Functional relevance of circRNA aberrant expression in pediatric acute leukemia with *KMT2A::AFF1* fusion

Caterina Tretti Parenzan[1], Anna Dal Molin[2], Giorgia Longo[1,2], Enrico Gaffo[2], Alessia Buratin[2,3], Alice Cani[1], Elena Boldrin[4,3], Valentina Serafin[5], Paola Guglielmelli[6], Alessandro M. Vannucchi[6], Giovanni Cazzaniga[7,8], Andrea Biondi[8,9], Franco Locatelli[10], Lueder H. Meyer[4], Barbara Buldini[1,5], Geertruij te Kronnie[2], Silvia Bresolin[1,5] and Stefania Bortoluzzi[2*]

[1]Pediatric Hematology, Oncology and Stem Cell Transplant Division, Women and Child Health Department, Padua University and Hospital, Padua, Italy; [2]Department of Molecular Medicine, University of Padova, Padova, Italy; [3]Department of Biology, University of Padova, Padova, Italy; [4]Ulm University Medical Center, Department of Pediatric and Adolescent Medicine, Ulm, Germany; [5]Onco-Hematology, Stem Cell Transplant and Gene Therapy, Istituto di Ricerca Pediatrica Foundation - Città della Speranza, Padua, Italy; [6]AOU Careggi, University of Florence, Florence, Italy; [7]Tettamanti Center, Fondazione IRCCS San Gerardo dei Tintori, Monza, Italia; [8]School of Medicine and Surgery, University of Milano-Bicocca , Italy; [9]Pediatrics, Fondazione IRCCS San Gerardo dei Tintori, Monza, Italia; and [10]Department of Paediatric Haematology and Oncology, IRCCS Ospedale Pediatrico Bambino Gesù, Catholic University of the Sacred Heart, Rome, Italy.

Supplementary Methods

Supplementary Methods

PDX samples

Patient BCP-ALL specimens were obtained at diagnosis from pediatric patients upon informed consent of patients and/or their legal guardians in accordance with the institution's ethical review board. Patient-derived xenograft (PDX) leukemia samples were generated by transplantation of patient cells into NOD/SCID mice (NOD.CB17-Prkdcscid, Charles River, Germany) as previously described[1,2] and approved by the appropriate authority (Regierungspräsidium Tübingen).

CircRNA detection and quantification from RNA-seq

CircRNAs were detected and quantified by CirComPara v0.6.3[3] using 9 backsplice detection methods (CIRI2 v2.0.2[4]; Findcirc v1.2[5,6]; circRNA_finder v.1.1; DCC v0.4.6[7]; CIRCexplorer2 v2.3.3[8] combined to each of BWA[9], STAR[10], Segemehl[4] and TopHat2[11] alignments) and default parameters, which selected only the circRNAs detected by at least two methods.

The version of other software tools included in CirComPara were: Bowtie2 v2.2.9[12], BWA v0.7.15-r1140[13], STAR v2.6.1d[10], Segemehl v0.3.4[14], and TopHat2 v2.1.0[11] with Bowtie v1.1.2[15].

CirComPara preprocessed raw reads with Trimmomatic v0.38[16] to remove residual adapters and select reads by quality and length. Read linear mapping to the human genome was performed with HISAT2 v2.0.4[17].

Circular to linear proportion (CLP) was calculated as described in[7]: $CLP = \frac{circular\ reads}{circular\ reads + linear\ reads}$, where *circular reads* are backspliced read count and *linear reads* are the count of linearly spliced reads at the backsplice junctions.

Loci without annotated genes and expressing one or more circRNAs (overlapping or not far than 5000 nt) defined new putative genes, called "CircClust".

CirComPara's non-default parameters used for analyses:
ADAPTER SEQUENCE = Trimmomatic file TruSeq3-PE-2.fa;
PREPROCESSOR = Trimmomatic;
TOGGLE_TRANSCRIPTOME_RECONSTRUCTION = 'False';
LINEAR_EXPRESSION_METHODS = 'stringtie';
CIRCRNA_METHODS = "testrealign, dcc, ciri, circexplorer2_star, findcirc, circexplorer2_segemehl, circexplorer2_bwa, circexplorer2_tophat, circrna_finder";
PREPROCESSOR_PARAMS = "MAXINFO:40:0.5 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:30 MINLEN:50 AVGQUAL:30";
HISAT2_EXTRA_PARAMS = "--rna-strandness RF ";
STRINGTIE_PARAMS = '--rf';
BWA_PARAMS = ['-T', '19', '-c', '1'];
SEGEMEHL_PARAMS = ['-M','1', '-D', '0'];

TOPHAT_PARAMS = ['--zpacker', 'pigz', '--max-multihits', '1', '--library-type', 'fr-firststrand'];
STAR_PARAMS = ['--outFilterMultimapNmax', '1', '--outSJfilterOverhangMin', '15', '15', '15', '15', '--alignSJoverhangMin', '15', '--alignSJDBoverhangMin', '15', '--seedSearchStartLmax', '30', '--outFilterScoreMin', '1', '--outFilterMatchNmin', '1', '--outFilterMismatchNmax', '2', '--chimSegmentMin', '15', '--chimScoreMin', '15', '--chimScoreSeparation', '10', '--chimJunctionOverhangMin', '15'];
MIN_READS = 2;
MIN_METHODS = 2;
DCC_EXTRA_PARAMS = ['-fg', '-M', '-F', '-Nr', 1, 1];
TESTREALIGN_PARAMS = ['-q', 'median_10'];
FINDCIRC_EXTRA_PARAMS = ['--best-qual', '40'];
FIX_READ_HEADER = 'True';
SAM_SORT_MM = '6G';
CIRC_PE_MAPPING = 'True'

Of circRNAs reported by at least two methods with a raw count of at least 5 reads in at least one sample, only circRNAs detected in at least 75% of samples for each cell type were considered expressed.

CircRNA functional predictions

The CRAFT tool[18] was applied to predict miRNA binding sites of circRNAs. GRCh38 human genome and annotation references were used to retrieve circRNA sequences. Parameters were set as follows: kind of prediction: ”M”; investigated species: ”hsa”; prefix of genome and indexes: ”hg38”; parameters for the graphical output: score_miRNA=120, energy_miRNA=-20, dGduplex_miRNA=-20, dGopen_miRNA=-10. For the other parameters, CRAFT default values were used.

MiRNAs were subsequently prioritized considering only those in the 75% of miRNAs with the highest average expression in patients with *KMT2A::AFF1* translocations according to in-house available microarray data (personal communication) and in the RS4;11 cell line. Recently identified miRNAs, not detected by the microarray, were all retained.

Validated miRNA target genes (“strong validation” categories of miRecords, mirTarBase and Tarbase) were retrieved by CRAFT.

Cell viability, proliferation and apoptosis assays

Cells were counted with Trypan blue (Sigma Solution) and prepared for EdU incorporation assay (Base Click GmbH EdU 488 cell proliferation kit, Munich, Germany), Annexin-PI assay (Roche, Basilea, Switzerland), and RNA extraction (Trizol™).

The proliferation of *KMT2A::AFF1* translocated cell lines was tested by EdU Incorporation assay (BaseClick GmbH EdU 488 cell proliferation kit, Neuried, Germany), adding 50mM of EdU to transfected cells (at 24h and 48h) and incubated at 37°C for 4 hours. After, cells were harvested, fixed with 3.7% of Formaldehyde 37% w/v (Carlo Erba, Milan, Italy), permeabilized using 0.5% of X-100 Triton (Applichem GmbH, Darmstadt, Germany) and

incubated with the fluorescent reaction mix. Data were detected by the FC500 cytometer (Beckman Coulter, Brea, CA, USA) and analyzed with the FlowJo software V7/8.
Apoptotic cells, after transfection with siRNA and negative control at 24h and 48h were detected by double staining with Annexin V-PI using Annexin-V-FLUOS Staining Kit, (Roche, Basilea, Switzerland) following the corresponding manufacturer's manuals, read at FC500 and analyzed by FlowJo Software V7/8.

Gene expression profiling of *KMT2A::AFF1* patients

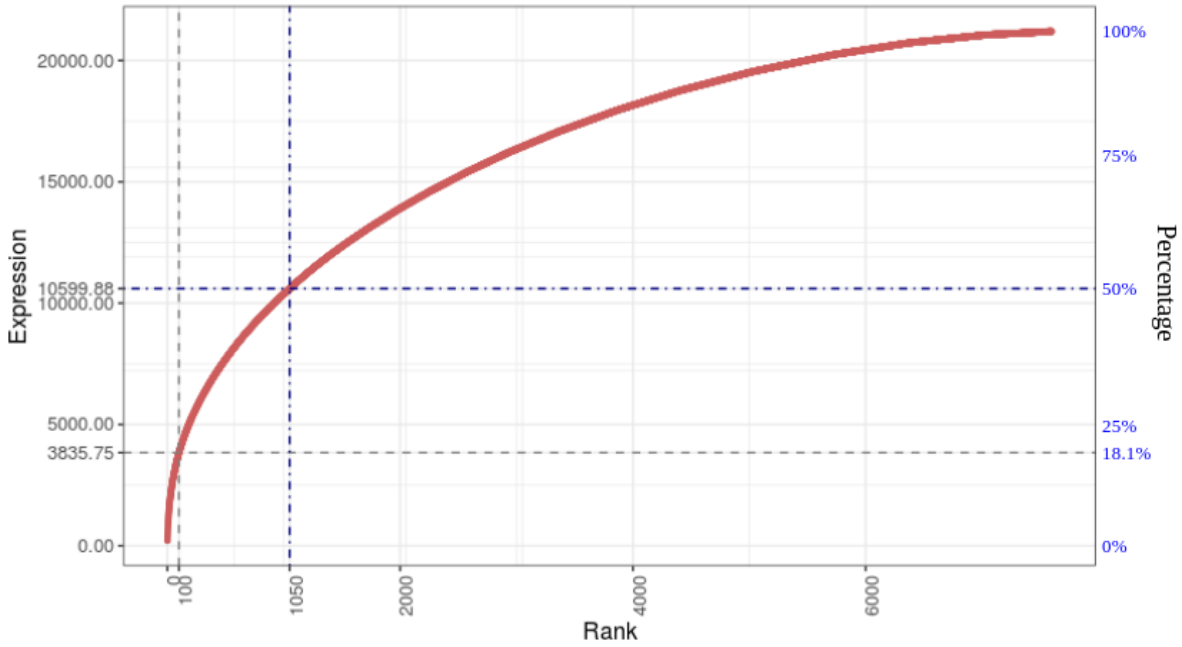
Microarray GEP data were processed by Robust Multiarray Average (RMA) method using Bioconductor R package (r-package.org). The ComBat function in the *sva* package was used to correct the batch effect of different protocols for microarray preparation. Gene Set Enrichment Analysis using C2CGP MSigDB was used to identify enrichment in *KMT2A::AFF1* subgroups according to the level of circAFF1 expression (i.e above or below the median of qRT-PCR). Gene Set Enrichment Analysis (GSEA) was done using the maximum probe set for each gene and 1000 permutations according to phenotype. The heatmap was generated using the top rank 50 up- and down-regulated genes according signal to noise measure.

Supplementary Results

Supplementary Figures

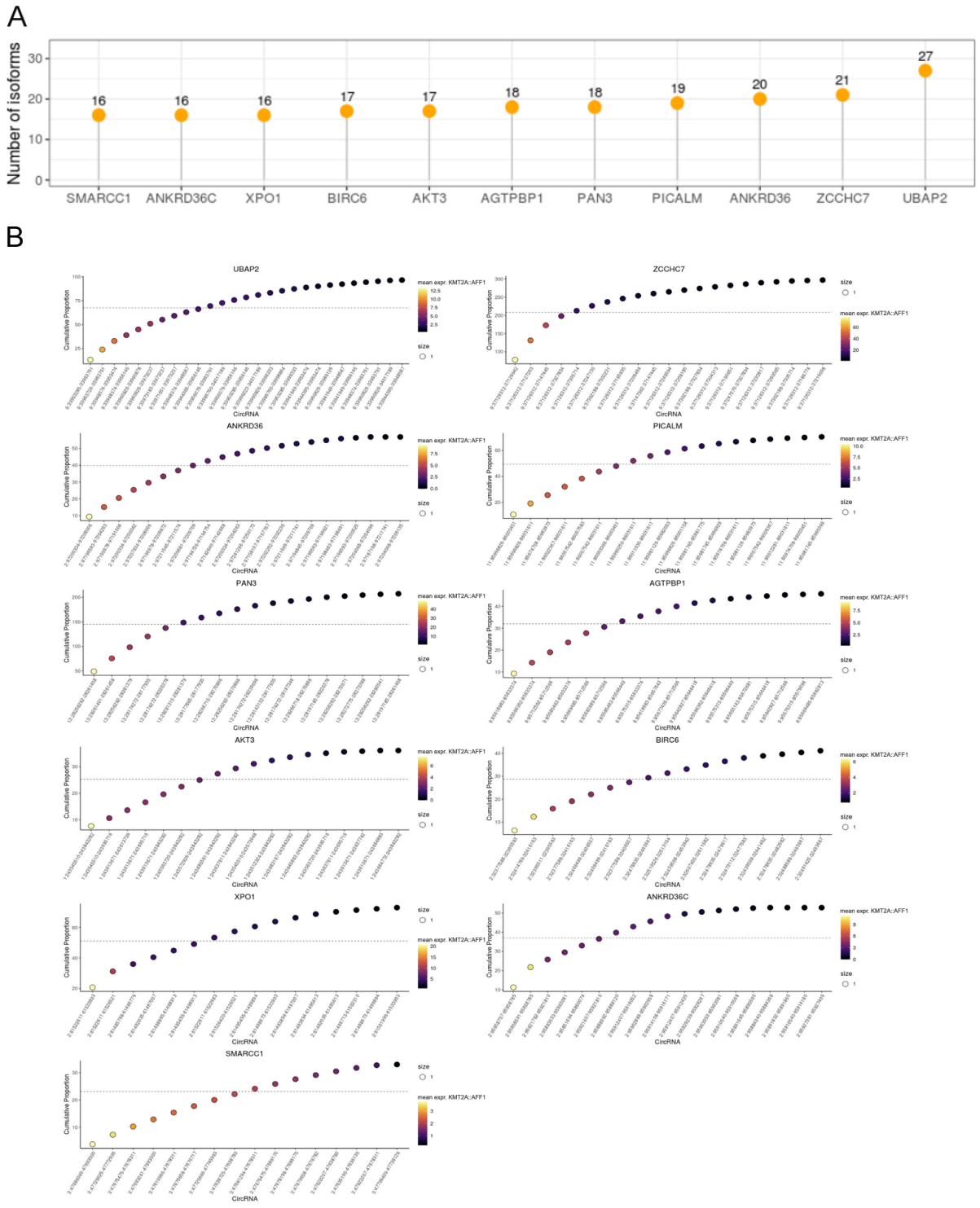
Supplementary Figure 1. Cumulative plot of circRNA expression in *KMT2A::AFF1* ALL.

Expression is given as normalized read counts. Dashed gray and blue lines indicate the 100 most expressed circRNAs and the 50% of expression, respectively.



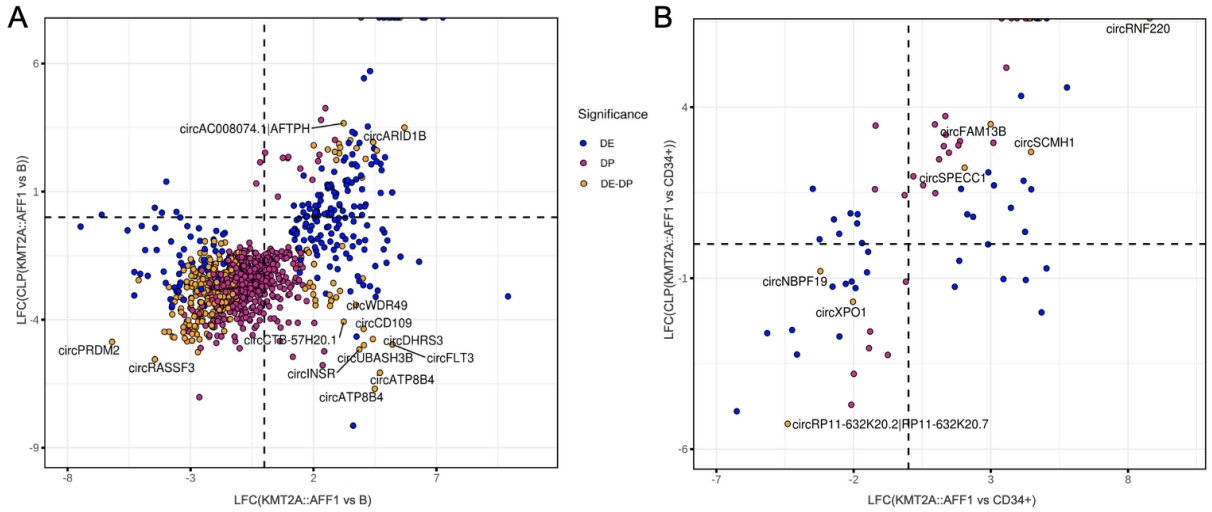
Supplementary Figure 2. CircRNA expressed from genes with at least 15 circular isoforms.

A) Number of isoforms for genes with more than 15 circRNAs each. B) Cumulative plot of circRNA expression per gene. Dot color represents the average expression of each circRNA in *KMT2A::AFF1* ALL patients; the dotted line indicates 70% of gene expression. Cumulative Proportion is measured in RPM.

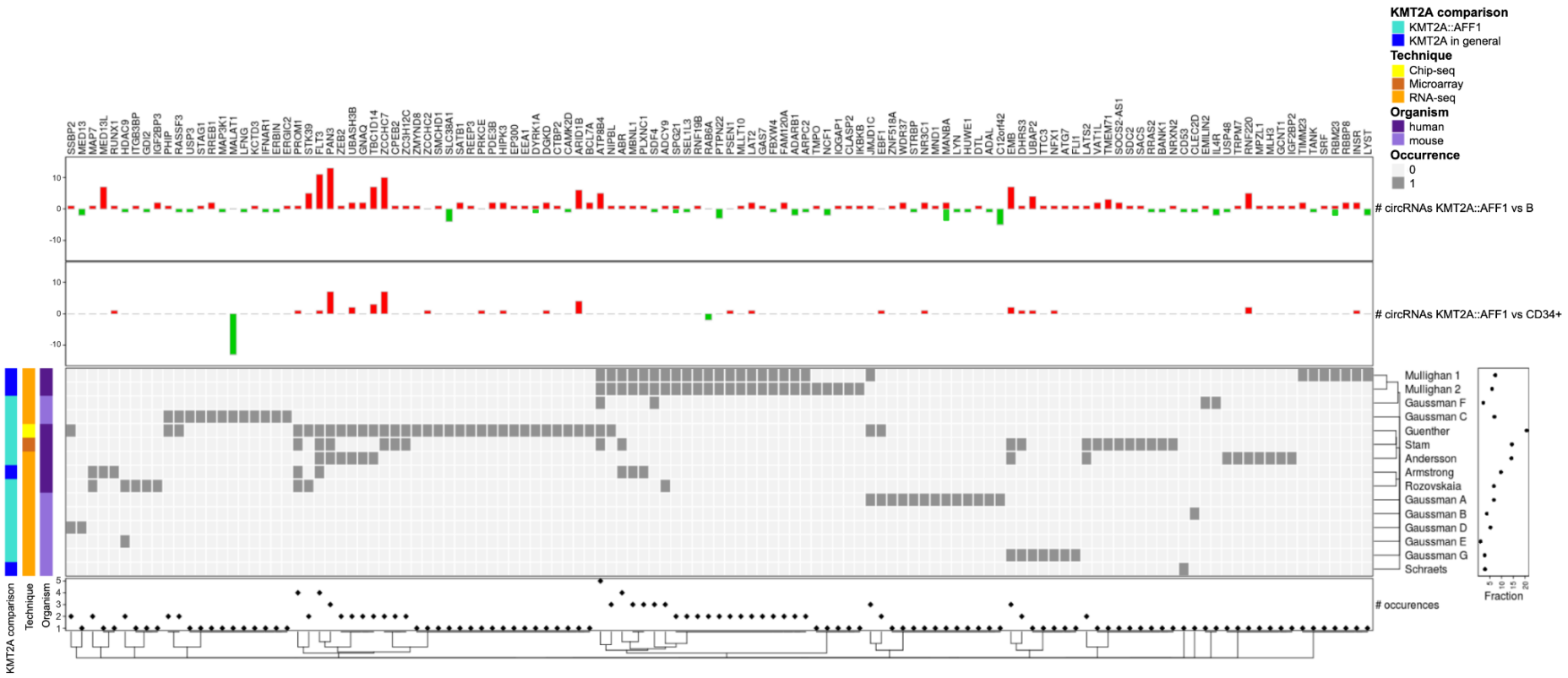


Supplementary Figure 3. Relation between dysregulation of circRNA expression and proportion in *KMT2A::AFF1* ALL.

CLP and circRNA expression variation for circRNAs with differential proportion (DP) and/or differentially expressed (DE) in *KMT2A::AFF1* ALL patients versus A) B-cells and B) versus CD34+ cells. Names of DE-DP circRNAs with |LFC(E)| and |LFC(CLP)| > 3 are shown. E, circRNA normalized expression.

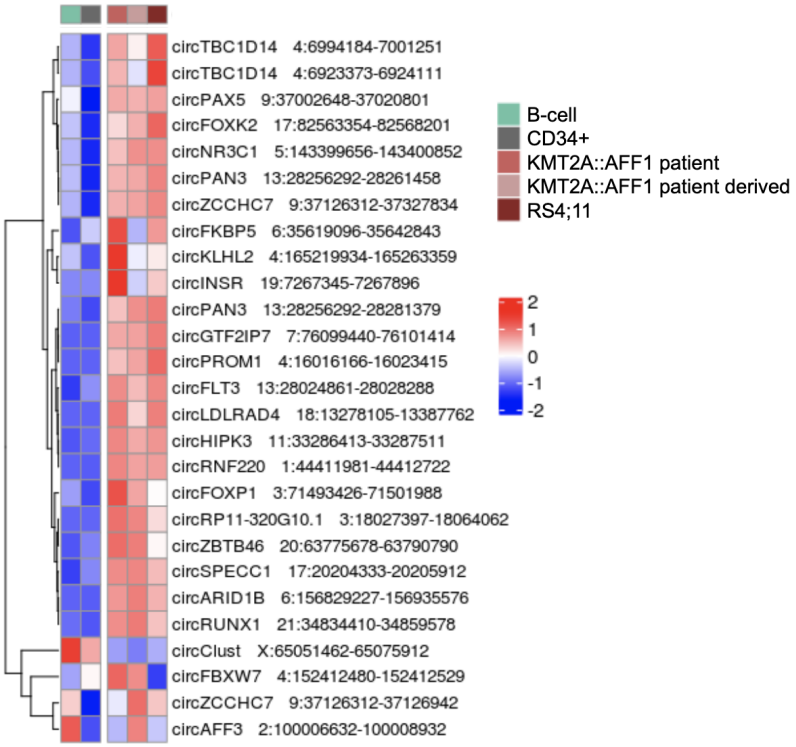


Supplementary Figure 4. Genes in the meta KMT2Are/*KMT2A::AFF1* signature with at least one dysregulated circRNA.
Barplot of the number of circRNAs dysregulated in *KMT2A::AFF1* expressed from genes belonging to the *KMT2Are/KMT2A::AFF1* meta-signature. The heatmap below shows the genes belonging to each original signature (dark gray), and is clustered according to the Dice distance on the columns and the Sokal & Michener distance on the rows. The marginal dot plots display the number signatures that contain each gene (bottom) and the fraction of genes of that signature that host at least one circRNA with respect to the number of genes of that signature (right). Information about the original signatures considered is given by the left bars.



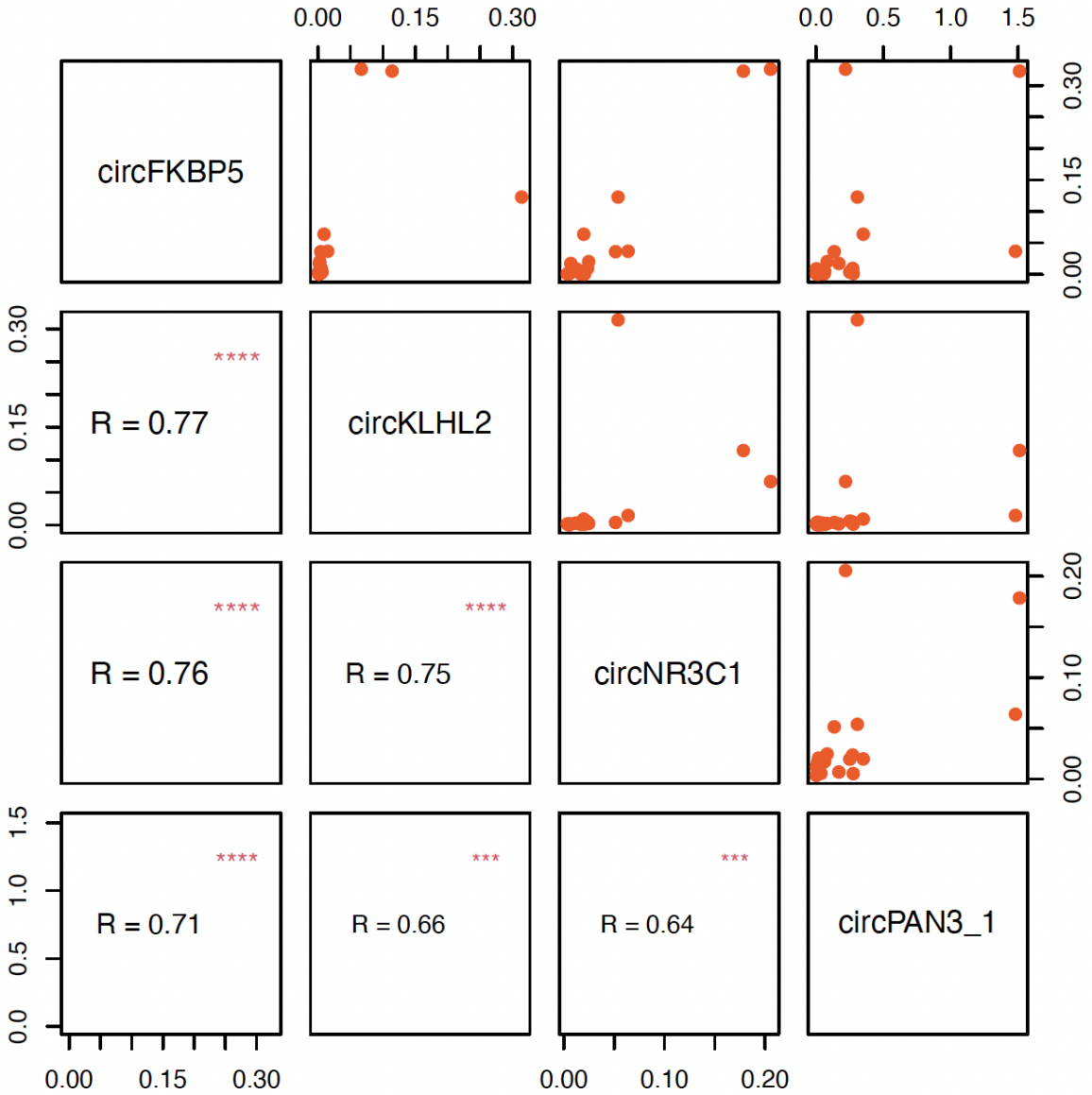
Supplementary Figure 5. Expression in *KMT2A::AFF1* patient derived samples and in the RS4;11 cell line of the 27 circRNAs aberrantly expressed in *KMT2A::AFF1* patients compared with both normal populations.

Heatmap of mean standardized expression per sample group with circRNAs are clustered according to the Euclidean distance.



Supplementary Figure 6. Pairwise correlation of circFKBP5, circKLHL2, circNR3C1 and circPAN3 expression levels in the extended cohort of 17 pediatric patients with *KMT2A::AFF1* ALL.

The plots in the top-right and down-left parts of the figure report respectively the scatterplots and the pairwise Spearman correlation values of the four considered circRNAs expression levels in patients.



Supplementary Tables

Supplementary Table 1. SiRNA sequences used for circRNA silencing.
The table reports the siRNA sequences targeting the circRNAs whose silencing effects were tested *in vitro* in SEM and RS4;11 cell lines.

SiRNA	Backplice	Sense sequence	Antisense sequence
circARID1B_sir1	6:156829227-156935576	CCAUCAAGUUUACCAGGCAtt	UGCCUGGUAAACUUGAUGGtc
circARID1B_sir2	6:156829227-156935576	GCAGCCCAAUGGAUCCAAUtt	AUUGGAUCCAUUGGGCUGCct
circFBXW7_sir1	4:152411303-152412529	AGAUUACUUCCUUAGGAUAtt	UAUCCUAAGGAAGUAAUCUtt
circFBXW7_sir2	4:152411303-152412529	ACAAAAGAUUACUUCCUUAtt	UAAGGAAGUAAUCUUUUGUtg
circFKBP5_sir1	6:35619096-35642843	GGAGCAACAGUAGAAAGUUtt	AACUUUCUACUGUUGCUCCtt
circFKBP5_sir2	6:35619096-35642843	CAACAGUAGAAAGUUCUCUtt	AGAGAACUUUCUACUGUUGct
circFLT3_sir1	13:28061867-28070612	UUUGAUUUACAAAACAGUUtt	AACUGUUUUGUAAAUCAAAat
circFLT3_sir2	13:28061867-28070612	AUUUACAAAACAGUUGUUUtt	AAACAACUGUUUUGUAAAUca
circHIPK3_sir1	11:33286413-33287511	GGUAUGGCCUCACAAGUCUtt	AGACUUGUGAGGCCAUACCtg
circHIPK3_sir2	11:33286413-33287511	CAAUCUCGGUACUACAGGUtt	ACCUGUAGUACCGAGAUUGta
circKLHL2_sir1	4:165219934-165263359	ACACCUAUGCAGAUGCACAtt	UGUGCAUCUGCAUAGGUGUtg
circKLHL2_sir2	4:165219934-165263359	CACCUAUGCAGAUGCACAAtt	UUGUGCAUCUGCAUAGGUGtt
circNR3C1_sir1	5:143399656-143400852	UGGCUAUUCAAGUUGAUAUtt	AUAUCAACUUGAAUAGCCAtt
circNR3C1_sir2	5:143399656-143400852	AAGUUGAUAUUCACUGAUGtt	CAUCAGUGAAUAUCAACUUga
circPAN3_sir1	13:28256292-28261458	CAGAUAUGCCAGGAAUGUCtt	GACAUUCCUGGCAUAUCUGct
circPAN3_sir2	13:28256292-28261458	CCAGGAAUGUCGUUGUCUGtt	CAGACAACGACAUUCCUGGca
circPROM1_sir1	4:16016166-16023415	AGGUAUCAAUUCAGUGCUAtt	UAGCACUGAAUUGAUACCUgt
circPROM1_sir2	4:16016166-16023415	GGUCCAACAGGUAUCAAUUtt	AAUUGAUACCUGUUGGACCag
circRNF220_sir1	1:44411981-44412722	GGAUCGGAAUGACAGAUGUtt	ACAUCUGUCAUUCCGAUCCtc
circRNF220_sir2	1:44411981-44412722	ACAGAUGUCUUAGGAGGGAtt	UCCCUCCUAAGACAUCUGUca
circTBC1D14_sir1	4:6923373-6924111	GUUUCUCCUUGGACCAAGAtt	UCUUGGUCCAAGGAGAAACtt
circTBC1D14_sir2	4:6923373-6924111	AGUUUCUCCUUGGACCAAGtt	CUUGGUCCAAGGAGAAACUtg
circTBC1D14_2_sir1	4:6994184-7001251	CCCACGGAAUAUGAAGACAtt	UGUCUUCAUAUUCCGUGGGtg
circTBC1D14_2_sir2	4:6994184-7001251	CCACGGAAUAUGAAGACAAtt	UUGUCUUCAUAUUCCGUGGgt
circZCCHC7_sir1	9:37126312-37327834	GCUUCAAGGUUACUGACUUtt	AAGUCAGUAACCUUGAAGCgt
circZCCHC7_sir2	9:37126312-37327834	ACCUAACGCUUCAAGGUUAtt	UAACCUUGAAGCGUUAGGUga

Supplementary Table 2. Divergent primers used to confirm circRNA presence and silencing efficacy in cell line and patient samples.

The table displays the divergent forward and reverse primer sequences used in RQ-PCR to obtain expression levels of circular RNAs, identified by backsplice end coordinates.

CircRNA	Backsplice	Forward primer sequence	Reverse primer sequence
circARIDB1	6:156829227-156935576	AGATCTCAACCTCCTCTGGC	CTGAGGCTGAGAATGAGGGT
circFBXW7	4:152411303-152412529	GGAGATGGACCAGGAGAGTG	TGACCCAGTAACTCCACTTCT
circFKBP5	6:35619096-35642843	CCAAACGGAAAGGAGAGGGA	GCTGTGGGGCTTTCTTCATT
circFLT3_2	13:28024861-28028288	TCAAGAGAAGTTCAGATACACCC	GGTTCACAATATTCTCGTGGT
circHIPK3	1:133286413-33287511	TCATGCTGATCTCAAGCCAGA	ACACAACTGCTTGGCTCTAC
circKLHL2	4:165219934-165263359	CAGCTTCACCCTGTCAACTG	AGGATTCACTGTCACTGGGC
circNR3C1	5:143399656-143400852	ACTCTGAACTTCCCTGGTCG	ACCTTCACAGTAGCTCCTCC
circPAN3_1	13:28256292-28261458	CGCACCTTCCTTCTTCATGG	TCTTCTTCTGGGAGCAGGAG
circPROM1	4:16016166-16023415	ACCCGTGGATGCAGAACTT	GGGTCTCAGTCGGTCAAGAA
circRNF220	1:44411981-44412722	ATCGGGAAGCCTCATCTAGC	TCGGAGTCTCTTTCTGTGGC
circTBC1D14	4:6923373-6924111	TTCATGAAGCTGAGGAGGGG	CGCCTTGAGATTGACGTGTT
circTBC1D14_2	4:6994184-7001251	AGAGGCAAAGTCTGGAGCTT	ATGAGTGCGGTGGTGGAAA
circZCCHC7	9:37126312-37327834	GAAATCTGGAGGCAGTATCACC	GCTCTTCCTCCCTGATGACA

Supplementary Table 3. Convergent primers used to quantify linear transcripts expressed from circRNA host genes.

Gene	Forward primer sequence	Reverse primer sequence
FKBP5	CTTTTTTGAGATTGAGCTCCTTGA	GCAGTCAAACATCCTTCCACCA
KLHL2	GATGGGAGTACAGGTTTGTCAT	GCTCCATCATAACCTCCTACAG
NR3C1	AGCAGTGAAATGGGCAAAGG	GGAGCAAAACACAGCAGGTT
PAN3	GAGGATACATGGTTTTCGTCTTG	CTCCAGCATGGAAATCATATGC
PROM1	ACCGACTGAGACCCAACATC	CAGGCTAGTTTTCACGCTGG

Supplementary Table 4. List of circRNAs differentially expressed in *KMT2A::AFF1* ALL compared with both normal cell populations.

The table reports the backsplice coordinates and the host gene names of the 86 circRNAs found to be differentially expressed (adjusted p-value<0.05) in the comparison between *KMT2A::AFF1* patients *versus* B and CD34+ cells, with the corresponding expression Log_2FoldChange (LFC).

CircRNA ID (backsplice ends)	Gene name	LFC (KMT2A::AFF1 vs B)	LFC (KMT2A::AFF1 vs CD34+)
1:12578718-12579412	*DHRS3*	4.41	4.5
1:44411981-44412722	*RNF220*	9.91	8.79
1:44411981-44416223	*RNF220*	8.25	5.45
1:45642427-45642499	*GPBP1L1*	2.82	4.72
2:24135119-24147086	*FAM228B|RP11-507M3.1*	-2.5	-2.29
2:40428473-40430304	*SLC8A1*	3.04	-2.11
2:61522611-61533903	*XPO1*	-1.37	-2.02
3:18013655-18064062	*RP11-320G10.1*	8.16	6.3
3:18027397-18049309	*RP11-320G10.1*	7.42	5.55
3:18027397-18064062	*RP11-320G10.1*	10.04	8.19
3:18027397-18072216	*RP11-320G10.1*	6.36	4.51
3:18027397-18121838	*RP11-320G10.1*	6.73	4.88
3:18027397-18187140	*RP11-320G10.1*	7.29	5.44
3:18041449-18064062	*RP11-320G10.1|TBC1D5*	9.72	7.88
3:18041449-18072216	*RP11-320G10.1*	6.97	5.11
3:18041449-18097732	*RP11-320G10.1*	7.77	5.93
3:18041449-18121838	*RP11-320G10.1|TBC1D5*	7.67	4.86
3:18041449-18187140	*RP11-320G10.1|TBC1D5*	7.66	4.85
3:71493426-71501988	*FOXP1*	4.46	5.79
3:183643480-183643542	*KLHL24*	3.62	4.71
4:152412480-152412529	*FBXW7*	4.43	2.35
4:16016166-16023415	*PROM1*	5.09	3.21
4:165219934-165263359	*KLHL2*	2.82	3.73
4:6923373-6924111	*TBC1D14*	2.16	3.46
4:6923373-6967424	*TBC1D14*	3.32	4.64
4:6994184-7001251	*TBC1D14*	2.41	4.57

4:77773081-77776392	*CNOT6L*	-1.45	4.45
4:87084120-87084164	*AFF1*	3.96	4.97
5:137988274-137988315	*FAM13B*	3.09	2.99
5:143399656-143400852	*NR3C1*	2.29	4.64
5:168488602-168494650	*RARS*	-0.91	-1.69
5:50274988-50411383	*EMB*	3.59	3.64
5:50399107-50411383	*EMB*	2.44	1.85
6:138943513-138943576	*REPS1*	3.39	4.69
6:156829227-156901525	*ARID1B*	3.38	4.96
6:156829227-156935576	*ARID1B*	5.7	5.78
6:157036835-157036876	*ARID1B*	1.94	4.4
6:35619096-35642843	*FKBP5*	5.93	3.27
7:155672867-155672915	*RBM33*	1.94	4.32
7:22317987-22318037	*RAPGEF5*	4.77	4.25
7:22991061-22991139	*FAM126A*	3.18	4.38
7:24623666-24668660	*MPP6*	-6.16	-6.36
7:74220704-74220734	*LAT2*	2.78	4.94
7:76088064-76101414	*GTF2IP7*	6.66	4.81
7:76099440-76101414	*GTF2IP7*	7.26	5.41
8:18765449-18804898	*PSD3*	6.51	4.67
9:33963726-33963791	*UBAP2*	3.17	5.23
9:36966549-37020801	*PAX5*	3.75	4.74
9:37002648-37020801	*PAX5*	1.62	5.53
9:37126312-37127263	*ZCCHC7*	1.48	3.9
9:37126312-37206494	*ZCCHC7*	2.57	4.69
9:37126312-37305714	*ZCCHC7*	1.98	5.65
9:37126312-37327834	*ZCCHC7*	3.18	6.47
9:37302188-37302231	*ZCCHC7*	2.53	5.21
10:124696158-124696212	*FAM53B|RP11-12J10.3*	4.37	4.41
10:32543300-32584304	*AL356053.1|CCDC7*	-2.24	-3.24
11:122776219-122776272	*UBASH3B*	6.51	4.65
11:122776219-122777210	*UBASH3B*	4.05	4.79

11:33286413-33287511	*HIPK3*	2.55	2.14
12:66203711-66217235	*IRAK3*	3.01	-2.18
12:95208843-95211267	*FGD6*	3.9	4.65
13:28024861-28028288	*FLT3*	7.39	3.7
13:28174272-28220378	*PAN3*	3.32	4.03
13:28174272-28239696	*PAN3*	5.12	4.66
13:28256292-28281379	*PAN3*	3.22	3.96
13:28256292-28261458	*PAN3*	2.08	4.19
13:28261401-28261458	*PAN3*	3.99	4.09
13:28281315-28281379	*PAN3*	5.25	5.3
13:99238427-99244624	*UBAC2*	-1.55	5.1
14:32090502-32117287	*ARHGAP5*	2.01	-2.52
15:30474356-30485267	*CTD-3092A11.1*	4.5	5.04
15:64499293-64500166	*ZNF609*	1.11	2.91
16:68121987-68123121	*NFATC3|RP11-67A1.2*	1.51	5.03
16:68121987-68126610	*NFATC3|RP11-67A1.2*	1.42	2.9
16:69370483-69372355	*TERF2*	-1.36	2.73
17:20204333-20205912	*SPECC1*	3.02	2.04
17:47402132-47414919	*CTD-2026D20.2|EFCAB13*	-2.01	-2.07
17:82563354-82568201	*FOXK2*	1.86	4.88
18:13278105-13387762	*LDLRAD4*	6.26	4.42
19:7267345-7267896	*INSR*	7.36	5.53
20:32366384-32366466	*ASXL1*	6.33	4.48
20:58438945-58441083	*VAPB*	-2.6	-2.77
20:63775678-63790790	*ZBTB46*	6.3	4.19
21:15014344-15043574	*AF127936.9|NRIP1*	3.93	-1.85
21:15762891-15762968	*USP25*	2.87	4.74
21:34834410-34859578	*RUNX1*	5.3	4.82

Supplementary Table 5. CircRNA with altered circular to linear proportion in *KMT2A::AFF1* ALL compared with both normal cell populations.

The table reports the coordinates and the host gene names of the 13 circRNAs found to have a differential circular to linear proportion (CLP) (adjusted p-value<0.05) in the comparison between *KMT2A::AFF1* patients *versus* B and CD34+ cells.

CircRNA ID	Gene name	LFC (KMT2A::AFF1 vs B)	LFC (KMT2A::AFF1 vs CD34+)
1:41070595-41075451	*SCMH1*	-1.76	2.69
2:61522611-61533903	*XPO1*	-3.26	-1.69
2:8908621-8958642	*MBOAT2*	-1.38	2.48
3:136398749-136502779	*STAG1*	2.64	3.5
3:138570318-138571356	*CEP70*	-2.7	-1.1
4:143543509-143543972	*SMARCA5*	-1.02	1.43
4:177353308-177353728	*NEIL3|RP11-376O6.2*	-3.13	-3.24
5:123545417-123575963	*CSNK1G3*	1.32	3.19
5:123557495-123557564	*CSNK1G3*	2.86	1.72
5:137985257-137988315	*FAM13B*	-2.24	2.88
5:137988274-137988315	*FAM13B*	2.72	3.5
10:7797047-7802854	*ATP5C1*	-3.08	-2.56
11:32927157-32935435	*QSER1*	1.55	1.49

Supplementary Table 6. KMT2Are and *KMT2A::AFF1* signatures.

Database	Year	Num. of genes	Comparison description	Comparison	Experiment	Technique	Organism
Mullighan 1	2007	384	Genes up-regulated in pediatric AML with rearranged KMT2A compared to all AML cases with the intact gene	KMT2Are	RNA-seq	Gene expression	Human
Mullighan 2	2007	420	Genes up-regulated in pediatric AML with rearranged KMT2A compared to the AML cases with intact KMT2A and NPM1	KMT2Are	RNA-seq	Gene expression	Human
Armstrong	2002	82	Top 100 genes that are relatively underexpressed in KMT2A as compared to ALL	KMT2Are	RNA-seq	Gene expression	Human
Rozovskaia	2003	119	Genes overexpressed in t(4;11) ALLs compared to all other ALLs	KMT2A::AFF1	RNA-seq	Gene expression	Human
Guenther	2008	169	SEM cells vs. REH control cells	KMT2A::AFF1	Chip-seq	Target genes	Human
Schraets_ down	2003	34	Genes down-regulated in fibroblasts from KMT2A knockout mice	KMT2Are	RNA-seq	Gene expression	Mouse
Gaussman_ A_up	2007	193	Up-regulated genes from the set A: specific to cells expressing KMT2A::AFF1 fusion protein alone	KMT2A::AFF1	RNA-seq	Gene expression	Mouse
Gaussman_ B_up	2007	27	Up-regulated genes from the set B: specific signature shared by cells expressing either AFF1::KMT2A or KMT2A::AFF1 fusion protein	KMT2A::AFF1	RNA-seq	Gene expression	Mouse
Gaussman_ C_up	2007	172	Up-regulated genes from the set C: specific to cells expressing AFF1::KMT2A fusion protein alone	KMT2A::AFF1	RNA-seq	Gene expression	Mouse
Gaussman_ D_up	2007	38	Up-regulated genes from the set D: specific signature shared by cells expressing KMT2A::AFF1 alone and those expressing both KMT2A::AFF1 and AFF1::KMT2A fusion proteins	KMT2A::AFF1	RNA-seq	Gene expression	Mouse
Gaussman_ E_up	2007	97	Up-regulated genes from the set E: specific signature shared by cells expressing either KMT2A::AFF1 or AFF1::KMT2A fusion proteins alone, and those expressing both fusion proteins	KMT2A::AFF1	RNA-seq	Gene expression	Mouse

Gaussman_F_up	2007	182	Up-regulated genes from the set F: specific signature shared by cells expressing AFF1::KMT2A alone and those expressing both AFF1::KMT2A and KMT2A::AFF1 fusion proteins	KMT2A::AFF1	RNA-seq	Gene expression	Mouse
Gaussman_G_up	2007	247	Up-regulated genes from the set G: specific to cells expressing both KMT2A::AFF1 and AFF1::KMT2A fusion proteins	KMT2A::AFF1	RNA-seq	Gene expression	Mouse
Stam	2010	132	KMT2A::AFF1 versus non-KMT2Are up	KMT2A::AFF1	Microarray	Gene expression	Human
Andersson	2015	106	KMT2A::AFF1 versus non-KMT2Are up	KMT2A::AFF1	RNA-seq	Gene expression	Human

Supplementary Table 7. Enrichment analysis on predicted circRNA target genes.
Table summarizing the miRNA target genes, of at least 3 out of 4 circRNAs, enriched in leukemia disease (adjusted p-value<0.05).

Pathway	Adj. P-value	Number of genes	Gene names
Leukemia	0.0010	9	*CDKN1A/FGFR1/IGF2/KMT2A/NF1/RUNX1/SMC1A/TET3/YWHAE*
Hematopoietic neoplasms	0.0055	11	*AKT2/CALM1/CTNNB1/DICER1/FGFR1/FOXP1/JARID2/KMT2A/RUNX1/SOX11/TRIM28*
Mixed Lineage Leukemia	0.0083	9	*ARID5B/CALM1/CTNNB1/FRYL/KDM2A/KMT2A/KMT2D/RUNX1/SETD1A*
Acute lymphoblastic leukemia	0.0200	4	*ARID5B/KMT2A/NSD2/RUNX1*
Childhood B acute lymphoblastic Leukemia	0.0225	5	*ARID5B/IGF2BP1/KMT2A/RUNX1/SPTBN1*

References

1. Meyer LH, Eckhoff SM, Queudeville M, et al. Early relapse in ALL is identified by time to leukemia in NOD/SCID mice and is characterized by a gene signature involving survival pathways. *Cancer Cell*. 2011;19(2):206–217.
2. Queudeville M, Seyfried F, Eckhoff SM, et al. Rapid engraftment of human ALL in NOD/SCID mice involves deficient apoptosis signaling. *Cell Death Dis.* 2012;3(8):e364.
3. Gaffo E, Bonizzato A, Kronnie G, Bortoluzzi S. CirComPara: A Multi‐Method Comparative Bioinformatics Pipeline to Detect and Study circRNAs from RNA‐seq Data. *Non-Coding RNA*. 2017;3(1):8.
4. Gao Y, Zhang J, Zhao F. Circular RNA identification based on multiple seed matching. *Brief. Bioinform.* 2018;19(5):803–810.
5. Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013;495(7441):333–338.
6. Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013;495(7441):333–338.
7. Cheng J, Metge F, Dieterich C. Specific identification and quantification of circular RNAs from sequencing data. *Bioinformatics*. 2016;32(7):1094–1096.
8. Zhang X-O, Dong R, Zhang Y, et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* 2016;26(9):1277–1287.
9. Du H, Liang C. Assembly of chromosome-scale contigs by efficiently resolving repetitive sequences with long reads.
10. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
11. Kim D, Pertea G, Trapnell C, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14(4):R36.
12. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. 2012;9(4):357–359.
13. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013;
14. Hoffmann S, Otto C, Doose G, et al. A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol.* 2014;15(2):R34.
15. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10(3):R25.
16. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120.
17. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*. 2015;12(4):357–360.
18. Dal Molin A, Gaffo E, Difilippo V, et al. CRAFT: a bioinformatics software for custom prediction of circular RNA functions. *Brief. Bioinform.* 2022;23(2.):