Leukemia (2011) 25, 1389–1399

www.nature.com/leu

SPOTLIGHT REVIEW

Discovery of new microRNAs by small RNAome deep sequencing in childhood acute lymphoblastic leukemia

D Schotte[1,4], F Akbari Moqadam[1,4], EAM Lange-Turenhout[1], C Chen[2], WFJ van IJcken[3], R Pieters[1] and ML den Boer[1]

[1]Department of Pediatric Oncology and Hematology, Erasmus MC/Sophia Children's Hospital, Erasmus University Medical Center, Rotterdam, The Netherlands; [2]Genomic Assays R&D, Applied Biosystems, Foster City, CA, USA and [3]Erasmus Center for Biomics, Erasmus MC, Rotterdam, The Netherlands

MicroRNAs (miRNAs) relevant to acute lymphoblastic leukemia (ALL) in children are hypothesized to be largely unknown as most miRNAs have been identified in non-leukemic tissues. In order to discover these miRNAs, we applied high-throughput sequencing to pooled fractions of leukemic cells obtained from 89 pediatric cases covering seven well-defined genetic types of ALL and normal hematopoietic cells. This resulted into 78 million small RNA reads representing 554 known, 28 novel and 431 candidate novel *miR* genes. In all, 153 known, 16 novel and 170 candidate novel mature miRNAs and miRNA-star strands were only expressed in ALL, whereas 140 known, 2 novel and 82 candidate novel mature miRNAs and miRNA-star strands were unique to normal hematopoietic cells. Stem-loop reverse transcriptase (RT)-quantitative PCR analyses confirmed the differential expression of selected mature miRNAs in ALL types and normal cells. Expression of 14 new miRNAs inversely correlated with expression of predicted target genes ($-0.49 \leqslant$ Spearman's correlation coefficients (Rs) $\leqslant -0.27$, $P \leqslant 0.05$); among others, low levels of novel sol-miR-23 associated with high levels of its predicted (antiapoptotic) target *BCL2* (*B-cell lymphoma 2*) in precursor B-ALL (Rs -0.36, $P = 0.007$). The identification of >1000 *miR* genes expressed in different types of ALL forms a comprehensive repository for further functional studies that address the role of miRNAs in the biology of ALL.
Leukemia (2011) **25,** 1389–1399; doi:10.1038/leu.2011.105; published online 24 May 2011
Keywords: novel microRNAs; deep sequencing; pediatric ALL

Introduction

In the early nineties, microRNAs (miRNAs) of ~21-nucleotide length were discovered in the nematode *Caenorhabditis elegans*.[1,2] As miRNAs were initially regarded to be specific for the worm, it took geneticists till the next decade to recognize that this novel gene-regulatory mechanism was also part of humans.[1] The importance of this highly conserved, non-protein-coding class of small RNA became evident upon studies showing that they regulate the activity of many protein-coding genes.[3] These miRNA-targeted genes include tumor suppressors and oncogenes that are regulated by base pairing of the mature miRNA with the complementary mRNA. This results in mRNA cleavage, translational repression or deadenylation.[3,4]

MiRNAs are expressed in a tissue-specific fashion.[5] Hematopoietic cells display other miRNAs than other tissues. For example, miR-142 is mainly expressed in hematopoietic cells,[6] whereas miR-192, miR-194 and miR-215 are abundantly present in the gut[5] and miR-372 is highly characteristic for testis.[7,8] In leukemia, both lineage (for example myeloid and lymphoid) and genetic type specific (for example, t(8;21), *MLL*-rearranged and *TEL-AML1*-positive), miRNA signatures have been found by us and others.[5,9–13]

As miRNAs have tumor-suppressor and oncogenic capacity, the discovery of leukemia-related miRNAs may give insight into the biology of disease. Detection of differences in expression levels of miRNAs is limited by the knowledge of previously identified miRNAs, often discovered in nonleukemic tissues. These types of expression analyses may miss miRNAs being relevant for leukemia. Today, over 900 miRNAs (www.mirbase.org) have been discovered—mostly by cloning followed by conventional sequencing and computational prediction.[14] High-throughput or deep sequencing of small RNA fractions may result in the discovery of many more miRNAs, as this technique has increased sequencing depth over the conventional method, and current estimates predict as many as 1000 to 25 000 miRNAs present in humans.[15–17]

In this study we applied Solexa high-throughput sequencing[18] on small RNA fractions isolated out of 70 cases covering seven different types of pediatric acute lymphoblastic leukemia (ALL) and 19 cases covering three types of normal hematopoietic cells. This technique yielded ~8 million small RNA reads in each leukemia type and control group. Bioinformatic analysis revealed that these reads correspond to 554 known, 28 novel and 431 candidate novel *miR* genes that have not been previously published in miRBase (www.mirbase.org). A selection of 22 novel and candidate novel miRNAs were validated by stem-loop reverse transcriptase (RT) real-time quantitative PCR (stem-loop RT-qPCR). Of these miRNAs, 17 were differentially expressed between genetic types of ALL and normal cells. Moreover, expression levels of new miRNAs inversely correlated to mRNA expression levels of predicted target genes. This study provides expression signatures of known, novel and candidate novel miRNAs per subtype of leukemia that can serve as repository for further functional studies in the type of cells they were discovered.

Materials and methods

Patient samples

Bone marrow and peripheral blood was collected from children with newly diagnosed ALL. Mononuclear cells of these samples were isolated and enriched as previously described.[19,20] All leukemia samples contained ⩾90% leukemic cells as determined by May–Grünwald–Giemsa (Merck, Darmstadt,

Correspondence: Dr ML den Boer, Department of Pediatric Oncology and Hematology, Erasmus MC/Sophia Children's Hospital, Room Sp2456, P.O. Box 2060, Rotterdam 3000 CB, The Netherlands.
E-mail: m.l.denboer@erasmusmc.nl
[4]These authors contributed equally to this work.
Received 30 November 2010; revised 28 February 2011; accepted 18 March 2011; published online 24 May 2011

SPOTLIGHT

Table 1 Features of the identified sequence used to determine the confidence level of novel *miR* genes

A	*Good features for precursor hairpin*	*Bad features for precursor hairpin*
	Dicer/Drosha overhang 1–3 nt	Dicer/Drosha overhang <1 or >3 nt
	Minimal 10 reads in $\geqslant 1$ library	High variability in sequence of 5′ region of mature miRNA[b]
	Folding-energy of permutated precursor with $P \leqslant 0.01$[a]	Distance from miRNA to stem base <5 nt
	—	Encoded on >10 loci in the human genome
	—	Length of mature miRNA <21 or >22 nt
	—	Overlap with other small RNA (for example, tRNA or rRNA) of non-human species
	—	miRNA sequence forms part of loop sequence
	—	$>90\%$ GC bonds within mature miRNA

B *Confidence level*	*Good feature for precursor hairpin*	*Bad feature for precursor hairpin*
Novel	$\geqslant 2$	0
Candidate novel	1	0
Candidate novel	$\geqslant 2$	1
Other	$\leqslant 1$	$\geqslant 1$

Abbreviations: miRNA, microRNA; rRNA, ribosomal RNA; tRNA, transfer RNA.
[a]In each of 1000 runs the precursor sequence was permutated and energy needed for folding into a hairpin was calculated. *P*-value was based on the number of permutated precursor hairpins with a folding-energy equal or smaller than that of the original precursor hairpin out of 1000 iterations.
[b]Definition of high variability: the top read accounts for $<30\%$ of total sequences representing a unique miRNA. The top read sequence is defined as the sequence variant for a unique miRNA with the highest read frequency.
Confidence levels for new *miR* genes were defined as follows: *Novel*: precursor hairpin with no bad features (bad = 0) AND at least 2 good features (good feature $\geqslant 2$); *Candidate novel*: bad = 0 AND good = 1 OR bad = 1 AND good $\geqslant 2$; *Other*: other miRNAs not belonging to the novel or candidate novel class.

Germany)-stained cytospins. Flow cytometry was used to determine the immunophenotype (precursor B-ALL or T-ALL). Fluorescence *in situ* hybridization and RT-PCR techniques allowed the screening of the genetic type, and conventional karyotyping was used to analyze the ploidy status of ALL cases. For each ALL type, 10 patients were included, that is, *MLL*-rearranged, *TEL-AML1*-positive, *BCR-ABL*-positive, *E2A-PBX1*-positive, hyperdiploid (>50 chromosomes), B-other (negative for the five mentioned genetic aberrations) and T-ALL. To study miRNAs in normal hematopoietic counterparts, sorted fractions representing different stages of B-cell and T-cell differentiation are preferred. However, these sub-populations are rare and require large amounts of starting material. Alternatively, sucrose-gradient processed normal bone marrow ($n=10$ children), CD34-positive cells ($>90\%$ purity) sorted from granulocyte-colony-stimulating factor-stimulated blood samples of children suffering from a solid tumor without bone marrow involvement ($n=4$) and thymocytes extracted from thymic lobes resected from children during surgery for their congenital heart disease ($n=6$) were included.[11,21] All samples were collected after approval of the institutional review board and informed consent from parents or legal guardians.

High-throughput sequencing of miRNAs

Total RNA was extracted with TRIzol reagent (Invitrogen, Leek, The Netherlands) according to the manufacturer's protocol. The quality of RNA samples was measured on the 2100 Bioanalyzer (Agilent, Amstelveen, The Netherlands). Only RNA samples with RNA Integrity Number $\geqslant 7.5$ were used for further processing. Each ALL subtype-representing library was constructed using pooled RNA extracted from 10 individual patients. For each of the normal hematopoietic tissues, RNA of 3 (CD34+-sorted cells), 10 (normal bone marrow) and 6 (thymocytes) fractions was pooled. A total of 10 μg of each RNA library was size fractioned on a 15% Tris-borate-EDTA gel. Small RNAs of 18–30 nucleotides were excised out of gel and each fraction was amplified by RT-PCR using the Small RNA Sample Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. DNA concentration and size was checked on a 2100 Bioanalyzer. Next, 2 to 3 pM of small DNA was loaded onto a flow cell and these fractions were sequenced using the Illumina genome analyzer GAII based on Solexa sequencing technology. Raw sequences of small RNAs were computational analyzed to identify miRNAs as described below.

Computational analysis of small RNA sequences

Following trimming of adaptor sequences, small RNA sequence reads were mapped to the genome of human and other species. Sequences were aligned to Ensembl (release 56/GRCh37 assembly), UCSC (GRCh37/hg 19) and miRBase version 14. If the small RNA sequence read appeared to be non-protein coding, flanking sequences of 100 nucleotides on either side of the small RNA were extracted for further computational analysis. The potential miRNA precursors were then computationally folded into hairpin structures and tested for a set of features derived from known *miR* genes in order to identify putative novel miRNAs (Table 1 and Figure 1). The features used to identify miRNAs were based on experience in the identification of miRNAs by the previous work of Berezikov *et al*.[22–24] Details can also be found on the website www.internagenomics.com.[22–24] We have normalized the read frequencies of miRNAs by dividing the number of absolute read sequences (numerator) by the sum of total miRNA sequence reads (denominator) per subtype. The total sum of sequence reads represents the reads of known miRNAs, novel miRNAs and candidate novel miRNAs. Raw data sets can be found at the Short Read Archive (SRA) of the NCBI website (www.ncbi.nlm.nih.gov/sra, accession number SRP005294). Novel and candidate novel miRNA reads are designated as sol-miR (of Solexa) followed by a sequential number throughout this paper. For new sol-miRs that have been approved by MiRBase, the official

SPOTLIGHT

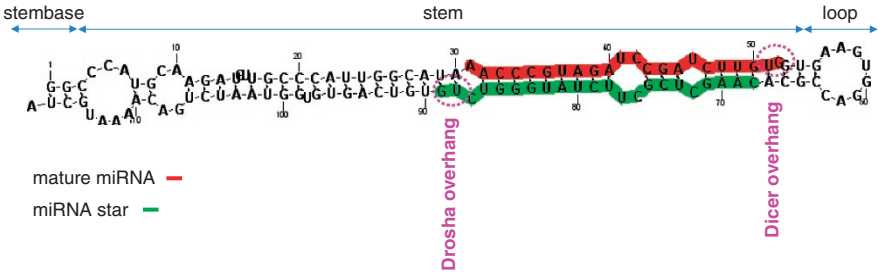

Figure 1 Parameters used to identify novel miRNAs. Small cloned RNA sequence reads are mapped to the human genome and the putative precursor is extracted by taking the miRNA sequence and the 100 nucleotides flanking sequence on either side of the miRNA. The precursor sequence is then computational folded into a hairpin structure, from which features are deduced as mentioned in Table 1 to identify the likelihood for a novel miRNA.

miRBase identification number has been added to Tables 2 and 3 and Supplementary Tables S2 and S5. For known miRNAs, the official miRBase identification number has been used.

Expression analysis of novel miRNAs and mRNA transcripts

The expression of miRNAs was validated by stem-loop RT-qPCR as described elsewhere.[11] Specific stem-loop RT-PCR primer and probe combinations were designed for 22 novel and candidate novel miRNAs by Applied Biosystems (Foster City, CA, USA).[25] Expression levels were validated in seven *MLL*-rearranged patients and eight cases of each of the following leukemia types: *TEL-AML1*-positive, *BCR-ABL*-positive, *E2A-PBX1*-positive, hyperdiploid, B-other and T-ALL cases. Expression levels in normal hematopoietic cells were determined in six normal bone marrow, four CD34+-sorted and six thymocyte fractions. Each RT-qPCR reaction was performed in duplicate with 5 ng of RNA as input.[11] Endogenous small nucleolar RNA 1 (RNU24) was used as reference for the RNA input as the expression of this reference showed limited variation among different types of ALL.[12] All RT-qPCR reactions were performed according to the manufacturer's instructions on an Applied Biosystems 7900HT system.

Previously published data sets of pediatric ALL cases (using Affymetrix U133 plus 2.0 GeneChips, Santa Clara, CA, USA) were used to determine the transcript levels of predicted target genes. This data set has been deposited at the NCBI's GEO[26] and is accessible through GEO series accession number GSE 13351. Data were extracted and normalized as described before.[27]

Statistics

Sequence reads were normalized by dividing the absolute read number per miRNA by the total number of miRNA (novel, candidate novel and known) reads per library to obtain a read frequency (%). Read frequencies and levels of miRNA expression were compared between types of ALL patients using the Mann–Whitney *U*-test. *P*-values were considered significant at $P \leqslant 0.05$ level (two tailed).

Results

MiRNAs identified by high-throughput sequencing in ALL and normal hematopoietic cells

High-throughput sequencing was performed on small RNA fractions representing seven types of pediatric ALL (T-ALL, *TEL-AML1*-positive, *MLL*-rearranged, *BCR-ABL*-positive, *E2A-PBX1*-positive, hyperdiploid and other precursor B-ALL patients negative for the major cytogenetic aberrations), and three normal tissues that is, normal bone marrow, CD34+-positive cells and thymocytes. The sequencing of these 10 small RNA libraries yielded a total number of 78 million sequence reads that entered a computational pipeline. This pipeline was used to distinguish miRNA sequences from other small RNAs. The criteria used to identify miRNA sequences and predict precursor hairpin structures were based on previous work by Berezikov *et al*.[22–24] In short, sequences were mapped to the human genome available in Ensembl (release 56/GRCh37 assembly) and UCSC (GRCh37/hg19). Sequences that did not map to protein-coding mRNA or to known small RNAs (including transfer RNA, ribosomal RNA, small nuclear and small nucleolar RNA) were further explored. Flanking sequences of 100 nt on both sides of a potential miRNA were retrieved from Ensembl (release 56/GRCh37 assembly). The small RNA and its flanking sequences were computationally folded into a hairpin structure, which was tested for a set of features that was previously used to define yet published and mainly evolutionary conserved *miR* genes (Table 1). The number of assigned good and bad features determined the likelihood (confidence level) of the small sequence to represent a true *miR* gene. The highest likelihood that a small RNA sequence represented a novel miRNA was obtained if the predicted precursor had at least two out of three good features and lacked bad features: these are depicted as novel miRNAs and fulfilled highly stringent criteria covering among others the length of mature miRNA, a minimal number of reads in at least one library, and predicted Drosha and Dicer cut sites (Table 1). Second in line are those predicted precursor sequences with one bad and at least two good features, as well as those without bad features but with one good feature: depicted as candidate novel miRNAs. All other hairpins may include potential (novel) miRNAs but do not fit these stringent criteria and are therefore called miRNA-other (Table 1).

Of the sequence reads, 80% were identified as known miRNAs, which indicates that the pre-fractioning of 18–30 nucleotide RNAs by gel electrophoresis is suitable to extract miRNAs from total RNA (Figure 2). In all, 14 866 reads (0.03%) represented novel and candidate novel miRNAs. Some reads (0.5%) were mapped to new locations on the genome while they encoded known miRNAs or their antisense transcript (complementary to the mature miRNA but encoded on the other DNA strand and therefore ruled out as star strand). These reads were categorized as homolog to known *miR* genes. Of the sequences, 2% was assigned as miRNA-other structures (lowest confi-

Table 2 Read frequency of 28 mature miRNAs and 12 miRNA-star forms (encoded by 28 novel *miR* genes) identified in different types of ALL and normal hematopoietic cells

Novel miRNA	*Annotation by miRBase*	*RT-qPCR*	*MLL* ($\% \times 10^{-4}$)	*TEL-AML1* ($\% \times 10^{-4}$)	*BCR-ABL* ($\% \times 10^{-4}$)	*E2A-PBX1* ($\% \times 10^{-4}$)	*Hyperdiploid* ($\% \times 10^{-4}$)	*B-other* ($\% \times 10^{-4}$)	*T-ALL* ($\% \times 10^{-4}$)	*CD34+* ($\% \times 10^{-4}$)	*nBM* ($\% \times 10^{-4}$)	*Thymocytes* ($\% \times 10^{-4}$)
Novel sol-miR-5	hsa-mir-3154	[a]	26.01	6.18	2.58	6.19	6.36	6.26	1.19	10.31	7.48	0.16
Novel sol-miR-6	hsa-mir-3150b	[a]	2.30	85.51	3.66	8.34	0.73	7.72	2.69	1.89	55.56	1.27
Novel sol-miR-6*	hsa-mir-3150b*		0	0.63	0.11	0.81	0.49	0	0	0	0	0
Novel sol-miR-11	hsa-mir-3136	[a]	3.65	9.22	2.05	5.38	0.98	4.94	5.69	1.26	6.73	2.69
Novel sol-miR-11*	hsa-mir-3136*		0	0.16	0	0	0	0.15	0.16	0	0	0
Novel sol-miR-13	hsa-mir-3117		0	39.55	0	0	0	0.62	0.47	0	0.19	2.85
Novel sol-miR-13*	hsa-mir-3117*		0	0.63	0	0	0	0	0	0	0	0
Novel sol-miR-14	hsa-mir-5187		2.97	2.50	1.51	2.15	0.98	2.01	1.58	5.04	13.09	1.11
Novel sol-miR-14*	hsa-mir-5187*		0	0	0	0	0	0	0.16	0	0	0.32
Novel sol-miR-15	hsa-mir-3151	[a]	9.86	1.41	4.63	0	1.96	1.08	0.95	6.92	0.75	0
Novel sol-miR-18	hsa-mir-3190	[a]	2.97	1.72	1.29	1.08	1.71	2.47	0.95	0.63	6.36	0.16
Novel sol-miR-18*	hsa-mir-3190*		0	0	0	0	0	0	0	0	0.19	0
Novel sol-miR-23	hsa-mir-4474	[a]	0.14	0.16	0.00	0	0	1.39	1.74	2.20	2.43	3.01
Novel sol-miR-24	hsa-mir-3177	[a]	0.54	2.50	0.32	0.27	0.24	1.85	0.63	0.16	3.18	1.27
Novel sol-miR-24*	hsa-mir-3177*		0	0	0.00	0	0	0.15	0	0	0	0
Novel sol-miR-27	hsa-mir-3942	[a]	0.68	0.63	0.22	0	0.24	0.77	2.06	0.79	1.50	2.53
Novel sol-miR-27*	hsa-mir-3942*		0.14	0	0	0.27	0	0.31	0.32	0	0	0
Novel sol-miR-35	hsa-mir-5186	[a]	0	2.50	0	0	0	0.00	0	0	0	0
Novel sol-miR-36	hsa-mir-5188		0	0	0	0	0	0.93	0.16	0	0.56	0
Novel sol-miR-37	hsa-mir-5006		0	0.16	0	0	0	0.15	1.11	0.63	1.87	0.79
Novel sol-miR-38	hsa-mir-5189		0	0	0.11	0	0	0.31	0.47	0.47	2.62	0.16
Novel sol-miR-39	hsa-mir-3183		0.41	0.31	0.11	0.54	0	0.31	0	0.47	1.50	0.32
Novel sol-miR-40	hsa-mir-5190		0.14	0.57	0.29	0.27	0	1.39	1.26	0	0	1.27
Novel sol-miR-41	hsa-mir-5191		0	0.16	1.08	1.08	0.86	1.54	0.32	0	0.19	0
Novel sol-miR-42	hsa-mir-5192		0.41	0.00	0.43	0.27	0	0.77	0.32	0	0	0
Novel sol-miR-43	hsa-mir-4774		0.81	1.25	0.22	0.81	0	1.08	0.47	0.47	0.94	0.47
Novel sol-miR-44	hsa-mir-5193		0	0.86	0	0	0	0.31	0.16	0.16	0.19	0
Novel sol-miR-45	hsa-mir-4637		0.41	0.16	0	0.54	0.49	0.62	0.47	0	0.37	0.47
Novel sol-miR-46	hsa-mir-3936		0	0.16	0.65	0.27	0.24	0.46	0.32	0.47	0.37	0.32
Novel sol-miR-47	hsa-mir-5194		1.22	0.78	0.11	0	0.49	0.15	0.47	0	1.12	0
Novel sol-miR-48	hsa-mir-5195		0.27	6.41	0.54	2.69	1.71	2.63	0	0	0.37	0
Novel sol-miR-48*	hsa-mir-5195*		0.14	0.78	0.22	1.35	0.49	0.93	0	0	0.19	0
Novel sol-miR-49	hsa-mir-5196		0	0	0	0	0	0.31	0	0	0	0
Novel sol-miR-49*	hsa-mir-5196*		0.27	0.99	0.43	1.35	0.73	0.51	0.47	0	0.06	0
Novel sol-miR-50	hsa-mir-5000		0.14	0.31	0	0	0	0.46	0.47	0	0.19	0.47
Novel sol-miR-50*	hsa-mir-5000*		0	0	0	0	0	0.15	0	0.16	0	0
Novel sol-miR-51	hsa-mir-3140		3.65	2.50	0.32	2.96	3.26	3.24	2.85	1.42	2.81	1.58
Novel sol-miR-51*	hsa-mir-3140*		0.14	0.31	0	0	0	0.15	0.16	0	0	0.16
Novel sol-miR-52	hsa-mir-5197		0.95	1.33	0	0.27	0.12	1.31	0	0	0	0
Novel sol-miR-52*	hsa-mir-5197*		0.54	1.25	0	0.81	0.73	0.77	0.16	0	0	0

Abbreviations: ALL, acute lymphoblastic leukemia; miRNA, microRNA; nBM, normal bone marrow; RT-qPCR, reverse transcriptase real-time quantitative PCR.
[a]Indicates the miRNAs of which the expression in ALL and control samples was validated by stem-loop RT-qPCR (Supplementary Figure S1).
Read frequency: the number encoding novel mature miRNAs and star miRNAs (miRNA*s) plotted as a percentage of total miRNA reads multiplied by a factor 10^4. Read frequencies that are shown have been normalized by dividing the absolute number of reads by the total number of miRNA reads for each specific subtype (total number equals the sum of known miRNA reads+novel miRNA reads+novel candidate miRNA reads).

Table 3 Top 10 of novel miRNAs with highest read frequency per type ALL

	Novel miRNA	*Annotation by miRbase*	*Read frequency % of total miRNAs × 10^{-4}*	*ALL type*
1	Novel sol-miR-5	hsa-mir-3154	26.0	
2	Novel sol-miR-15	hsa-mir-3151	9.9	
3	Novel sol-miR-11	hsa-mir-3136	3.6	
4	Novel sol-miR-51	hsa-mir-3140	3.6	
5	Novel sol-miR-14	hsa-mir-5187	3.0	***MLL***
6	Novel sol-miR-18	hsa-mir-3190	3.0	
7	Novel sol-miR-6	hsa-mir-3150b	2.3	
8	Novel sol-miR-47	hsa-mir-5194	1.2	
9	Novel sol-miR-52	hsa-mir-5197	0.9	
10	Novel sol-miR-43	hsa-mir-4774	0.8	
1	Novel sol-miR-6	hsa-mir-3150b	85.5	
2	Novel sol-miR-13	hsa-mir-3117	40.0	
3	Novel sol-miR-11	hsa-mir-3136	9.2	
4	Novel sol-miR-48	hsa-mir-5195	6.4	
5	Novel sol-miR-5	hsa-mir-3154	6.2	***TEL-AML1***
6	Novel sol-mir-51	hsa-mir-3140	2.5	
7	Novel sol-miR-14	hsa-mir-5187	2.5	
8	Novel sol-miR-24	hsa-mir-3177	2.5	
9	Novel sol-miR-35	hsa-mir-5186	2.5	
10	Novel sol-miR-18	hsa-mir-3190	1.7	
1	Novel sol-miR-15	hsa-mir-3151	4.6	
2	Novel sol-miR-6	hsa-mir-3150b	3.7	
3	Novel sol-miR-5	hsa-mir-3154	2.6	
4	Novel sol-miR-11	hsa-mir-3136	2.0	
5	Novel sol-miR-14	hsa-mir-5187	1.5	***BCR-ABL***
6	Novel sol-miR-18	hsa-mir-3190	1.3	
7	Novel sol-miR-41	hsa-mir-5191	1.1	
8	Novel sol-miR-46	hsa-mir-3936	0.6	
9	Novel sol-miR-48	hsa-mir-5195	0.5	
10	Novel sol-miR-49*	hsa-mir-5196*	0.4	
1	Novel sol-miR-6	hsa-mir-3150b	8.3	
2	Novel sol-miR-5	hsa-mir-3154	6.2	
3	Novel sol-miR-11	hsa-mir-3136	5.4	
4	Novel sol-miR-51	hsa-mir-3140	3.0	
5	Novel sol-miR-48	hsa-mir-5195	2.7	***E2A-PBX1***
6	Novel sol-miR-14	hsa-mir-5187	2.2	
7	Novel sol-miR-48*	hsa-mir-5195*	1.3	
8	Novel sol-miR-49*	hsa-mir-5196*	1.3	
9	Novel sol-miR-41	hsa-mir-5191	1.1	
10	Novel sol-miR-18	hsa-mir-3190	1.1	
1	Novel sol-miR-5	hsa-mir-3154	6.4	
2	Novel sol-miR-51	hsa-mir-3140	3.3	
3	Novel sol-miR-15	hsa-mir-3151	2.0	
4	Novel sol-miR-48	hsa-mir-5195	1.7	
5	Novel sol-miR-18	hsa-mir-3190	1.7	**Hyperdiploid**
6	Novel sol-miR-11	hsa-mir-3136	1.0	
7	Novel sol-miR-14	hsa-mir-5187	1.0	
8	Novel sol-miR-41	hsa-mir-5191	0.9	
9	Novel sol-miR-6	hsa-mir-3150b	0.7	
10	Novel sol-miR-49*	hsa-mir-5196*	0.7	
1	Novel sol-miR-11	hsa-mir-3136	5.7	
2	Novel sol-miR-51	hsa-mir-3140	2.8	
3	Novel sol-miR-6	hsa-mir-3150b	2.7	
4	Novel sol-miR-27	hsa-mir-3942	2.1	
5	Novel sol-miR-23	hsa-mir-4474	1.7	**T-ALL**
6	Novel sol-miR-14	hsa-mir-5187	1.6	
7	Novel sol-miR-40	hsa-mir-5190	1.3	
8	Novel sol-miR-5	hsa-mir-3154	1.2	
9	Novel sol-miR-37	hsa-mir-5006	1.1	
10	Novel sol-miR-18	hsa-mir-3190	0.9	

Abbreviations: ALL, acute lymphoblastic leukemia; miRNA, microRNA.
Ten novel miRNAs with the highest read frequency are shown for each ALL type. Read frequencies are presented as a percentage of the total number of identified miRNAs per ALL type multiplied by a factor 10^4.
*Indicates the miRNA star strand.

dence level). Other non-coding RNAs including transfer RNAs, ribosomal RNAs, small nuclear RNAs and small nucleolar RNAs represented 17% of all sequences.

In each ALL type, 5–10 million small RNA sequence reads were analyzed. Figure 3 summarizes the number of reads corresponding to known and newly identified miRNA sequences in different ALL types, for example, 9 246 296 out of 10 315 285 of small RNA reads in *BCR-ABL*-positive ALL patients were representing known miRNA sequences, whereas 194 and 2008 reads represented novel and candidate novel miRNAs, respectively. Strikingly, *TEL-AML1*-positive ALL patients displayed ~10-fold more reads of novel miRNAs than other types of ALL. In total, high-throughput sequencing revealed 895 unique miRNA sequences representing 470 known, 28 novel and 397 candidate novel mature miRNAs. Moreover, 372 known, 12 novel and 39 candidate novel miRNA-star forms (miRNA*) were detected (Figure 3). Both novel and candidate novel sequences have not yet been reported in miRBase version 14 (www.mirbase.org). These mature miRNAs were encoded by 554 known (Supplementary Table S1), 28 novel (Supplementary Table S2) and 431 candidate novel *miR* genes (Supplementary Table S3), respectively.

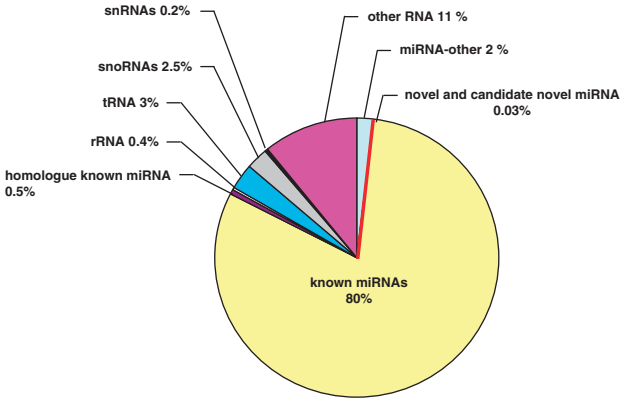

Figure 2 Composition of the small RNAome in leukemic cells of pediatric ALL patients. Frequency of non-coding small RNAs identified by high-throughput sequencing of ALL patient samples is shown. Frequencies of novel and candidate novel miRNAs are taken together. rRNA, ribosomal RNA; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA; tRNA, transfer RNA. 'Homolog to known' miRNAs refers to sequences that map to novel genomic loci of yet known miRNAs. The miRNA-other category reflects new miRNAs with lower confidence level than those in category 'novel' and 'candidate novel'. Other RNA category represents all other small RNAs not belonging to the categories mentioned above. See Table 1 for details of used features to determine the confidence level of predicted miRNAs.

The read frequency of the 28 novel mature miRNAs and 12 novel miRNA-star forms in ALL and normal hematopoietic cells is illustrated by Table 2 and Supplementary Table S2. The novel mature sol-miR-35 was detected in *TEL-AML1*-positive but not in other ALL types nor in normal hematopoietic cells, and therefore may be unique for this type of ALL (Table 2). Most miRNAs, however, are not restricted to one specific type of ALL (Table 2). In all, 4 out of 28 (14%) novel mature miRNAs and 6 out of 12 (50%) miRNA-star forms were detected in ALL but not in normal bone marrow and CD34+ fractions (Table 2 and Supplementary Table S5). Sol-miR-14* was exclusively found in T-lineage cells (T-ALL and thymocytes, Table 2). Six mature miRNAs and four miRNA-star forms were found in T-ALL but not in thymocytes and one mature miRNA (novel sol-miR-39) was present in thymocytes but not in T-ALL. Overall, 16 novel, 170 candidate novel and 153 known mature miRNAs/miRNA-star strands were uniquely found in ALL, whereas 2 novel, 82 candidate novel and 140 known mature miRNAs/miRNA-star forms were unique for normal hematopoietic counterparts (Supplementary Tables S5–S7). Hence, this indicates that miRNAs and miRNA-star forms are differentially expressed in ALL compared with normal hematopoietic cells.

Novel miRNAs with highest read frequencies are shown in Table 3 for each ALL type. Overall, the read frequency of novel miRNAs was tenfold lower than of known miRNAs (Figure 4a, $P<0.0001$). Novel miRNAs also differ in evolutionary conservation from yet reported miRNAs. Supplementary Tables

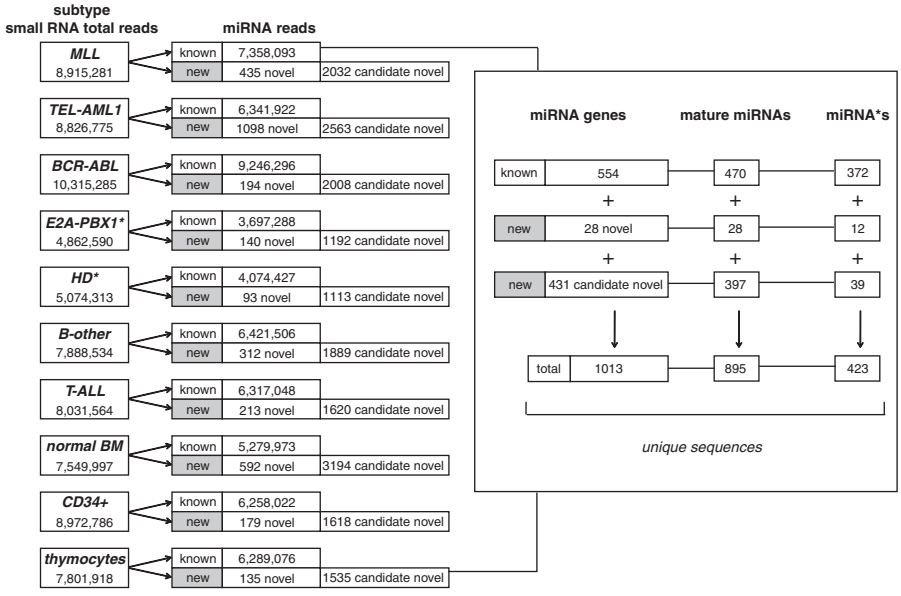

Figure 3 Overview of known and novel miRNAs in ALL types and normal hematopoietic cells. At the left, the total numbers of small RNA sequence reads are listed. At the right, the numbers of unique *miR* genes encoding mature miRNAs and miRNA-star forms (miRNA*s) are depicted. New miRNAs in both panels represent novel and candidate novel sequences (see Materials and methods section for explanation of used features). *For *E2A-PBX1* and HD (hyperdiploid), 2 pM was loaded, whereas for other libraries 3 pM was used.

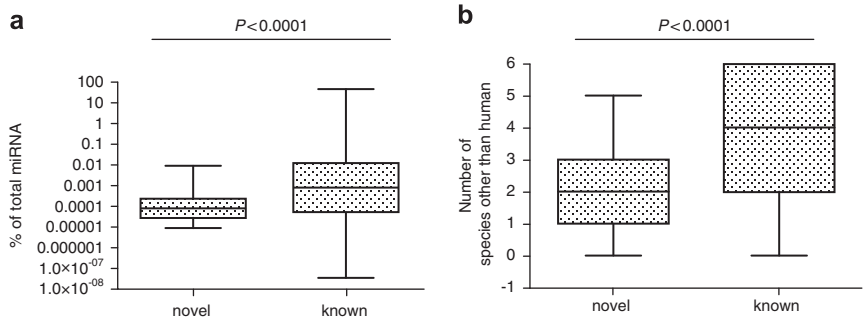

Figure 4 Abundance and evolutionary conservation of novel and known miRNAs in ALL. (**a**) Box plots represent the read frequency of novel and known miRNAs cloned from the seven different ALL types. Read frequency represents the percentage of total miRNA reads. (**b**) Box plots reflect the number of species other than humans in which novel and known miRNAs are present. Mann–Whitney *U-test* (MWU-test) was used to compare the abundance and number of non-human species for which the miRNA sequence can also be aligned.

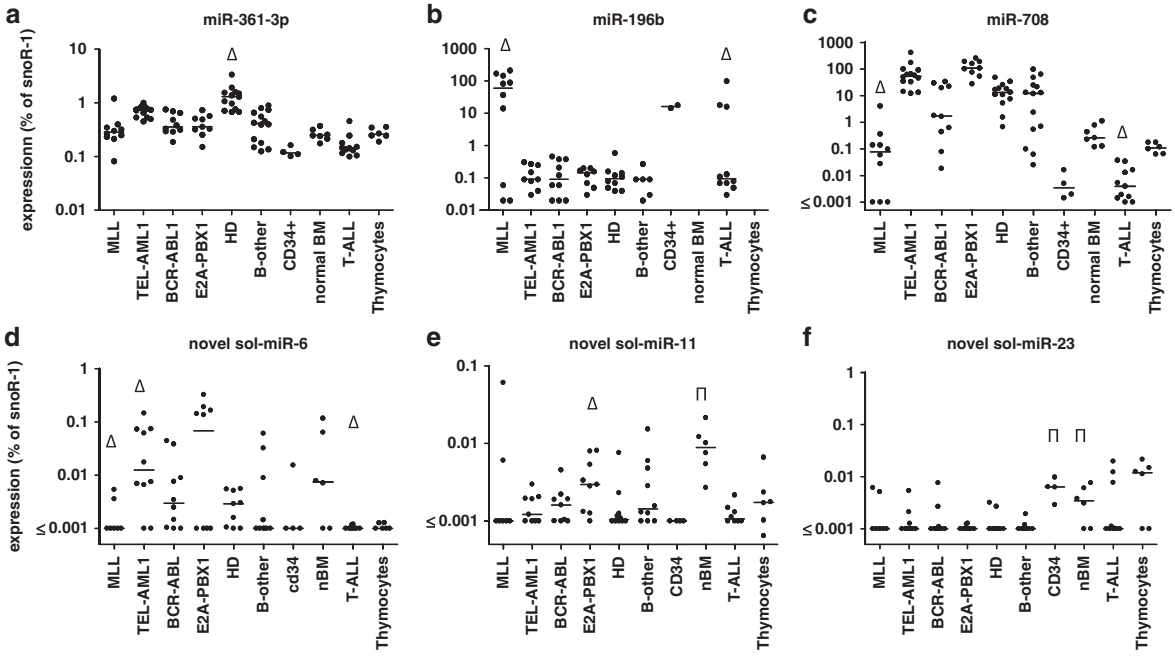

Figure 5 Expression levels of known and novel miRNAs in types of ALL and normal hematopoietic cells. Expression levels of miR-361-3p (**a**), miR-196b (**b**), miR-708 (**c**), novel sol (of Solexa)-miR-6 (**d**), novel sol-miR-11 (**e**) and novel sol-miR-23 (**f**) were measured by stem-loop RT-qPCR. Dots represent the expression level in individual patients as a percentage of the expression level of the endogenous reference snoRNA-1. Levels below 0.001% are undetermined ($C_t > 40$). The median in each group is indicated by a line. HD, hyperdiploid. $^{\Delta}P \leqslant 0.05$, based on the comparison between expression levels of the indicated type and the remaining ALL cases; $^{\Pi}P \leqslant 0.05$, based on comparison between normal hematopoietic cell populations and ALL.

S1 and S2 report the detection of novel and known miRNAs in different species. Whereas novel miRNAs are often present in two other species besides humans (for example, chimpanzee, macaque, mouse, rat or zebrafish), yet known miRNAs are frequently present in four of these species (Figure 4b, Supplementary Tables S1 and S2, $P<0.0001$).

Newly discovered miRNAs are aberrantly expressed in different types of ALL

Stem-loop RT-qPCR confirmed that known miRNAs with high read frequencies, for example, miR-361-3p, miR-196b and miR-708, were abundantly expressed in ALL. Moreover, similar to the read frequencies, these miRNAs were differentially expressed between genetic types of ALL (Figure 5a–c); miR-361-3p was threefold higher ($P<0.001$) expressed in hyperdiploid cases than in other precursor B-ALL cases, miR-196b was ~500-fold higher ($P<0.001$) expressed in *MLL*-rearranged compared with non-*MLL*-rearranged precursor B-ALL cases and miR-708 was ~300-fold and 3000-fold downregulated ($P \leqslant 0.002$) in *MLL*-rearranged and T-ALL cases, respectively. The expression level of 10 novel and 12 candidate novel miRNAs (Supplementary Tables S2 and S3) was also measured with stem-loop RT-qPCR (Figures 5d–f and Supplementary Figure S1). The expression levels of 5 out of 22 selected new miRNAs were below detection limits of real-time qPCR (that is, comparative cycle threshold (C_t) >40, corresponding to <0.001% of snoR-1 levels).

The remaining 17 (candidate) novel miRNAs were differentially expressed between ALL types (Supplementary Figure S1). Although novel sol-miR-6 had a relative low read frequency of 0.7×10^{-4} to 86×10^{-4}% of total miRNAs in ALL (Table 2),

SPOTLIGHT

Table 4 Correlation between the expression level of the newly identified miRNAs and the expression level of their predicted targets

Novel sol-miR-5	hsa-mir-3154	GTF2I Rs −0.46**	UBE3A Rs −0.45**	PPP1R12A Rs −0.45**	ZFAND5 Rs −0.43**	RAD23B Rs −0.37*
Novel sol-miR-6	hsa-mir-3150b	SIT1 Rs −0.36*	RLBP1 Rs −0.29*	PAX2 Rs −0.28*	CPLX2 Rs −0.27*	—
Novel sol-miR-11	hsa-mir-3136	MBNL1 Rs −0.29*	ARL1 Rs −0.27*	—	—	—
Novel sol-miR-14	hsa-mir-5187	—	—	—	—	—
Novel sol-miR-15	hsa-mir-3151	CAMK2G Rs −0.38**	ABCE1 Rs −0.36*	HNRNPA −0.32*	RNF125 Rs −0.32*	EIF5B Rs −0.31*
Novel sol-miR-18	hsa-mir-3190	AP3B1 Rs −0.49**	FGF12 Rs −0.37*	TMEM97 Rs −0.32*	NLK Rs −0.3*	ABCF2 Rs −0.3*
Novel sol-miR-23	hsa-mir-4474	NDRG4 Rs −0.4**	MAGED1 Rs −0.39**	BCL2 Rs −0.36**	SLC16A2 Rs −0.29*	IMPAD1 Rs −0.29*
Novel sol-miR-27	hsa-mir-3942	TNPO1 Rs −0.38**	CLTC Rs −0.33*	FBXW11 Rs −0.3*	APC Rs −0.29*	KPNA3 Rs −0.29*
Candidate novel sol-miR-9	—	ZNF576 Rs −0.33*	GOSR2 Rs −0.31*	RNF8 Rs −0.31*	WDTC1 Rs −0.3*	CASP2 Rs −0.29*
Candidate novel sol-miR-16	—	—	—	—	—	—
Candidate novel sol-miR-19	—	RPGRIP1L Rs −0.43**	VRK3 Rs −0.39**	HS2ST1 Rs −0.37**	CALML4 Rs −0.36**	TAOK3 Rs −0.32*
Candidate novel sol-miR-22	—	SRP72 Rs −0.33*	AAK1 Rs −0.29*	—	—	—
Candidate novel sol-miR-28	—	PHF15 Rs −0.45**	LRP4 Rs −0.38**	CACNA1A Rs −0.35**	PARD6B Rs −0.34*	BTBD3 Rs −0.3*
Candidate novel sol-miR-30	—	CCNL2 Rs −0.37**	CUL1 Rs −0.37*	BRCA2 Rs −0.33*	PBX1 Rs −0.32*	SNX4 Rs −0.31*
Candidate novel sol-miR-31	—	RNF4 Rs −0.32*	RANBP17 Rs −0.30*	HSPA9 Rs −0.27*	—	—
Candidate novel sol-miR-33	—	—	—	—	—	—
Candidate novel sol-miR-34	—	FUT9 Rs −0.43**	BCAT1 Rs −0.42**	LARP4 Rs −0.34*	PTPLAD1 Rs −0.33*	HNRNPC Rs −0.33*

Abbreviations: miRNA, microRNA; RT-qPCR, reverse transcriptase real-time quantitative PCR.
*P-value $\leqslant 0.05$.
**P-value $\leqslant 0.01$.
Target genes were predicted for novel and candidate novel miRNAs by Targetscan 5.1. MiRNA expression levels were determined by stem-loop RT-qPCR and compared with the expression of target genes as determined by Affymetrix U133 plus 2.0 GeneChips in the same patients (see Materials and methods section for more details). The table summarizes the Spearman's correlation coefficients (Rs) of mRNA expression levels of a maximum of top 5 predicted target genes that inversely correlated with the expression level of the indicated miRNA.
Annotation of novel miRNAs by miRBase is shown in the second column.

RT-qPCR analysis showed that this miRNA was median ninefold higher expressed in *TEL-AML1*-positive patients than in precursor B-ALL cases without this translocation ($P=0.02$, Figure 5d). Sol-miR-11 was threefold upregulated in *E2A-PBX1*-positive cases compared with other precursor B-ALL patients, whereas this miRNA was undetectable (that is, $<0.001\%$ of snoRNA-1 input) in *MLL*-rearranged cases ($P=0.02$, Figure 5e).

RT-qPCR analysis of novel miRNAs also confirmed the differential expression of miRNAs between ALL and normal hematopoietic cells: novel sol-miR-14 and sol-miR-23 were undetectable in most precursor B-ALL cases but expressed in normal bone marrow and CD34+ cells ($P\leqslant 0.02$, Figure 5f and Supplementary Figures S1-D). Similarly, sol-miR-30 was 4- to 17-fold lower expressed in precursor B-ALL than in CD34+-sorted cells and normal bone marrow ($P<0.001$, Supplementary Figure S1-N). In T-ALL, sol-miR-18 and sol-miR-16 were up to fivefold lower expressed than in healthy thymocytes ($P<0.01$, Supplementary Figures S1-F and -J).

As miRNAs may inhibit translation of proteins by cleaving mRNA, aberrant expression of new miRNAs may affect the expression level of their mRNA targets. Targetscan 5.1 (www.targetscan.org) was used to predict these targets based on their homology to the 2–8 nucleotide seed sequence of each miRNA. Table 4 shows that the expression levels of 14 out of total 17 differentially expressed (candidate) novel miRNAs inversely correlated with the mRNA levels of predicted target genes for these miRNAs ($-0.49 \leqslant$ Spearman's correlation coefficients (Rs) $\leqslant -0.27$, $P\leqslant 0.05$). Novel sol-miR-23 was predicted to target *BCL2* (*B-cell lymphoma 2*) and a lower expression level of this miRNA correlated with higher expression level of *BCL2* in patients (Rs −0.36, $P=0.007$, Table 4, Supplementary Figure S2).

Discussion

High-throughput Solexa deep sequencing followed by computational analyses identified 554 known, 28 novel and 431 candidate novel *miR* genes being expressed in seven different types of childhood ALL and three types of normal hematopoietic cells. Validation of selected miRNAs by stem-loop RT-qPCR confirmed aberrant expression patterns in subtypes of ALL and normal cells. Expression levels of 14 newly discovered miRNAs were inversely correlated to the transcript level of predicted target genes. This points to new miRNAs that may contribute to the biology of ALL and are therefore plausible candidates for more detailed functional studies.

MiRNA expression levels and function highly depend on the cellular context in which they are studied, including type of tissue, hematopoietic lineage and/or the presence of genomic translocations. For example, miR-221 and miR-222 are downregulated in erythroblastic leukemia but overexpressed in chronic lymphocytic leukemia.[28] MiR-221 and miR-222 inhibit growth of erythroblastic leukemia cells by targeting the oncogene *c-KIT*, suggesting a tumor-suppressor function for both miRNAs. However, the same miRNAs were reported to stimulate proliferation in thyroid and hepatocellular carcinomas through dowregulation of the tumor-suppressor genes *PTEN* and *p27*, implying that both miRNAs can also serve a oncogenic role.[28] As the function of miRNAs is cell-type dependent and most known miRNAs have been discovered in non-leukemic/non-hematopoietic cell types, we hypothesized that many miRNAs of interest to ALL are yet unknown. Because array-based expression techniques are limited to known miRNAs, we chose for sequencing of expressed miRNAs to address this hypothesis in well-defined types of pediatric ALL and normal hematopoietic cells. The contemporary deep sequencing technique enables simultaneous sequencing of millions of small RNA reads and is by far more sensitive to identify miRNAs than Sanger-based sequencing of conventional small RNA concatemer-cloning products like we and others previously used.[11,29–31] For example, 10 μg of total RNA input results in up to 10 million reads by high-throughput sequencing in contrast to ~1100 reads by conventional cloning methodology.[11] Therefore, high-throughput sequencing of small RNAs expressed in leukemic cells is currently the most sensitive approach to discover novel miRNAs that may be relevant to ALL.

In the present study we used well-defined and stringent criteria to define the confidence levels of identified miRNA sequences (Table 1 and Figure 1). In general, 16 novel, 170 candidate novel and 153 known mature miRNA/miRNA-star

strands were only expressed in ALL, whereas 2 novel, 82 candidate novel and 140 known mature miRNA/miRNA-star strands were unique for normal hematopoietic cells (Supplementary Tables S5–S7). The number of novel and candidate novel miRNAs identified in this study is in correspondence with other high-throughput studies that used similar stringency criteria in melanoma, ovarian tissues and acute myeloid cell lines.[32–34] In addition, high-throughput sequencing of two libraries compiled of 3 pediatric (genotypically not defined) ALL cases and 2 normal bone marrow samples revealed 42 novel miRNAs of which 5 were unique to ALL and 22 exclusively detected in normal donor bone marrow cells.[35] In a recent study, seven new miRNAs were cloned from AML with normal karyotype.[36] These studies suggest that the number of novel miRNAs being identified is not expected to drastically increase upon additional high-throughput sequencing analyses of similar samples unless the criteria for identification of miRNAs are being altered upon new scientific insights into the structure of miRNAs. The ultimate proof for a genuine miRNA is given by the experimental evidence that a (candidate) novel miRNA precursor is being processed into a mature miRNA by an active Dicer machinery.[37] The stem-loop RT-qPCR can be used for this purpose as it selectively detects expression levels of processed mature miRNAs.[25] In the present study we confirmed 17 out of 22 tested novel miRNAs as being genuine. The other five tested were below detection limits of stem-loop RT-qPCR, which can be indicative for a less efficient (stem-loop) primer design and/or the fact that a predicted miRNA is not a true miRNA.

In general, novel and candidate novel miRNAs were expressed at lower levels than known miRNAs (Figure 4). Despite this reduced average of expression, individual cases and/or specific subtypes can display relative high levels of newly identified miRNAs; for example, sol-miR-6 in a subset of *E2A-PBX1*-positive and *TEL-AML1*-positive ALL cases (Figure 5 and Supplementary Figure S1). The fact that novel and candidate novel miRNAs were not previously detected in other tissue types may point to *miR* genes being selectively expressed in genetic subtypes of leukemia and/or normal hematopoietic cells. A similar heterogeneity among patients was observed for the known miRNA-196b. Among *MLL*-rearranged and T-lineage ALL cases, miR-196b expression was specifically upregulated in cases with genetic lesions that affect HOXA-cluster gene activities that is an important leukemogenic event in these subtypes.[13]

In correspondence to the relative low expression levels, the read frequency of novel miRNAs was much lower than those of known miRNAs (Figure 4). The most abundantly expressed *miR* genes across all studied types of ALL include let-7 family members. These let-7 family miRNAs were also highly expressed in normal bone marrow cells, CD34+ hematopoietic precursor cells and thymocytes (see Supplementary Table S1). Also, in other tissues the let-7 family is abundantly expressed, suggesting a general, non-cell type-specific function of let-7 miRNAs in gene transcription.[5,38,39] The fine tuning may come from less abundantly but more cell type-specific miRNAs. Strikingly, we identified a higher frequency of sequence reads for novel miRNAs in *TEL-AML1*-positive patients compared with other precursor B-ALL types. This may suggest that miRNA-regulated gene expression is more active in *TEL-AML1*-positive patients compared with the other types of B-lineage ALL. Of particular interest are the known miR-125b, miR-126* and miR-383 (Supplementary Tables S4) and the newly identified sol-miR-6 in *TEL-AML1*-positive ALL, the latter being recently also found in ovarian tissue and that is now called miR-3150b (Tables 2 and 3).[32]

Based on the seed sequence of the newly identified miRNAs (for example, UGUGGCU for sol-miR-23), Target scan 5.1 was used to predict the target genes of these miRNAs (www.targetscan.org). The expression of 14 newly identified miRNAs negatively correlated with the expression of 61 target genes, which may point to the functionality of these newly identified miRNAs (Table 4). Novel sol-miR-23 (recently annotated by miRBase as hsa-mir-4474 (www.mirbase.org) was 6- and 10-fold lower expressed in precursor B-ALL than in normal bone marrow and CD34+-sorted cells, respectively (Figure 5f). A decrease in sol-miR-23 was linked to an increase of mRNA levels of its predicted target *BCL2* (Table 4 and Supplementary Figure S2). High expression level of *BCL2* is shown to have potential oncogenic effects at critical stages of differentiation[40] and may affect resistance to cytotoxic drugs.[41] Although high level of *BCL2* expression may not cause resistance to chemotherapy in leukemia,[42] *BCL2* expression may promote *BCR-ABL1*-dependent leukemogenesis.[43]

MiRNAs can reduce protein expression levels, both by mRNA degradation and by translational silencing (without need for mRNA degradation).[44] Therefore, a lack of an inverse correlation between miRNA and mRNA levels of predicted target genes does not exclude an inhibitory role for the miRNA in protein translation. To discover targeted genes, proteome-wide screens of altered protein levels upon miRNA manipulation are wishful, but unfortunately these technologies are still in its infancy. Alternative methods based on interaction between miRNA/mRNA and/or an RNA-induced silencing complex such as Ribonucleoprotein ImmunoPrecipitation–gene chip (RIP-chip)[45,46] and 3′-untranslated region-binding luciferase assays[47] may be informative. However, these techniques often yield false positive information because of artificial (binding) conditions in the experimental procedures.

In conclusion, high-throughput sequencing of 7 well-characterized ALL types and 3 normal hematopoietic cell fractions (representing 70 and 19 cases, respectively) resulted in the discovery of 28 novel and 431 candidate novel *miR* genes besides 554 yet described *miR* genes. Subsequent stem-loop RT-qPCR confirmed aberrant expression levels of newly discovered miRNAs in ALL types and normal hematopoietic cells. Hence, the presented data form a comprehensive basis for further functional studies in order to understand the role of miRNAs in pediatric ALL.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This research was financially supported by The Netherlands Organization for Scientific Research (NWO-Vidi Grant; MLdB), the Quality of Life Foundation (MLdB/RP) and the Pediatric Oncology Foundation Rotterdam, KOCR (MLdB/RP). We thank Christel Kockx and Michael Moorhouse (core-facility high-throughput sequencing, Erasmus Center for Biomics, Rotterdam) for assistance with the Solexa high-throughput sequencing reactions and data extraction, respectively. We are grateful to Dr Eugene Berezikov (InteRNA genomics BV and the Hubrecht Institute, Utrecht) for the bioinformatic analysis of the small RNA sequence reads and helpful discussions. We thank Dr Leila Mohammadi Khankahdani (Department of Pediatric Oncology, Erasmus MC/Sophia Children's Hospital) for her help in the integration of miRNA and mRNA expression data.

SPOTLIGHT

Author contributions

DS performed research, analyzed data and wrote manuscript; FAM and EAML-T performed research, analyzed data and revised the manuscript; CC provided custom stem-loop based TaqMan assays for miRNA detection and revised the manuscript; WFJvIJ performed Solexa sequencing and revised the manuscript; RP and MLdB supervised research, analyzed data, wrote and revised the manuscript.

References

1 Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004; **116**: 281–297.
2 Ventura A, Jacks T. MicroRNAs and cancer: short RNAs go a long way. *Cell* 2009; **136**: 586–591.
3 Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009; **136**: 215–233.
4 Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* 2009; **11**: 228–234.
5 Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D *et al.* MicroRNA expression profiles classify human cancers. *Nature* 2005; **435**: 834–838.
6 Chen CZ, Lodish HF. MicroRNAs as regulators of mammalian hematopoiesis. *Semin Immunol* 2005; **17**: 155–165.
7 Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, Zepeniuk M *et al.* MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol* 2008; **26**: 462–469.
8 Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, Nagel R *et al.* A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell* 2006; **124**: 1169–1181.
9 Jongen-Lavrencic M, Sun SM, Dijkstra MK, Valk PJ, Lowenberg B. MicroRNA expression profiling in relation to the genetic heterogeneity of acute myeloid leukemia. *Blood* 2008; **111**: 5078–5085.
10 Mi S, Lu J, Sun M, Li Z, Zhang H, Neilly MB *et al.* MicroRNA expression signatures accurately discriminate acute lymphoblastic leukemia from acute myeloid leukemia. *Proc Natl Acad Sci USA* 2007; **104**: 19971–19976.
11 Schotte D, Chau JC, Sylvester G, Liu G, Chen C, van der Velden VH *et al.* Identification of new microRNA genes and aberrant microRNA profiles in childhood acute lymphoblastic leukemia. *Leukemia* 2009; **23**: 313–322.
12 Schotte D, De Menezes RX, Akbari Moqadam F, Mohammadi Khankahdani L, Lange-Turenhout EA, Chen C *et al.* MicroRNAs characterize genetic diversity and drug resistance in pediatric acute lymphoblastic leukemia. *Haematologica* 2011; **96**: 703–711.
13 Schotte D, Lange-Turenhout EA, Stumpel DJ, Stam RW, Buijs-Gladdines JG, Meijerink JP *et al.* Expression of miR-196b is not exclusively MLL-driven but is especially linked to activation of HOXA genes in pediatric acute lymphoblastic leukemia. *Haematologica* 2010; **95**: 1675–1682.
14 Berezikov E, Cuppen E, Plasterk RH. Approaches to microRNA discovery. *Nat Genet* 2006; **38**(Suppl) : S2–S7.
15 Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O *et al.* Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* 2005; **37**: 766–770.
16 Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K *et al.* Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* 2005; **434**: 338–345.
17 Ahmed FE. Role of miRNA in carcinogenesis and biomarker selection: a methodological view. *Expert Rev Mol Diagn* 2007; **7**: 569–603.
18 Whiteford N, Skelly T, Curtis C, Ritchie ME, Lohr A, Zaranek AW *et al.* Swift: primary data analysis for the Illumina Solexa sequencing platform. *Bioinformatics* 2009; **25**: 2194–2199.
19 Stam RW, den Boer ML, Schneider P, Nollau P, Horstmann M, Beverloo HB *et al.* Targeting FLT3 in primary MLL-gene-rearranged infant acute lymphoblastic leukemia. *Blood* 2005; **106**: 2484–2490.
20 Den Boer ML, Harms DO, Pieters R, Kazemier KM, Gobel U, Korholz D *et al.* Patient stratification based on prednisolone-vincristine-asparaginase resistance profiles in children with acute lymphoblastic leukemia. *J Clin Oncol* 2003; **21**: 3262–3268.
21 Weerkamp F, de Haas EF, Naber BA, Comans-Bitter WM, Bogers AJ, van Dongen JJ *et al.* Age-related changes in the cellular composition of the thymus in children. *J Allergy Clin Immunol* 2005; **115**: 834–840.
22 Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* 2005; **120**: 21–24.
23 Berezikov E, Liu N, Flynt AS, Hodges E, Rooks M, Hannon GJ *et al.* Evolutionary flux of canonical microRNAs and mirtrons in Drosophila. *Nat Genet* 2010; **42**: 6–9; author reply -10.
24 Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E *et al.* Diversity of microRNAs in human and chimpanzee brain. *Nat Genet* 2006; **38**: 1375–1377.
25 Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT *et al.* Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res* 2005; **33**: e179.
26 Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002; **30**: 207–210.
27 Den Boer ML, van Slegtenhorst M, De Menezes RX, Cheok MH, Buijs-Gladdines JG, Peters ST *et al.* A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *Lancet Oncol* 2009; **10**: 125–134.
28 Croce CM. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* 2009; **10**: 704–714.
29 Basso K, Sumazin P, Morozov P, Schneider C, Maute RL, Kitagawa Y *et al.* Identification of the human mature B cell miRNome. *Immunity* 2009; **30**: 744–752.
30 Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz Jr LA, Sjoblom T *et al.* The colorectal microRNAome. *Proc Natl Acad Sci USA* 2006; **103**: 3687–3692.
31 Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 2007; **129**: 1401–1414.
32 Creighton CJ, Benham AL, Zhu H, Khan MF, Reid JG, Nagaraja AK *et al.* Discovery of novel microRNAs in female reproductive tract using next generation sequencing. *PLoS One* 2010; **5**: e9637.
33 Stark MS, Tyagi S, Nancarrow DJ, Boyle GM, Cook AL, Whiteman DC *et al.* Characterization of the melanoma miRNAome by deep sequencing. *PLoS One* 2010; **5**: e9685.
34 Vaz C, Ahmad HM, Sharma P, Gupta R, Kumar L, Kulshreshtha R *et al.* Analysis of microRNA transcriptome by deep sequencing of small RNA libraries of peripheral blood. *BMC Genomics* 2010; **11**: 288.
35 Zhang H, Yang JH, Zheng YS, Zhang P, Chen X, Wu J *et al.* Genome-wide analysis of small RNA and novel microRNA discovery in human acute lymphoblastic leukemia based on extensive sequencing approach. *PLoS One* 2009; **4**: e6849.
36 Ramsingh G, Koboldt DC, Trissal M, Chiappinelli KB, Wylie T, Koul S *et al.* Complete characterization of the microRNAome in a patient with acute myeloid leukemia. *Blood* 2010; **116**: 5316–5326.
37 Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D *et al.* Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev* 2010; **24**: 992–1009.
38 Lan FF, Wang H, Chen YC, Chan CY, Ng SS, Li K *et al.* Hsa-let-7 g inhibits proliferation of hepatocellular carcinoma cells by down-regulation of c-Myc and up-regulation of p16(INK4A). *Int J Cancer* 2011; **128**: 319–331.
39 Zhao C, Sun G, Li S, Lang MF, Yang S, Li W *et al.* MicroRNA let-7b regulates neural stem cell proliferation and differentiation by targeting nuclear receptor TLX signaling. *Proc Natl Acad Sci USA* 2010; **107**: 1876–1881.
40 Dong C, Ji M, Ji C. microRNAs and their potential target genes in leukemia pathogenesis. *Cancer Biol Ther* 2009; **8**: 200–205.
41 Reed JC. Bcl-2: prevention of apoptosis as a mechanism of drug resistance. *Hematol Oncol Clin North Am* 1995; **9**: 451–473.
42 Coustan-Smith E, Kitanaka A, Pui CH, McNinch L, Evans WE, Raimondi SC *et al.* Clinical relevance of BCL-2 overexpression in childhood acute lymphoblastic leukemia. *Blood* 1996; **87**: 1140–1146.

43 Cirinna M, Trotta R, Salomoni P, Kossev P, Wasik M, Perrotti D *et al*. Bcl-2 expression restores the leukemogenic potential of a BCR/ABL mutant defective in transformation. *Blood* 2000; **96**: 3915–3921.
44 Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. *Nature* 2008; **455**: 64–71.
45 Keene JD, Komisarow JM, Friedersdorf MB. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* 2006; **1**: 302–307.
46 Tan LP, Seinen E, Duns G, de Jong D, Sibon OC, Poppema S *et al*. A high throughput experimental approach to identify miRNA targets in human cells. *Nucleic Acids Res* 2009; **37**: e137.
47 Felli N, Fontana L, Pelosi E, Botta R, Bonci D, Facchiano F *et al*. MicroRNAs 221 and 222 inhibit normal erythropoiesis and erythroleukemic cell growth via kit receptor down-modulation. *Proc Natl Acad Sci USA* 2005; **102**: 18081–18086.

Supplementary Information accompanies the paper on the Leukemia website (http://www.nature.com/leu)

SPOTLIGHT