LETTERS

# MicroRNA expression profiles classify human cancers

Jun Lu[1,4]*, Gad Getz[1]*, Eric A. Miska[2]*†, Ezequiel Alvarez-Saavedra[2], Justin Lamb[1], David Peck[1], Alejandro Sweet-Cordero[3,4], Benjamin L. Ebert[1,4], Raymond H. Mak[1,4], Adolfo A. Ferrando[4], James R. Downing[5], Tyler Jacks[2,3], H. Robert Horvitz[2] & Todd R. Golub[1,4,6]

**Recent work has revealed the existence of a class of small non-coding RNA species, known as microRNAs (miRNAs), which have critical functions across various biological processes[1,2]. Here we use a new, bead-based flow cytometric miRNA expression profiling method to present a systematic expression analysis of 217 mammalian miRNAs from 334 samples, including multiple human cancers. The miRNA profiles are surprisingly informative, reflecting the developmental lineage and differentiation state of the tumours. We observe a general downregulation of miRNAs in tumours compared with normal tissues. Furthermore, we were able to successfully classify poorly differentiated tumours using miRNA expression profiles, whereas messenger RNA profiles were highly inaccurate when applied to the same samples. These findings highlight the potential of miRNA profiling in cancer diagnosis.**

Much progress has been made over the last decade in developing a molecular taxonomy of cancer (see ref. 3). In particular, it has become clear that among the ~22,000 protein-coding transcripts are mRNAs that can be used to classify a wide variety of human cancers[4]. Recently, hundreds of small, non-coding miRNAs have been discovered (see ref. 1). The first identified miRNAs, the products of the *C. elegans* genes *lin-4* and *let-7*, have important roles in controlling developmental timing and probably act by regulating mRNA translation[5–7]. When *lin-4* or *let-7* is inactivated, specific epithelial cells undergo additional cell divisions instead of their normal differentiation. Because abnormal cell proliferation is a hallmark of human cancers, it seems possible that miRNA expression patterns might denote the malignant state. Indeed, altered expression of a few miRNAs has been found in some tumour types[8–11]. However, the potential for miRNA expression to inform cancer diagnosis has not been systematically explored.

To determine the expression pattern of all known miRNAs, we first needed to develop an accurate and inexpensive profiling method. This goal is challenging, because of the short size of miRNAs (about 21 nucleotides) and the sequence similarity between miRNA family members. Glass-slide microarrays have been used for miRNA profiling[12–18], but cross-hybridization of related miRNAs has been problematic. We therefore developed a bead-based profiling method. Oligonucleotide-capture probes complementary to miRNAs of interest were coupled to carboxylated 5-micron polystyrene beads impregnated with variable mixtures of two fluorescent dyes (that can yield up to 100 colours), each representing a single miRNA. Following adaptor ligations which use both the 5′-phosphate and the 3′-hydroxyl groups of miRNAs[13], reverse-transcribed miRNAs were (1) amplified by polymerase chain reaction (PCR) using a common biotinylated primer, (2) hybridized to the capture beads, and (3) stained with streptavidin-phycoerythrin. The beads were then analysed using a flow cytometer capable of measuring bead colour (denoting miRNA identity) and phycoerythrin intensity (denoting miRNA abundance) (see Supplementary Fig. 1).

Bead-based hybridization has the theoretical advantage that it might more closely approximate hybridization in solution, and as such, we might expect the specificity to be superior to glass microarray hybridization. Indeed, a spiking experiment involving 11 related sequences showed increased specificity of bead-based detection compared with microarray-based detection, even for single base-pair mismatches (Fig. 1a, b). In addition, the bead method showed linear detection over a hundred-fold range of expression (see Supplementary Information). The expression patterns of eight miRNAs in seven different cell lines were validated by northern blotting. In all cases, bead-based detection paralleled the data from northern blotting (Fig. 1c). These results demonstrate that bead-based miRNA detection is feasible, and has the attractive properties of improved accuracy, high speed and low cost. The bead-based detection platform also provides flexibility, in that additional miRNA capture beads can be added to the mixture, allowing detection of newly discovered miRNAs.

We then set out to determine the expression pattern of all known miRNAs across a large panel of samples representing diverse human tissues and tumour types. Although miRNA expression has been explored in small sets of tissues[12,14–18] or isolated cell types (for example, chronic lymphocytic leukaemia[19]), the extent of differential miRNA expression across cancers has not been determined. Indeed, we might not expect that miRNA expression patterns could be informative with respect to cancer diagnosis, because of the relatively small number of miRNAs encoded in the genome. However, we observed differential expression of nearly all miRNAs across cancer types (Fig. 2a). Moreover, hierarchical clustering of the samples using miRNA profiles paralleled the developmental origins of the tissues. For example, samples of epithelial origin were located on a single branch of the dendrogram, whereas the other major branch was predominantly populated with samples of haematopoietic malignancies.

Furthermore, the miRNAs partitioned tumours within a single lineage. For example, we examined the miRNA profiles of 73 bone

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02141, USA. [2]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, and [3]MIT Center for Cancer Research, Cambridge, Massachusetts 02139, USA. [4]Department of Pediatric Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts 02115, USA. [5]Department of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [6]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. †Present address: Wellcome Trust/Cancer Research UK, Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, UK.
*These authors contributed equally to this work.

marrow samples obtained from patients with acute lymphoblastic leukaemia (ALL). As shown in Fig. 2b, hierarchical clustering revealed non-random partitioning of the samples into three major branches: one containing all five t(9;22) *BCR/ABL*-positive samples and 10 out of 11 t(12;21) *TEL/AML1* samples; a second branch containing 15 out of 19 T-cell ALL samples; and a third branch containing all but one of the samples with an *MLL* gene rearrangement. These experiments demonstrate that even within a single developmental lineage, distinct patterns of miRNA expression can be observed that reflect mechanisms of transformation, and further support the idea that miRNA expression patterns encode the developmental history of human cancers.

Among the epithelial samples, those of the gastrointestinal tract were of particular interest. Samples from colon, liver, pancreas and stomach all clustered together (Fig. 2a), reflecting their common derivation from tissues of embryonic endoderm. We suggest that sample clustering in miRNA space is predominantly driven by developmental history. In contrast, when the same samples were profiled in the space of ~16,000 mRNAs, the coherence of gut-derived samples was not observed in hierarchical clustering (Fig. 2c). This observation might result from the large amount of noise and unrelated signals that are embedded in the high-dimensional mRNA data. Whether or not the miRNAs that are highly expressed in the gut-associated cluster (miR-192, miR-194 and miR-215) have a functional role in the specification of gut development or gut-derived tumours remains to be investigated.

Having determined that miRNA expression distinguishes tumours of different developmental origin, we next asked whether miRNAs could be used to distinguish tumours from normal tissues. We have previously reported that there are no robust mRNA markers that show consistent differential expression between tumours and normal tissues of different lineages[4]. It was therefore striking to observe that despite the fact that some miRNA expression levels were upregulated or unchanged, most of the miRNAs (129 out of 217, $P < 0.05$ after correction for multiple hypothesis testing) had lower expression levels in tumours compared with normal tissues, irrespective of cell type (Fig. 3a). Cancer cell lines also showed low miRNA expression relative to normal tissues (see Supplementary Fig. 4).

To exclude any possibility that this differential miRNA expression might be related to differences in the collection of tumour samples versus normal samples, we studied a mouse model of *K-Ras*-induced lung cancer[20]. We isolated miRNAs from normal lung or lung adenocarcinomas from different individual mice, thereby precluding any differences in collection procedure. Owing to miRNA sequence conservation between human and mouse, the same miRNA capture beads could be used to profile the murine samples. As shown in Fig. 3b, the same distinction between normal and tumour samples is observed in mouse. Accordingly, a tumour/normal classifier constructed with human samples had 100% accuracy when tested in the mouse. Taken together, these studies indicate that miRNAs are unexpectedly rich in information content with respect to cancer.

Our observation that miRNA expression seems globally higher in normal tissues compared with tumours led us to the hypothesis that global miRNA expression reflects the state of cellular differentiation. To test this hypothesis, we explored an experimental model in which we treated the myeloid leukaemia cell line HL-60 with *all-trans* retinoic acid, a potent inducer of neutrophilic differentiation[21]. As predicted, miRNA profiling revealed the induction of many miRNAs coincident with differentiation (Fig. 3c). In primary human haematopoietic progenitor cells undergoing erythroid differentiation *in vitro*, we observed a similar increase in miRNA expression at a stage in differentiation when the cells continued to proliferate (see

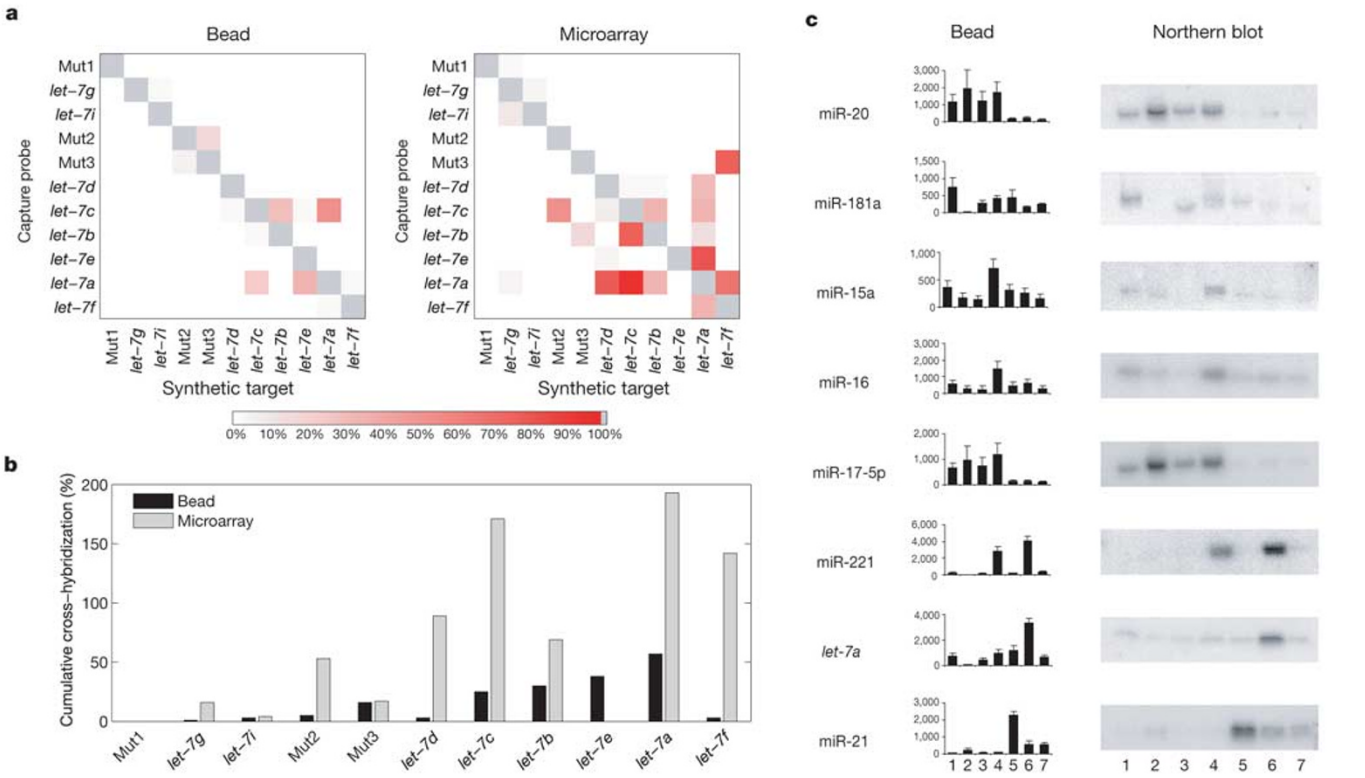


**Figure 1 | Specificity and accuracy of bead-based miRNA detection.** **a**, Synthetic oligonucleotides corresponding to the *let-7* family and mutants (mut1–3) (see Supplementary Information for sequence similarity) were PCR-labelled and hybridized separately on beads and a glass microarray. Synthetic targets are indicated on the horizontal axis, capture probes on the vertical axis. Values represent the proportion of signal relative to correct probe (set to 100%). **b**, Cumulative cross-hybridization on capture probes. **c**, Northern blot versus bead detection of cell line samples (lanes 1–7: HEL, K562, TF-1, 293, MCF-7, PC-3, SKMEL-5 cells). Bead results shown on the left are averages from three independent experiments for HEL, TF-1, 293, MCF-7 and PC-3 cells, and two experiments for K562 and SKMEL-5 cells. Error bars indicate s.d.

Supplementary Information). These experiments support the hypothesis that global changes in miRNA expression are associated with differentiation, the abrogation of which is a hallmark of all human cancers. These findings are also consistent with the recent observation that mouse embryonic stem cells lacking Dicer, an enzyme required for miRNA maturation, fail to differentiate normally[22].

We next turned to a more challenging diagnostic distinction: that of tumours of histologically uncertain cellular origin. It is estimated that 2–4% of all cancer diagnoses represent cancers of unknown origin or diagnostic uncertainty (see ref. 23). To address this, we analysed 17 poorly differentiated tumours, the histological appearance of which was non-diagnostic, but for which clinical diagnosis was established by anatomical context, either directly (for example, a primary tumour arising in the colon) or indirectly (a metastasis of a previously identified primary tumour). A training set of 68 more-differentiated tumours, representing 11 tumour types and for which both mRNA and miRNA profiles were available, was used to generate a classifier. This classifier was then used without modification to classify the 17 poorly differentiated test samples. As a group, poorly differentiated tumours had lower global levels of miRNA expression compared with the more-differentiated training set

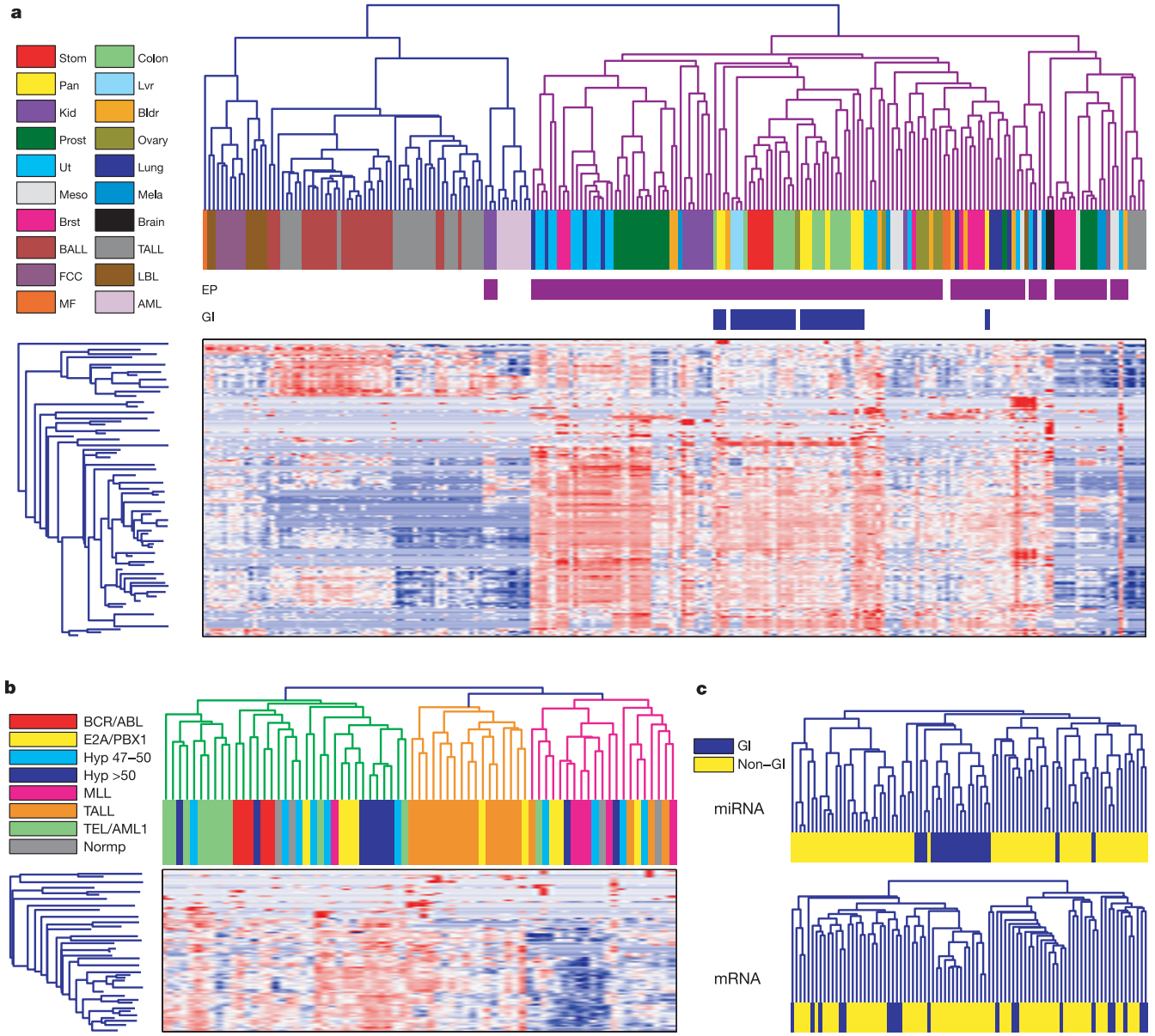


**Figure 2 | Hierarchical clustering of miRNA expression.** **a**, miRNA profiles of 218 samples from several different tissues were clustered (average linkage, correlation similarity). Samples are in columns, miRNAs in rows. Samples of epithelial (EP) origin or derived from the gastrointestinal tract (GI) are indicated. More detail is shown in Supplementary Fig. 4. **b**, Clustering of 73 bone marrow samples from patients with acute lymphoblastic leukaemia (ALL). Coloured bars indicate the different ALL subtypes. **c**, Comparison of miRNA data and mRNA data. For 89 epithelial samples from **a** that had mRNA expression data, hierarchical clustering was performed. Samples of GI origin are shown in blue. GI-derived samples largely cluster together in miRNA expression space, but not in mRNA expression space. Abbreviations used: Bldr, bladder; Brst, breast; Fcc, follicular lymphoma; Kid, kidney; Lvr, liver; Mela, melanoma; Meso, mesothelioma; Pan, pancreas; Prost, prostate; Stom: stomach; Ut, uterus; AML: acute myelogenous leukaemia; BALL, B-cell ALL; LBL, diffuse large-B cell lymphoma; MF, mycosis fungoides; MLL, mixed lineage leukaemia; TALL, T-cell ALL; Hyper 47–50, hyperdiploid with 47–50 chromosomes; Hyper >50, hyperdiploid with over 50 chromosomes; Normp, normal ploidy. Further details can be found in Supplementary Information.

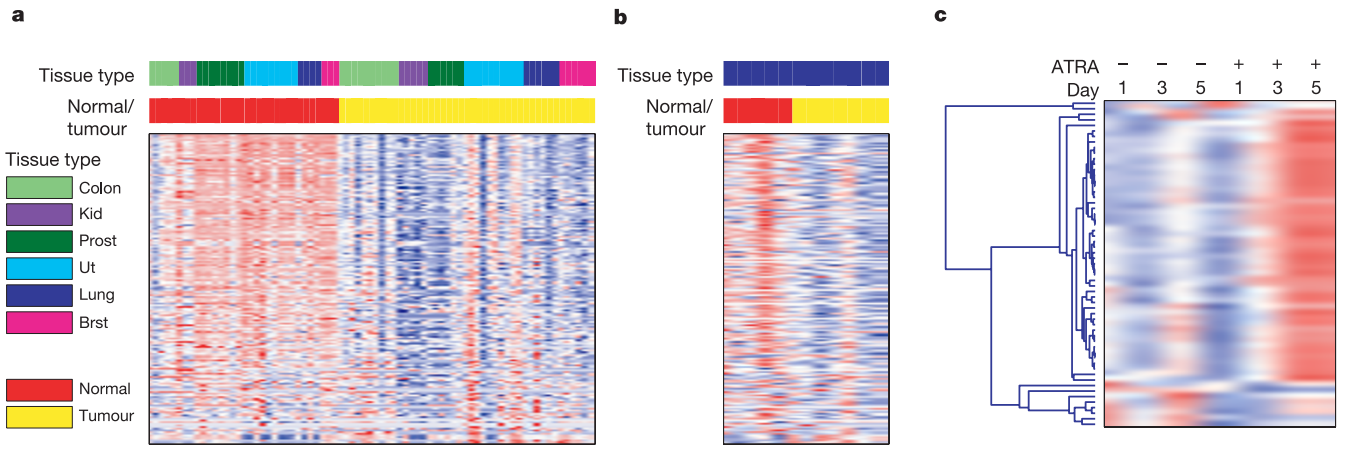


**Figure 3 | Comparison between normal and tumour samples reveals global changes in miRNA expression.** **a**, Markers were selected to correlate with the normal versus tumour distinction. A heatmap of miRNA expression is shown, with miRNAs sorted according to the variance-thresholded *t*-test score. **b**, miRNA markers of normal versus tumour distinction in human tissues from **a**, applied to normal lungs and lung adenocarcinomas of *K-Ras*$^{LA1}$ mice. A *k*-nearest neighbour (*k*NN) classifier based on human sample-derived markers yielded a perfect classification of the mouse samples (euclidean distance, $k = 3$). Mouse tumour T_MLUNG_5 (third column from right) was occasionally classified as normal using other *k*NN parameters (see Supplementary Information). **c**, HL-60 cells were treated with *all-trans* retinoic acid (ATRA + ) or vehicle (−) for the indicated number of days. A heatmap of miRNA expression from a representative experiment is shown.

samples (see Supplementary Fig. 5), consistent with the notion that miRNA expression is closely linked to differentiation. Despite the overall low level of miRNA expression, the miRNA-based classifier established the correct diagnosis of the poorly differentiated samples with far greater accuracy than would be expected by chance for an 11-class classifier (12 out of 17 correct; $P < 5 \times 10^{-11}$). In contrast, the mRNA-based classifier was highly inaccurate (1 out of 17 correct; $P = 0.47$), as we previously reported[4].

The experiments reported here demonstrate the feasibility and utility of monitoring the expression of miRNAs in human cancer. Our unexpected findings are the extraordinary level of diversity in miRNA expression across cancers, and the large amount of diagnostic information encoded in a relatively small number of miRNAs. The implication is that, unlike with mRNA expression, a modest number of miRNAs (~200 in total) might be sufficient to classify human cancers. Moreover, the bead-based miRNA detection method has the attractive property of being not only accurate and specific, but also easy to implement in a routine clinical setting. In addition, unlike mRNAs, miRNAs remain largely intact in routinely collected, formalin-fixed, paraffin-embedded clinical tissues[15]. More work is required to establish the clinical utility of miRNA expression in cancer diagnosis, but the work described here indicates that miRNA profiling has unexpected diagnostic potential. The mechanism by which miRNAs are under-expressed in cancer remains unknown. We did not observe substantial decreases in the mRNAs encoding components of the miRNA processing machinery (Dicer, Drosha, Argonaute2 or DGCR8 (ref. 24); see Supplementary Information), but clearly other mechanisms of miRNA regulation are possible.

The findings reported here are consistent with the hypothesis that in mammals, as in *C. elegans*, miRNAs can function to prevent cell division and drive terminal differentiation. An implication of this hypothesis is that downregulation of some miRNAs might play a causal role in the generation or maintenance of tumours. Epithelial cells affected in *C. elegans lin-4* and *let-7* miRNA mutants generate a stem-cell-like lineage, dividing to produce daughters that, like themselves, divide rather than differentiate[5,7]. We speculate that abnormalities in miRNA expression might similarly contribute to the generation or maintenance of 'cancer stem cells', recently proposed to be responsible for cancerous growth in both leukaemias and solid tumours[25–28].

## METHODS

**Samples.** Information regarding samples is available in Supplementary Table 2. Total RNAs were prepared from tissues or cell lines using TRIzol (Invitrogen), as described[4], and in compliance with IRB protocols. Leukaemia bone marrow mononuclear cells were collected from patients treated at St. Jude Children's Research Hospital and at the Dana-Farber Cancer Institute, and their immunophenotype and genotype were determined as previously described[29,30]. Normal mouse lung and mouse lung cancer samples were collected from *K-Ras*$^{LA1}$ mice and genotyped as described[20]. Lungs from four- to five-month-old mice were inflated with phosphate-buffered saline before removal. Individual lung tumours and normal lungs were dissected and immediately frozen on dry ice before RNA preparation. HL-60 cells were plated at $1.5 \times 10^5$ cells ml$^{-1}$ and induced to differentiate using 1 μM *all-trans* retinoic acid (Sigma; in ethanol). Cells were harvested after 1, 3 and 5 days. Culturing conditions for other cells are detailed in the Supplementary Information.

**miRNA labelling.** Target preparation from total RNA followed the described procedure[13], with modifications. Briefly, two synthetic pre-labelling-control RNA oligonucleotides (5′-pCAGUCAGUCAGUCAGUCAGUCAG-3′, and 5′-pGACCUCCAUGUAAACGUACAA-3′, Dharmacon) were used to control for target preparation efficiency. They were each spiked at 3 fmoles per μg total RNA. Small RNAs (18–26 nucleotides) were recovered from 1–10 μg total RNA through denaturing polyacrylamide gel purification. Small RNAs were adaptor-ligated sequentially on the 3′-end and 5′-end using T4 RNA ligase (Amersham Biosciences). After reverse-transcription using adaptor-specific primers, products were PCR-amplified (95 °C for 40 s; 50 °C for 30 s; 72 °C for 30 s; 18 cycles for 10 μg starting total RNA) using a 3′-primer 5′-TACTGG AATTCGCGGTTA-3′ and 5′ primer 5′-biotin-CAACGGAATTCCTCACTAAA-3′ (IDT). For side-by-side comparison of bead detection and the glass microarray, a 5′-Alexa-532-modified primer was used for compatibility with the glass microarray. PCR products were precipitated and dissolved in 66 μl TE buffer (10 mM TrisHCl pH 8.0, 1 mM EDTA) containing two biotinylated post-labelling-control oligonucleotides (100 fmoles of FVR506, 25 fmoles PTG20210; see Supplementary Table 1).

**Bead-based detection.** miRNA capture probes were 5′-amino-modified oligonucleotides with a 6-carbon linker (IDT). Capture probes for miRNAs and controls were divided into three sets (see Supplementary Table 1). To profile a sample with all probes, three assays were performed on the sample, each using one of the three probe sets. Probes were conjugated to carboxylated xMAP beads (Luminex Corporation) in 96-well plates, following the manufacturer's protocol. For each probe set, 3 μl of every probe–bead conjugate was mixed into 1 ml of 1.5 × TMAC (4.5 M tetramethylammonium chloride, 0.15% sarkosyl, 75 mM Tris-HCl pH 8.0, 6 mM EDTA). Samples were hybridized in a 96-well plate, with two mock PCR samples (using water as template) in each plate as a background control. Hybridization was carried out overnight at 50 °C with 33 μl of the bead

mixture and 15 μl of labelled material. Beads were spun down, resuspended in 1 × TMAC containing 10 μg ml$^{-1}$ streptavidin-phycoerythrin (Molecular Probes) and incubated at 50 °C for 10 min before data acquisition on a Luminex 100IS machine. Median fluorescence intensity values were measured.

**Computational analyses.** Profiling data were first scaled according to the post-labelling-controls and then the pre-labelling-controls, in order to normalize readings from different probe/bead sets for the same sample, and to normalize for the labelling efficiency, respectively, as detailed in Supplementary Methods. Data were thresholded at 32 and $\log_2$-transformed. Hierarchical clustering was performed with average linkage and Pearson correlation. Before clustering, data were filtered to eliminate genes with expression lower than 7.25 (on a $\log_2$ scale) in all samples. Next, all features were centred and normalized to a mean of 0 and a standard deviation of 1. *k*-nearest-neighbour classification of normal versus tumour samples was performed with $k = 3$ in the selected feature space using euclidean distance measure. Note that different metrics were used for clustering and normal/tumour classification. Features were selected for the distinction between all normal samples versus all tumours (for colon, kidney, prostate, uterus, lung and breast; $P < 0.05$ after Bonferroni correction). *P* values were calculated using a variance-thresholded *t*-test with a minimal standard deviation of 0.75, while treating the tissue type as a confounding variable. Multi-class predictions of poorly differentiated tumours were performed using the probabilistic neural network algorithm, a gaussian-weighted nearest neighbour method. For each test sample, the tissue type that had the highest probability in multiple one-tissue-versus-the-rest predictions was assigned. Feature number and the gaussian width were optimized on the basis of leave-one-out cross-validations on the training data set. Features were selected on the basis of the variance-thresholded *t*-test score, requiring equal numbers of up- and down-regulated features. Distances were based on the cosine in the selected feature space.

**Expression data.** miRNA expression data have been submitted to the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo) with the series accession number GSE2564. mRNA expression data were published previously[4], and are available together with miRNA expression data at http://www.broad.mit.edu/cancer/pub/miGCM.

**Received 2 February; accepted 5 May 2005.**


1. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116,** 281–297 (2004).
2. Ambros, V. The functions of animal microRNAs. *Nature* **431,** 350–355 (2004).
3. Chung, C. H., Bernard, P. S. & Perou, C. M. Molecular portraits and the family tree of cancer. *Nature Genet.* **32,** 533–540 (2002).
4. Ramaswamy, S. *et al.* Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl Acad. Sci. USA* **98,** 15149–15154 (2001).
5. Ambros, V. & Horvitz, H. R. Heterochronic mutants of the nematode *Caenorhabditis elegans*. *Science* **226,** 409–416 (1984).
6. Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75,** 843–854 (1993).
7. Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403,** 901–906 (2000).
8. Michael, M. Z., O'Connor, S. M., van Holst Pellekaan, N. G., Young, G. P. & James, R. J. Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol. Cancer Res.* **1,** 882–891 (2003).
9. Calin, G. A. *et al.* Frequent deletions and down-regulation of micro-RNA genes *miR15* and *miR16* at 13q14 in chronic lymphocytic leukemia. *Proc. Natl Acad. Sci. USA* **99,** 15524–15529 (2002).
10. Eis, P. S. *et al.* Accumulation of miR-155 and *BIC* RNA in human B cell lymphomas. *Proc. Natl Acad. Sci. USA* **102,** 3627–3632 (2005).
11. Johnson, S. M. *et al. RAS* is regulated by the *let-7* microRNA family. *Cell* **120,** 635–647 (2005).
12. Liu, C. G. *et al.* An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc. Natl Acad. Sci. USA* **101,** 9740–9744 (2004).
13. Miska, E. A. *et al.* Microarray analysis of microRNA expression in the developing mammalian brain. *Genome Biol.* **5,** R68 (2004).
14. Thomson, J. M., Parker, J., Perou, C. M. & Hammond, S. M. A custom microarray platform for analysis of microRNA gene expression. *Nature Methods* **1,** 47–53 (2004).
15. Nelson, P. T. *et al.* Microarray-based, high-throughput gene expression profiling of microRNAs. *Nature Methods* **1,** 155–161 (2004).
16. Babak, T., Zhang, W., Morris, Q., Blencowe, B. J. & Hughes, T. R. Probing microRNAs with microarrays: tissue specificity and functional inference. *RNA* **10,** 1813–1819 (2004).
17. Sun, Y. *et al.* Development of a micro-array to detect human and mouse microRNAs and characterization of expression in human organs. *Nucleic Acids Res.* **32,** e188 (2004).
18. Barad, O. *et al.* MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. *Genome Res.* **14,** 2486–2494 (2004).
19. Calin, G. A. *et al.* MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc. Natl Acad. Sci. USA* **101,** 11755–11760 (2004).
20. Johnson, L. *et al.* Somatic activation of the *K-ras* oncogene causes early onset lung cancer in mice. *Nature* **410,** 1111–1116 (2001).
21. Stegmaier, K. *et al.* Gene expression-based high-throughput screening (GE-HTS) and application to leukemia differentiation. *Nature Genet.* **36,** 257–263 (2004).
22. Kanellopoulou, C. *et al.* Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.* **19,** 489–501 (2005).
23. Pavlidis, N., Briasoulis, E., Hainsworth, J. & Greco, F. A. Diagnostic and therapeutic management of cancer of an unknown primary. *Eur. J. Cancer* **39,** 1990–2005 (2003).
24. Cullen, B. R. Transcription and processing of human microRNA precursors. *Mol. Cell* **16,** 861–865 (2004).
25. Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367,** 645–648 (1994).
26. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414,** 105–111 (2001).
27. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl Acad. Sci. USA* **100,** 3983–3988 (2003).
28. Singh, S. K. *et al.* Identification of human brain tumour initiating cells. *Nature* **432,** 396–401 (2004).
29. Yeoh, E. J. *et al.* Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* **1,** 133–143 (2002).
30. Ferrando, A. A. *et al.* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* **1,** 75–87 (2002).


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank E. Lander for critical review of the manuscript, S. Ramaswamy for discussions, and J.-P. Brunet, S. Monti, C. Ladd-Acosta and S. Shurtleff for computational help and technical assistance. We also thank J. Jacobson and Luminex Corporation for advice and technical support. E.A.M. was supported by the Howard Hughes Medical Institute. H.R.H., T.J. and T.R.G. are Investigators of the Howard Hughes Medical Institute.

**Author Information** miRNA expression data have been submitted to the Gene Expression Omnibus under the series accession number GSE2564. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.R.G. (golub@broad.mit.edu).