JS, Lin S, et al. MicroRNA expression profiles for the NCI-60 cancer cell panel. Molecular Cancer Therapeutics 2007;6:1483. [PubMed: 17483436]

24. Shankavaram UT, 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. [PubMed: 17339364]

25. 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. [PubMed: 17088437]

26. 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. [PubMed: 17088436]

27. Okabe M, Szakacs G, Reimers MA, et al. Profiling SLCO and SLC22 genes in the NCI-60 cancer cell lines to identify drug uptake transporters. Mol Cancer Ther 2008;7:3081–91. [PubMed: 18790787]

28. Orina JN, Calcagno AM, Wu CP, et al. Evaluation of current methods used to analyze the expression profiles of ATP-binding cassette transporters yields an improved drug-discovery database. Mol Cancer Ther 2009;8:2057–66. [PubMed: 19584229]

29. Stevens EV, Nishizuka S, Antony S, et al. Predicting cisplatin and trabectedin drug sensitivity in ovarian and colon cancers. Mol Cancer Ther 2008;7:10–8. [PubMed: 18187810]

30. Weinstein JN, Kohn KW, Grever MR, et al. Neural computing in cancer drug development: predicting mechanism of action. Science 1992;258:447–51. [PubMed: 1411538]

31. Weinstein JN, Myers TG, O'Connor PM, et al. An information-intensive approach to the molecular pharmacology of cancer. Science 1997;275:343–9. [PubMed: 8994024]

32. Annereau JP, Szakacs G, Tucker CJ, et al. Analysis of ATP-Binding Cassette Transporter Expression in Drug-Selected Cell Lines by a Microarray Dedicated to Multidrug Resistance &boxs. Molecular pharmacology 2004;66:1397–405. [PubMed: 15342794]

33. DTP. DTP Human Tumor Cell Line Screen. [cited 2009 May 21, 2009]; Standard mechanism ]. Available from: http://dtp.nci.nih.gov/docs/cancer/searches/standard_mechanism_list.html

34. Blower PE, Yang C, Fligner MA, et al. Pharmacogenomic analysis: correlating molecular substructure classes with microarray gene expression data. The Pharmacogenomics Journal 2002;2:259–71. [PubMed: 12196914]

35. Ramsköld D, Wang ET, Burge CB, Sandberg R, Jensen LJ, Embl G. An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data. PLoS Comput Biol 2009;5:e1000598. [PubMed: 20011106]

36. Pradervand S, Weber J, Thomas J, et al. Impact of normalization on miRNA microarray expression profiling. Rna 2009;15:493. [PubMed: 19176604]

37. Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. nature genetics 2000;24:227–35. [PubMed: 10700174]

38. Stinson SF, Alley MC, Kopp WC, et al. Morphological and immunocytochemical characteristics of human tumor cell lines for use in a disease-oriented anticancer drug screen. Anticancer research 12:1035. [PubMed: 1503399]

39. Ikediobi ON, Davies H, Bignell G, et al. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. Molecular Cancer Therapeutics 2006;5:2606. [PubMed: 17088437]

40. Okuda S, Yamada T, Hamajima M, et al. KEGG Atlas mapping for global analysis of metabolic pathways. Nucleic Acids Res 2008;36:W423–6. [PubMed: 18477636]

41. Blower PE, Chung JH, Verducci JS, et al. MicroRNAs modulate the chemosensitivity of tumor cells. Molecular Cancer Therapeutics 2008;7:1. [PubMed: 18187804]

**Figure 1.**
Pearson correlation coefficients of expression levels of 41,000 mRNA probes and 365 microRNAs, organized by tissue of origin (not clustered) in the NCI-60. **A.** Heat map of correlation coefficients of replicates (diagonal) and cell-cell correlation coefficients for mRNA. **B.** Tissue-of-origin mean correlation coefficients for mRNA. **C.** Heat map of correlation coefficients of replicates (diagonal) and cell-cell correlation coefficients for microRNA. **D.** Tissue-of-origin mean correlation coefficients for microRNA. For the mRNA comparisons, all 41,000 probes were used; for the microRNA comparisons, the subset of 365 microRNAs expressed in at least 10% of the cell lines was used. In each case, all possible array cross-comparisons were done for the designated cell line or tissue-of-origin pairs. Averages were taken when there were multiple values.
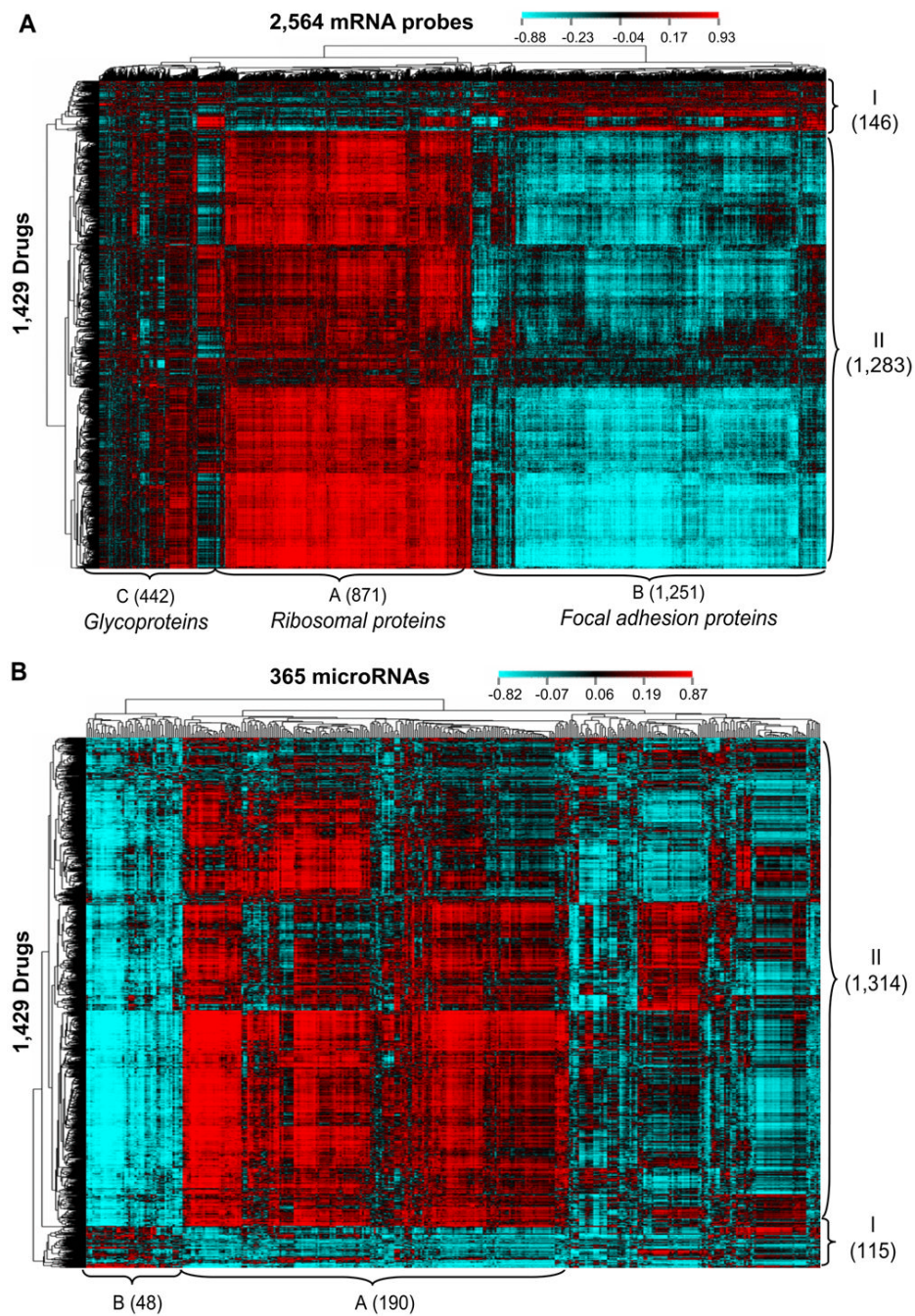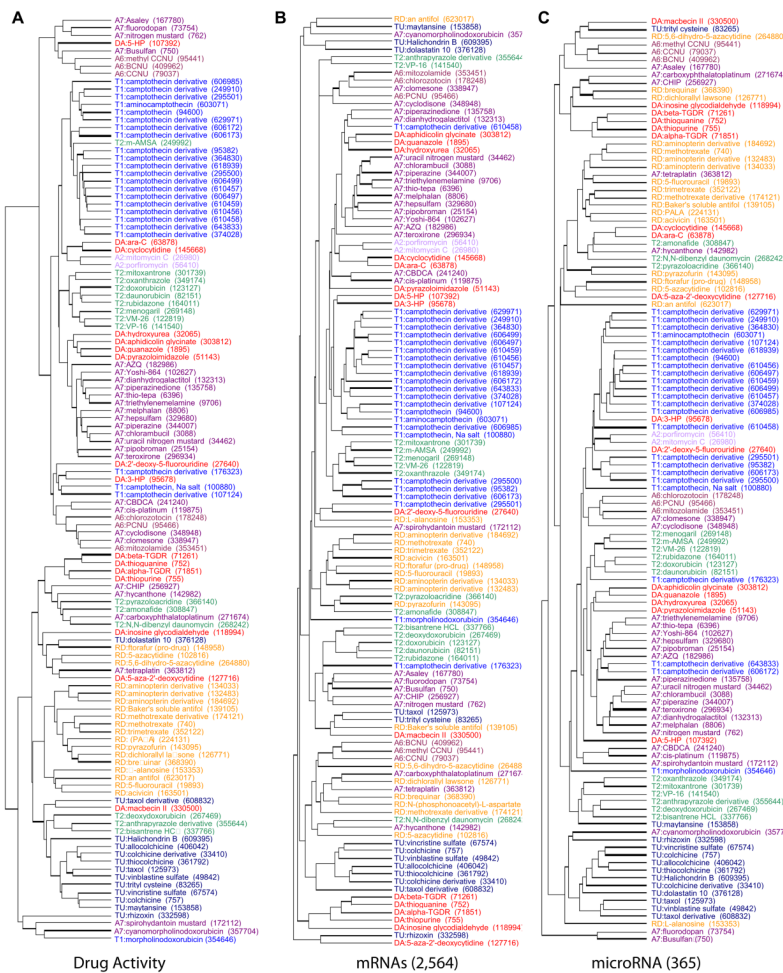
**Figure 2.**
Clustered image maps (CIMs) of mRNA and microRNA expression levels in the NCI-60. **A.** ClM of mRNA expression levels (X-axis) for 3,032 mRNA probes with both high and varied expression versus the NCI-60 cell lines (Y-axis). Selection was done as follows: for each probe (p), we first calculated max(p) and IQR(p) where max(p) is the maximum expression level of p across NCI-60 and IQR(p) is the interquartile range of expression level of P across NCI-60. The probes with both top-quartile max(p) value and the top-quartile IQR(p) values were selected to generate the CIM. Red indicates high expression and blue indicates low expression. **B.** CIM of 365 microRNA probe expression levels (X-axis) versus the NCI-60 (Y-axis). The microRNAs were selected according to the legend of Figure 1. Red indicates high expression and blue indicates low expression.
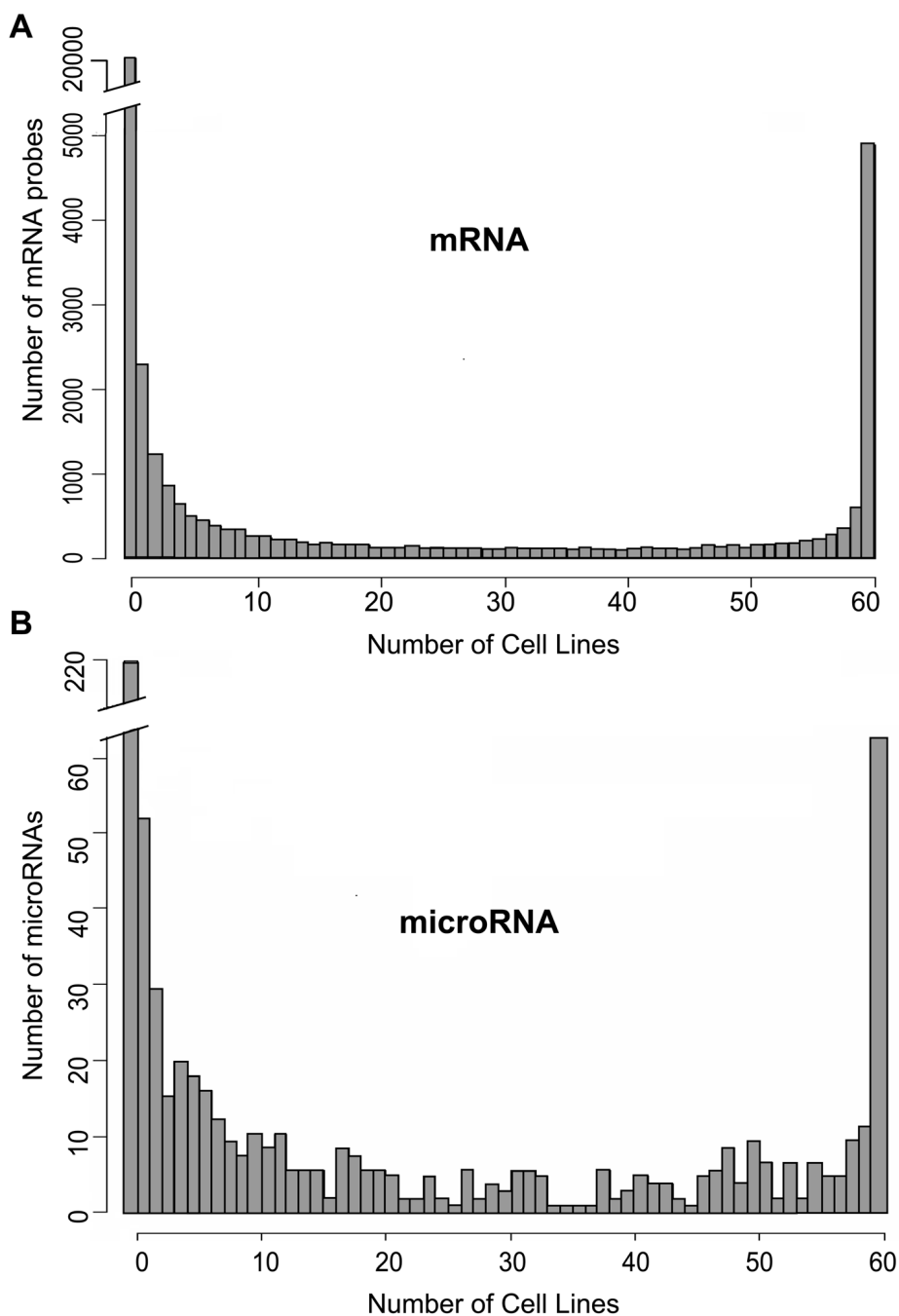
**Figure 3.**
Histograms of mRNA and microRNA expression levels in the NCI-60. **A.** Histogram of the number of cell lines expressed with mRNA expression levels greater than or equal to 0 (i.e., with the top quartile expression level). **B.** Histogram of the number of cell lines (X-axis) in which a microRNA has detectable expression.

**Figure 4.**
Clustered image maps (CIMs) of mRNA and microRNA expression levels versus drug activity in the NCI-60. **A.** CIM of the correlations of 2,564 mRNA expression levels (X-axis) versus 1,429 drug activities (Y-axis) for the NCI-60. Red color indicates high positive correlation and blue color indicates high negative correlation. The 2,566 mRNA probes were selected as having the 6.25% largest absolute correlations to the activities of the 1,429 drugs. **B**. CIM of the correlations of 365 microRNA expression levels (X-axis) versus 1,429 drug activities (Y-axis) in the NCI-60. Red color indicates high positive correlation and blue color indicates high negative correlation.

**Figure 5.**
Dendograms of the set of 121 drugs with putatively known mechanisms of action. **A.** Clustering on the basis of the drug activity data alone. **B.** Clustering on the basis of correlations with the expression levels of 2,564 mRNAs from Figure 4A. **C.** Clustering on the basis of correlations with the expression levels of 365 microRNAs from Figure 4B. Color coding is by drug mechanism of action. Included are A2, A6, and A7 alkylating agents (alkylating at the N-2 position of guanine, alkylating at the O-6 position of guanine, and alkylating at the N-7 position of guanine, respectively), tubulin-active antimitotic agents (Tu), topoisomerase 1 inhibitors (T1), topoisomerase2 inhibitors (T2), DNA antimetobolites (DA), RNA/DNA antimetobolites (RD).

**Table 1**

Pearson correlations of microRNA expression of let-7i, mir-16 and mir-21 and −log10(GI50) values of eight out of eleven drugs used in Blower's drug potency paper (41).

| NSC | let-7i | mir-16 | mir-21 |
|---|---|---|---|
| 63878 | 0.091 | −0.135 | −0.052 |
| 107124 | 0.084 | −0.136 | 0.056 |
| 125973 | 0.061 | 0.144 | −0.240 |
| 265450 | −0.116 | 0.064 | −0.158 |
| 359449 | −0.117 | 0.202 | −0.244 |
| 621888 | −0.274 | 0.244 | −0.408 |
| 622700 | −0.331 | 0.361 | −0.445 |
| 665076 | −0.340 | 0.244 | −0.585 |