Genome-wide identification of human microRNAs located in leukemia-associated genomic alterations

Daniel T Starczynowski\textsuperscript{1,2}, Ryan Morin\textsuperscript{1}, Andrew McPherson\textsuperscript{1}, Jeff Lam\textsuperscript{1,2}, Raj Chari\textsuperscript{1,2}, Joanna Wegrzyn\textsuperscript{1,2}, Florian Kuchenbauer\textsuperscript{1}, Martin Hirst\textsuperscript{1}, Kaoru Tohyama\textsuperscript{3}, R Keith Humphries\textsuperscript{1}, Wan L Lam\textsuperscript{1,2}, Marco Marra\textsuperscript{1}, Aly Karsan\textsuperscript{1,2}

\textsuperscript{1}British Columbia Cancer Agency Research Centre, 675 West 10\textsuperscript{th} Avenue, Vancouver, BC, Canada V5Z 1L3; \textsuperscript{2}Department of Pathology and Laboratory Medicine, University of British Columbia; \textsuperscript{3}Clinical Pathology and Laboratory Medicine, Kawasaki Medical School, Kurashiki, Okayama, Japan

Correspondence to:

Aly Karsan
British Columbia Cancer Agency Research Centre
675 West 10\textsuperscript{th} Avenue
Vancouver, BC V5Z 1L3
Canada

e-mail: akarsan@bccrc.ca
Tel: 604-675-8033
Fax: 604-675-8049
Abstract

Cytogenetic alterations, such as amplifications, deletions, or translocations, contribute to myeloid malignancies. MicroRNAs (miRNAs) have emerged as critical regulators of hematopoiesis and their aberrant expression has been associated with leukemia. Genomic regions containing sequence alterations and fragile sites in cancers are enriched with miRNAs, however the relevant miRNAs within these regions have not been evaluated on a global basis. Here we investigated miRNAs relevant to acute myeloid leukemia (AML) by; 1) mapping miRNAs within leukemia-associated genomic alterations in human AML cell lines by high-resolution genome arrays, and 2) evaluating absolute expression of these miRNAs by massively parallel small RNA sequencing. Seventy-seven percent (542/706) of miRNAs mapped to leukemia-associated copy-number alterations (CNA) in the cell lines, however, only 18% (99/542) of these miRNAs are expressed above background levels. As evidence that this subset of miRNAs is relevant to leukemia, we show that loss of two miRNAs identified in our analysis, miR-145 and miR-146a, results in leukemia in a mouse model. Small RNA sequencing identified 28 putative novel miRNAs, 18 of which map to leukemia-associated CNA. This detailed genomic and small RNA analysis of human leukemic cell lines points to a subset of miRNAs that may play a role in myeloid malignancies.
Introduction

Small noncoding RNAs (ncRNAs) are conserved and encoded in the genomes of invertebrates, vertebrates, and plants\(^1\). The largest subset of naturally occurring small RNAs are microRNAs (miRNAs). Mature miRNAs are 19 to 24 nucleotide transcripts processed from precursor hairpin intermediate RNAs by endonuclease-mediated reactions\(^1\). miRNAs have been implicated in critical hematopoietic processes and their deregulation is associated with leukemogenesis. Functional validation of deregulated miRNAs in hematopoiesis has been shown for several miRNAs\(^2\), however, direct evidence that leukemogenesis is specifically mediated by miRNAs is lacking. In one example, overexpression of miR-155 results in a fatal and aggressive myeloproliferative disorder in mice\(^3\). In other examples, knockout or overexpression of an miRNA have been shown to deregulate hematopoietic processes related to one or more steps of leukemogenesis\(^2\).

Hematological malignancies often exhibit genomic alterations. Balanced and unbalanced chromosome alterations are found in AML patients. Many miRNA expression studies have been performed to identify differentially expressed miRNAs between normal and leukemic samples\(^4,6\). These studies are extremely valuable, but have not examined the direct relationship between genome-wide alterations and the miRNAs within these alterations. Furthermore, these studies, which rely on the annotated sequence information about the miRNAs, are not capable of identifying novel or variant miRNAs.

Analysis of human and mouse genomes reveals that miRNA genes are frequently located at fragile sites and regions of copy number alteration (CNA) associated with cancer\(^7\). The current approximation states that 50% of miRNAs are within cancer-
associated genomic regions or in fragile sites. These findings suggest that a mechanism of miRNA deregulation in oncogenesis is due to genomic instability. Although miRNAs have been mapped to common leukemia-associated alterations by utilizing database searches, the potentially relevant miRNAs within these alterations have not been thoroughly investigated. In this study we addressed two main objectives: 1) identification of miRNAs that map to common leukemia-associated genomic alterations; and 2) identification of relevant miRNAs within the leukemia-associated genomic alterations. We investigated miRNAs relevant to leukemogenesis by first mapping miRNAs within common leukemia-associated genomic alterations in six human AML cell lines by high-resolution array comparative genomic hybridization (CGH). We then determined absolute expression of these miRNAs by massively parallel small RNA sequencing. Although 77% (542/706) of miRNAs mapped to leukemia-associated CNA in the 6 cell lines, only 18% (99/542) of these miRNAs are expressed at levels above background. In support of our conclusion, knockdown of miR-145 and miR-146a, two miRNAs mapping to a commonly deleted region in AML, results in a long-latency myeloid leukemia in mice. This study provides a detailed genomic and miRNA expression analysis of human leukemic cell lines to identify an enriched subset of leukemia-associated miRNAs, and facilitates selection of miRNAs to investigate in human leukemogenesis or normal hematopoiesis.
Material and Methods

Cell lines and CD34+ cells

Acute myeloid leukemic cell lines, KG-1, KG-1a, UT-7, HL-60, and THP-1 were purchased from ATCC. The myelodysplastic cell line, MDS-L, was generated in the lab of one of the coauthors (K.T.). CD34+ cells were positively selected from cryopreserved marrow or peripheral blood cells by immunomagnetic separation (Stem Cell Technologies).

Whole genome tiling-path aCGH analysis

Details of whole genome array construction and probe prime labeling and hybridization have been described previously8,9. The submegabase-resolution tiling set (SMRT) array contains 32,433 overlapping BAC-derived DNA segments that provide tiling coverage over the entire human genome with a theoretical resolution of 50-100 kb8. All clones were spotted in duplicate. Sample (cell line) and reference (normal diploid) genomic DNA (50-200 ng each) were separately labeled using Cyanine 3 and Cyanine 5 dCTPs fluorescence markers, respectively. The images were captured with a charge-coupled device camera, and analyzed using an ArrayWorx scanner and SoftWorx Tracker Spot Analysis software (Applied Precision). SeeGH custom software was used to visualize all data as log ratio plots10. Clones with standard deviation between duplicate spots of > 0.1 were filtered from the raw data. In order to avoid false-positives due to hybridization noise, a minimum of two overlapping consecutive clones showing change was required for a region to be considered altered. Breakpoints of genomic alterations were identified using a hidden Markov model algorithm, and verified by visual inspection.
Small RNA library preparation

Leukemia cell lines were individually harvested and RNA extracted with TRIzol (Invitrogen). The extracted RNA was subjected to miRNA library construction using the Illumina sequencing platform according to a published protocol\textsuperscript{11}. The sequencing library of small RNAs generated from human CD34\textsuperscript{+} bone marrow cells was obtained from a recently published analysis\textsuperscript{12}.

microRNA isolation and expression analysis

The small RNA fraction was isolated using the mirVana Paris Isolation kit (Ambion). miRNA expression was quantified using miRNA-specific stem-loop primers. The small RNA fraction was used for reverse transcription using reverse transcription primers specific for each independent miRNA. The cDNA synthesis reaction was subsequently used for quantitative PCR using miRNA-specific primers.

Annotation and prediction of novel miRNAs

The annotation procedure was performed as described but employed annotations from miRBase version 14 and the human genome (NCBI build 36.3). Novel miRNAs were predicted as previously described\textsuperscript{11}. Predicted target genes were identified using TargetScan (v5.1). For novel or edited miRNAs, the target genes were predicted by TargetScan Custom (v5.1). Binding energies for the miRNA to its mRNA targets were calculated with IntaRNA\textsuperscript{13}.

MicroRNA decoy retroviral vectors, packaging cell lines, and bone marrow transplantation

MiRNA decoy sequences (tandem repeats each complementary to miR-145 and miR-146a, respectively) were fused to the 3’UTR of the YFP cDNA and then cloned into the
dual promoter phosphoglycerate (pGK) retroviral vector as previously reported\textsuperscript{14}. Virus packaging and infection of ecotropic packaging cell lines (GP’E86) was performed as previously described\textsuperscript{14}. Marrow transplantation studies were carried out using protocols approved by the University of British Columbia Animal Care Committee.

**Peripheral blood and bone marrow analysis**

Donor-derived engraftment and reconstitution were monitored by flow cytometric analysis of YFP expression in the peripheral blood. For immunophenotypic analysis, bone marrow cells or peripheral blood were washed and resuspended in PBS containing 4\% goat serum, followed by primary mAb (PE- or APC-labeled) staining overnight. Details on antibodies are described elsewhere\textsuperscript{14}. Samples were run on a FACScalibur flow cytometer (Beckman Coulter), and data analyzed using FlowJo (version 8.7). Complete blood counts were performed on the peripheral blood using Scil Vet ABC Hematology Analyzer (Scil Animal Care Company). Blood counts were obtained at time of death. Organs and tissues were fixed in PBS with 4\% paraformaldehyde, paraffin-embedded, sectioned and stained with hematoxylin and eosin (H&E).

**Microarray Analysis of miRNAs**

The expression data analyzed in this study was accessed at ArrayExpress (www.ebi.ac.uk/microarray-as/ae). The accession numbers for controls (n = 11) and AML patient samples (n = 54, normal karyotype; n = 4, del(5q)) are E-TABM-970 and E-TABM-405, respectively. Analyses were performed using BRB-ArrayTools developed by Dr. Richard Simon and BRB-ArrayTools Development Team\textsuperscript{15}.

**Retroviral transduction of miR-145 and miR-146a into HL-60 myeloid leukemia cells.**
Amphophoenix cells were transfected with MSCV-IRES-GFP (MIG), MIG-miR-145, or MIG-miR-146a constructs. Constructs have been previously described\textsuperscript{14}. Virus was harvested and used to infect HL-60 cells. Transduced cell lines were sorted for GFP expression and cultured in DMEM/10% FBS.

**Apoptosis analysis**

Approximately $1 \times 10^5$ HL-60 cells were washed in PBS, and resuspended in AnnexinV-binding buffer (10 mM HEPES; 140 mM NaCl; 2.5 mM CaCl\textsubscript{2}; pH 7.4) and AnnexinV-conjugated antibody (1:20). Following a 15 min incubation, an additional 500 $\mu$l AnnexinV-binding buffer was added and analyzed by flow cytometry.

**Luciferase Assay**

hsa-miR-1, hsa-miR-1 (4A:G) mutant, hMEIS1-3’UTR, and hTP53-3’UTR were synthesized and cloned into pcDNA3.1 or pMirReport, respectively. A 95 bp fragment (134-233) of the MEIS1 3’UTR encompassing the miR-1 target site or 95 bp fragment (634-735) of the TP53 3’UTR encompassing the miR-1(4A:G) predicted target site were inserted downstream of the open reading frame of the luciferase reporter gene. For the 3’UTR-luciferase assays, 50 ng of pMirReport-3’UTR, 250 ng of pcDNA or pcDNA-miR, and 7.5 ng of Thymidine kinase-driven *Renilla* luciferase were cotransfected into HEK293 cells (24-well format) using TransiT transfection reagent (Mirus).

**5-azacytidine (5-aza) treatment of myeloid leukemia cell lines**

HL-60, KG-1a, and THP-1 cells were cultured in DMEM/10% FBS in 35 mm plates for 24 hours. Following plating, cells were treated with 10 $\mu$M 5-aza (company) or DMSO for 24 hours. Small RNAs were collected with miRVANA Paris (Ambion) isolation kit, and miR-145 and miR-146a were measured by quantitative PCR.
Statistics

Results are depicted as the mean ± standard error of the mean. Statistical analyses were performed using Student’s t-test. Comparison of survival between different groups was done by the Kaplan-Meier test and P value calculated by the log rank test. Box-and-whisker plots were used in Figure 5c to depict the range and percentiles of values obtained. The box indicates the 75th (top), 50th (middle line - median), and 25th (bottom) percentiles of the values obtained. GraphPad Prism (v4, GraphPad) was used for statistical analysis.
Results

Mapping of copy number alterations in human leukemic cell lines by whole genome high-resolution array CGH.

We obtained six human cell lines representing *de novo* and secondary acute myeloid malignancies: THP-1 (acute monocytic leukemia); UT-7 (acute megakaryoblastic leukemia); KG-1 (acute myeloid leukemia with myeloblast-like features); KG-1a (a subline of KG-1 that has acquired new karyotypic markers); MDSL (refractory anemia with excess blasts); and HL-60 (promyelocytic leukemia). These cell lines were chosen as models of common AML subtypes and for their large number of cytogenetic abnormalities16. To precisely define genomic breakpoints in these leukemic model cell lines, we utilized high-resolution aCGH8. Comprehensive genomic profiles of copy number gains and losses obtained from six leukemic cell lines are summarized in Supplemental Table 1 and Figure 1. The median number of genomic alterations per cell line is 22 (13 to 29). As expected, all cell lines exhibit genomic instability, including CNA of all or part of chromosomal arms (Figure 1 and Supplemental Table 1). UT-7 exhibited the most genomic alterations (29 CNA, and 1156 Mb of total alterations), while HL-60 had the least disrupted genome (429 Mb of total alterations) and KG-1 had the fewest distinct alterations (13 high level CNA). The most recurrent alterations include gains of chromosomes 1, 6q, 8, 9p, 11q, and 19, and deletions of chromosomes 5q, 7q, 12p, and 17p. Recurring submegabase genomic alterations were also detected and include a deletion at 12q13.11-q13.12 and a gain at 12p12.1-p11.23 (Figure 1 and Supplemental Table 1). Most of these CNA have been reported in AML patients and are thought to contribute to leukemic initiation or progression17. Application of high-resolution aCGH
provides a detailed and comprehensive analysis of genomic copy number alterations in
six model leukemic cell lines.

**Discovery of expressed microRNAs mapping to leukemia-associated genomic
alterations by massively parallel sequencing of small RNAs.**

After determining the breakpoints and regions of copy number alterations in the
leukemic cell lines, we next sought to identify miRNAs within the identified CNA. To
map miRNAs to the DNA CNA in our leukemic cell lines, we utilizing the most recent
human genome build (UCSC Human Mar. 2006) and miRNA database (miRBase 14.0)
(Supplemental Table 2). According to the current miRNA database, there are 706 human
miRNAs\(^{18}\). Of these, 542 (77\%) are found within leukemia-associated CNA in the six
cell lines we studied (see further discussion below and note Figure 3b).

To determine which of the 542 miRNAs are expressed and potentially relevant to
AML, we applied massively parallel sequencing of small RNAs to each of the six cell
lines. We selected this approach because all variants of mature miRNA transcripts can be
identified\(^{19}\). In contrast, hybridization-based miRNA arrays rely on annotated miRNA
sequences and therefore are not adequate to detect all miRNA variants. Small RNA
fractions were isolated from each cell line and prepared for sequencing on the Illumina
platform. Sequencing of small RNA libraries resulted in an average of 7.2 million reads
(range: 4.5 to 9.5 million) from each of the cell lines. Sequences were mapped according
to their overlap with available genomic annotations of miRNAs, resulting in an average
of 0.9 million filtered reads from each cell line. The miRNA reads within the small RNA
fraction of each cell line represent 40-90\% of all filtered reads (Supplemental Tables 3,4).
Because the cell lines exhibited numerous miRNAs with single reads, we decided to
primarily focus on the miRNAs above an arbitrary background expression level, with a miRNA frequency above 0.001 (1 read per 1000 total reads) (Figure 2). By this cutoff, approximately 98% of all miRNA reads and 20% of unique miRNAs are included in our analysis from each cell line (Figure 2). miRNAs with fewer reads (less than a frequency of 0.001) were excluded from further analysis.

To include miRNAs relevant only to myeloid leukemic cells, we used the cell lines as controls to cross-compare absolute miRNA levels within genomic alterations, rather than non-leukemic controls (Figures 3a). By this approach, abundantly expressed miRNAs within genomic alterations that are common and relevant only to myeloid leukemia cells are identified (Figure 3a). miRNAs mapping to the leukemia-associated genomic alterations (n = 542) (Supplemental Table 2) were next investigated for their absolute expression levels. Of these 542 miRNAs, 18% (n = 99) were expressed at levels above background (>1 per 1000 expressed tags) in at least one of the cell lines (Figure 3b, Table 1). Nearly an equal distribution of miRNAs were mapped within deleted (51%) and amplified (49%) CNA (Figure 3c,d). These findings suggest that, although many miRNAs map within CNA, only a subset are expressed at notable levels in myeloid leukemic cell lines. We next wanted to confirm the relevance of this subset of miRNAs in primary AML samples. We utilized published gene expression data on microRNAs isolated from CD34+ cells from 58 AML patients and 11 controls. Out of 99 expressed miRNAs, 43 were evaluated by microarray analysis, of which 72% (n = 31) were significantly differentially expressed in AML patients as compared to controls (Table 1).

It is possible that tumor suppressor miRNAs may be transcriptionally silenced by mechanisms other than genomic deletions, and unintentionally omitted in our analysis.
above. Therefore we evaluated miRNAs within deleted regions in the leukemic cell lines and compared their expression levels to nonleukemic human CD34+ cells. We utilized a recently published sequencing library of small RNAs generated from human CD34+ bone marrow cells12 (Supplemental Table 4). Of the miRNAs mapping to deleted regions, 29 are expressed in human CD34+ cells (Figure 4) and were further analyzed. We compared the expression of these 29 miRNAs between normal CD34+ cells and each leukemic cell line, and found that 65-86% of miRNAs that map to leukemia-associated deleted regions were expressed lower in the cell lines as compared to CD34+ cells. A similar pattern of miRNA repression was observed in primary AML samples (n = 58) when compared to control CD34+ cells (n = 11) (Figure 4). Collectively, these findings suggest that a subset of miRNAs within deleted regions in leukemic cells correlates with reduced expression (Figure 4). Interestingly, a few miRNAs (e.g., miR-378, miR-103-1, miR-25, and let-7f-2) are consistently overexpressed in the leukemic cell lines despite residing within a hemizygous deletion (Figure 4), potentially representing genes essential to cell maintenance.

MicroRNAs on chromosome 5q may be important in AML.

To further explore the role of miRNAs within deleted regions in myeloid malignancies we focused on miRNAs on chromosome 5 (Supplemental Figure 1). Deletion of chromosome 5q is the most common cytogenetic alteration in AML and MDS (Figure 1)20. To evaluate the potential impact of del (5q) on miRNA expression, we compared the relative tag counts for two leukemic cell lines, one with a deletion of chromosome 5q (KG-1a) and one diploid at chromosome 5q (THP-1). Impact of the chromosome 5q deletion on miRNA expression is evident by the lower expression levels
of miRNAs in KG-1a as compared to THP-1 (Figure 5a). As an example, miR-145 and miR-146a, two miRNAs within a commonly deleted region in del (5q) myeloid malignancies (Supplemental Figure 1), are expressed lower in cell lines with the chromosome 5q deletion or diploid at this locus as compared to CD34+ cells (Figures 4 and 5a). To independently confirm this observation, we performed quantitative PCR to access miR-145 and miR-146a expression levels in CD34+ (n = 3), THP-1, and KG-1a cells. As shown in Figure 5a (inset), miR-145 and miR-146a levels are highest in CD34+ cells and lower in leukemic cell lines with del (5q) or diploid (5q). Further, miR-145 and miR-146a are expressed significantly lower in diploid chromosome 5q leukemic cell lines as compared to CD34+ cells (P = 0.009 and P = 0.02, respectively; Figures 5a, inset). This suggests that miR-145 and/or miR-146a may be silenced in leukemia with a diploid chromosome 5q and further depressed in leukemias with del (5q). To investigate a potential epigenetic suppression (such as by DNA methylation) of miR-145 and miR-146a in AML, we treated 3 AML cell lines with a demethylating agent (5-azacytidine) and examined miRNA expression. As shown in Figure 5b, expression of miR-145 and miR-146a increase at least 3-fold following 5-azacytidine treatment of THP-1 and HL-60, but not in KG-1a. To further confirm the importance of these miRNAs in primary AML patient samples, we evaluated the expression of miR-145 and miR-146a in CD34+ cells isolated from AML patients with a normal karyotype (NK; n = 54) and with del(5q) (n = 4), and controls (n = 11) from previously published expression array data (Figure 5c). Expression of miR-145 is significantly reduced in AML patients with a normal karyotype as compared to controls (P = 0.019), but did not reach significance in the group with del(5q) (P = 0.14). Expression of miR-146a is significantly reduced in AML patients with
a normal karyotype \( (P = 0.0006) \) and further reduced in patients with del(5q) \( (P = 0.0055) \) as compared to control samples (Figure 5c). That we observe reduced expression of miRNAs that are associated with deleted regions in AML, even in cell lines or patient samples that may not harbor the specific deletion, suggests that this subset of miRNAs may play a leukemia suppressor role.

**Reduced expression of miR-145 and miR-146 results in acute myeloid leukemia in vivo, while re-expression of miR-145 or miR-146a suppresses AML cells in vitro.**

To determine whether miR-145 and miR-146a function as leukemic suppressors, restoration of their expression in AML cells harboring a del(5q) should therefore negatively impact survival and/or proliferation. We overexpressed miR-145 or miR-146a in HL-60 cells using a retroviral vector coexpressing GFP as a marker. HL-60 were chosen because miR-145 and miR-146a are both expressed at low levels (Figure 4) and are also potentially suppressed by epigenetic mechanisms (Figure 5b). Following infection of HL-60 cells and sorting for the transduced population, a noticeable impairment in growth and loss of transduced cells (GFP+) was observed in HL-60 cells expressing miR-145 or miR-146a (Supplemental Figure 2, and not shown). To determine whether miR-145 and miR-146a effect the survival of HL-60 cells, we measured AnnexinV staining. As shown in Figure 5c, approximately 40% and 50% of HL-60 cells transduced with miR-145 or miR-146a, respectively, stained positively for AnnexinV. In contrast, vector control cells or non-transduced parental cells grew robustly following infection and exhibited approximately 10% AnnexinV staining (Figure 5c). Re-expression of miR-145 or miR-146a inhibits survival and growth of leukemic cells, therefore implicating them in the pathogenesis of AML.
In a recent study, we showed the knockdown of miR-145 and miR-146a results in non-fatal myelodysplastic syndrome in vivo\textsuperscript{14}. Because MDS often precedes evolution to leukemia, we wanted to determine whether miR-145 and miR-146a contribute to an eventual acute myeloid leukemia by extending the analysis of mice for up to 14 months. Stably-knocked down miR-145 and miR-146a in mouse HSPC was achieved using retroviral-mediated overexpression of miR-145 and miR-146 target sequences (miR decoy) engineered into the 3’-UTR of YFP\textsuperscript{14}. Lethally-irradiated C57Bl/6 mice were transplanted with $7.5 \times 10^5$ marrow cells transduced with the miR decoy construct or vector alone, along with $2 \times 10^5$ wild-type competitor marrow cells. We found that starting at 8 months, approximately 30\% ($4/14$) of the miR decoy mice began to succumb to hematological myeloid malignancies ($P = 0.049$) (Figure 5e). Diseased mice displayed features of bone marrow failure, myeloproliferation, or acute leukemia, as supported by histological findings and blood counts (Figure 5d, and data not shown). Depending on the diseased mouse, evidence of hypercellular marrows, dys hematopoiesis, anemia, immature blast-like cells, and splenomegaly was observed (Figure 5f). To investigate previously defined targets of miR-145 and miR-146a in the mice, we measured protein expression of IRAK1 (miR-146a target), TRAF6 (miR-146a target), and Mal/TIRAP (miR-145 target). We confirmed by Western blot analysis that IRAK1, TRAF6, and TIRAP protein were, generally increased in miR decoy marrow cells prior to transplantation, after transplantation, and in a leukemic miR-decoy mouse (Supplemental Figure 3). Reduced expression of miR-145 and miR-146a, two leukemia-associated miRNAs identified in our analysis, results in myeloid diseases preceded by an MDS-like phenotype \textit{in vivo}.

\textbf{Novel miRNAs mapping to leukemia-associated genomic alterations}
Unclassified sequence reads represented 10-60% of the filtered libraries (Supplemental Table 3). To identify novel candidate miRNAs among the unclassified reads, we utilized available algorithms\textsuperscript{11}. Novel miRNAs from the six cell line libraries include 28 unique miRNA sequences (Table 2 and Supplemental Table 5). Five novel miRNAs were expressed at levels near the background cutoff (>0.001), including miR-478 and miR-486 (Supplemental Table 5). Since we are most interested in miRNAs that map to CNA, we determined which of the novel miRNA sequences are encoded from leukemia-associated alterations in the six cell lines. Eighteen (of 28) of the novel miRNAs are encoded from genomic regions of CNA, distributed evenly between deletions and amplifications (Figure 3d, Table 2, and Supplemental Table 5). For example, miR-468 and miR-484 both map to a region of chromosome 11q, a recurring region of amplification in our cell lines (Figure 1 and Table 2).

Identification of miRNA variants and isomiRs

Use of small RNA sequencing approaches allowed identification of miRNAs that exhibit variations in nucleotides as compared to their miRBase reference sequence. In the six libraries, mature mRNA variants (isomiRs) were found more abundant as compared to the miRBase reference sequence in some instances. For example, the most abundant sequence of miR-140-3p in the HL60 library did not match the miRBase reference sequence. The miRBase sequence (TACCACAGGGTAGAACCACGG) was expressed at low levels in this library, but instead an isomiR of miR-140-3p (ACCACAGGGTAGAACCACCGGAC) with a 5’ nucleotide deletion and 3’ nucleotide additions (underlined) was abundantly detected (913 vs 3 reads). Examination of the libraries revealed that eighteen highly expressed miRNAs are overrepresented by an
isomiR rather than the reference miRNA sequence (Table 3). Modifications among the
isomiRs included nucleotide 5’ and 3’ additions or deletions. The functional impact of the
various miRNA modifications has not been studied in depth, but theoretically could
impact the function of miRNAs as well as detection by standard real-time PCR
approaches.

Another less frequent subset of miRNAs are ones that show nucleotide
discrepancies of the mature miRNA, other than single nucleotide extensions or
eliminations. These nucleotide discrepancies are inconsistent with sequencing errors, but
instead likely represent post-transcriptional RNA editing of miRNA mature sequences.
We detected seven miRNAs undergoing adenine to guanine substitutions with a
frequency greater than 1% of the respective isomiR reads. Adenine to guanine (A:G)
substitutions are the most commonly identified nucleotide substitutions in our libraries,
potentially explained by spontaneous or enzymatic deamination of adenine to guanine. Of
the A:G edits, miR-423-5p represented 6% of all miR-423 isomiRs (Table 4). Six other
miRNAs (miR-1, miR-30d, let-7d, miR-103, miR-25*, and miR-24 contained edits in
more than 1% of their respective isomiR pools (Table 4). The majority of edits occurred
between positions 17 and 20. However, editing events observed in miR-1 were in the 4th
position, which includes the seed region (2-8 nucleotides) (Figure 6a). As shown in
Figure 6a, a substitution of an A to a G (position 4) changed the repertoire of predicted
miRNA targets.

To demonstrate that the editing event in the seed region of miR-1 (4A:G) alters
the binding properties of the miR-1, we measured miRNA binding using 3’-UTR
sequence fragments containing the predicted targets of miR-1 and miR-1(4A:G) inserted
downstream of a luciferase reporter. MEIS1, an AML-associated cofactor\textsuperscript{21}, is predicted to be a target of miR-1 but not of miR-1(4A:G) (Figure 6b). In contrast, a tumor suppressor associated with AML\textsuperscript{22,23}, TP53, is predicted to be a target of miR-1(4A:G) but not of miR-1 (Figure 6b). As expected, transient cotransfection of miR-1 or miR-1(4A:G) and MEIS1 3’-UTR into HEK293 cells yielded a 20% decrease in reporter activity by miR-1 but not miR-1(4A:G) (Figure 6c). Transient cotransfection of miR-1(4A:G) and TP53 3’-UTR into HEK293 cells yielded a 40% decrease ($P = 0.027$) in reporter activity, but surprisingly, miR-1 expression also suppressed luciferase activity of TP53 3’-UTR reporter but to a lesser extent (20%; $P = 0.89$). Further evaluation of the miRNA binding sites revealed that miR-1 is predicted to bind to 2 sites in the TP53 UTR albeit weakly (-3.7 and –2.9 kcal/mol), however, editing of A to G at position 4 increases the binding affinity of miR-1 to one these sites (-3.7 versus –10.4 kcal/mol) (Figure 6b).
Discussion

Cytogenetic abnormalities represent a large proportion of de novo and therapy-related AML patients. Identification of AML-relevant genes within cytogenetic alterations is the foremost objective and a daunting challenge to better understand AML. Previous work has shown that miRNAs are frequently located at cancer-associated fragile sites, and firmly implicated them as drivers of leukemogenesis. We performed an integrative analysis to identify relevant miRNAs located in leukemia-associated cytogenetic changes. We further corroborated our observations by examining relative expression of the leukemia-associated miRNAs in AML patient samples from published gene expression data (Table 1 and Figure 4). Collectively, our finding suggests that, although many miRNAs are located in regions of leukemia-associated cytogenetic changes (~70%), only a subset (~20%) of these miRNAs are expressed and likely relevant myeloid malignancies.

To determine whether our subset of miRNAs have been previously implicated in leukemogenesis or hematopoiesis, we searched PubMed for relevant findings. A search revealed that miR-143, miR-145, miR-146a, miR-155, miR-181, miR-221, and miR-222 are implicated in cellular processes relevant to AML and thus are consistent with our conclusion that the refined subset of leukemia-associated miRNAs are potentially important2 (Table 1). miRNAs within the commonly deleted region on chromosome 5q in MDS have recently been evaluated by our group14. Previously we have shown that deletion of miR-145 and miR-146a disrupts hematopoiesis resulting a phenotype similar to del (5q) MDS (e.g., 5q- syndrome) in mice14. Depletion of miR-145 and miR-146a also results in a clonal advantage and enhanced survival of hematopoietic stem/progenitor
cells. Further analysis revealed that a subset of these mice go onto develop myeloid
diseases consistent with leukemia (Figure 5); therefore, deletion of miR-145 and miR-
146a, two miRNAs identified in our analysis as potentially relevant to human leukemia,
results in a long-latency myeloid disease in mice. In addition, re-introduction of miR-145
or miR-146a into AML cells, significantly induced cell death (Figure 5d) and prevented
growth in vitro (Supplemental Figure 2). We also confirmed that miR-145 and miR-146a
are downregulated in AML patient samples and potentially suppressed by epigenetic
mechanisms (Figure 5b,c). Although miR-146a was down regulated in patients with a
normal karyotype and further suppressed in patients with del(5q), surprisingly miR-145
was significantly downregulated only in AML patients with a normal karyotype, but not
in AML patients with del(5q) (Figure 5c). This discrepancy may be explained by the
relatively small cohort of del(5q) patients or that low levels of miR-145 may be less
critical for maintaining the AML phenotype and more important for initiation of a
premalignant state conducive to development of AML. In support of the latter
hypothesis, we previously reported that miR-145 is significantly downregulated in MDS
patients with del(5q).

Use of a massively parallel sequencing platform revealed a proportion of
alternative miRNA species. Identification of miRNA sequence variants, miRNA*, and
post-transcriptional editing adds to the complexity of the miRNA transcriptome. The
relevance of alternative miRNA species is not precisely known but has also been recently
described in a leukemia progression mouse model. Our best understanding of the
impact of nucleotide modifications comes from studies on single nucleotide
polymorphisms (SNP) within miRNAs. These SNPs have been identified and are
associated with altered miRNA biogenesis, stem-loop formation, and strand preference\textsuperscript{24}. The functional consequence of nucleotide deletions and additions remains to be known. Although rare occurrences in our libraries, we did identify miRNAs that exhibited editing within sequences of the miRNA:mRNA duplex. miRNA editing has also been reported in mouse embryo, \textit{Oryza sativa}, and \textit{Arabidopsis thaliana} libraries\textsuperscript{25,26}. We have shown that an edit within the seed region of miR-1 (4A:G) may alter binding to mRNA targets. In our example, we specifically found that wild-type miR-1, but not the edited version, can bind and repress the 3’-UTR of MEIS1 (Figure 6b,c). This observation is particularly relevant to AML because MEIS1 is a key cofactor of the Hox cluster and critical to leukomogenesis in mice and humans\textsuperscript{21}. Not only did we show that the edited version of miR-1 does not bind MEIS1, but it now binds and suppresses the 3’UTR of TP53, a gene which is mutated in human AML and its loss induces AML in mice \textsuperscript{22,23}. Therefore, it is compelling to speculate editing of miR-1 at position 4 from an adenine to guanine makes miR-1 pro-leukemic by derepressing MEIS1 and simultaneously suppressing TP53.

Massively parallel sequencing also facilitated identification of novel miRNAs. Because this is the first reported attempt to perform massively parallel sequencing on human leukemia cells, we identified 28 novel miRNAs. Approximately 65\% of the novel miRNAs are located in leukemia-associated genomic alterations. Notably, miR-481, which is within a deleted region on chromosome 7q is predicted to target meningioma 1 (Mn1). Elevated expression of Mn1 is a predictor of poor outcome in AML patients, and expression of a Mn1 transgene in mouse hematopoietic cells results in aggressive and rapid leukemia\textsuperscript{27,28}. As such, deletion of miR-481 provides a compelling mechanism to
increase expression of Mn1 in AML. Collectively, we have delineated the genomic alterations and identified leukemia-associated miRNAs.
Acknowledgements

We thank Megan Fuller for animal husbandry. This work was supported by Canadian Institutes of Health Research (CIHR, MOP 89976) and Leukemia and Lymphoma Society of Canada grants to A.K., Genome Canada and CIHR grants to W.L.L., and CIHR and Michael Smith Foundation for Health Research (MSFHR) fellowships to D.T.S. A.K. is a Senior Scholar of the MSFHR.

Authorship Contributions

D.T.S. and A.K participated in designing the research and drafting the manuscript; D.T.S. J.L, and J.W. performed experiments; F.K. and R.K.H. provided technical assistance and assisted in the analysis of data; W.L and R.C. provided valuable reagents and expertise in array CGH; R.M., A.M., M.H. and M.M. performed analysis and provided expertise in small RNA sequencing on the Illumina platform; K.T. provided the MDS-L cell line.

Conflict of interest disclosure:

The authors declare no competing financial interests.
References

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>miR</th>
<th>bp start</th>
<th>bp end</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr 1</td>
<td>miR-30e</td>
<td>40683565</td>
<td>40683656</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>miR-30c-1</td>
<td>40686494</td>
<td>40686582</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-101-1</td>
<td>65296705</td>
<td>65296779</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>miR-186</td>
<td>71305902</td>
<td>71305987</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>miR-9-1</td>
<td>15465675</td>
<td>15465685</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>miR-181b-1</td>
<td>197094625</td>
<td>197094734</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>miR-181a-1</td>
<td>197094796</td>
<td>197094905</td>
<td>0.0084</td>
</tr>
<tr>
<td></td>
<td>miR-29c</td>
<td>206041820</td>
<td>206041907</td>
<td>na</td>
</tr>
<tr>
<td>Chr 3</td>
<td>miR-128-2</td>
<td>35760972</td>
<td>35761055</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>miR-26a-1</td>
<td>37985899</td>
<td>37985975</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>miR-138-1</td>
<td>44130708</td>
<td>44130806</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>miR-425</td>
<td>49032585</td>
<td>49033146</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>let-7g</td>
<td>52273734</td>
<td>52277417</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>miR-15b</td>
<td>161605070</td>
<td>161605167</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-16-2</td>
<td>161605227</td>
<td>161605307</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>miR-28</td>
<td>189689263</td>
<td>189689348</td>
<td>na</td>
</tr>
<tr>
<td>Chr 5</td>
<td>miR-9-2</td>
<td>87998427</td>
<td>87998513</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-886</td>
<td>135444076</td>
<td>135444196</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-143</td>
<td>148788674</td>
<td>148788779</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-378</td>
<td>149092581</td>
<td>149092646</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-146a</td>
<td>159844937</td>
<td>159845035</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>miR-103-1</td>
<td>167920487</td>
<td>167920548</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-340</td>
<td>179374909</td>
<td>179375003</td>
<td>0.0001</td>
</tr>
<tr>
<td>Chr 6</td>
<td>miR-30a</td>
<td>72169975</td>
<td>72170047</td>
<td>na</td>
</tr>
<tr>
<td>Chr 7</td>
<td>miR-148a</td>
<td>25956064</td>
<td>25956131</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>miR-25</td>
<td>99529119</td>
<td>99529202</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-93</td>
<td>99529327</td>
<td>99529406</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-106b</td>
<td>99529552</td>
<td>99529633</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-129-1</td>
<td>127635161</td>
<td>127635232</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-182</td>
<td>12997459</td>
<td>12997568</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>miR-183</td>
<td>129201981</td>
<td>129202090</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td>miR-29a</td>
<td>130212046</td>
<td>130212109</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-26b-1</td>
<td>130212758</td>
<td>130212838</td>
<td>na</td>
</tr>
<tr>
<td>Chr 8</td>
<td>miR-32a</td>
<td>22158420</td>
<td>22158501</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-486</td>
<td>41674116</td>
<td>41673183</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-30d</td>
<td>13586301</td>
<td>13586370</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>miR-151</td>
<td>141811845</td>
<td>141811934</td>
<td>0.057</td>
</tr>
<tr>
<td>Chr 9</td>
<td>miR-101-2</td>
<td>4802097</td>
<td>4804375</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>miR-199b</td>
<td>130046821</td>
<td>130046930</td>
<td>na</td>
</tr>
<tr>
<td>Chr 10</td>
<td>miR-146b</td>
<td>104186259</td>
<td>104186331</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>miR-130b</td>
<td>105144000</td>
<td>105144148</td>
<td>na</td>
</tr>
<tr>
<td>Chr 11</td>
<td>miR-130a</td>
<td>57165247</td>
<td>57165335</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>miR-192</td>
<td>64415185</td>
<td>64415294</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>miR-34c</td>
<td>110889374</td>
<td>110889450</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>let-7a-2</td>
<td>121522440</td>
<td>121522511</td>
<td>0.0001</td>
</tr>
<tr>
<td>Chr 12</td>
<td>miR-200c</td>
<td>6943123</td>
<td>6943190</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>miR-26a-2</td>
<td>56504659</td>
<td>56504742</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>let-7i</td>
<td>61283733</td>
<td>61283816</td>
<td>0.045</td>
</tr>
</tbody>
</table>

bp, base pair; as, antisense; na, not available

* P-value was generated from microarray data when comparing miRNA expression changes between AML patient (n = 58) and control CD34+ samples (n = 11).
<table>
<thead>
<tr>
<th>miRNA name</th>
<th>chromosome</th>
<th>bp start</th>
<th>bp end</th>
<th>deletion/amplification</th>
<th>mature miRNA sequence</th>
<th>Target Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-472</td>
<td>10</td>
<td>115041356</td>
<td>115041465</td>
<td>Del</td>
<td>AAATGAATCAGTTGGGCCTGT</td>
<td>ADIPOQ</td>
</tr>
<tr>
<td>miR-463</td>
<td>10</td>
<td>103351146</td>
<td>103351255</td>
<td>Del</td>
<td>AAGGGCCTCCTCTGAGGAC</td>
<td>SYNGR1</td>
</tr>
<tr>
<td>miR-431</td>
<td>13</td>
<td>49468528</td>
<td>49468637</td>
<td>N</td>
<td>ACAAAAAAAACGCACACCT</td>
<td></td>
</tr>
<tr>
<td>miR-442</td>
<td>16</td>
<td>14902849</td>
<td>14902958</td>
<td>N</td>
<td>AGAGGGGGTGAATTAAAGCG</td>
<td>SYNGAP1</td>
</tr>
<tr>
<td>miR-474</td>
<td>1</td>
<td>159683046</td>
<td>159683155</td>
<td>Amp</td>
<td>AGGCGUAACTTGAAGCGGAG</td>
<td>MEK3A</td>
</tr>
<tr>
<td>miR-460</td>
<td>4</td>
<td>153629916</td>
<td>153630025</td>
<td>Del</td>
<td>AGCTTTTGGAATTACGTTAG</td>
<td>PTPRT</td>
</tr>
<tr>
<td>miR-468</td>
<td>11</td>
<td>61032625</td>
<td>61032734</td>
<td>Amp</td>
<td>AGGGGGCGGGCTCCGCGC</td>
<td>ZNF385A, AAK1, NFIX</td>
</tr>
<tr>
<td>miR-478</td>
<td>2</td>
<td>96827730</td>
<td>96827839</td>
<td>N</td>
<td>ATCAGGGCTTGTGAATGGGAG</td>
<td>ADAM11</td>
</tr>
<tr>
<td>miR-432</td>
<td>19</td>
<td>6367420</td>
<td>6367529</td>
<td>Amp</td>
<td>CAGCCCGATCCCCAGCCCAC</td>
<td>FBX19</td>
</tr>
<tr>
<td>miR-457</td>
<td>16</td>
<td>2260666</td>
<td>2260775</td>
<td>N</td>
<td>CTCGTTGGGCTCTGCGACGCC</td>
<td>MN1</td>
</tr>
<tr>
<td>miR-442</td>
<td>3</td>
<td>6918077</td>
<td>69180887</td>
<td>Del</td>
<td>CGAATGATAGGCTAGGTCAT</td>
<td>CADM2</td>
</tr>
<tr>
<td>miR-422</td>
<td>15</td>
<td>63798621</td>
<td>63798730</td>
<td>Amp</td>
<td>GAAAGAATCTTGTACCTGCC</td>
<td>MSL21L, ERG, MPL, TRAF7</td>
</tr>
<tr>
<td>miR-425</td>
<td>22</td>
<td>18331270</td>
<td>18331379</td>
<td>Amp</td>
<td>GAGGCGATCGGACTGTGCTTC</td>
<td>RC3H1</td>
</tr>
<tr>
<td>miR-427</td>
<td>3</td>
<td>50687498</td>
<td>50687607</td>
<td>Del</td>
<td>GATGGCGGAGGCTGCGCCG</td>
<td>IMPAD1</td>
</tr>
<tr>
<td>miR-437</td>
<td>2</td>
<td>207365179</td>
<td>20736628</td>
<td>N</td>
<td>GCTGCACGGAGACTGGTAA</td>
<td>SP8, MAF, TET3</td>
</tr>
<tr>
<td>miR-486</td>
<td>22</td>
<td>29886043</td>
<td>29886152</td>
<td>Amp</td>
<td>GAGAGGAACCTTGGAGCTTCGG</td>
<td>SPTBN1, NFIX, NF2, ADAR</td>
</tr>
<tr>
<td>miR-418</td>
<td>19</td>
<td>12675394</td>
<td>12675503</td>
<td>Amp</td>
<td>GTATGCGTATCGTCGTAGGG</td>
<td></td>
</tr>
<tr>
<td>miR-416</td>
<td>2</td>
<td>12256683</td>
<td>12256792</td>
<td>Del</td>
<td>TACTGAGATGCAGCTGTGCG</td>
<td>ABCA5, JAK1</td>
</tr>
<tr>
<td>miR-428</td>
<td>8</td>
<td>96154304</td>
<td>96154413</td>
<td>Amp</td>
<td>TGAAGGAGATCGTGAGGTTTCG</td>
<td>LUZP1</td>
</tr>
<tr>
<td>miR-419</td>
<td>16</td>
<td>1724971</td>
<td>1725080</td>
<td>N</td>
<td>TGCACGGCACTGGGACACGT</td>
<td>RELT</td>
</tr>
<tr>
<td>miR-404</td>
<td>16</td>
<td>14912572</td>
<td>14912681</td>
<td>N</td>
<td>TGGGCGGAGCTTCCGCCAGGCC</td>
<td>TGRB1, NFIX</td>
</tr>
<tr>
<td>miR-408</td>
<td>16</td>
<td>84505724</td>
<td>84505833</td>
<td>Del</td>
<td>TGTGGTGAAGTCTAGGGCTG</td>
<td>FBX041</td>
</tr>
<tr>
<td>miR-481</td>
<td>7</td>
<td>29668666</td>
<td>29668675</td>
<td>Del</td>
<td>TGCTTTACTCCCTCGAGGACAT</td>
<td>CUGBP2, MN1</td>
</tr>
<tr>
<td>miR-450</td>
<td>9</td>
<td>136881134</td>
<td>136881243</td>
<td>N</td>
<td>TGTGATATCATGTTCTGGG</td>
<td>NLGN4X, ROCK1, ITCH</td>
</tr>
<tr>
<td>miR-402</td>
<td>9</td>
<td>96888098</td>
<td>96888207</td>
<td>N</td>
<td>TGGTCCCTGCTAAGCTGACAG</td>
<td>KLHL9, src</td>
</tr>
<tr>
<td>miR-421</td>
<td>19</td>
<td>12912277</td>
<td>12912386</td>
<td>Amp</td>
<td>TTGGAGGCTGAGGAGACATC</td>
<td>ACT1, SPARC</td>
</tr>
<tr>
<td>miR-414</td>
<td>19</td>
<td>764560</td>
<td>764669</td>
<td>N</td>
<td>TGGCAGATGGGCTGCGGGG</td>
<td>KRAS</td>
</tr>
</tbody>
</table>

bp, base pair; Del, deletion; Amp, amplification; N, copy number neutral

1Target Scan v5.1 was used to identify predicted target genes
Table 3. Abundantly expressed and overrepresented isomiRNAs

<table>
<thead>
<tr>
<th>Cell line</th>
<th>miRNA name</th>
<th>miRBase counts</th>
<th>miRBase sequence</th>
<th>isomiR sequence</th>
<th>isomiR counts</th>
<th>Ratio (isomR: miRbase)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL60</td>
<td>miR-107</td>
<td>1136</td>
<td>AGCAGCATTGTACAGGGCTATCA</td>
<td>AGCAGCATTGTACAGGGCTAT</td>
<td>3915</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>miR-181a</td>
<td>840</td>
<td>AACATTCAACGGCTGCAGGTAGT</td>
<td>AACATTCAACGGCTGCAGGTAGT</td>
<td>3481</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>miR-199a-3p</td>
<td>0</td>
<td>ACAGTACTCAGCACATTGGTTA</td>
<td>ACAGTACTCAGCACATTGGTT</td>
<td>2896</td>
<td>2896</td>
</tr>
<tr>
<td></td>
<td>miR-181b</td>
<td>296</td>
<td>AACATTCAACGGCTGCAGGTAGT</td>
<td>AACATTCAACGGCTGCAGGTAGT</td>
<td>2455</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>miR-140-3p</td>
<td>3</td>
<td>TACCACAGGTAGAACCACGG</td>
<td>ACCACAGGTAGAACCACGGAC</td>
<td>913</td>
<td>304.3</td>
</tr>
<tr>
<td></td>
<td>miR-101</td>
<td>74</td>
<td>TACAGTACTGATAACTGAA</td>
<td>TACAGTACTGATAACTGAA</td>
<td>624</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td>miR-222</td>
<td>24</td>
<td>AGCTACATCTGGCTACTGTTGTT</td>
<td>AGCTACATCTGGCTACTGTTGTT</td>
<td>391</td>
<td>16.3</td>
</tr>
<tr>
<td></td>
<td>miR-23a</td>
<td>79</td>
<td>ATCACATTGCCAGGGATTCC</td>
<td>ATCACATTGCCAGGGATTCCAC</td>
<td>295</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>miR-30d</td>
<td>7</td>
<td>TGTAACATCCCCGACTGGAAG</td>
<td>TGTAACATCCCCGACTGGAAGCT</td>
<td>193</td>
<td>27.6</td>
</tr>
<tr>
<td>THP1</td>
<td>miR-101</td>
<td>1859</td>
<td>TACAGTACTGATAACTGAA</td>
<td>TACAGTACTGATAACTGAA</td>
<td>7710</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>miR-140-3p</td>
<td>249</td>
<td>TACCACAGGTAGAACCACGG</td>
<td>TACCACAGGTAGAACCACGGAC</td>
<td>6943</td>
<td>27.9</td>
</tr>
<tr>
<td>UT-7</td>
<td>miR-140-3p</td>
<td>15</td>
<td>TACCACAGGTAGAACCACGG</td>
<td>TACCACAGGTAGAACCACGGAC</td>
<td>642</td>
<td>42.8</td>
</tr>
<tr>
<td></td>
<td>miR-101</td>
<td>10</td>
<td>TACAGTACTGATAACTGAA</td>
<td>TACAGTACTGATAACTGAA</td>
<td>72</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>miR-142-5p</td>
<td>23</td>
<td>CATAAAGTAGAAGACCTACT</td>
<td>CATAAAGTAGAAGACCTACTAA</td>
<td>68</td>
<td>3.0</td>
</tr>
<tr>
<td>MDSL</td>
<td>miR-183</td>
<td>133</td>
<td>TATGGCACTGGTAGAATTCACT</td>
<td>ATGGCACTGGTAGAATTCACTG</td>
<td>367</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>miR-140-3p</td>
<td>0</td>
<td>TACCACAGGTAGAACCACGG</td>
<td>TACCACAGGTAGAACCACGGAC</td>
<td>148</td>
<td>148</td>
</tr>
<tr>
<td>KG-1</td>
<td>miR-101</td>
<td>877</td>
<td>TACAGTACTGATAACTGAA</td>
<td>TACAGTACTGATAACTGAA</td>
<td>2001</td>
<td>2.3</td>
</tr>
<tr>
<td>KG-1a</td>
<td>miR-101</td>
<td>671</td>
<td>TACAGTACTGATAACTGAA</td>
<td>TACAGTACTGATAACTGAA</td>
<td>1573</td>
<td>2.3</td>
</tr>
</tbody>
</table>
Table 4. miRNA editing of adenine to guanine

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Cell line</th>
<th>Position</th>
<th>Count</th>
<th>Total</th>
<th>Ratio&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-1</td>
<td>UT7</td>
<td>4</td>
<td>23</td>
<td>2306</td>
<td>0.0100</td>
</tr>
<tr>
<td>miR-30d</td>
<td>KG1a</td>
<td>20</td>
<td>11</td>
<td>1077</td>
<td>0.0102</td>
</tr>
<tr>
<td>let-7d</td>
<td>UT7</td>
<td>1</td>
<td>12</td>
<td>999</td>
<td>0.0120</td>
</tr>
<tr>
<td>miR-103</td>
<td>UT7</td>
<td>20</td>
<td>15</td>
<td>1160</td>
<td>0.0129</td>
</tr>
<tr>
<td>miR-25*</td>
<td>KG1a</td>
<td>17</td>
<td>13</td>
<td>953</td>
<td>0.0136</td>
</tr>
<tr>
<td>miR-24</td>
<td>KG1a</td>
<td>18</td>
<td>32</td>
<td>1774</td>
<td>0.0180</td>
</tr>
<tr>
<td>miR-423-5p</td>
<td>KG1a</td>
<td>19</td>
<td>770</td>
<td>11628</td>
<td>0.0662</td>
</tr>
</tbody>
</table>

A:G edits occurring more than 1% of reads

<sup>1</sup> Ratio was calculated from the edited miRNA per all of the miRNA reads
Figure Legends

Figure 1. Summary of cytogenetic alterations in six human leukemia cell lines by aCGH. Whole genome frequency distribution of chromosomal alterations in 6 human leukemia cell lines as detected by array CGH and visualized by SeeGH. Lines to the right of the chromosome indicate gain of chromosomal material. Lines to the left of the chromosome indicate loss of chromosomal material. Line patterns correspond to the indicated cell line.

Figure 2. MicroRNA expression in six human leukemia cell lines.
(a-f) Proportion (in percent) of the most abundantly expressed miRNAs in 6 human leukemia cell lines: (a) THP-1, (b) UT-7, (c) KG-1, (d) HL-60, (e) MDSL, and (f) KG-1a. White portion of the pie chart represents miRNAs reads that are expressed below background.

Figure 3. miRNA mapping to copy number alterations.
(a) Schematic representation of the approach to identify miRNAs within copy number alterations. Each of the 6 cell lines were simultaneously analyzed for copy number alterations by aCGH and small RNA expression by massively parallel sequencing on the Illumina platform. The copy number alterations for the 6 cell lines were annotated (Table 1) and subsequently merged. Similarly, miRNA expression data from all 6 cell lines was used to identify expressed miRNAs that map to the copy number alterations (Table 2). By this approach, we are able to identify miRNAs that map to a homozygous deletion in one cell line by utilizing miRNA expression values from the pooled data. (b) Distribution summary of miRNAs mapping to leukemia-associated genomic alterations found in 6 human leukemia cell lines. According to the current miRNA database (miRBase v13.0),
there are 702 human miRNAs. Of these, 542 miRNAs map to copy number alterations identified in the 6 leukemia cell lines. Following small RNA sequencing, 99 miRNAs (of 542) are expressed above background and map to copy number alterations. (c-d) Proportion (in percent) of the 99 expressed miRNAs within each cell line mapping to genomic regions of amplifications or deletions.

**Figure 4. miRNA expression comparison to normal human CD34+ cells.**

Tag counts of miRNAs that encoded from deleted regions for each cell line are compared to normal CD34+ cells. Shown are the log expression levels for individual cell lines relative to CD34+ cells. The sequencing library of small RNAs generated from human CD34+ bone marrow cells was obtained from a recently published analysis12. miRNA expression levels from AML patient samples are shown from CD34+ marrow cells (n = 58) and shown relative to CD34+ cells evaluated from non-diseased individuals (n = 11). Data is adapted from published microarray data (E-TABM-970 and E-TABM-405).

**Figure 5. Reduced expression of miR-145 and miR-146a results in myelodysplastic/myeloproliferative features.**

(a) miRNA tag counts (log2) are shown for two cell lines, one with a chromosome 5q deletion (KG-1a) and one diploid at chromosome 5q (THP-1). (Inset) Relative levels of miR-145 and miR-146a were determined by qPCR in human CD34+ cells (n = 3), KG-1a, and THP-1 cells. (b) KG-1a, THP-1, and HL-60 cells were treated with 10 μM 5-azacytidine (5-aza) or DMSO for 48 hours. Expression of miR-145 and miR-146a was evaluated by qRT-PCR and shown relative to DMSO following normalization to 5S. (c) miR-145 (top) and miR-146a (bottom) expression levels are shown for CD34+ marrow cells isolated from AML patients with a normal karyotype (NK; n = 54) or deletion of
chr 5q (del 5q; n = 4). As controls, CD34+ cells were evaluated from non-diseased individuals (n = 11). Data is adapted from published microarray data (E-TABM-970 and E-TABM-405). Normalized expression levels are visualized as box-and-whisker plots.

(d) HL-60 cells were retrovirally transduced with empty vector (MIG), miR-145, or miR-146a. Transduced cells were isolated by FACS and analyzed for AnnexinV binding after 1 week in culture. Shown is a representative analysis from 2 independent transductions. Gating strategy is provided in Supplemental Figure 2. (e) Kaplan-Meier survival curves for mice reconstituted with marrow transduced with vector (n = 11) or miR-145/miR-146a decoy (n = 15) from three independent transplants. (f) Hematoxylin and eosin-stained femur (i) and spleen sections (ii) from a myeloproliferative-like diseased mouse. Wright-Giemsa-stained bone marrow cytospins (iii, v) and blood smears (iv) from a bone marrow failure mouse (iii, iv) and a leukemic mouse (v). Spleen image was obtained at time of death from a leukemic mouse (587 mg) and compared to a spleen from a control mouse (100 mg) (vi).

**Figure 6. Editing of miR-1 occurs within the seed region and impacts binding to mRNA targets.**

(a) Nucleotide modification at position 4 is shown for miR-1. An adenine to guanine (A:G) change was observed in UT-7, representing 23 of 2306 total miR-1 reads (Table 4). The impact of A:G edit at position 4 (4A:G) on mRNA target specificity is illustrated by comparing select miR-1 and miR-1 (4A:G) targets. The predicted targets for miR-1 (4A:G) are entirely different as compared to the unedited miR-1. (b) Shown is a sequence alignment of miR-1 (wild-type and edited) and its binding sites in the 3’UTR of MEIS1 (position 160-181) and TP53 (position 654-672). The miRNA-mRNA binding energy
score (kcal/mol) was determined for each pairing and shown on the right. *, indicates the binding energy score for a second miR-1 binding site in the TP53 UTR (position 689-698). (c) Luciferase activity was measured for MEIS1-UTR and TP53-UTR luciferase reporters in the presence of miR-1, miR-1 (4A:G), or vector control. Shown are relative values of 2 independent experiments performed in triplicate.
Figure 1. aCGH summary of 6 human cell lines.
Figure 2. Illumina sequencing of 6 human cell lines.

(a) THP1
(b) UT-7
(c) KG-1
(d) HL60
(e) MDSL
(f) KG-1a

miR-223
miR-23a
miR-320a
miR-320b
miR-320c
let-7a
let-7b
let-7c
let-7d
let-7e
let-7f
miR-21
miR-103
miR-128
miR-144
miR-150
miR-155
miR-16
miR-17
miR-17*+miR-106b
miR-130a
miR-191
miR-221
miR-222
miR-223a
miR-224
miR-225
let-7a
let-7b
let-7c
let-7d
let-7e
let-7f
miR-21
miR-103
miR-128
miR-144
miR-150
miR-155
miR-16
miR-17
miR-17*+miR-106b
miR-130a
miR-191
miR-221
miR-222
miR-223a
miR-224
miR-225
let-7a
let-7b
let-7c
let-7d
let-7e
let-7f
Figure 3. miRNA mapping to copy number alterations.
Figure 4. miRNA expression comparison to normal human CD34+ cells
Figure 5. Reduced expression of miR-145 and miR-146a results in myelodysplastic/myeloproliferative features.

(a) Log_{10}(fold change) of miRNA expression in CD34+ cells, THP-1 (N 5q), and KG-1a (del 5q).

(b) Relative miRNA expression in THP-1, KG-1a, and HL-60 cells treated with DMSO or 5-aza.

(c) Normalized expression of miR-145 and miR-146a in CD34+, NK, and del(5q) cells.

(d) Flow cytometry analysis of parental, MIG, MIG-miR-145, and MIG-miR-146a cells.

(e) Percent survival over 14 months for vector and miR decoy groups.

(f) Histological images of AML cells.
Figure 6. miR-1 editing within the seed region changes mRNA target specificity and impact binding to mRNA targets.

**a**

<table>
<thead>
<tr>
<th>miR-1</th>
<th>miR-1 (4A:G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGGAAUGUAAAGAAAGUAUGUAU</td>
<td>UGGgAUGUAAAGAAAGUAUGUAU</td>
</tr>
</tbody>
</table>

**Predicted miR-1 Targets**

- TNKS2: TRF1-interacting ankyrin-related ADP-ribose poly 2
- UST: uronyl-2-sulfotransferase
- SLC44A1: solute carrier family 44, member 1
- GJA1: gap junction protein, alpha 1, 43kDa
- PAX3: paired box 3
- MEIS1: myeloid ecotropic viral integration site 1

**Predicted miR-1 (4A:G) Targets**

- WDR42A: WD repeat domain 42A
- APLP1: amyloid beta (A4) precursor-like protein 1
- ATN1: atrophin 1
- BTBD7: BTB (POZ) domain containing 7
- C18orf25: chromosome 18 open reading frame 25
- TP53: tumor protein p53

**b**

- miR-1 (3’-5’): UAUGUAUGAAGAAUGUAAGGU
- MEIS UTR (160-181): CCCCAUGCAUACGUACAUCUCC
- miR-1 (4A:G) (3’-5’): UAUGUAUGAAGAAUGUAAGGu
- TP53 UTR (654-672): AAGGAAAUCUCACCCAUCCCCCC
- miR-1 (4A:G) (3’-5’): UAUGUAUGAAGAAUGUAAGGu

**Energy (kcal/mol):**

- miR-1 (3’-5’): -7.5
- MEIS UTR (160-181): -7.5
- miR-1 (4A:G) (3’-5’): -6.9
- TP53 UTR (654-672): -10.4
- miR-1 (4A:G) (3’-5’): -3.7, -2.9*

**c**

Bar graph showing relative luciferase activity:

- MEIS1-UTR: vector, miR-1, miR-1 (4A:G)
- TP53-UTR: vector, miR-1, miR-1 (4A:G)

Significance levels:

- 0.005
- 0.027
- 0.089
Genome-wide identification of human microRNAs located in leukemia-associated genomic alterations

Daniel T Starczynowski, Ryan Morin, Andrew McPherson, Jeff Lam, Raj Chari, Joanna Wegrzyn, Florian Kuchenbauer, Martin Hirst, Kaoru Tohyama, R Keith Humphries, Wan L Lam, Marco Marra and Aly Karsan

Information about reproducing this article in parts or in its entirety may be found online at: http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at: http://www.bloodjournal.org/site/misc/rights.xhtml#reprints

Information about subscriptions and ASH membership may be found online at: http://www.bloodjournal.org/site/subscriptions/index.xhtml

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include digital object identifier (DOIs) and date of initial publication.