An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues

Chang-Gong Liu*†, George Adrian Calin*†, Brian Meloon‡, Nir Gamliel‡, Cinzia Sevignani*, Manuela Ferracin§, Calin Dan Dumitru*, Masayoshi Shimizu*, Simona Zupo¶, Mariella Dono‖, Hansjuerg Alder*, Florencia Bullrich*, Massimo Negrini*§, and Carlo M. Croce*,

*Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA 19107; ‡Compugen USA Inc., 7 Center Drive, Suite 9, Jamesburg, NJ 08831; §Department of Experimental and Diagnostic Medicine and Interdepartment Center for Cancer Research, University of Ferrara, Ferrara 44700, Italy; ¶Diagnostics of Lymphoproliferative Diseases, National Institute of Cancer, Genoa 16123, Italy; and ‖Laboratorio di Analisi Cliniche, Ospedale Civile di la Spezia, La Spezia 19126, Italy

Contributed by Carlo M. Croce, May 18, 2004

MicroRNAs (miRNAs) are a class of small noncoding RNA genes recently found to be abnormally expressed in several types of cancer. Here, we describe a recently developed methodology for miRNA gene expression profiling based on the development of a microchip containing oligonucleotides corresponding to 245 miRNAs from human and mouse genomes. We used these microarrays to obtain highly reproducible results that revealed tissue-specific miRNA expression signatures, data that were confirmed by assessment of expression by Northern blots, real-time RT-PCR, and literature search. The microchip oligolibrary can be expanded to include an increasing number of miRNAs discovered in various species and is useful for the analysis of normal and disease states.

A recently discovered class of small noncoding RNAs, named microRNAs (miRNAs), has been identified in plants and animals (1, 2). The 19- to 22-nt active product is processed from a 60- to 110-nt pre-miRNA hairpin transcript thought to derive from a longer pre-miRNA product (3). It is believed that miRNAs act to regulate gene expression during development and differentiation, at the transcriptional and/or translational level, although targets are still elusive (1, 4). We have previously shown frequent deletions and down-regulation of *miR-15a* and *miR-16-1* miRNAs genes at 13q14 in B-cell chronic lymphocytic leukemia, the most common adult leukemia in the Western world (5). We also reported that human miRNA genes are frequently located at fragile sites and genomic regions involved in cancers (6), suggesting that the role of miRNA in human cancer may involve more than a few genes. In fact, two recent papers reported reduced accumulation of *miR-145* and *miR-143* in colorectal neoplasia (7) and high expression of precursor miRNA-155/BIC RNA in children with Burkitt's lymphoma (8). Assessing cancer-specific expression levels for hundreds of miRNA genes is time consuming, and requires a large amount of total RNA (at least 20 μg for each Northern blot) and autoradiographic techniques that require radioactive isotopes. To overcome these limitations and investigate alterations in expression of all known miRNAs in human cancer, we developed a miRNA microarray and established detection methodology for miRNA expression that overcomes the size limitation imposed by their length (18–22 nt).

Methods

miRNA Oligo Probe Design. A total of 281 miRNA precursor sequences (190 *Homo sapiens*, 88 *Mus musculus*, and 3 *Arabidopsis thaliana*) with annotated active sites were selected for oligonucleotide design. These correspond to human and mouse miRNAs found in the miRNA Registry (www.sanger.ac.uk/Software/Rfam/mirna; accessed June 2003) or collected from published papers (9–11). All of the sequences were confirmed by BLAST alignment with the corresponding genome at www.ncbi.nlm.nih.gov and the hairpin structures were analyzed at www.bioinfo.rpi.edu/applications/mfold/old/rna. When two precursors with different length or slightly different base composition for the same miRNAs were found, both sequences were included in the database, and the one that satisfied the highest number of design criteria was used. The sequences were clustered by organism using the LEADS platform (12), resulting in 248 clusters (84 mouse, 161 human, and 3 *Arabidopsis*). For each cluster, we evaluated all 40-mer oligos for their cross-homology to all genes of the relevant organism, the number of bases in alignment to a repetitive element, the amount of low-complexity sequence, the maximum homopolymeric stretch, global and local G+C content, and potential hairpins (self 5-mers). We then selected the best oligo that contained each active site of each miRNA. This produced a total of 259 oligos; there were 11 clusters with multiple annotated active sites. Next, we attempted to design an oligo that did not contain the active site for each cluster, when it was possible to choose such an oligo that did not overlap the selected oligo(s) by >10 nt. To design each of these additional oligos, we required <75% global cross-homology and <20 bases in any 100% alignment to the relevant organism, <16 bases in alignments to repetitive elements, <16 bases of low complexity, homopolymeric stretches of no more than six bases, G+C content between 30% and 70%, no more than 11 windows of size 10 with G+C content outside 30–70%, and no self 5-mers. A total of 76 additional oligos were designed. In addition, we designed oligos for seven mouse and eight human tRNAs by using similar design criteria. We selected a single oligo for each, with the exception of the human and mouse tRNAs for initiator methionine (Met-tRNA-i), for which we selected two oligos each (Table 1, which is published as supporting information on the PNAS web site).

miRNA Microarray Fabrication. Forty-mer 5′ amine modified C6 oligos were resuspended in 50 mM phosphate buffer, pH 8.0, at 20 μM concentration. The individual oligo-probe was printed in triplicate on Amersham Bioscience CodeLink-activated slides under 45% humidity by GeneMachine OmniGrid 100 Microarrayer in 2 × 2 pin configuration and 20 × 20 spot configuration of each subarray. The spot diameter was 100 μm, and distance from center to center was 200 μm. The printed miRNA microarrays were further chemically covalently coupled under 70% humidity overnight. The miRNA microarrays were ready for sample hybridization after additional blocking and washing steps.

Target Preparation. Five micrograms of total RNA were separately added to reaction mix in a final volume of 12 μl, containing 1 μg of [3′-(N)8-(A)12-biotin-(A)12-biotin-5′] oligonucleotide primer. The mixture was incubated for 10 min at 70°C and chilled on ice. With the mixture remaining on ice, 4 μl of 5× first-strand

Abbreviation: miRNA, microRNA.

Data deposition: The sequences reported in this paper have been deposited in the ArrayExpress database (accession nos. SAMPLE169150SUB621–SAMPLE169193SIUB621).

†C.-G.L. and G.A.C. contributed equally to this work.

**To whom correspondence should be addressed. E-mail: c_croce@mail.jci.tju.edu

buffer, 2 μl of 0.1 M DTT, 1 μl of 10 mM dNTP mix, and 1 μl of SuperScript II RNaseH$^-$ reverse transcriptase (200 units/μl) was added to a final volume of 20 μl, and the mixture was incubated for 90 min in a 37°C water bath. After incubation for first-strand cDNA synthesis, 3.5 μl of 0.5 M NaOH/50 mM EDTA was added into 20 μl of first-strand reaction mix and incubated at 65°C for 15 min to denature the RNA/DNA hybrids and degrade RNA templates. Then, 5 μl of 1 M Tris·HCl (pH 7.6, Sigma) was added to neutralize the reaction mix, and labeled targets were stored in 28.5 μl at −80°C until chip hybridization.

Array Hybridization. Labeled targets from 5 μg of total RNA was used for hybridization on each Kimmel Cancer Center/Thomas Jefferson University miRNA microarray containing 368 probes in triplicate, corresponding to 245 human and mouse miRNA genes. All probes on these microarrays are 40-mer oligonucleotides spotted by contacting technologies and covalently attached to a polymeric matrix. The microarrays were hybridized in 6× SSPE (0.9 M sodium chloride/60 mM sodium phosphate/8 mM EDTA, pH 7.4)/30% formamide at 25°C for 18 h, washed in 0.75× TNT (Tris·HCl/sodium chloride/Tween 20) at 37°C for 40 min, and processed by using direct detection of the biotin-containing transcripts by Streptavidin–Alexa647 conjugate. Processed slides were scanned by using a PerkinElmer ScanArray XL5K Scanner with the laser set to 635 nm, at power 80 and PMT 70 settings, and a scan resolution of 10 μm.

Data Analysis. Images were quantified by QUANTARRAY software (PerkinElmer). Signal intensities for each spot were calculated by subtracting local background (based on the median intensity of the area surrounding each spot) from total intensities. Raw data were normalized and analyzed by using the GENESPRING software (Version 6.1.1, Silicon Genetics, Redwood City, CA). GENESPRING generates an average value of the three spot replicates of each miRNA. After data transformation (to convert any negative value to 0.01), normalization was performed by using a per-chip 50th percentile method that normalizes each chip on its median, allowing comparison among chips. Hierarchical clustering for both genes and conditions were then generated by using standard correlation as a measure of similarity. To highlight genes that characterize each tissue, a per-gene on median normalization was performed, which normalizes the expression of every miRNA on its median among samples.

Samples. HeLa cells were purchased from American Type Culture Collection and grown as recommended. Mouse macrophage cell line RAW264.7 (established from BALB/c mice) was also used (13). RNA from 20 normal human tissues, including 18 of adult origin (seven hematopoietic: bone marrow, lymphocytes B, T, and CD5+ cells from two individuals, peripheral blood leukocytes derived from three healthy donors, spleen, and thymus; and 11 solid tissues, including brain, breast, ovary, testis, prostate, lung, heart, kidney, liver, skeletal muscle, and placenta) and 2 of fetal origin (fetal liver and fetal brain) were assessed for miRNA expression. Each RNA was labeled and hybridized in duplicate, and the average expression was calculated. For all of the normal tissues except lymphocytes B, T, and CD5+ cells, total RNA was purchased from Ambion (Austin, TX).

Cell Preparation. Mononuclear cells from peripheral blood of normal donors were separated by Ficoll–Hypaque density gradients. T cells were purified from these mononuclear cells by rosetting with neuraminidase-treated sheep erythrocyte (SRBC), and depletion of contaminant monocytes (Cd11b+) and natural killer cells (CD16+); B lymphocytes (CD19+) were purified by using magnetic beads (Dynabeads, Unipath, Milan, Italy) and specific mAbs (Becton Dickinson). Total B cells and CD5+ B cells were prepared from tonsils as described (14). Briefly, tonsils were obtained from patients in the pediatric age group that underwent routine tonsillectomies after informed consent. Purified B cells were prepared by rosetting T cells from mononuclear cells with neuraminidase-treated SRBC. To obtain CD5+ B cells, purified B cells were incubated with anti-CD5 mAb followed by goat anti-mouse Ig conjugated with magnetic microbeads. CD5+ B cells were positively selected by collecting the cells retained on magnetic columns by Mini MACS system (Miltenyi Biotec, Auburn, CA). The degree of purification of the cell preparations was >95%, as assessed by flow cytometry.

RNA Extraction and Northern Blots. Total RNA isolation and blots were performed as described (5). After RNA isolation, the washing step with ethanol was not performed, or, if performed, the tube walls were rinsed with 75% ethanol without perturbing the RNA pellet (9). For reuse, blots were stripped by boiling in 0.1% aqueous SDS/0.1× SSC for 10 min and reprobed. 5S rRNA stained with ethidium bromide served as loading control.

Quantitative RT-PCR for miRNA Precursors. Quantitative RT-PCR was performed as described (15). Briefly, RNA was reverse transcribed to cDNA with gene-specific primers and Thermoscript, and the relative amount of each miRNA to both U6 RNA and tRNA for initiator methionine was described by using the equation 2^{-dC_T}, where $dC_T = (C_{T\text{miRNA}} - C_{T\text{U6 or HUMTMI RNA}})$. The miRNAs analyzed included *miR-15a*, *miR-16-1*, *miR-18*, *miR-20*, *miR-21*, *miR-28-2*, *miR-30d*, *miR-93-1*, *miR-105*, *miR-124a-2*, *miR-147*, *miR-216*, *miR-219*, and *miR-224*. The primers used were as published (15).

Microarray Data Submission. All data were submitted by using MIAMEXPRESS to ArrayExpress database, and each of the 44 samples described here received an ID number ranging from SAMPLE169150SUB621 to SAMPLE169193SIUB621.

Results and Discussions

Microarray Construction. The microchip microarray contains 368 gene-specific oligonucleotide probes generated from 248 miRNAs (161 human, 84 mouse, and 3 *Arabidopsis*) and 15 tRNAs (eight human and seven mouse). These sequences correspond to human and mouse miRNAs found in the miRNA Registry (www.sanger.ac.uk/Software/Rfam/mirna, accessed June 2003) (16) or collected from published papers (9–11). A detailed description of oligo sequence design and chip preparation is provided in *Methods* and Table 1. For 76 miRNAs, two different oligo probes were designed, one containing the active sequence and the other specific for the precursor. Using these distinct sequences, we were able to separately analyze the expression of miRNA and pre-miRNA transcripts for the same gene.

Various specificity controls were used to validate data. For intraassay validation, individual oligo-probes were printed in triplicate. Fourteen oligos had a total of six replicates because of identical mouse and human sequences and therefore were spotted on both human and mouse sections of the array. Several mouse and human orthologs differ only in few bases, serving as controls for the hybridization stringency conditions. tRNAs from both species were also printed on the microchip, providing an internal, relatively stable positive control for specific hybridization, whereas *Arabidopsis* sequences were selected based on the absence of any homology with known miRNAs from other species, and used as controls for nonspecific hybridization.

The hybridization sensitivity of our miRNA microarray was tested by using various quantities of total RNA from HeLa cells, starting from 2.5 μg up to 20 μg (Fig. 1*a*). The coefficients of correlation between the 5-μg experiment versus the 2.5-, 10-, and 20-μg experiments were 0.98, 0.99, and 0.97, respectively. These results clearly showed the high interassay reproducibility, even in

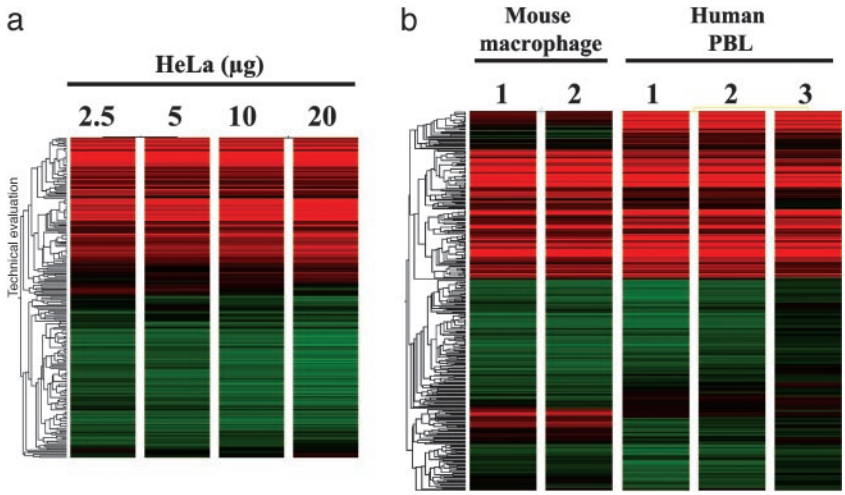

Fig. 1. Microarray expression data. (*a*) miRNA expression in HeLa cells by using various quantities of total RNA. (*b*) miRNA expression in human and mouse samples: human peripheral blood leucocytes (PBL) from three healthy donors and two samples of mouse RAW 264.7 macrophages. Five micrograms of total RNA were used and hybridized as described.

the presence of large differences in RNA quantities. In addition, standard deviation calculated for miRNA triplicates was <10% for the vast majority (>95%) of oligos. We choose to perform all of the other experiments described here with 5 μg of total RNA.

Microarray Specificity. To test the specificity of the microchip, we analyzed miRNA expression in human blood leukocytes from three healthy donors and two samples of mouse macrophages. As shown in Fig. 1*b*, samples derived from the same type of tissue present homogenous patterns of miRNA expression. Furthermore, the pattern of hybridization is different for the two species. To confirm microarray results, the same RNA samples from mouse macrophages and HeLa cells were also analyzed by quantitative RT-PCR for a randomly selected set of 14 miRNAs (15). When we were able to amplify a miRNA precursor for which a correspondent oligo was present on the chip (*hsa-miR-15a*, *hsa-mir-30d*, *mmu-miR-219*, and *mmu-miR-224*), the concordance between the two techniques was 100%. Furthermore, it has been reported that expression levels of the active miRNA and the precursor pre-miRNA are different in the same sample (5, 11, 17); in fact, for another 10 miRNAs for which only the oligo corresponding to the active version was present on the chip, no concordance with quantitative real-time PCR results was observed for the precursor.

The stringency of hybridization was, in several instances, sufficient to distinguish nucleotide mismatches for members of closely related miRNA families, and very similar sequences gave distinct expression profiles (for example *let-7a-1* and *let-7f-2*, which are 89% similar in 88 nucleotides) (Fig. 4, which is published as supporting information on the PNAS web site). Therefore, each quantified result represents the specific expression of a single miRNA member and not the combined expression of the entire family. In other cases, when a portion of oligo was 100% identical for two probes (for example, the 23-mer of active molecule present in the 40-mer oligos for both *mir-16*

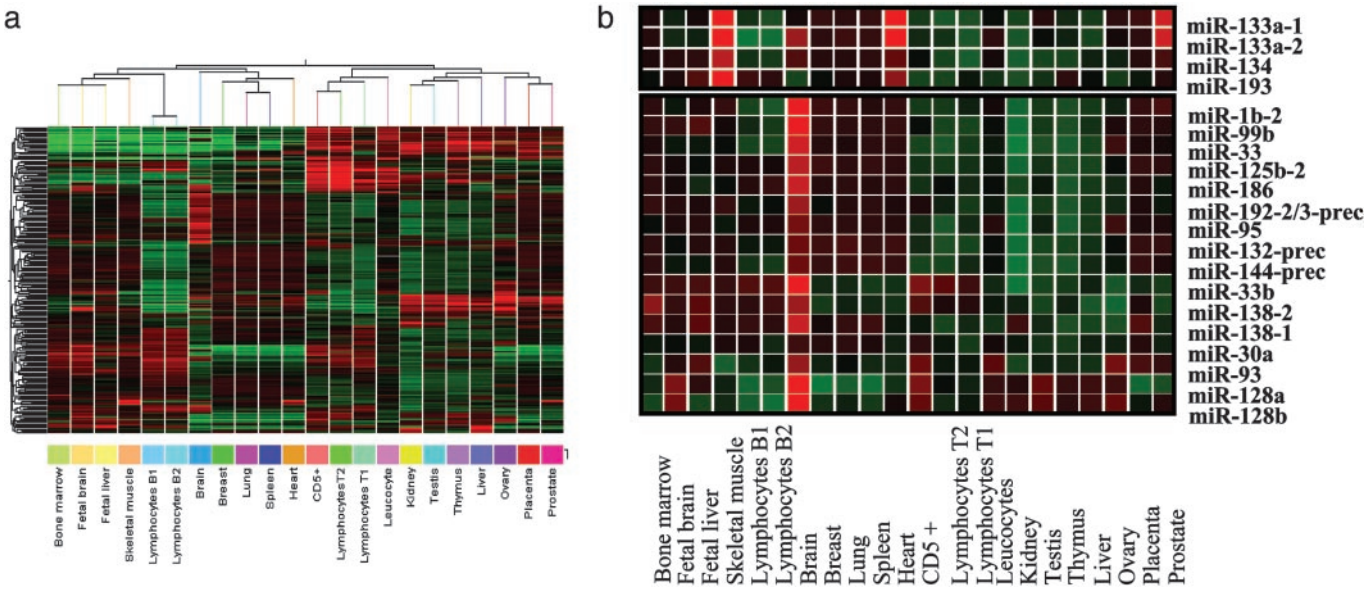

Fig. 2. The human miRNome expression profiles corresponding to 161 miRNAs. (*a*) Distinct patterns of miRNA expression found in human adult and fetal tissues. (*b*) Examples of miRNA signatures: specific overexpression of various miRNAs in skeletal muscle, heart, prostate, and brain. All of the data represent the average of at least two replicates or two samples from the same tissue.

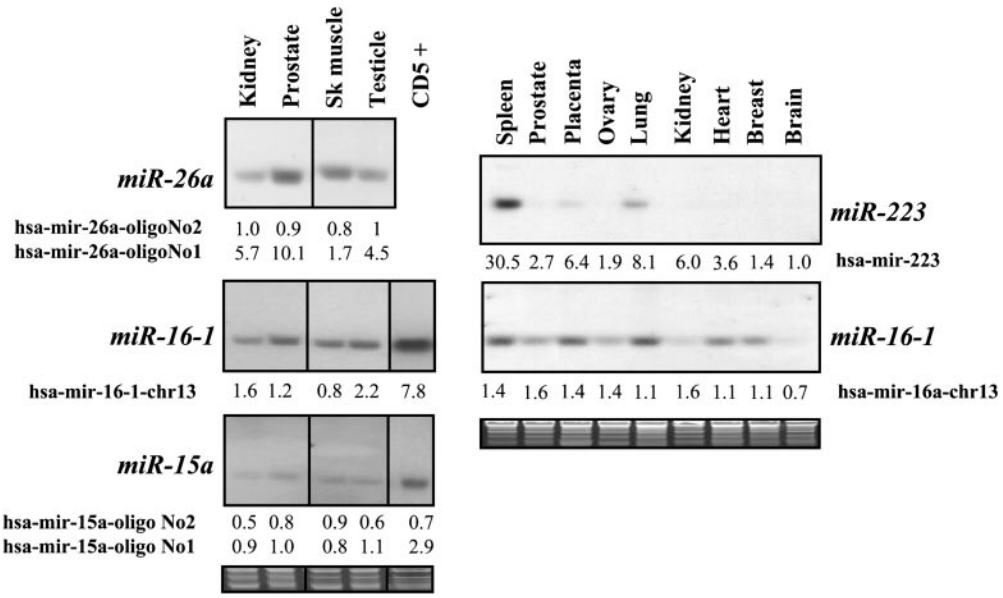

Fig. 3. Northern blot confirmation of the microarray data. The names of miRNAs and specific oligonucleotides spotted on the array are presented. The numbers correspond to absolute expression value of each miRNA (determined by a per-chip on median normalization) and, therefore, can be compared with the band intensity on Northern blots. 5S rRNA stained with ethidium bromide was used as loading control.

sequences from chromosome 13 and chromosome 3), very similar profiles were observed. Therefore, both sequence similarity and secondary structure influence the cross-hybridization between different molecules on this type of microarray.

miRNA Expression in Normal Human Tissues. To further validate reliability of the microarray, we analyzed a panel of 20 RNAs from human normal tissues, including 18 of adult origin (seven hematopoietic and 11 solid tissues) and two of fetal origin (fetal liver and brain). For 15 of them, at least two different RNA samples or two replicates from the same preparation were used (for a detailed list of samples, see *Methods*) (Fig. 2). Several observations were made from these experiments. First, different tissues have distinctive patterns of miRNome expression (defined as the full complement of miRNAs in a cell) with each tissue presenting a specific signature (Fig. 2*a* and Fig. 5 and Table 2, which are published as supporting information on the PNAS web site). When unsupervised hierarchical clustering was used, the same types of tissue from different individuals clustered together. The hematopoietic tissues presented two distinct clusters, the first one containing CD5+ cells, T lymphocytes, and leucocytes, and the second cluster including bone marrow, fetal liver, and B lymphocytes. Of note, RNA of fetal or adult type from the same tissue origin (brain) present different miRNA expression patterns. Second, some miRNAs are highly expressed in only one or few tissues, such as *miR-1b-2* or *miR-99b* in brain, and the closely related members *miR-133a* and *miR-133b* in skeletal muscle, heart, and prostate (Fig. 2*b*). The types of normalization of the GENESPRING software (on 50% percentile with or without a per-gene on median normalization) did not influence these results (see Fig. 5 *a* and *b*).

To verify these data, we used several approaches. We have performed Northern blot analysis on total RNA used in the microarray experiments by using four miRNA probes: *miR-15a*, *miR-16-1*, *miR-26a*, and *miR-223* (Fig. 3). In each case, the concordance between the two techniques was high: in all instances, the highest and the lowest expression levels were concordant (Fig. 3). For example, high levels of *miR-223* expression were found by both techniques in spleen, for *miR-16-1* in CD5+ cells, whereas very low levels were found in brain for both miRNAs (Fig. 3). Moreover, in several instances (for example, *miR-15a*), we were able to identify the same pattern of expression for the precursor and the active miRNA with both microchip and Northern blots.

We also compared the published expression data for cloned human and mouse miRNAs by Northern blot analyses against the microarray results. We found again that the concordance with the chip data are high for both pattern and intensity of expression. For example, *miR-133* was reported to be strongly expressed only in the skeletal muscle and heart (18), precisely as was found with the microarray, whereas *miR-125* and *mir-128* were reported to be highly expressed in brain (18), a finding confirmed on the microchip (Fig. 2*b*).

All these data show that the miRNA microchip is a robust tool, not only for detection of the pattern of miRNA expression, but also for expression quantification. The use of the array has several advantages for miRNA expression detection. First, the global expression of several hundred genes can be identified in the same sample at one time point. Second, through careful design of the oligonucleotide probes, expression of both mature and precursor molecules can be identified. Third, in comparison with Northern blot analysis, the chip requires a small amount of RNA and provides reproducible results using 2.5 μg of total RNA. The relatively limited number of miRNAs (a few hundred per species) (10) offers the possibility to build a common microarray for several species, with distinct oligo probes for each. Such a tool would be of great scientific value, providing opportunities to analyze transspecies expression for each known miRNA under various conditions. Deciphering the miRNome expression in normal and disease states will be useful for the identification of miRNA targets, and alterations in the pattern of miRNA expression may disclose new pathogenetic pathways in human tumorigenesis.

We thank Dr. Richard Davidson for his administrative support of the Kimmel Cancer Center Microarray Facility and Andrew Olson and Hanqing Xie (Compugen USA) for technical support. This work was supported by National Institutes of Health Grant P01-CA81534 for the Chronic Lymphocytic Leukemia Research Consortium, by the Italian Ministero dell' Istruzione, dell' Università e della Ricerca, and by the Italian Ministero della Salute.

1. Bartel, D. P. (2004) *Cell* **116,** 281–297.
2. Pasquinelli, A. E. (2002) *Trends Genet.* **18,** 171–173.
3. Nelson, P., Kiriakidou, M., Sharma, A., Maniataki, E. & Mourelatos, Z. (2003) *Trends Biochem. Sci.* **28,** 534–540.
4. Ambros, V. (2003) *Cell* **113,** 673–676.
5. Calin, G. A., Dumitru, C. D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 15524–15529.
6. Calin, G. A., Sevignani, C., Dumitru, C. D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M. & Croce, C. M. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 2999–3004.
7. Michael, M. Z., O' Connor, S. M., van Holst Pellekaan, N. G., Young, G. P. & James, R. J. (2003) *Mol. Cancer Res.* **1,** 882–891.
8. Metzler, M., Wilda, M., Busch, K., Viehmann, S. & Borkhardt, A. (2004) *Genes Chromosomes Cancer* **39,** 167–169.
9. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. (2001) *Science* **294,** 853–858.
10. Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B. & Bartel, D. P. (2003) *Science* **299,** 1540.
11. Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M. & Dreyfuss, G. (2002) *Genes Dev.* **16,** 720–728.
12. Sorek, R., Ast, G. & Graur, D. (2002) *Genome Res.* **12,** 1060–1067.
13. Dumitru, C. D., Ceci, J. D., Tsatsanis, C., Kontoyiannis, D., Stamatakis, K., Lin, J. H., Patriotis, C., Jenkins, N. A., Copeland, N. G., Kollias, G. & Tsichlis, P. N. (2000) *Cell* **103,** 1071–1083.
14. Dono, M., Zupo, S., Leanza, N., Melioli, G., Fogli, M., Melagrana, A., Chiorazzi, N. & Ferrarini, M. (2000) *J. Immunol.* **164,** 5596–5604.
15. Schmittgen, T. D., Jiang, J., Liu, Q. & Yang, L. (2004) *Nucleic Acids Res.* **32,** 43–53.
16. Griffiths-Jones, S. (2004) *Nucleic Acids Res.* **32,** D109–D111.
17. Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. & Tuschl, T. (2002) *Curr. Biol.* **12,** 735–739.
18. Sempere, L. F., Freemantle, S., Pitha-Rowe, I., Moss, E. G., Dmitrovsky, E. & Ambros, V. (2003) *Genome Biol.* **5,** R13.1–R13.11.