MicroRNA expression signatures accurately discriminate acute lymphoblastic leukemia from acute myeloid leukemia

Shuangli Mi[†], Jun Lu[‡§], Miao Sun[†], Zejuan Li[†], Hao Zhang[‡], Mary Beth Neilly[†], Yungui Wang[¶], Zhijian Qian[†], Jie Jin[¶], Yanming Zhang[†], Stefan K. Bohlander[‖], Michelle M. Le Beau[†], Richard A. Larson[†], Todd R. Golub[‡§††‡‡], Janet D. Rowley[†‡‡], and Jianjun Chen[†‡‡]

[†]Department of Medicine, University of Chicago, Chicago, IL 60637; [‡]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02141; [§]Department of Pediatric Oncology, Dana–Farber Cancer Institute, Boston, MA 02115; [¶]Institute of Hematology, First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310003, People's Republic of China; [‖]Department of Medicine III, University of Munich Hospital Grosshadern and Clinical Cooperative Group ''Leukemia,'' GSF–National Research Center for Environment and Health, 81377 Munich, Germany; and [††]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115

Contributed by Janet D. Rowley, October 8, 2007 (sent for review August 30, 2007)

Acute lymphoblastic leukemia (ALL) is the most common childhood cancer, whereas acute myeloid leukemia (AML) is the most common acute leukemia in adults. In general, ALL has a better prognosis than AML. To understand the distinct mechanisms in leukemogenesis between ALL and AML and to identify markers for diagnosis and treatment, we performed a large-scale genome-wide microRNA (miRNA, miR) expression profiling assay and identified 27 miRNAs that are differentially expressed between ALL and AML. Among them, miR-128a and -128b are significantly overexpressed, whereas let-7b and miR-223 are significantly downregulated in ALL compared with AML. They are the most discriminatory miRNAs between ALL and AML. Using the expression signatures of a minimum of two of these miRNAs resulted in an accuracy rate of >95% in the diagnosis of ALL and AML. The differential expression patterns of these four miRNAs were validated further through large-scale real-time PCR on 98 acute leukemia samples covering most of the common cytogenetic subtypes, along with 10 normal control samples. Furthermore, we found that overexpression of miR-128 in ALL was at least partly associated with promoter hypomethylation and not with an amplification of its genomic locus. Taken together, we showed that expression signatures of as few as two miRNAs could accurately discriminate ALL from AML, and that epigenetic regulation might play an important role in the regulation of expression of miRNAs in acute leukemias.

expression profiling | lineage classification | diagnosis | prediction | DNA copy number

Human acute leukemias, including acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML), are genetically very diverse, arising from blood cell progenitors developing in the lymphoid or myeloid pathway or from primitive stem cells with multilineage potential (1, 2). Although ALL can originate from either B or T lymphocyte progenitors, most cases are B-lineage ALL, which represents >85% of childhood and >75% of adult ALL (3, 4). In the United States from 2000 to 2004, ≈80% of childhood and adolescent acute leukemia patients have ALL, whereas >85% of adult acute leukemia patients have AML. It is estimated that 5,200 and 13,410 cases will be diagnosed with, and 1,420 and 8,990 patients will die of, ALL and AML, respectively, in the United States in 2007 (http://seer.cancer.gov).

Acquired ("somatic") clonal karyotype abnormalities are detected in 60–80% of patients with ALL or AML, whereas the remaining 20–40% have a normal karyotype (2, 4–8). In addition to those with a normal karyotype, *TEL-AML1* (*ETV6-RUNX1*)/t(12;21), *BCR-ABL* (*BCR-ABL1*)/t(9;22), *MLL* rearrangements/t(11q23), *E2A-PBX1* (*TCF3-PBX1*)/t(1;19), *MYC-IGH*/[t(8;14), t(2;8), or t(8;22)], and hyperdiploidy/(>50 chromosomes) are among the most common cytogenetic subtypes in ALL, whereas *AML1-ETO* (*RUNX1-RUNX1T1*)/t(8;21), *CBFB-MYH11*/inv(16), *PML-RARA*/t(15;17), *MLL* rearrangements/t(11q23), trisomy 8, and monosomy 7 are among the most common cytogenetic subtypes in AML (2, 4, 5, 8). Interestingly, *MLL* rearrangements/t(11q23) are frequent cytogenetic abnormalities found in both ALL and AML, occurring in 5–6% of AML and 7–10% of ALL patients (1, 9). Recurrent genetic abnormalities have prognostic and therapeutic implications and also provide insights into the mechanisms of leukemogenesis (4, 5, 8).

Because the therapeutic strategies and prognosis vary considerably between ALL and AML (10, 11) (see also http://seer.cancer.gov), ALL must be distinguished from AML at diagnosis. Although ALL can be distinguished from AML through appropriate use of morphologic, immunohistochemical, and immunologic methods (12), the conventional clinical practice requires experienced personnel, and no single test is currently sufficient to establish the diagnosis. In a pioneer study to find a more efficient diagnostic approach, Golub *et al.* (13) showed that ALL could be distinguished from AML based on gene expression profiles. Since then, messenger RNA (protein-coding gene) expression profiling has been widely used in classification of subtypes of AML and ALL, as well as in prediction of prognosis/outcome of leukemias [see reviews (14–16)]. However, the precise genes and pathways that exert critical control over determination of lineage fate during leukemia development remain unclear.

One of the most exciting findings in the life sciences in recent years has been the discovery of an abundant class of small (≈22 nucleotides), non-(protein-)coding RNAs, called microRNAs (miRNAs, miRs), which can play important regulatory roles in development, cell proliferation, differentiation, and apoptosis (17–19). Evidence is emerging that miRNAs can function as oncogenes and tumor suppressors (20–24). Using a bead-based flow cytometric method, Lu *et al.* (25) found that miRNA expression profiles could accurately classify human cancers; furthermore, they were able to classify successfully poorly differentiated tumors by using miRNA expression profiles, whereas mRNA profiles were highly inaccurate when applied to the same samples. To identify markers

Author contributions: S.M., J.L., M.S., and Z.L. contributed equally to this work; J.D.R. and J.C. designed research; S.M., J.L., M.S., Z.L., H.Z., M.B.N., and Y.W. performed research; J.L., Z.Q., J.J., Y.Z., S.K.B., M.M.L.B., R.A.L., and T.R.G. contributed new reagents/analytic tools; S.M., J.L., M.S., Z.L., and J.C. analyzed data; and J.D.R. and J.C. wrote the paper.

The authors declare no conflict of interest.

[‡‡]To whom correspondence may be addressed. E-mail: golub@broad.mit.edu, jrowley@medicine.bsd.uchicago.edu, or jchen@medicine.bsd.uchicago.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0709313104/DC1.

MEDICAL SCIENCES

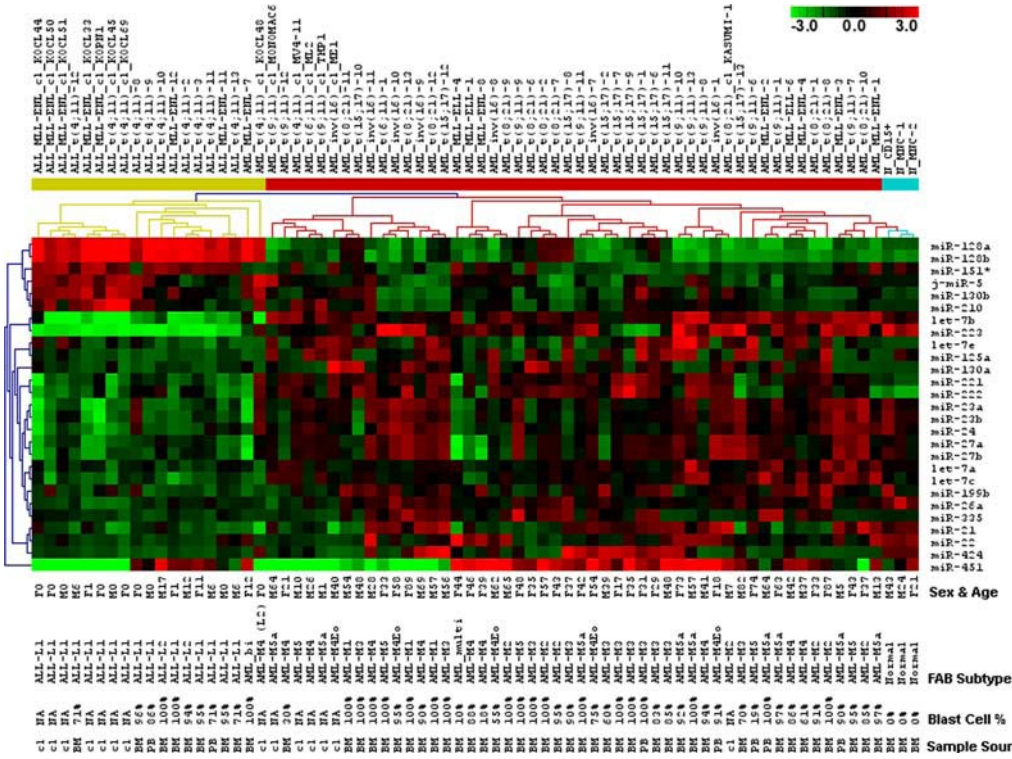

Fig. 1. Twenty-seven miRNAs differentially expressed between ALL and AML. Unsupervised average linkage hierarchical clustering was performed. C1_,cell line; N_, normal control; MNC_, mononuclear cells; CD15+, CD15+ myeloid progenitor cells; F, female; M, male; AML_multi, AML with multiple lineages, AML_bi, biphenotypic AML; NA, not available; BM, bone marrow; PB, peripheral blood; cl, cell line.

for diagnosis and treatment of ALL and AML and to shed light on the distinct mechanisms of leukemogenesis between ALL and AML, we performed a genome-wide miRNA expression analysis using the bead-based flow cytometric method (25) on 18 ALL and 54 AML samples, along with three normal controls. In both unsupervised and supervised cluster analyses, ALL and AML samples segregated clearly according to their lineage. The result was validated by large-scale quantitative RT-PCR. Possible mechanisms underlying differential expression of some miRNAs were also studied.

Results

miRNA Expression Signatures Discriminate ALL from AML. We performed a systematic miRNA expression profiling analysis of 435 mammalian miRNAs on 72 acute leukemia samples, including 18 ALL and 54 AML, along with three normal bone marrow control samples [see supporting information (SI) Table 2]. All 18 ALL samples were B-lineage, including11 primary ALL specimens [i.e., nine t(4;11)/*MLL-AF4* (*MLL-AFF1*) and two t(11;19)(q23;p13.3)/*MLL-ENL* (*MLL-MLLT1*)] and seven ALL cell lines [i.e., two t(4;11) and five *MLL-ENL*]. The 54 AML samples included 47 primary AML specimens [i.e., 10 t(15;17)/*PML-RARA*, 10 t(8;21)/*AML1-ETO* (*RUNX1-RUNX1T1*), 7 inv(16)/*CBFB-MYH11*, and 20 *MLL*-rearrangement AMLs, including 9 t(9;11)/*MLL-AF9* (*MLL-MLLT3*), 7 t(11;19)(q23;p13.3)/*MLL-ENL*, 3 t(11;19)(q23;p13.1)/*MLL-ELL*, and one t(6;11)/*MLL-AF6* (*MLL-MLLT4*)] and seven AML cell lines [one t(8;21), one inv (16), and five *MLL*-rearrangement AMLs, including two t(4;11), one t(6;11)/*MLL-AF6*, and two t(9;11)]. Note that *MLL* rearrangements were present in all 18 ALL samples and in 25/54 AML samples. Three normal bone marrow samples from healthy donors were used as normal controls, which included two mononuclear cell (MNC) samples and one CD15+ myeloid progenitor cell sample (see *Materials and Methods*). Bead-based miRNA expression profiling detection was performed as described (25). To control for data quality, three samples with total fluorescence <15,000 were discarded as unsuccessful labeling/sample quality. In addition, after normalization, only probes with maximum expression in all samples ≥7.25 (regarded as confidently expressed) were retained for further analyses. Probes specific for mouse or rat miRNAs were excluded (see *Materials and Methods*). Finally, after filtering, a total of 72 human samples (including 17 ALL, 52 AML, and 3 normal controls) and 112 human miRNA genes with confidently detectable expression levels were selected for further analyses (SI Table 2).

We performed an unsupervised two-way (genes against samples), hierarchical cluster analysis (HCA) (26) using the expression profiling of the 112 human miRNA genes in the 72 samples. As shown in SI Fig. 3*a*, leukemia samples clearly grouped into two clusters: (*i*) all of the ALL samples grouped together as Cluster 1 and (*ii*) all of the AML samples, except for *MLL-ENL*-7, grouped together as Cluster 2. The three normal control samples also grouped together and formed a subcluster under the cluster of AMLs (SI Fig. 3*a*). Interestingly, *MLL-ENL*-7 is an AML sample with biphenotypic features (see SI Table 2). Thus, that *MLL-ENL*-7 clustered together with ALL samples may reflect its intrinsic ALL-related genetic factors. To visualize the clustering pattern more effectively, we also carried out a principal component analysis (PCA) using the gene expression profiles. As shown in SI Fig. 3*b*, ALL and AML samples separate clearly. A similar pattern was observed when we analyzed MLL-rearrangement cases alone (data not shown). In the analysis of AML samples alone, we observed that t(15;17) samples grouped together as one cluster, as did the MLL-rearrangement AMLs despite different partner genes fused to the *MLL* gene; interestingly, t(8;21) and inv(16), both core-binding factor (CBF) AMLs, grouped together as a unique cluster (Z.L., J.L., M.S., S.M., and H.Z., unpublished work).

miRNAs Differentially Expressed Between ALL and AML. To identify miRNAs differentially expressed between ALL and AML samples,

Table 1. Prediction of ALL and AML with PAM using expression signatures of part of or the whole set of the four most discriminatory miRNAs (i.e., miR-128a, miR-128b, let-7b, and miR-223)

Predictors used	Overall accuracy, %[†]	Sensitivity, %[‡]	Specificity, %[§]
All the four miRNAs	97 (67/69)	ALL: 100 (17/17) AML: 96 (50/52)	ALL: 96 (50/52) AML: 100 (17/17)
Any three of the four miRNAs	97 (67/69)	ALL: 100 (17/17) AML: 96 (50/52)	ALL: 96 (50/52) AML: 100 (17/17)
miR-128a and -128b	97 (67/69)	ALL: 100 (17/17) AML: 96 (50/52)	ALL: 96 (50/52) AML: 100 (17/17)
miR-128a or -128b and let-7b	97 (67/69)	ALL: 100 (17/17) AML: 96 (50/52)	ALL: 96 (50/52) AML: 100 (17/17)
miR-128a or -128b and miR-223	99 (68/69)	ALL: 100 (17/17) AML: 98 (51/52)	ALL: 98 (51/52) AML: 100 (17/17)
let-7b and miR-223	97 (67/69)	ALL: 100 (17/17) AML: 96 (50/52)	ALL: 96 (50/52) AML: 100 (17/17)

[†]Prediction accuracy was determined by 10-fold cross-validation on the 69 leukemia samples consisting of 17 ALL and 52 AML samples. If only one sample was "misclassified," it is *MLL-ENL*-7; if two samples were misclassified, they are *MLL-ENL*-7 and KOCL-48. However, given that both are biphenotypic leukemia, the overall prediction accuracy as well as the sensitivity and specificity probably could be considered 100%.
[‡]Sensitivity = (the number of positive samples predicted)/(the number of true positives).
[§]Specificity = (the number of negative samples predicted)/(the number of true negatives).

we used significance analysis of microarray (SAM) (27) and permutation tests (10,000 permutations) in a comparison between ALL and AML (normal controls were not included in the SAM analysis). We identified 27 miRNAs differentially expressed between ALL and AML samples (see Fig. 1). Six (i.e., miR-128a, miR-128b, miR-151*, j-miR-5, miR-130b, and miR-210) were expressed at a significantly higher level in ALL than in AML. In contrast, the remaining 21 (i.e., let-7b, miR-223, let-7e, miR-125a, miR-130a, miR-221, miR-222, miR-23a, miR-23b, miR-24, miR-27a, miR-27b, let-7a, let-7c, miR-199b, miR-26a, miR-335, miR-21, miR-22, miR-424, and miR-451) were expressed at a significantly higher level in AML compared with ALL (Fig. 1). miR-151* is the counterpart of miR-151, and both were derived from the same miRNA precursor; miR151* is thought to be the minor product of miRNA biogenesis. j-miR-5 is an miRNA identified by J.L., H.Z., and T.R.G. (unpublished data). Each differentially expressed miRNA has at least a 2-fold difference in expression between ALL and AML, along with a q value <0.01. The q value is a measure of the false discovery rate (FDR) (28). FDR is defined as the expected proportion of false positives among the significant genes (29). The overall FDR of the set of 27 differentially expressed miRNAs is $<1\%$, which means that any of the 27 miRNAs is unlikely to be false positive. MiR-128a (at 2q21), miR-128b (at 3p22), miR-223 (at Xq12), and let-7b (at 22q13) are the genes with the greatest expression difference, because each has >5-fold, and actually the first three have >10-fold, difference in expression between ALL and AML.

As expected, ALL and AML samples separate clearly in both HCA (Fig. 1) and PCA (data not shown) analyses using the expression profiles of the 27 differentially expressed miRNAs, with two samples [i.e., *MLL-ENL*-7 and t(4;11) cell line KOCL-48] as exceptions, which are AML but cluster together with ALL. *MLL-ENL*-7 is a biphenotypic leukemia, whereas KOCL-48 was derived from a patient who had been diagnosed as ALL-L2 but relapsed as AML-M4 and the cell line has biphenotypic features (M4-L2) (30). Thus, both samples had biphenotypic features, and it is not surprising they clustered together with ALL samples. The 27 miRNAs have a similar expression pattern between the cell lines and the relevant primary leukemia specimens in each lineage (Fig. 1), suggesting that the relevant critical regulatory pathways remain conserved in the cell lines despite numerous passages. The expression patterns of these 27 miRNAs in three normal controls were largely similar to those in the AML samples, except for miR-221, miR-222, miR-130a, and let-7e (Fig. 1).

As shown in both Figs. 1 and SI Fig. 3*a*, the clustering is not associated with the patients' gender, the blast cell percentages, or sample source [i.e., bone marrow (BM), peripheral blood (PB), or cell line (cl)]. Except for *MLL-ENL-7* and KOCL-48, there are seven AML patients (or patients from whom the relevant cell lines were derived) who were diagnosed before age 20, none of whom clustered together with ALL patients of similar ages, suggesting that clustering might be independent of the patients' ages. Clustering of cases with *MLL* rearrangement did not correlate completely with either *MLL* partner genes or French–American–British (FAB) subtypes; instead, they segregated according to the lineage from which they were derived.

Accurate Diagnosis of ALL and AML with Expression Signatures of a Minimal Number of Two miRNAs. We further used the prediction analysis of microarrays (PAM) method (31) to determine the minimal number of miRNAs that can be used to diagnose and discriminate ALL and AML cases accurately. Four miRNAs (i.e., miR-128a, miR-128b, let-7b, and miR-223) were the most discriminatory. As shown in Table 1, a combination of any two of these four miRNAs can discriminate ALL from AML cases with an overall diagnostic accuracy of 97–99%. Prediction accuracy was estimated by 10-fold cross-validation using PAM (see *Materials and Methods*). The detailed cross-validated probabilities are shown in SI Fig. 4. The sensitivity for ALL is 100%, whereas for AML, it is 96–98%; the specificity for ALL is 96–98%, whereas for AML, it is 100% (Table 1). The two samples, which are AML but were usually classified as ALL, are *MLL-ENL*-7 and KOCL-48. Given that both *MLL-ENL*-7 and KOCL-48 are biphenotypic, if we do not take them into account, the overall prediction accuracy, as well as the sensitivity and specificity of the remaining samples, probably can be considered as 100%. If using any three or all four miRNAs for prediction, a similar accuracy rate was obtained (Table 1). If using a single one from these four miRNAs for prediction, the overall accuracy varies from 84% to 94% (data not shown). A similar result was achieved by using a different supervised learning algorithm, namely Support Vector Machine (SVM) (32) (data not shown). Thus, as few as two of the most discriminatory miRNAs can diagnose and accurately discriminate all ALL from all AML cases, excluding the two biphenotypic samples.

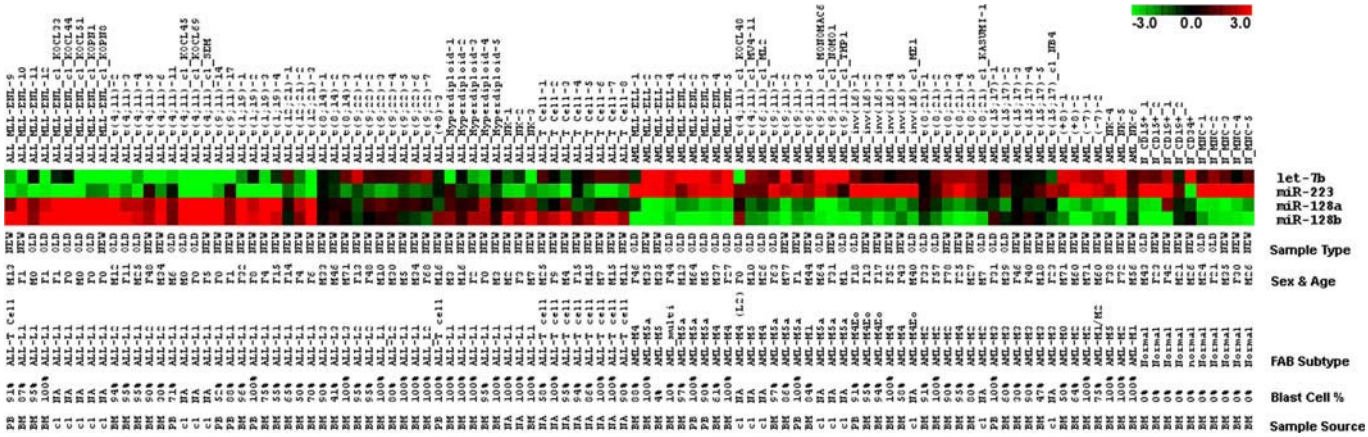

Fig. 2. Expression profiling of miR-128a, miR-128b, let-7b, and miR-223 in the 108 leukemic and normal samples as determined by TaqMan quantitative real-time PCR. Data are presented as ΔCt, which refers here to the difference in threshold cycles for a miRNA and U6 RNA. Expression data were mean-centered. Unsupervised average linkage hierarchical clustering was performed. Annotation information of this plot is similar to that in the legend of Fig. 1, in addition to the following: CD19+ means CD19+ lymphoblastic progenitor cells; OLD samples are those included in the bead-based miRNA expression assay, whereas NEW samples are independent samples not included in that assay.

The Differential Expression Pattern of the Four Most Discriminatory miRNAs Was Confirmed in a Large-Scale Real-Time PCR Validation Assay in 108 Leukemia and Normal Samples. We performed a TaqMan real-time PCR (33) assay to validate the differential expression pattern of miR-128a, miR-128b, let-7b, and miR-223 in 108 samples, including 98 leukemia samples (54 ALL and 44 AML) and 10 normal bone marrow controls (SI Table 2). The 10 normal bone marrow controls included two CD15+ (myeloid progenitor), two CD19+ (lymphoblastic progenitor), one CD34+ (hematopoietic progenitor), and five mononuclear cell (MNC) samples (see *Materials and Methods*). The 98 leukemia samples covered most common cytogenetic subtypes of acute leukemia (see Fig. 2). Among 54 ALL samples, 44 are B-cell ALL, including 19 *MLL* rearrangements, four *E2A-PBX1* (*TCF3-PBX1*)/t(1;19), three *TEL-AML1 (ETV6-RUNX1)*/t(12;21), three *MYC-IGH*/t(8;14), seven *BCR-ABL (BCR-ABL1)*/t(9;22), five hyperdiploidy/(>50 chromosomes), and three normal karyotype (NK) samples; the remaining 10 ALL samples are T cell ALL, including one *MLL* rearrangement [i.e., *MLL-ENL*/t(11;19)(q23;p13.3)], one trisomy 8 (i.e., +8), and eight with other abnormalities (for details, see SI Table 2). The 44 AML samples included 19 *MLL* rearrangements, six *CBFB-MYH11*/inv(16), six *AML1-ETO*/t(8;21), six *PML-RARA*/t(15;17), two trisomy 8 (i.e., +8), two monosomy 7 (i.e., −7), and three NK samples. Among the 108 samples, 73 (68%) were new independent samples that were not used for the bead-based expression assay, whereas the remaining 35 (32%) were included in the bead-based expression assay (see SI Table 2).

As shown in Fig. 2, miR-128a and -128b were expressed at a significantly higher level ($P < 10^{-4}$, t test; $q = 0$, SAM; 10,000 permutations) in the group of ALL samples compared with the group of AMLs together with normal controls, with a difference of expression on average >14- and 29-fold for miR-128a and -128b, respectively, between the two groups (i.e., "ALLs" vs. "AMLs + normal controls"). In contrast, miR-223 and let-7b were expressed at a significantly higher level (for miR-223: $P < 10^{-4}$, $q = 0$, and >28-fold; for let-7b: $P < 10^{-4}$, $q < 0.01$, and >6-fold; 10,000 permutations) in the AML and normal control samples compared with the ALL samples. The relevant differential expression patterns of these miRNAs existed in most of the samples we tested, with some samples such as those in t(9;22) ALLs, T cell ALLs, and t(15;17) AMLs as exceptions in which let-7b, miR-128a, and/or miR-128b exhibited an atypical expression pattern compared with that in the majority of ALL or AML samples. Whether the different expression patterns of these miRNAs reflect some influence of the abnormalities such as the *BCR-ABL* fusion in t(9;22) ALL samples needs to be defined further. The expression pattern of these miRNAs in the KOCL-48 cell line is closer to that in ALL rather than in AML samples (Fig. 2).

To determine whether the expression profiling of these miRNAs obtained using quantitative real-time PCR can accurately discriminate ALL from AML samples as well, we performed PAM analysis on these data. As shown in SI Table 3, using various combinations of these four miRNAs, the lineage was correctly predicted in 87% (85/98 with miR-128a and let-7b) to 96% (94/98 with miR-128a and -223). Clearly, miR-223 is the best, whereas let-7b is the worst discriminator among these four miRNAs. As expected, the biphenotypic cell line KOCL-48 was predicted as ALL by using all of the combinations except for that using miR-128a and -223. If we do not take KOCL-48 into account, the accuracy of prediction will be 96% (93/97) for the combination of miR-223 together with any of the other three miRNAs. The prediction accuracy was estimated by 10-fold cross-validation. The detailed cross-validated probabilities were shown in SI Fig. 5. The expression pattern of these four miRNAs in the 10 normal controls (i.e., two CD15+ myeloid progenitor, two CD19+ lymphoblastic progenitor, one CD34+ hematopoietic progenitor, and five MNC cell samples) is largely similar to that in the AML samples (Fig. 2), which is consistent with what we observed in bead-based miRNA expression assay (see Fig. 1). Because the normal controls used for the real-time PCR assay included both myeloid (i.e., CD15+) and lymphoid (i.e., CD19+) as well as two lineage mixture cells (i.e., MNC) and hematopoietic progenitor cells (i.e., CD34+), this result may suggest that these four miRNAs are deregulated mainly in ALL rather than in AML, relative to the normal controls. Taken together, this large-scale real-time PCR validation assay confirmed the differential expression pattern of the miRNAs observed in the bead-based miRNA expression assay, suggesting that expression signatures of as few as two discriminatory miRNAs could accurately discriminate ALL from AML samples using either bead-based or real-time PCR-based expression profiling platform.

Overexpression of miR-128 in ALL Is Not Associated with DNA Locus Amplification. Amplification of DNA copy number is a common mechanism by which genes achieve overexpression in cancers (34). To investigate whether the overexpression of miR-128a and -128b in ALL compared with AML and normal controls was associated with an amplification of their genomic loci (miR-128a at 2q21 and miR-128b at 3p22), we performed a TaqMan real-time PCR study

to determine DNA copy number of their genomic loci in 45 samples (13 ALL, 30 AML, and two normal control samples; see *Materials and Methods*). As shown in SI Fig. 6, there was no amplification of the genomic loci of miR-128a or -128b in ALL cases, compared with either normal controls or AML cases. Moreover, there was no deletion of genomic locus of miR-128a or -128b in AML cases, relative to either normal controls or ALL cases. Thus, the differential expression of miR-128 was not a consequence of alteration of genomic DNA copy number.

Differential Expression of miR-128 Is associated with Epigenetic Regulation. Epigenetic regulation such as DNA methylation and histone modification play critical roles in chromatin remodeling and general regulation of protein-coding gene expression in mammalian development and human cancer (35). Indeed, CpG-island methylation has been shown to be a possible mechanism for the down-regulation of miRNAs (36–38). Thus, we further investigated whether the differential expression of miR-128b between ALL and AML (and normal controls) is associated with CpG-island methylation. We determined the DNA methylation status of the CpG island region of miR-128b in 10 ALL, 14 AML, and three normal control samples using bisulfite genomic sequencing (see *Materials and Methods*). As shown in SI Fig. 7*a*, the CpG islands of miR-128b much less methylated in ALL, as compared with AML. Indeed, in the sequenced 20 CpG dinucleotides in the promoter region of miR-128b, the average methylation rate is <5% in all 10 ALL samples tested, whereas it is >10% in nine and >20% in five of the 14 AML samples tested (SI Fig. 7*a*). The average methylation rate of the ALL group (2.7%) is significantly lower (independent *t* test; two-tailed $P < 0.005$) than that (17.1%) of the AML group. The detailed methylation information for the 20 CpG sites of miR-128b in each sample was shown in SI Table 4. Although miR-128b was expressed at a similar level between normal controls and AML samples (Figs. 1 and 2), the average methylation rates in normal controls were lower than those in the majority of AML samples (SI Fig. 7*a*), suggesting that the expression of miR-128b might also be under regulation of some mechanism(s) other than methylation regulation. Nonetheless, in the analysis of 20 samples with both methylation and expression data for miR-128b (see SI Table 5), we observed that the expression level of the miRNA is significantly negatively correlated (two-tailed $P < 0.001$; $rs = -0.69$; Spearman's Rank Correlation test) with the degree of methylation of the CpG islands among the tested samples (see SI Fig. 7*b*). Thus, the differential expression of miR-128b, at least partly, is associated with epigenetic regulation, particularly, methylation regulation of the CpG islands in its promoter region.

Discussion

In the present study, we showed that in both unsupervised and supervised cluster analyses using expression profiles of miRNAs, ALL, and AML samples segregated clearly according to their lineage (SI Fig. 3). Among 27 miRNAs that were differentially expressed between ALL and AML, miR-128a, miR-128b, let-7b, and miR-223 were the most significant and discriminatory. MiR-128a and -128b were expressed at significantly higher levels in ALL, whereas miR-223 and let-7b were expressed at significantly higher levels in AML (Fig. 1). In accord with our findings, miR-223 was previously reported as a "myeloid gene" that plays a critical roles in myeloid functions and differentiation (39–41). Differential expression of miR-128a and -128b and let-7b in ALL compared with AML has not been reported. Using the expression signature of any two of these four miRNAs in the diagnosis of ALL and AML cases could result in an accuracy rate as high as 97–100% (Table 1). The differential expression pattern of the four most discriminatory miRNAs was confirmed further in a large-scale real-time PCR assay in 108 samples, which included most of the common subtypes of AML and B-cell ALL, as well as T cell, ALL (Fig. 2). Although the overall accuracy of predictions using the real-time PCR data is a little lower than that using the bead-based expression profiling data, which may be attributed to a larger number of samples with more types of genetic abnormalities, using expression signatures of as few as two miRNAs such as miR-223 and -128a or -128b could also result in prediction/diagnosis accuracy of >95% (SI Table 3 and SI Fig. 5). Notably, the *MLL* rearrangements are frequently found in both ALL and AML, but the lineage discriminatory miRNAs can accurately discriminate the relevant ALL samples from their counterpart AML samples even though they had exactly the same translocations/fusions such as t(11;19)(q23;p13.3)/*MLL-ENL* (see Figs. 1 and 2). Thus, our study provides several potential markers for the classification and diagnosis of ALL and AML.

It is of great interest to understand how the expression of the miRNAs is regulated in acute leukemia. Because most miRNAs, like protein-coding genes, are transcribed by RNA polymerase II (42), the regulation of miRNA expression is probably similar to that of protein-coding gene expression (24). In the present study, we found that overexpression of miR-128 in ALL compared with AML and normal control samples was not related to genomic locus amplification (SI Fig. 7). Instead, we observed that the degree of methylation of the CpG islands in the promoter of miR-128b is significantly lower in ALL samples than in AML samples, and that there was a significantly negative correlation between the expression level and the degree of methylation of the CpG islands of the miRNA (see SI Fig. 7). Therefore, the overexpression of miR-128 in ALLs compared with AMLs was at least partly associated with epigenetic regulation, particularly, hypomethylation of the CpG islands in the promoter region. Notably, although miR-128 was expressed at a significantly lower level in almost all AML and normal control samples than in ALL samples (Fig. 2), the degrees of methylation of miR-128 promoter in normal controls and in a subset of AML samples were almost similar to that in ALL samples (SI Fig. 7*a*), suggesting there is another mechanism controlling expression that will require additional analysis. Nonetheless, our results indicated that epigenetic regulation might be an important mechanism underlying regulation of miRNA expression in acute leukemia.

Although remarkable progress has been achieved in the past decades in the treatment and understanding of the biology of acute leukemias, with contemporary improved risk assessment, chemotherapy, hematopoietic stem cell transplantation, and supportive care, the 5-year overall survival rate of patients with AML is only ≈22%, which is much lower than that of patients with ALL (65%) (10, 11) (http://seer.cancer.gov). Indeed, ALL exhibits a better response to standard chemotherapy than AML in almost every age group. For instance, ≈98% of children and 85% of adults with ALL achieve complete remission and ≈80% of children and 30–40% of adults with ALL can be cured to date (4, 10, 43, 44). In contrast, ≈35–50% of children and 20–25% of adults with AML can be cured to date, although ≈80–90% of children and 50–70% of adults with AML achieve complete remission (10, 45–48). Among patients older than 65 years, the 5-year overall survival rate for AML is ≈4%, whereas for ALL, it is 9% (http://seer.cancer.gov). Therefore, accurately diagnosis of ALL and AML is very critical for the selection of the appropriate therapy. Although a combinational use of morphologic, immunohistochemical, and immunologic methods can already accurately diagnose these two types of acute leukemias (12), such methods require experienced personnel, and no single test is currently sufficient to establish the diagnosis. Our finding that expression signatures of as few as two miRNAs could accurately discriminate ALL from AML raises the possibility of using such lineage-discriminatory miRNAs to develop a rapid and accurate diagnostic test of ALL vs. AML in the future. Indeed, compared with the mRNA profiling that has been proposed for diagnosis of various cancers (49, 50), miRNA profiling has several advantages (25): (*i*) in contrast to mRNA expression, a modest number of miRNAs might be sufficient to classify human cancers; and (*ii*) unlike mRNAs, miRNAs remain largely intact in routinely col-

lected, formalin-fixed, paraffin-embedded clinical tissues (51). Nonetheless, further validation of these lineage-discriminatory miRNAs in large cohorts and in independent studies is required before clinical application becomes feasible.

Materials and Methods

See *SI Text* for more details.

Patient and Control Samples. The total number of samples we used for this study is 164, including 154 leukemic (136 patient samples and 18 cell lines) and 10 normal bone marrow controls [two CD15+, two CD19+, one CD34+, and five mononuclear cell (MNC) samples] (see SI Table 2). All patient samples were obtained at the time of diagnosis or relapse and with informed consent at the University of Chicago or other hospitals and were stored in liquid nitrogen until used.

Bead-Based miRNA Expression Profiling Assay. A total of 58 acute leukemia patient samples (11 ALL and 47 AML) and 14 cell lines (7 ALL and 7 AML), along with three normal bone marrow controls, were used in the bead-based expression assay as described (25). To control for data quality, only samples with total miRNA signals $\geq$15,000 were further analyzed. After normalization, only probes with maximum expression in all samples $\geq$7.25 were retained for further analyses. Probes specific for mouse or rat miRNAs were excluded. Values for each miRNA and for each sample were further mean-centered to minimize potential print-run-specific bias (52, 53). TIGR Mutiple Array Viewer software package (TMeV version 4.0) (54) was used to perform data analysis and visualize the results.

Prediction of ALL and AML. Class prediction analysis was performed by using PAM software in Excel Add-in (Version 2.1) (31). The method of the nearest shrunken centroids was used to identify a subgroup of genes that best characterizes a predefined class. The prediction accuracy was estimated by 10-fold cross-validation. We have also used the Support Vector Machine (SVM) program (32) in the TMeV version 4.0 package (54) for the prediction.

TaqMan Quantitative Real-Time PCR Assay of the miRNAs. We used a TaqMan real-time PCR method (33) to validate the differential expression patterns of four miRNAs (i.e., miR-128a, miR-128b, let-7b, and miR-223) in 54 ALL, 44 AML, and 10 normal bone marrow control samples. The miRNA TaqMan real-time PCR kit was purchased from Applied Biosystems, and PCRs were performed according to the manufacturer's recommendation.

Analysis of DNA Copy Number of the miR-128a and -128b Locus in Leukemia Cells. Following a method described by He *et al.* (20) with modification, we performed a TaqMan real-time PCR study to determine DNA copy number of the miR-128a and -128b in 13 ALL, 30 AML, and two normal control samples. The reported values represent the ratios of DNA copy number at the miR-128a or -128b locus to the normal reference probe (β-actin or 6p22).

Quantitative Bisulfite Sequencing Using Pyrosequencing Technology. Genomic DNA (1 μg) was first converted with sodium bisulfite (55). Then, the CpG island region of miR-128b promoter was PCR-amplified, and DNA methylation levels were analyzed by bisulfite genomic sequencing using the pyrosequencing technology (56).

We thank Y. Sato and K. Sugita for providing eight cell lines. This work was supported in part by the G. Harold and Leila Y. Mathers Charitable Foundation (J.C.), a Cancer Research Foundation Young Investigator Award (J.C.), a Leukemia and Lymphoma Society Translational Research Grant (J.D.R.), the Spastic Paralysis Foundation of Illinois, the Eastern Iowa Branch of Kiwanis International (J.D.R.), and Grant Public Health Service (PHS) CA40046 (to M.M.L.B. and R.A.L). T.R.G. is an Investigator of the Howard Hughes Medical Institute.

1. Rowley JD (2000) *Leukemia* 14:513–517.
2. Look AT (1997) *Science* 278:1059–1064.
3. Pui CH, Relling MV, Downing JR (2004) *N Engl J Med* 350:1535–1548.
4. Pui CH, Jeha S (2007) *Nat Rev Drug Discov* 6:149–165.
5. Rowley JD (2001) *Nat Rev Cancer* 1:245–250.
6. Haferlach T, Bacher U, Kern W, Schnittger S, Haferlach C (2007) *Ann Hematol* 86:311–327.
7. Deschler B, Lubbert M (2006) *Cancer* 107:2099–2107.
8. Armstrong SA, Look AT (2005) *J Clin Oncol* 23:6306–6315.
9. Rowley JD, Olney HJ (2002) *Genes Chromosomes Cancer* 33:331–345.
10. Pui CH, Schrappe M, Ribeiro RC, Niemeyer CM (2004) *Hematology Am Soc Hematol Educ Program* 118–45.
11. Randolph TR (2004) *Clin Lab Sci* 17:235–245.
12. Lowenberg B, Downing JR, Burnett A (1999) *N Engl J Med* 341:1051–1062.
13. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, *et al.* (1999) *Science* 286:531–537.
14. Ebert BL, Golub TR (2004) *Blood* 104:923–932.
15. Bullinger L, Valk PJ (2005) *J Clin Oncol* 23:6296–6305.
16. Haferlach T, Kohlmann A, Bacher U, Schnittger S, Haferlach C, Kern W (2007) *Br J Cancer* 96:535–540.
17. Bartel DP (2004) *Cell* 116:281–297.
18. Ambros V (2004) *Nature* 431:350–355.
19. He L, Hannon GJ (2004) *Nat Rev Genet* 5:522–531.
20. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, *et al.* (2005) *Nature* 435:828–833.
21. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ (2005) *Cell* 120:635–647.
22. Esquela-Kerscher A, Slack FJ (2006) *Nat Rev Cancer* 6:259–269.
23. Calin GA, Croce CM (2006) *Nat Rev Cancer* 6:857–866.
24. Wu W, Sun M, Zou GM, Chen J (2007) *Int J Cancer* 120:953–960.
25. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, *et al.* (2005) *Nature* 435:834–838.
26. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) *Proc Natl Acad Sci USA* 95:14863–14868.
27. Tusher VG, Tibshirani R, Chu G (2001) *Proc Natl Acad Sci USA* 98:5116–5121.
28. Storey JD, Tibshirani R (2003) *Proc Natl Acad Sci USA* 100:9440–9445.
29. Benjamini Y, Hochberg Y (1995) *J R Stat Soc Ser B* 57:289–300.
30. Iida S, Saito M, Okazaki T, Seto M, Yamamoto K, Akao Y, Ogura M, Suzuki H, Ariyoshi Y, Koike K, *et al.* (1992) *Leuk Res* 16:1155–1163.
31. Tibshirani R, Hastie T, Narasimhan B, Chu G (2002) *Proc Natl Acad Sci USA* 99:6567–6572.
32. Vapnik V (1998) *Statistical Learning Theory* (Wiley, New York).
33. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, *et al.* (2005) *Nucleic Acids Res* 33:e179.
34. Zhang L, Coukos G (2006) *Cell Cycle* 5:2216–2219.
35. Egger G, Liang G, Aparicio A, Jones PA (2004) *Nature* 429:457–463.
36. Scott GK, Mattie MD, Berger CE, Benz SC, Benz CC (2006) *Cancer Res* 66:1277–1281.
37. Saito Y, Jones PA (2006) *Cell Cycle* 5:2220–2222.
38. Saito Y, Liang G, Egger G, Friedman JM, Chuang JC, Coetzee GA, Jones PA (2006) *Cancer Cell* 9:435–443.
39. Chen CZ, Li L, Lodish HF, Bartel DP (2004) *Science* 303:83–86.
40. Fazi F, Rosa A, Fatica A, Gelmetti V, De Marchis ML, Nervi C, Bozzoni I (2005) *Cell* 123:819–831.
41. Fukao T, Fukuda Y, Kiga K, Sharif J, Hino K, Enomoto Y, Kawamura A, Nakamura K, Takeuchi T, Tanabe M (2007) *Cell* 129:617–631.
42. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN (2004) *EMBO J* 23:4051–4060.
43. Pui CH, Evans WE (2006) *N Engl J Med* 354:166–178.
44. Rowe JM, Buck G, Burnett AK, Chopra R, Wiernik PH, Richards SM, Lazarus HM, Franklin IM, Litzow MR, Ciobanu N, *et al.* (2005) *Blood* 106:3760–3767.
45. Ravindranath Y (2003) *Curr Opin Oncol* 15:23–35.
46. Tallman MS, Gilliland DG, Rowe JM (2005) *Blood* 106:1154–1163.
47. Morgan MA, Reuter CW (2006) *Ann Hematol* 85:139–163.
48. Stone RM (2007) *Exp Hematol* 35:163–166.
49. Bertucci F, Viens P, Tagett R, Nguyen C, Houlgatte R, Birnbaum D (2003) *Lab Invest* 83:305–316.
50. Sotiriou C, Piccart MJ (2007) *Nat Rev Cancer* 7:545–553.
51. Nelson PT, Baldwin DA, Scearce LM, Oberholtzer JC, Tobias JW, Mourelatos Z (2004) *Nat Methods* 1:155–161.
52. Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, O'Connell JX, Zhu S, Fero M, Sherlock G, Pollack JR, *et al.* (2002) *Lancet* 359:1301–1307.
53. Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, Dohner H, Pollack JR (2004) *N Engl J Med* 350:1605–1616.
54. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J (2006) *Methods Enzymol* 411:134–193.
55. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992) *Proc Natl Acad Sci USA* 89:1827–1831.
56. Yang AS, Estecio MR, Doshi K, Kondo Y, Tajara EH, Issa JP (2004) *Nucleic Acids Res* 32:e38.