REGULAR ARTICLE

blood advances

Clinical and functional significance of circular RNAs in cytogenetically normal AML

Dimitrios Papaioannou,[1,*] Stefano Volinia,[2,*] Deedra Nicolet,[1,3] Michał Świerniak,[4] Andreas Petri,[5] Krzysztof Mrózek,[1] Marius Bill,[1] Felice Pepe,[1] Christopher J. Walker,[1] Allison E. Walker,[1] Andrew J. Carroll,[6] Jessica Kohlschmidt,[1,3] Ann-Kathrin Eisfeld,[1] Bayard L. Powell,[7] Geoffrey L. Uy,[8] Jonathan E. Kolitz,[9] Eunice S. Wang,[10] Sakari Kauppinen,[5] Adrienne Dorrance,[1] Richard M. Stone,[11] John C. Byrd,[1] Clara D. Bloomfield,[1,†] and Ramiro Garzon[1,†]

[1]The Ohio State University Comprehensive Cancer Center, Columbus, OH; [2]Department of Morphology, Surgery and Experimental Medicine, University of Ferrara, Ferrara, Italy; [3]Alliance Statistics and Data Center, The Ohio State University, Columbus, OH; [4]Centre of New Technologies, University of Warsaw, Warsaw, Poland; [5]Center for RNA Medicine, Department of Clinical Medicine, Aalborg University, Copenhagen, Denmark; [6]Department of Genetics, University of Alabama at Birmingham, Birmingham, AL; [7]The Comprehensive Cancer Center of Wake Forest University, Winston-Salem, NC; [8]Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO; [9]Monter Cancer Center, Hofstra Northwell School of Medicine, Lake Success, NY; [10]Department of Medicine, Roswell Park Cancer Institute, Buffalo, NY; and [11]Dana-Farber Cancer Institute, Harvard University, Boston, MA

Key Points

- Distinctive circRNA expression profiles are associated with recurrent mutations and clinical features of CN-AML patients.
- Individual circRNAs are associated with outcome and are functionally relevant in CN-AML.

Circular RNAs (circRNAs) are noncoding RNA molecules that display a perturbed arrangement of exons, called backsplicing. To examine the prognostic and biologic significance of circRNA expression in cytogenetically normal acute myeloid leukemia (CN-AML), we conducted whole-transcriptome profiling in 365 younger adults (age 18-60 years) with CN-AML. We applied a novel pipeline, called Massive Scan for circRNA, to identify and quantify circRNA expression. We validated the high sensitivity and specificity of our pipeline by performing RNase R treatment and RNA sequencing in samples of AML patients and cell lines. Unsupervised clustering analyses identified 3 distinct circRNA expression–based clusters with different frequencies of clinical and molecular features. After dividing our cohort into training and validation data sets, we identified 4 circRNAs (*circCFLAR*, *circKLHL8*, *circSMC1A*, and *circFCHO2*) that were prognostic in both data sets; high expression of each prognostic circRNA was associated with longer disease-free, overall, and event-free survival. In multivariable analyses, high *circKLHL8* and high *circFCHO2* expression were independently associated with better clinical outcome of CN-AML patients, after adjusting for other covariates. To examine the biologic relevance of circRNA expression, we performed knockdown screening experiments in a subset of prognostic and gene mutation–related candidate circRNAs. We identified *circFBXW7*, but not its linear messenger RNA, as a regulator of the proliferative capacity of AML blasts. In summary, our findings underscore the molecular associations, prognostic significance, and functional relevance of circRNA expression in CN-AML.

Introduction

Aberrations in the noncoding transcriptome are emerging as important molecular mechanisms that contribute to cancer pathogenesis.[1-3] The advent of next-generation sequencing has enabled the global, in-depth study of noncoding RNA species and has intensified the interest in their role in health and

Submitted 10 June 2019; accepted 14 November 2019; published online 16 January 2020. DOI 10.1182/bloodadvances.2019000568.

*D.P. and S.V. contributed equally to this study.

†C.D.B. and R.G. contributed equally to this study as senior authors.

Presented in abstract form at the 23rd Annual Meeting of the European Hematology Association, Stockholm, Sweden, 14-18 June 2018.

Data sharing requests can be e-mailed to the corresponding author, Ramiro Garzon (ramiro.garzon@osumc.edu).

The full-text version of this article contains a data supplement.

disease. Circular RNAs (circRNAs) are a class of noncoding, single-stranded RNA molecules that form covalently closed circles and are characterized by a perturbed arrangement of exons caused by aberrant splicing (referred to as backsplicing).[4] CircRNAs were first identified almost 5 decades ago in plants, protozoans, and viruses[5-7] and subsequently in mammals.[8-10] Additional studies revealed a number of features that distinguish circRNAs from linear protein-coding RNA transcripts, such as the lack of 5′ or 3′ ends, the lack of polyadenylated tails, and increased stability and extended half-life.[11] Although not yet fully elucidated, the biogenesis of circRNAs has been shown to depend on the spliceosome machinery,[12-14] the presence of repeat sequences in the genome,[9,10] and RNA binding proteins.[15-17] The application of total RNA sequencing (RNA-seq) and novel bioinformatic approaches, designed to detect backspliced junctions, has uncovered the expression of a high number of circRNAs in humans, mice, metazoans, and plants.[18] Recent reports have indicated that circRNAs are produced from ~10% of all expressed genes, that a subset of them is evolutionarily conserved, and that circRNAs are often expressed in a tissue-specific manner.[4,19,20] Although initially considered to represent inherent errors of the splicing process,[2,8] recent studies have unraveled the biologic relevance of circRNAs.[21-25]

With regard to the functional significance of circRNAs, recent studies have demonstrated that circRNAs can sequester microRNAs (miRs) and inhibit their function. Two examples of such circRNAs are CDR1as/ciRS-7, which contains repetitive binding sites that capture miR-7,[11,21] and circular *SRY*, which captures miR-138.[11] Although other circRNAs have been proposed to function in the same manner,[25-28] bioinformatic studies have indicated that only a small fraction of circular RNAs contain target sequences for miRs and can function as miR regulators.[29,30] CircRNAs have also been shown to interact with RNA binding proteins or RNA polymerases and regulate gene transcription.[31] In addition, recent studies have suggested that individual circRNAs can interact with ribosomes and be translated into peptides.[32-36] However, the functional relevance and mechanism of action of the majority of the identified circRNAs remain unknown.

In cancer, expression levels of individual circRNAs have been identified as candidate biomarkers of disease progression.[37-39] With regard to acute myeloid leukemia (AML), it was recently reported that chromosomal translocations, such as t(15;17)/*PML-RARA* or t(9;11)/*MLLT3-KMT2A*, give rise to fusion circRNAs that contribute to cellular transformation and survival of the leukemic blasts.[40] However, the prognostic and functional significance of circRNA expression in AML has not been extensively studied in large patient cohorts. Herein, we performed total RNA-seq in younger adults (age <60 years) with de novo cytogenetically normal AML (CN-AML) and applied a novel bioinformatic algorithm to identify and quantify circRNAs. Our goals were to determine whether circRNA expression is associated with clinical and molecular features of CN-AML patients and whether circRNAs have functional relevance in this disease.

Patients, materials, and methods

Patients

Pretreatment bone marrow (BM) or blood samples were obtained from 365 younger adult patients (age 17-59 years) with de novo CN-AML. All analyzed patients were treated with cytarabine/anthracycline-based frontline chemotherapy on Cancer and Leukemia Group B (CALGB)/Alliance trials and were alive 30 days after initiation of treatment. No patient received allogeneic stem cell transplantation in first complete remission. Details regarding treatment protocols are provided in the supplemental Data. All patients were also enrolled onto companion protocols CALGB 8461 (cytogenetic studies), CALGB 9665 (Leukemia Tissue Bank) and CALGB 20202 (molecular analyses) (registered at www.clinicaltrials.gov as #NCT00048958 [CALGB 8461], #NCT00899223 [CALGB 9665] and #NCT00900224 [CALGB 20202]). All patients provided written informed consent, and all study protocols were in compliance with the Declaration of Helsinki and approved by institutional review boards.

Cytogenetic and gene mutation analyses

Cytogenetic analyses were performed in CALGB/Alliance-approved institutional laboratories and results confirmed by central karyotype review.[41] The diagnosis of normal karyotype was based on the absence of clonal chromosome abnormalities in ≥20 metaphases obtained from BM samples subjected to 24- to 48-hour unstimulated cultures.[41]

Targeted amplicon sequencing using the Miseq platform (Illumina) was used to analyze DNA samples for the presence of gene mutations that are established prognosticators in CN-AML (ie, mutations in the *ASXL1*,[42] *DNMT3A* [R882 and non-R882],[43] *IDH1*, *IDH2* [R140 and R172],[44] *NPM1*,[45] *RUNX1*,[46] *TET2*,[47] and *WT1* genes[48] and *FLT3*–tyrosine kinase domain [*FLT3*-TKD] mutations[49]) or that occur recurrently in CN-AML (ie, *SF1*, *SF3A1*, *SF3B1*, *SRSF2*, *U2AF1*, *U2AF2*, and *ZRSR2*), as described previously.[50,51] A variant allele frequency of ≥10% was used as the cutoff to distinguish between mutated vs wild-type alleles. The presence of biallelic mutations in the *CEBPA* gene and *FLT3*–internal tandem duplications (*FLT3*-ITDs) were evaluated using Sanger sequencing and fragment analysis, respectively, as previously described.[52,53]

Transcriptome analyses

RNA samples of all patients (N = 365) were analyzed with RNA-seq (after depletion of ribosomal and mitochondrial RNA) using the Illumina HiSeq 2500 platform. Details regarding library generation protocols and sequencing are provided in the supplemental Data. To detect and quantify circRNA expression, we used a novel bioinformatic pipeline, called Massive Scan for circRNAs (MScircRNA), the features of which are presented in the Results section and in supplemental Data.

To determine the status of patients (ie, high vs low expressers) with regard to prognostic expression markers (ie, expression of the *BAALC*,[54] *MN1*,[55] and *ERG*[56] genes and expression of miR-181a[57] and miR-155[58]), the median values of normalized RNA-seq reads were used as the cutoff. With regard to miR-3151,[59] patients expressing this miR were compared with those who did not. Data from small RNAs (miR-181a, miR-155, and miR-3151) were obtained using small RNA-seq.

Statistical analyses

Clinical end point definitions are given in the supplemental Data. Baseline demographic, clinical, and molecular features were compared between the training and validation sets, circRNA-expression

based clusters, and low and high circRNA expressers using the Kruskal-Wallis and Fisher's exact tests for continuous and categorical variables, respectively. The estimated probabilities of disease-free (DFS) and overall survival (OS) were calculated using the Kaplan-Meier method, and the log-rank test evaluated differences between survival distributions. Cox proportional hazards models were used to calculate hazard ratios (HRs) for DFS, OS, and event-free survival (EFS). Multivariable proportional hazards models were constructed for DFS, OS, and EFS using a backward selection procedure.[60] Variables significant at $\alpha = 0.20$ from the univariable analyses were considered for multivariable analyses (supplemental Data provides variables considered in model inclusion). For the time-to-event end points, the proportional hazards assumption was checked for each variable individually. All statistical analyses were performed by the Alliance Statistics and Data Center using SAS 9.4 and TIBCO Spotfire S+ 8.2 software. For laboratory in vitro experiments, 2-tailed Student *t* tests were performed. $P < .05$ was considered significant.

Results

MScircRNA: a novel pipeline for the detection and quantification of circular RNAs

To detect and quantify bona fide circRNAs, we developed a novel computational method called MScircRNA. As a starting point, we assembled a list with all backsplicing events detectable in our RNA-seq data set. The list was further filtered for backsplicing events compatible with perturbed arrangement of exons of the same gene. We compiled a FASTA file containing the sequences of the putative backsplicing junctions, as well as those annotated in the circBase database[61] (supplemental Table 1). To quantify the expression levels of the detected circRNAs, we removed all linear reads and aligned the remaining reads to the curated list of candidate circRNA sequences. Fragments per kilobase pair of transcript per million mapped reads (FPKMs) were obtained by normalizing the fragment counts against the sequencing space across the backsplicing junction (supplemental Figure 1). Additional details regarding MScircRNA are provided in the supplemental Data.

Biochemical validation of the MScircRNA pipeline

To experimentally test our novel pipeline, we performed a sequencing experiment using samples from 7 CN-AML patients and 3 AML cell lines (Eol-1, OCI-AML3, and K-562) that were either mock-treated or treated with the RNase R exonuclease. RNase R preferentially degrades linear transcripts but has minimal effect on circularized transcripts and can therefore be used to differentiate between true circular RNAs and sequencing artifacts.[62,63]

We generated circRNA profiles of the mock and RNase R–treated samples using MScircRNA and identified 3031 circRNAs that were robustly expressed (FPKMs $\geq$1) in the 10 samples (supplemental Table 2). We found that RNase R treatment led to a significant FPKM increase for 2380 (78.5%) of the 3031 predicted circRNAs with respect to the untreated samples. Only 1 candidate circRNA (SNORD10) had significantly lower FPKMs in RNase R–treated AML samples (<0.01%). Regarding the remaining 650 circRNAs in which difference in expression between the 2 groups did not reach statistical significance, the trend was still that of enrichment because of resistance to RNase R for all except 10 circRNAs (10 [1.5%] of 650). Finally, in the case of 2323 circRNAs for which a related linear transcript could be detected and quantified, we calculated the ratio of the circular transcript FPKMs to the total FPKMs for the gene (circular FPKMs + linear FPKMs; ie, the circular fraction [CF]). It was recently proposed that the CF can be used as a parameter to distinguish circRNAs with functional significance from transcriptional noise.[30] We found that for 2321 of these circRNAs, RNase R treatment increased the calculated CF (supplemental Table 3; supplemental Figure 2A-B). These results indicate that a vast majority of the circRNAs identified by MScircRNA were true circularized RNA transcripts and not artifacts of sequencing.

Circular RNA expression in AML and association with host linear gene expression

To study the prognostic and biologic significance of circRNA expression in CN-AML, we conducted analyses in a cohort of 365 younger CN-AML patients using MScircRNA (Table 1 lists the patients' clinical features; supplemental Table 4 lists the RNA-seq characteristics). First, we detected and measured a total of 34 269 candidate circRNAs. To focus on the circular transcripts that were robustly and differentially expressed among the CN-AML samples, we applied stringent criteria. We retained the circRNAs, which were expressed above a threshold level (FPKM $\geq$1) and displayed variance of at least a 1.5-fold change in either direction from their corresponding median expression value. Additionally, we selected the circRNAs that had a CF >0.1 (which is the CF ratio threshold proposed to detect bona fide true circRNAs).[30] These combined filters yielded a curated list of 180 circRNAs that were used in additional analyses (supplemental Table 5).

Associations between circRNA expression and molecular features of CN-AML patients

To examine whether distinctive patterns of circRNA expression were present in the data set of younger adults with CN-AML, we performed unsupervised clustering analysis. Specifically, we used a consensus nonnegative matrix factorization method to evaluate the segregation of the patients in circRNA-based clusters, according to circular expression. Cophenetic correlation coefficients indicated that the separation of patients into 3 distinct clusters generated the most robust consensus clustering (Figure 1A). Cluster 1 consisted of 115 patients, cluster 2 of 106, and cluster 3 of 144 (Figure 1B).

Overall, there were significant differences with regard to the distribution of clinical and molecular features among patients in the 3 circRNA-based clusters. Patients in cluster 1 were older than patients in clusters 2 and 3 ($P = .02$). Furthermore, cluster 1 patients had higher white blood cell ($P < .001$) and platelet counts ($P < .001$) but lower percentages of blasts in blood ($P < .001$) than patients in the other 2 clusters. Patients in cluster 1 were also more likely to present with extramedullary manifestations of their disease ($P = .005$). With regard to recurrent prognostic gene mutations, cluster 3 was enriched for the presence of *FLT3*-ITD. Concerning composite genotypes, cluster 3 patients were more likely to harbor the *FLT3*-ITD/*NPM1*–mutated genotype (Figure 1B). Patients in cluster 1 showed enrichment for *NPM1* and *DNMT3A* mutations. The *FLT3* wild-type/*NPM1*-mutated genotype was more frequent in patients in cluster 1. Cluster 2 showed a strong enrichment for the presence of mutations in transcription factors, such as biallelic *CEBPA* and *RUNX1* mutations. *NPM1* mutations, *DNMT3A*

Table 1. Clinical and molecular features of younger adult patients with cytogenetically normal AML (N = 365)

Characteristic	Summary statistics
Age, y	
Median	46
Range	17-59
Sex, n (%)	
Male	188 (52)
Female	177 (48)
Race, n (%)	
White	324 (91)
Nonwhite	34 (9)
Hemoglobin, g/dL	
Median	9.2
Range	4.2-25.1
Platelet count, $\times 10^9$/L	
Median	56
Range	8-445
WBC count, $\times 10^9$/L	
Median	28.8
Range	0.6-475.0
Blood blasts, %	
Median	60
Range	0-97
BM blasts, %	
Median	68
Range	18-96
Extramedullary involvement, n (%)	
Present	106 (30)
Absent	250 (70)
***ASXL1*, n (%)**	
Mutated	12 (3)
Wild type	334 (97)
***CEBPA*, n (%)**	
Double mutated	54 (16)
Wild type	294 (84)
***DNMT3A*, n (%)**	
Mutated	139 (39)
R882	105
Non-R882	34
Wild type	213 (61)
***FLT3*-ITD, n (%)**	
Present	137 (38)
Absent	221 (62)
***FLT3*-TKD, n (%)**	
Present	36 (10)
Absent	314 (90)
***IDH1*, n (%)**	
Mutated	27 (8)
Wild type	325 (92)
***IDH2*, n (%)**	
Mutated	34 (10)
Wild type	318 (90)
***NPM1*, n (%)**	
Mutated	201 (58)
Wild type	143 (42)
***RUNX1*, n (%)**	
Mutated	19 (5)
Wild type	333 (95)
***SF1*, n (%)**	
Mutated	2 (1)
Wild type	350 (99)
***SF3A1*, n (%)**	
Mutated	1 (0)
Wild type	351 (100)
***SF3B1*, n (%)**	
Mutated	10 (3)
Wild type	342 (97)
***SRSF2*, n (%)**	
Mutated	11 (3)
Wild type	339 (97)
***TET2*, n (%)**	
Mutated	40 (11)
Wild type	312 (89)
***U2AF1*, n (%)**	
Mutated	4 (1)
Wild type	348 (99)
***U2AF2*, n (%)**	
Mutated	0 (0)
Wild type	352 (100)
***WT1*, n (%)**	
Mutated	39 (11)
Wild type	313 (89)
***ZRSR2*, n (%)**	
Mutated	15 (4)
Wild type	337 (96)
ELN genetic group*, n (%)	
Favorable	199 (59)
Intermediate	88 (26)
Adverse	52 (15)

ELN, European LeukemiaNet; WBC, white blood cell count.

*Among patients with CN-AML, the ELN favorable-risk category comprises patients with double-mutated *CEBPA* and patients with mutated *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}. The ELN intermediate-risk category includes patients with wild-type or single-mutated *CEBPA*, patients with wild-type *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}, and/or patients with mutated *NPM1* and *FLT3*-ITD^{high}. The ELN adverse-risk category comprises patients with wild-type or single-mutated *CEBPA* and wild-type *NPM1* with *FLT3*-ITD^{high}, patients with mutated *RUNX1* and/or *ASXL1* (if these mutations do not co-occur with favorable-risk subtypes), and/or patients with mutated *TP53*. *FLT3*-ITD^{low} is defined by an *FLT3*-ITD/*FLT3* wild-type allelic ratio of <0.5, and *FLT3*-ITD^{high} is defined by an *FLT3*-ITD/*FLT3* wild-type allelic ratio of ≥ 0.5.

†The median expression value was used as a cut point.

Table 1. (continued)

Characteristic	Summary statistics
***NPM1* and *DNMT3A*, n (%)**	
Either or neither mutated	239 (69)
Both mutated	105 (31)
***NPM1* and *FLT3*-ITD, n (%)**	
Either or neither mutated	252 (74)
Both mutated	89 (26)
***ERG* expression group†, n (%)**	
High	184 (51)
Low	178 (49)
***BAALC* expression group†, n (%)**	
High	168 (50)
Low	166 (50)
***MN1* expression group†, n (%)**	
High	176 (50)
Low	173 (50)
***miR-181a* expression group†, n (%)**	
High	148 (50)
Low	147 (50)
***miR-3151*, n (%)**	
Expressed	51 (17)
Not expressed	244 (83)
***miR-155* expression group†, n (%)**	
High	147 (50)
Low	148 (50)

ELN, European LeukemiaNet; WBC, white blood cell count.

*Among patients with CN-AML, the ELN favorable-risk category comprises patients with double-mutated *CEBPA* and patients with mutated *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}. The ELN intermediate-risk category includes patients with wild-type or single-mutated *CEBPA*, patients with wild-type *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}, and/or patients with mutated *NPM1* and *FLT3*-ITD^{high}. The ELN adverse-risk category comprises patients with wild-type or single-mutated *CEBPA* and wild-type *NPM1* with *FLT3*-ITD^{high}, patients with mutated *RUNX1* and/or *ASXL1* (if these mutations do not co-occur with favorable-risk subtypes), and/or patients with mutated *TP53*. *FLT3*-ITD^{low} is defined by an *FLT3*-ITD/*FLT3* wild-type allelic ratio of <0.5, and *FLT3*-ITD^{high} is defined by an *FLT3*-ITD/*FLT3* wild-type allelic ratio of ≥ 0.5.

†The median expression value was used as a cut point.

mutations, and *FLT3*-ITD were markedly underrepresented in cluster 2 (Table 2; Figure 1B).

The patients in the circRNA clusters also showed notable differences in expression levels of messenger RNAs (mRNAs) and miRs prognostic in CN-AML. Patients in cluster 3 were more frequently high expressers of *ERG* and miR-155, and those in cluster 2 of ERG *BAALC*, *MN1*, miR-181a, and miR-3151. In contrast, patients with low expression of prognostic mRNAs and miRNAs were significantly overrepresented in cluster 1 (Table 2; Figure 1B).

Prognostic significance of circRNA expression in younger adults with CN-AML

To evaluate the prognostic significance of circRNA expression in CN-AML, we randomly divided our cohort of younger CN-AML patients into a training set, used for exploratory analysis (n = 254), and a validation set (n = 111). There were no significant differences in clinical and molecular features between the 2 groups, except for percentages of blasts in blood (higher in training set patients; $P = .03$), frequency of *FLT3*-TKD mutations (more frequent in the training set; $P = .02$), and *ERG* ($P = .01$) and *BAALC* ($P = .002$) expression levels (for both genes, training set patients were more frequently high expressers; supplemental Table 6).

Next, we evaluated the association of circRNA expression with EFS in the training set using univariate Cox regression. EFS was chosen because it is a composite outcome end point that comprehensively evaluates response to chemotherapy, probability of relapse, and probability of survival. Twelve circRNAs were found to be significantly associated with EFS in the training set (threshold $P < .05$; supplemental Table 7). Four of these 12 circRNAs were also significantly associated with clinical outcome in the validation data set (*circCFLAR*, *circKLHL8*, *circSMC1A*, and *circFCHO2*). *circKLHL8* and *circSMC1A* showed the strongest association with clinical outcome; patients with high *circKLHL8* expression had longer DFS, OS, and EFS than patients with low *circKLHL8* expression ($P < .001$ for all 3 end points; Figure 2A-C). Patients with high *circSMC1A* had longer DFS ($P = .002$), OS ($P = .004$), and EFS ($P = .02$) than patients with low *circSMC1A* expression (Figure 2D-F). Similarly, patients with either high *circFCHO2* expression or high *circCFLAR* expression had better outcomes than those of patients with low *circFCHO2* or *circCFLAR* expression, respectively (supplemental Figure 3A-F). We did not detect an association between circRNA expression and complete remission rates in either of the data sets (data not shown).

Validation of the prognostic circRNA expression status of CN-AML

Before further studying the circRNAs, which were found to be associated with clinical outcome of younger adult CN-AML patients, we sought to validate our circRNA expression measurements and the distinction of patients in high vs low expressers with additional transcriptome profiling methods. To this end, we designed real-time quantitative polymerase chain reaction (RT-qPCR) assays that target the backsplicing regions of circRNA transcripts and can quantify their expression levels. We then measured circRNA expression in subsets of patients who had been classified as high or low expressers of the respective prognostic circRNAs based on MScircRNA analysis. There was overall high concordance between the MScircRNA and RT-PCR measurements in these patients, as indicated by the Pearson's correlation coefficient values (Pearson's *r*, 0.94-0.99; supplemental Figure 4A-D).

Associations of prognostic circRNAs and pretreatment clinical and molecular features of CN-AML patients

Next, we investigated whether the expression of the prognostic circRNAs in CN-AML patients were associated with distinctive clinical or molecular features. Patients with high *circKLHL8* expression had higher platelet counts ($P = .05$) and lower percentages of blasts in blood ($P = .002$) and BM ($P = .05$) than patients with low *circKLHL8* expression (supplemental Table 8). High *circKLHL8* expressers were also more likely to have *FLT3*-TKD ($P = .05$) and less likely to have *FLT3*-ITD ($P < .001$) or high expression of miR-155 ($P < .001$) or *ERG* ($P < .001$; supplemental Table 8). Patients with high *circSMC1A* were more likely to be female ($P = .002$) and less likely to have *FLT3*-ITD ($P = .04$) than patients

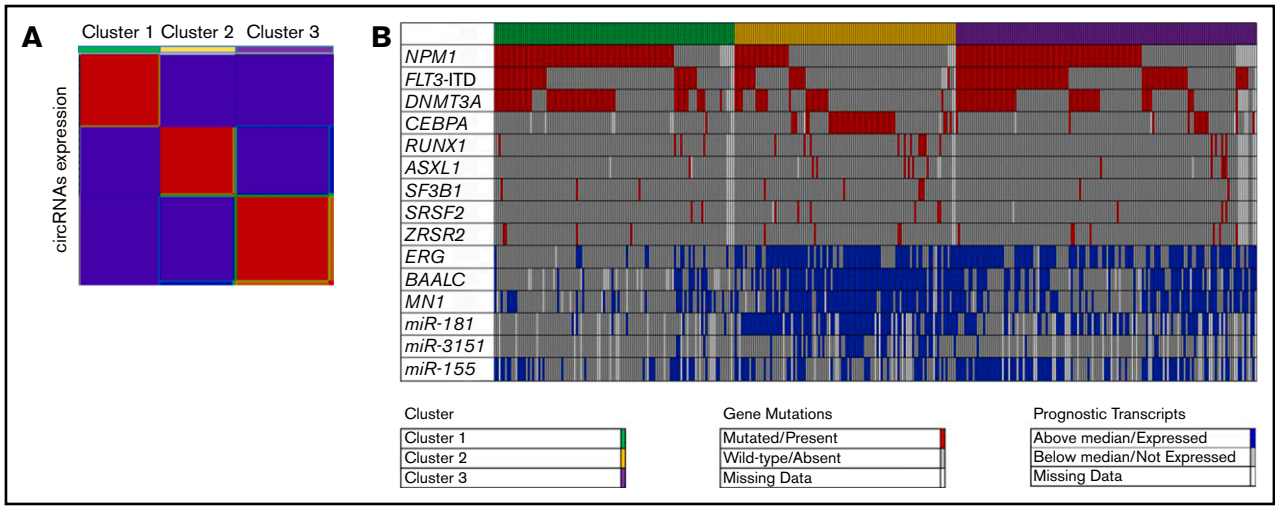

Figure 1. Unsupervised clustering of CN-AML patients based on expression of circRNAs. (A) Heatmap of circRNA expression in the 3 distinct circRNA expression–based clusters. Red indicates high and blue low expression. (B) Distribution of prognostic molecular features (ie, gene mutations, gene expression, and miR expression) across the 3 distinct circRNA expression–based clusters of younger adult patients with CN-AML.

with low *circSMC1A* expression (supplemental Table 9). The clinical and molecular features associated with high expression of *circCFLAR* and *circFCHO2* are provided in supplemental Tables 10 and 11, respectively.

Association of prognostic circRNA expression with corresponding linear transcripts

To ensure that our observations on prognostic circRNAs were not merely reflecting the prognostic significance of the corresponding linear transcripts, we examined the correlation of circular and linear RNA expression for the 4 circRNAs with validated prognostic significance. With the exception of *circFCHO2*, we found the correlation between circular and linear transcripts to be weak (range of Pearson's *r*, −0.05 to 0.29). In keeping with these findings, linear *CFLAR*, *KLHL8*, and *SMC1A* mRNAs were not prognostic in the validation set of younger adults with CN-AML. In contrast, *circFCHO2* and linear *FCHO2* were moderately correlated in the validation set (Pearson's *r*, 0.69). High linear *FCHO2* expression also was associated with better clinical outcome and longer DFS, although not as strongly as *circFCHO2* expression (supplemental Figure 5A-L).

Multivariable analyses

To evaluate the prognostic significance of circRNAs in the context of established prognostic markers, we constructed multivariable models, concomitantly analyzing expression of the 4 prognostic circRNAs. With regard to DFS, high *circKLHL8* expression remained significantly associated with longer DFS (HR, 0.53; $P = .02$) after adjusting for presence of *FLT3*-ITD and *MN1* expression status. High *circKLHL8* (HR, 0.53; $P = .01$) and high *circFCHO2* (HR, 0.45; $P = .002$) were significantly associated with longer OS after adjusting for *MN1* expression status. Finally, high *circKLHL8* and high *circFCHO2* expression were significantly associated with longer EFS (HR, 0.46; $P = .002$ and HR, 0.56; $P = .02$, respectively) after adjusting for the presence of *DNMT3A* mutations and *MN1* expression status (Table 3).

Molecular pathways associated with expression of prognostic circRNAs

To gain mechanistic insights and detect the molecular pathways associated with the expression levels of the prognostic circRNAs, we performed transcriptome and molecular pathway analyses and compared patients with high with those with low expression of the prognostic circRNAs.

We found *circCFLAR* and *circKLHL8* expression status to be associated with a high number of differentially expressed mRNAs (supplemental Tables 12 and 13, respectively). High *circCFLAR* expression was associated with increased abundance of transcripts involved in intracellular signaling (*SGK1* and *RGS2*) and regulation of gene expression (*SMAD7*, *KMD7A*, and *ID2*). In pathway analysis, high *circCFLAR* expression was found to be associated with immune system activation, leukocyte chemotaxis, and cytokine excretion (supplemental Table 14). High *circKLHL8* was associated with increased expression of cell cycle, apoptosis, or immune response regulators (*CDKN1*, *CDKN2*, *BCL6*, and *TLR4*) as well as transcription factors involved in macrophage differentiation (*CEBPD* and *CEBPB*). In pathway analysis, high *circKLHL8* expression was associated with activation of the differentiation and apoptosis pathways, as well as with cytokine production and secretion (supplemental Table 15). *CircFCHO2* and *circSMC1A* expression were not associated with distinct gene expression patterns.

Because circRNAs have been shown to function as decoys that sequester and inhibit the function of miRs, we used an in silico approach and publicly available databases (miRBase[64] and TargetScan[65]) to examine whether the prognostic circRNAs contained miR-binding regions. We did not identify repeat regions that could function as miR-sequestering sites in the nucleotide sequence of any of the prognostic circRNAs.

Biologic significance of circRNA expression in AML

To evaluate the biologic significance of circRNA expression in AML, we performed loss-of-function in vitro experiments focusing on 4

Table 2. Comparison of clinical and molecular features among 3 distinct, circRNA expression–based clusters of younger adults with CN-AML

Characteristic	Cluster 1 (n = 115)	Cluster 2 (n = 106)	Cluster 3 (n = 144)	*P*
Age, y				.02
Median	50	43	46	
Range	19-59	17-59	18-59	
Sex, n (%)				.06
Male	64 (56)	61 (58)	63 (44)	
Female	51 (44)	45 (42)	81 (56)	
Race, n (%)				.75
White	102 (90)	97 (92)	125 (89)	
Nonwhite	11 (10)	8 (8)	15 (11)	
Hemoglobin, g/dL				.17
Median	9.1	9.5	9.1	
Range	4.8-25.1	4.2-13.6	4.9-14.4	
Platelet count, $\times 10^9$/L				<.001
Median	76	51	47	
Range	9-433	8-445	8-270	
WBC count, $\times 10^9$/L				<.001
Median	42.8	15.3	29.0	
Range	1.9-308.8	0.8-223.8	0.6-475.0	
Blood blasts, %				<.001
Median	44	64	68	
Range	0-95	0-97	0-97	
BM blasts, %				.31
Median	70	65	70	
Range	18-96	19-95	21-95	
Extramedullary involvement, n (%)				.005
Present	47 (41)	23 (22)	36 (26)	
Absent	67 (59)	80 (78)	103 (74)	
***ASXL1*, n (%)**				.12
Mutated	1 (1)	6 (6)	5 (4)	
Wild type	108 (99)	95 (94)	131 (96)	
***CEBPA*, n (%)**				<.001
Double mutated	1 (1)	40 (39)	13 (9)	
Wild type	106 (99)	63 (61)	125 (91)	
***DNMT3A*, n (%)**				<.001
Mutated	63 (57)	23 (22)	53 (39)	
Wild type	48 (43)	81 (78)	84 (61)	
***FLT3*-ITD, n (%)**				<.001
Present	36 (32)	19 (19)	82 (57)	
Absent	77 (68)	83 (81)	61 (43)	
***FLT3*-TKD, n (%)**				.02
Present	18 (16)	5 (5)	13 (9)	
Absent	77 (68)	83 (81)	61 (43)	
***IDH1*, n (%)**				.11
Mutated	6 (5)	13 (13)	8 (6)	
Wild type	105 (95)	91 (88)	129 (94)	
***IDH2*, n (%)**				.20
Mutated	7 (6)	14 (13)	13 (9)	
Wild type	104 (94)	90 (87)	124 (91)	
***NPM1*, n (%)**				<.001
Mutated	86 (80)	26 (25)	89 (66)	
Wild type	22 (20)	76 (75)	45 (34)	
***RUNX1*, n (%)**				.03
Mutated	4 (4)	11 (11)	4 (3)	
Wild type	107 (96)	93 (89)	133 (97)	
***SF1*, n (%)**				.52
Mutated	1 (1)	1 (1)	0 (0)	
Wild type	110 (99)	103 (99)	137 (100)	
***SF3A1*, n (%)**				1.00
Mutated	0 (0)	0 (0)	1 (1)	
Wild type	111 (100)	104 (100)	136 (99)	
***SF3B1*, n (%)**				.32
Mutated	3 (3)	5 (5)	2 (1)	
Wild type	108 (97)	99 (95	135 (99)	
***SRSF2*, n (%)**				.22
Mutated	2 (2)	6 (6)	3 (2)	
Wild type	109 (98)	97 (94)	133 (98)	
***TET2*, n (%)**				.61
Mutated	15 (14)	12 (12)	13 (9)	
Wild type	96 (86)	92 (88)	124 (91)	
***U2AF1*, n (%)**				.11
Mutated	3 (3)	1 (1)	0 (0)	
Wild type	108 (97)	103 (99)	137 (100)	
***U2AF2*, n (%)**				NA*
Mutated	0 (0)	0 (0)	0 (0)	
Wild type	111 (100)	104 (100)	137 (100)	
***WT1*, n (%)**				.007
Mutated	6 (5)	9 (9)	24 (18)	
Wild type	105 (95)	95 (91)	113 (82)	
***ZRSR2*, n (%)**				.85
Mutated	4 (4)	4 (4)	7 (5)	
Wild type	107 (96)	100 (96)	130 (95)	
ELN genetic group†, n (%)				<.001
Favorable	75 (69)	65 (65)	59 (45)	

NA, not applicable.

**P* value could not be calculated because of 0 cell counts.

†Among patients with CN-AML, the ELN favorable-risk category comprises patients with double-mutated *CEBPA* and patients with mutated *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}. The ELN intermediate-risk category includes patients with wild-type or single-mutated *CEBPA*, patients with wild-type *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}, and/or patients with mutated *NPM1* and *FLT3*-ITD^{high}. The ELN adverse-risk category comprises patients with wild-type or single-mutated *CEBPA* and wild-type *NPM1* with *FLT3*-ITD^{high}, patients with mutated *RUNX1* and/or *ASXL1* (if these mutations do not co-occur with favorable-risk subtypes), and/or patients with mutated *TP53*. *FLT3*-ITD^{low} is defined by an *FLT3*-ITD/*FLT3* wild-type allelic ratio of <0.5, and *FLT3*-ITD^{high} is defined by an *FLT3*-ITD/*FLT3* wild-type allelic ratio of equal to or more than 0.5.

‡The median expression value was used as a cut point.

Table 2. (continued)

Characteristic	Cluster 1 (n = 115)	Cluster 2 (n = 106)	Cluster 3 (n = 144)	*P*
Intermediate	20 (18)	17 (17)	51 (39)	
Adverse	14 (13)	18 (18)	20 (15)	
***NPM1* and *DNMT3A*, n (%)**				<.001
Either or neither mutated	57 (53)	92 (90)	90 (67)	
Both mutated	51 (47)	10 (10)	44 (33)	
***NPM1* and *FLT3*-ITD, n (%)**				<.001
Either or neither mutated	83 (77)	89 (90)	80 (60)	
Both mutated	25 (23)	10 (10)	54 (40)	
***ERG* expression group‡, n (%)**				<.001
High	14 (12)	75 (71)	95 (66)	
Low	100 (88)	30 (29)	48 (34)	
***BAALC* expression group‡, n (%)**				<.001
High	15 (16)	85 (82)	68 (51)	
Low	81 (84)	19 (18)	66 (49)	
***MN1* expression group‡, n (%)**				<.001
High	29 (27)	77 (76)	70 (50)	
Low	79 (73)	24 (24)	70 (50)	
***miR-181a* expression group‡, n (%)**				<.001
High	8 (8)	81 (90)	59 (54)	
Low	87 (92)	9 (10)	51 (46)	
***miR-3151* expression group‡, n (%)**				<.001
Expressed	3 (3)	36 (40)	12 (11)	
Not expressed	92 (97)	54 (60)	98 (89)	
***miR-155* expression group‡, n (%)**				<.001
High	33 (35)	37 (41)	77 (70)	
Low	62 (65)	53 (59)	33 (30)	

NA, not applicable.

**P* value could not be calculated because of 0 cell counts.

†Among patients with CN-AML, the ELN favorable-risk category comprises patients with double-mutated *CEBPA* and patients with mutated *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}. The ELN intermediate-risk category includes patients with wild-type or single-mutated *CEBPA*, patients with wild-type *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}, and/or patients with mutated *NPM1* and *FLT3*-ITD^{high}. The ELN adverse-risk category comprises patients with wild-type or single-mutated *CEBPA* and wild-type *NPM1* with *FLT3*-ITD^{high}, patients with mutated *RUNX1* and/or *ASXL1* (if these mutations do not co-occur with favorable-risk subtypes), and/or patients with mutated *TP53*. *FLT3*-ITD^{low} is defined by an *FLT3*-ITD/*FLT3* wild-type allelic ratio of <0.5, and *FLT3*-ITD^{high} is defined by an *FLT3*-ITD/*FLT3* wild-type allelic ratio of equal to or more than 0.5.

‡The median expression value was used as a cut point.

circRNAs (*circPCMTD1*, *circKLHL8*, *circCFLAR*, and *circFBXW7*) that were either prognostic or associated with established, recurrent gene mutations in this disease. Selection of candidate circRNAs was also based on our predicted capacity to generate oligonucleotides that would target the circRNAs without perturbing the expression of the corresponding linear transcripts. We first profiled the expression of these 4 circRNAs and the corresponding linear RNA transcripts with RT-qPCR assays in 2 AML cell lines (KG-1a and OCI-AML3). In both cell lines, we found that *circKLHL8* and *circCFLAR* represented small fractions of the total amount of *KLHL8* and *CFLAR* transcripts, whereas *circFBXW7* and *circPCMTD1* were expressed at similar or higher levels than linear *PCMTD1* and linear *FBXW7*, respectively (Figure 3A-B). We confirmed the circular nature of the transcripts we studied by demonstrating their resistance to RNase R and the subsequent increase of the circular/linear transcript ratio upon RNase R treatment (Figure 3C-D). In addition, we performed PCR experiments using divergent primers that would bind to DNA regions adjacent to the backsplicing sites of the circRNAs (Figure 3E). These assays generated amplicons on reverse transcribed RNA (ie, complementary DNA) but not genomic DNA of KG-1a cells (Figure 3F; supplemental Table 16). Cloning of the amplicons into bacterial vectors and Sanger sequencing confirmed the RNAseq-predicted sequences of the circular transcripts and showed that they consisted of fused exons.

For the knockdown (KD) experiments, we used RNase H-recruiting LNA-modified oligonucleotides (gapmers) that specifically targeted the backsplicing regions of the circRNAs (supplemental Table 17). Of the 4 tested circRNAs, depletion of *circFBXW7* affected the phenotype of the leukemic cells. More specifically, *circFBXW7* KD led to an increase in the proliferative capacity of KG-1a and OCI-AML3 cells as measured by WST1 degradation ($P < .001$ and $P = .001$, respectively; Figure 3G-H). In addition, cell proliferation analysis, using BrdU, confirmed an increase in the fraction of proliferating cells upon depletion of *circFBXW7* in both KG-1a and OCI-AML3 cells ($P = .01$ in both cases; Figure 3I-J). RT-qPCR showed that the gapmers specifically targeted *circFBXW7* without affecting the expression level of linear *FBXW7* mRNA (Figure 3K-L).

To further study the functional significance of *circFBXW7*, we performed KD experiments in blasts of 3 AML patients. RT-qPCR confirmed that gapmers also induced specific depletion of *circFBXW7* in the blasts, with no effect on the level of the linear transcript (Figure 3M). In all 3 patients, *circFBXW7* KD generated a similar phenotype and led to an increase in the number of colonies formed by the blasts in methylcellulose-based assays (Figure 3N).

To evaluate whether miR sequestration could be the mechanistic basis of *circFBXW7* function, we used a bioinformatics-based approach and searched the miRBase and TargetScan databases, as described in "Molecular pathways associated with expression of prognostic circRNAs." There were no detectable miR-binding sites contained within the *cirFBXW7* sequence, arguing against an miR-dependent mechanism of action for this circRNA.

To study the molecular pathways that are associated with the expression levels of *circFBXW7*, we performed transcriptome analysis in high vs low circ*FBXW7* expressers in our cohort of younger adults with CN-AML. We found that high expression of *circFBXW7* was associated with a distinct gene expression signature and high expression of genes involved in the regulation of chromatin state and transcription (*HIST1H4F*, *HIST1H2BG*, *HISTH2AE*, and *MLLT3*) as well as key transcription factors and signal transduction mediators implicated in leukocyte activation and differentiation (*IKZF2*, *AKT3*, and *RHOBT3*; supplemental Table 18). Low expression of *circFBXW7* was associated with increased expression of known regulators of stem cell properties, such as multiple Homeobox genes (*HOXA2*, *HOXA7*, *HOXA9*, *HOXB3*, and *MEIS1*) as well as genes involved in the *WNT*/β-catenin and *NOTCH* signaling pathways (*PRICKLE1* and *JAG1*; supplemental Table 18; supplemental Figure 6). In pathway analysis, high expression of *circFBXW7* was found to be associated with positive regulation of signal transduction and leukocyte differentiation (supplemental Table 19). In contrast, decreased expression of *circFBXW7* was associated with the process of hemopoiesis and positive

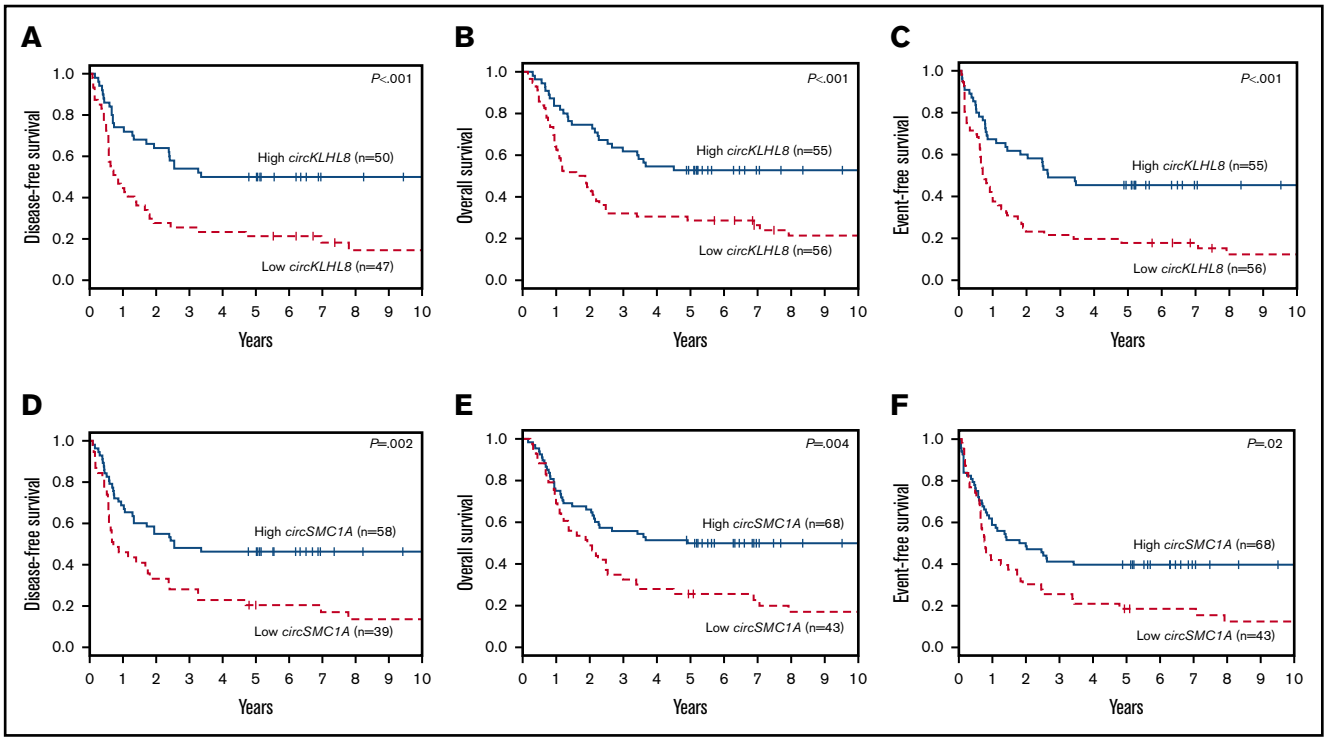

Figure 2. Clinical outcome of 111 younger adult patients with CN-AML in the validation set with either high or low expression of individual circRNAs. DFS (A), OS (B), and EFS (C) of high vs low expressers of *circKLHL8*. DFS (D), OS (E), and EFS (F) of high vs low expressers of *circSMC1A*.

regulation of RNA biosynthesis (supplemental Table 20). Collectively, these data show that circRNA expression, in addition to its association with clinical outcome and prognostic mutations, also has biologic significance in CN-AML.

Discussion

CircRNAs are a novel class of noncoding RNA molecules that are gaining increasing recognition for their functional significance in health and disease.[21-25] Although circRNAs have previously been reported to stem from chromosomal translocations that are recurrent in AML and to be involved in malignant transformation,[40] their prognostic and biologic significance in AML has not been extensively studied.

In this work, we have profiled a large number of samples from CN-AML patients using ribosomal RNA-depleted RNA-seq protocols to capture RNA transcripts independent of polyadenylation status. Most of the available RNA-seq data sets from AML patients have been generated using polyadenosine tail-based selection, because this has been shown to be the most cost-effective approach for the study of protein-coding transcripts. In contrast to these data sets, our approach

Table 3. Multivariable analyses for clinical outcome in the validation set of younger adult patients with CN-AML

	DFS		OS		EFS	
Variable in final model	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
circKLHL8, high vs low*	0.53 (0.30-0.91)	.02	0.53 (0.32-0.88)	.01	0.46 (0.28-0.76)	.002
circFCHO2, high vs low*	—	—	0.45 (0.27-0.75)	.002	0.56 (0.34-0.92)	.02
FLT3-ITD, present vs absent	2.53 (1.46-4.39)	<.001	—	—	—	—
DNMT3A, mutated vs wild type	—	—	—	—	1.93 (1.18-3.16)	.009
MN1, high vs low*	2.23 (1.31-3.80)	.003	2.63 (1.59-4.36)	<.001	2.15 (1.30-3.56)	.003

Variables considered for model inclusion were: *circKLHL8* (high vs low), *circFCHO2* (high vs low), *circCFLAR* (high vs low), *circSMC1A* (high vs low), age (as a continuous variable, in 10-y increments), sex (male vs female), race (white vs nonwhite), white blood cell count (as a continuous variable, in 50-unit increments), hemoglobin (as a continuous variable, in 1-unit increments), platelet count (as a continuous variable, in 50-unit increments), extramedullary involvement (present vs absent), *ASXL1* mutations (mutated vs wild type), *CEBPA* mutations (double mutated vs single mutated or wild type), *DNMT3A* mutations (mutated vs wild type), *FLT3*-ITD (present vs absent), *FLT3*-TKD (present vs absent), *IDH1* mutations (mutated vs wild type), *IDH2* mutations (mutated vs wild type), *NPM1* mutations (mutated vs wild type), *RUNX1* mutations (mutated vs wild type), *SF1* (mutated vs wild type), *SF3A1* (mutated vs wild type), *SF3B1* (mutated vs wild type), *SRSF2* (mutated vs wild type), *TET2* mutations (mutated vs wild type), *U2AF1* (mutated vs wild type), *U2AF2* (mutated vs wild type), *WT1* mutations (mutated vs wild type), *ZRSR2* (mutated vs wild type), *ERG* expression levels (high vs low), *BAALC* expression levels (high vs low), *MN1* expression levels (high vs low), miR-181a expression levels (high vs low), miR-3151 (expressed vs not expressed), and miR-155 expression levels (high vs low).

CI, confidence interval.

*The median expression value was used as the cut point.

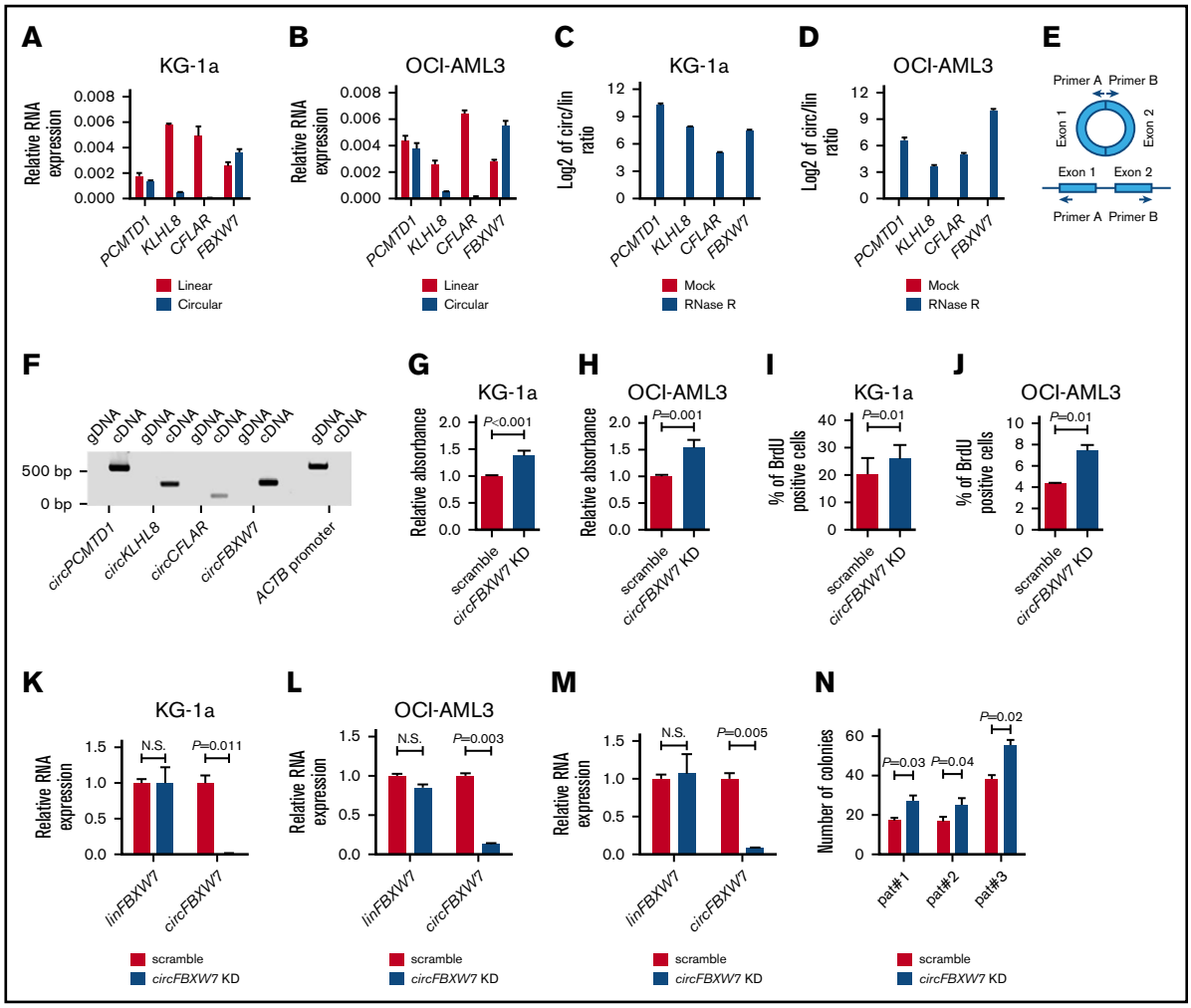

Figure 3. Functional significance of circRNA expression in AML. (A-B) Relative expression of *linPCMTD1*, *circPCMTD1*, *linKLHL8*, *circKLHL8*, *linCFLAR*, *circCFLAR*, *linFBXW7*, and *circFBXW7* in KG-1a (A) and OCI-AML3 (B) cells. (C-D) Circular-to-linear (circ/lin) ratio of the *PCMTD1, KLHL8*, *CFLAR*, and *FBXW7* transcripts in mock vs RNase R–treated samples of KG1a (C) and OCI-AML3 (D) cells. The log2 value of the circ/lin ratio is depicted in the figures. (E) Schematic description of the PCR experiments with divergent primers. The reverse primer (primer A) binds to the 5′ end of exon 1 and the forward primer (primer B) binds to the 3′ end of exon 2. With this design, amplicons can be generated after hybridization of the primers on circular RNA transcripts but not on genomic DNA. (F) *circPCMTD1*, *circKLHL8*, *circCFLAR*, and *circFBXW7* PCR results using complementary DNA and genomic DNA of KG-1a cells. The integrity of the DNA was tested with primers specific for the *ACTB* DNA promoter region. (G-H) Proliferation analyses of scramble vs *circFBXW7* KD-treated KG-1a (G) and OCI-AML3 (H) cells, as measured by a colorimetric assay, based on degradation of the WST1 reagent. (I-J) Proliferation analyses of scramble vs *circFBXW7* KD-treated KG-1a (I) and OCI-AML3 (J) cells, based on BrdU incorporation. (K-M) Relative RNA expression of *linFBXW7* and *circFBXW7* in scramble vs *circFBXW7* KD-treated KG-1a cells (K), OCI-AML3 cells (L), and AML patient blasts (M). (M) Samples of 1 patient are depicted as an example. (N) Colony-forming unit assays in methylcellulose-based media with scramble vs *circFBXW7* KD-treated AML patient blasts of 3 different patients. N.S., not significant.

allows profiling of a wide variety of noncoding transcripts in addition to mRNAs. Thus, the availability of whole-transcriptome sequencing data of CN-AML patients with detailed clinical and molecular information is an important step toward understanding noncoding RNA biology and discovering novel biomarkers and therapeutic targets.

We studied the expression of circRNAs and their association with clinical features and outcome of patients with CN-AML and performed experiments to investigate whether circRNAs are functionally relevant in this disease. We developed a new algorithm (MScircRNA) to detect and quantify circRNA transcripts from RNA-seq. We further validated in vitro the performance of our pipeline by conducting analysis on samples treated with RNase R, the gold-standard method for identifying true circular transcripts. Altogether, our results confirmed that MScircRNA is a reliable and sensitive method for circRNA detection and quantification.

Global circRNA expression profiling classified CN-AML samples of our data set into 3 groups, each with distinctive enrichment of recurrent mutations and clinical features. By using a robust training/validation approach, we identified 4 circRNAs that were associated with outcome. In multivariable analyses, high expression of 2 of these circRNAs (*circKLHL8* and *circFCHO2*) were independently associated with EFS, DFS, and OS after adjusting for other covariates.

Lastly, we conducted proof-of-principle experiments to evaluate the biologic role of circRNAs in AML. We performed a small-scale KD screening, focusing on 4 candidate circRNAs and using oligonucleotides, which targeted the circRNAs without perturbing the corresponding linear transcripts. Depletion of 1 of the candidate circRNAs, *circFBXW7*, affected the phenotype of the AML cell lines and primary AML samples that were tested and led to an increase in their proliferative capacity. On the basis of the lack of miR binding sites in *circFBXW7*, it is unlikely that these effects are mediated via sequestration of miRs. Although more experiments will be needed to understand how *circFBXW7* regulates proliferation of the leukemic blasts, our current data lend support to the functional relevance of circRNA expression in AML.

In summary, we have performed the first comprehensive circRNA profiling of a large CN-AML patient cohort. We identified distinctive clusters of circRNA expression, which are associated with recurrent mutations and clinical features of CN-AML patients. In addition, we provide evidence that 1 of the analyzed circRNAs, *circFBXW7*, acts as a tumor suppressor in AML.

Acknowledgments

The authors thank the Alliance Hematology Malignancy Biorepository and the Alliance National Cancer Institute National Clinical Trials Network (NCTN) Biorepository and Biospecimen Resource for sample processing and storage services and Lisa J. Sterling and Christine Finks (The Ohio State University Comprehensive Cancer Center, Columbus, OH) for data management.

This work was supported by the National Cancer Institute (NCI), National Institutes of Health (NIH), under awards U10CA180821 and U10CA180882 (to the Alliance for Clinical Trials in Oncology), UG1CA233338, U10CA077658, U10CA180850, U10CA180861, CA140158, CA16058, and R35 CA197734. This work was also supported in part by the Leukemia Clinical Research Foundation, D. Warren Brown Foundation, and Pelotonia Fellowship Program. S.V. was supported by the Associazione Italiana Ricerca sul Cancro. R.G. was supported by a Scholar in Clinical Research award from the Leukemia & Lymphoma Society. The Alliance Hematology Malignancy Biorepository was supported by Washington University subcontract WU-15-398/WU-16-51, and the Alliance NCTN Biorepository and Biospecimen Resource was supported by NCI NIH award U24CA196171.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Authorship

Contribution: D.P., S.V., M.Ś., C.D.B., and R.G. conceived and designed the study; D.P., S.V., D.N., K.M., C.D.B., and R.G. drafted the manuscript; D.P., S.V., D.N., M.Ś. and J.K. analyzed data; S.V. and M.Ś. contributed vital new analytical tools; A.P. and S.K. contributed vital new reagents; J.C.B., C.D.B., and R.G. obtained funding for this study; C.D.B. and R.G. supervised the study; and all authors participated in the acquisition, analysis, and interpretation of the data and in the critical revision of the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: D.N., 0000-0001-7796-0251; A.P., 0000-0002-0425-5794; F.P., 0000-0001-6507-9993; G.L.U., 0000-0002-7809-0996.

Correspondence: Ramiro Garzon, The Ohio State University Comprehensive Cancer Center, 460 West 12th Ave, Columbus, OH 43210; e-mail: ramiro.garzon@osumc.edu.

References

1. Yan X, Hu Z, Feng Y, et al. Comprehensive genomic characterization of long non-coding RNAs across human cancers. *Cancer Cell*. 2015;28(4):529-540.
2. Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet*. 2015;47(3):199-208.
3. Schwarzer A, Emmrich S, Schmidt F, et al. The non-coding RNA landscape of human hematopoiesis and leukemia. *Nat Commun*. 2017;8(1):218.
4. Jeck WR, Sorrentino JA, Wang K, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*. 2013;19(2):141-157.
5. Sanger HL, Klotz G, Riesner D, Gross HJ, Kleinschmidt AK. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc Natl Acad Sci USA*. 1976;73(11):3852-3856.
6. Grabowski PJ, Zaug AJ, Cech TR. The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of Tetrahymena. *Cell*. 1981;23(2):467-476.
7. Kos A, Dijkema R, Arnberg AC, van der Meide PH, Schellekens H. The hepatitis delta (δ) virus possesses a circular RNA. *Nature*. 1986;323(6088):558-560.
8. Capel B, Swain A, Nicolis S, et al. Circular transcripts of the testis-determining gene *Sry* in adult mouse testis. *Cell*. 1993;73(5):1019-1030.
9. Nigro JM, Cho KR, Fearon ER, et al. Scrambled exons. *Cell*. 1991;64(3):607-613.
10. Cocquerelle C, Daubersies P, Majérus MA, Kerckaert JP, Bailleul B. Splicing with inverted order of exons occurs proximal to large introns. *EMBO J*. 1992;11(3):1095-1098.
11. Hansen TB, Jensen TI, Clausen BH, et al. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013;495(7441):384-388.
12. Braun S, Domdey H, Wiebauer K. Inverse splicing of a discontinuous pre-mRNA intron generates a circular exon in a HeLa cell nuclear extract. *Nucleic Acids Res*. 1996;24(21):4152-4157.
13. Ivanov A, Memczak S, Wyler E, et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Reports*. 2015;10(2):170-177.
14. Zhang X-O, Wang H-B, Zhang Y, Lu X, Chen L-L, Yang L. Complementary sequence-mediated exon circularization. *Cell*. 2014;159(1):134-147.

15. Ashwal-Fluss R, Meyer M, Pamudurti NR, et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell*. 2014;56(1):55-66.

16. Conn SJ, Pillman KA, Toubia J, et al. The RNA binding protein quaking regulates formation of circRNAs. *Cell*. 2015;160(6):1125-1134.

17. Aktaş T, Avşar Ilık İ, Maticzka D, et al. DHX9 suppresses RNA processing defects originating from the *Alu* invasion of the human genome. *Nature*. 2017; 544(7648):115-119.

18. Wang PL, Bao Y, Yee M-C, et al. Circular RNA is expressed across the eukaryotic tree of life [published correction appears in *PLoS One*. 2014;9(4): e95116]. *PLoS One*. 2014;9(6):e90859.

19. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression [published correction appears in *PLoS Genet*. 2013;9(12)]. *PLoS Genet*. 2013;9(9):e1003777.

20. Rybak-Wolf A, Stottmeister C, Glažar P, et al. Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol Cell*. 2015;58(5):870-885.

21. 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.

22. Holdt LM, Kohlmaier A, Teupser D. Molecular roles and function of circular RNAs in eukaryotic cells. *Cell Mol Life Sci*. 2018;75(6):1071-1098.

23. Armakola M, Higgins MJ, Figley MD, et al. Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. *Nat Genet*. 2012;44(12):1302-1309.

24. Greene J, Baird A-M, Brady L, et al. Circular RNAs: biogenesis, function and role in human diseases. *Front Mol Biosci*. 2017;4(4):38.

25. Yu C-Y, Li T-C, Wu Y-Y, et al. The circular RNA *circBIRC6* participates in the molecular circuitry controlling human pluripotency. *Nat Commun*. 2017; 8(1):1149.

26. Zheng Q, Bao C, Guo W, et al. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat Commun*. 2016;7:11215.

27. Wei X, Li H, Yang J, et al. Circular RNA profiling reveals an abundant circLMO7 that regulates myoblasts differentiation and survival by sponging miR-378a-3p. *Cell Death Dis*. 2017;8(10):e3153.

28. Li Y, Zheng F, Xiao X, et al. CircHIPK3 sponges miR-558 to suppress heparanase expression in bladder cancer cells. *EMBO Rep*. 2017;18(9): 1646-1659.

29. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol*. 2014;32(5):453-461.

30. Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol*. 2014;15(7):409.

31. Li Z, Huang C, Bao C, et al. Exon-intron circular RNAs regulate transcription in the nucleus [published correction appears in *Nat Struct Mol Biol*. 2017; 24(2):194]. *Nat Struct Mol Biol*. 2015;22(3):256-264.

32. Legnini I, Di Timoteo G, Rossi F, et al. Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol Cell*. 2017;66(1):22-37.e9.

33. Pamudurti NR, Bartok O, Jens M, et al. Translation of circRNAs. *Mol Cell*. 2017;66(1):9-21.e7.

34. Abe N, Matsumoto K, Nishihara M, et al. Rolling circle translation of circular RNA in living human cells. *Sci Rep*. 2015;5:16435.

35. Yang Y, Fan X, Mao M, et al. Extensive translation of circular RNAs driven by N^6-methyladenosine. *Cell Res*. 2017;27(5):626-641.

36. Ng WL, Mohd Mohidin TB, Shukla K. Functional role of circular RNAs in cancer development and progression. *RNA Biol*. 2018;15(8):995-1005.

37. Kristensen LS, Hansen TB, Venø MT, Kjems J. Circular RNAs in cancer: opportunities and challenges in the field. *Oncogene*. 2018;37(5):555-565.

38. Li Y, Zeng X, He J, et al. Circular RNA as a biomarker for cancer: a systematic meta-analysis. *Oncol Lett*. 2018;16(3):4078-4084.

39. Du WW, Yang W, Liu E, Yang Z, Dhaliwal P, Yang BB. Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res*. 2016;44(6):2846-2858.

40. Guarnerio J, Bezzi M, Jeong JC, et al. Oncogenic role of fusion-circRNAs derived from cancer-associated chromosomal translocations [published correction appears in *Cell*. 2016;166(4):1055-1056]. *Cell*. 2016;165(2):289-302.

41. Mrózek K, Carroll AJ, Maharry K, et al. Central review of cytogenetics is necessary for cooperative group correlative and clinical studies of adult acute leukemia: the Cancer and Leukemia Group B experience. *Int J Oncol*. 2008;33(2):239-244.

42. Metzeler KH, Becker H, Maharry K, et al. *ASXL1* mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN Favorable genetic category. *Blood*. 2011;118(26):6920-6929.

43. Marcucci G, Metzeler KH, Schwind S, et al. Age-related prognostic impact of different types of *DNMT3A* mutations in adults with primary cytogenetically normal acute myeloid leukemia. *J Clin Oncol*. 2012;30(7):742-750.

44. Marcucci G, Maharry K, Wu Y-Z, et al. *IDH1* and *IDH2* gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2010;28(14):2348-2355.

45. Döhner K, Schlenk RF, Habdank M, et al. Mutant nucleophosmin (*NPM1*) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations. *Blood*. 2005;106(12):3740-3746.

46. Mendler JH, Maharry K, Radmacher MD, et al. *RUNX1* mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and microRNA expression signatures. *J Clin Oncol*. 2012;30(25):3109-3118.

47. Metzeler KH, Maharry K, Radmacher MD, et al. *TET2* mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2011;29(10):1373-1381.

48. Paschka P, Marcucci G, Ruppert AS, et al. Wilms' tumor 1 gene mutations independently predict poor outcome in adults with cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2008;26(28):4595-4602.

49. Whitman SP, Ruppert AS, Radmacher MD, et al. *FLT3* D835/I836 mutations are associated with poor disease-free survival and a distinct gene-expression signature among younger adults with de novo cytogenetically normal acute myeloid leukemia lacking *FLT3* internal tandem duplications. *Blood*. 2008;111(3):1552-1559.

50. Kroll KW, Eisfeld A-K, Lozanski G, Bloomfield CD, Byrd JC, Blachly JS. MuCor: mutation aggregation and correlation. *Bioinformatics*. 2016;32(10): 1557-1558.

51. Eisfeld A-K, Mrózek K, Kohlschmidt J, et al. The mutational oncoprint of recurrent cytogenetic abnormalities in adult patients with de novo acute myeloid leukemia. *Leukemia*. 2017;31(10):2211-2218.

52. Marcucci G, Maharry K, Radmacher MD, et al. Prognostic significance of, and gene and microRNA expression signatures associated with, *CEBPA* mutations in cytogenetically normal acute myeloid leukemia with high-risk molecular features: a Cancer and Leukemia Group B Study. *J Clin Oncol*. 2008;26(31):5078-5087.

53. Whitman SP, Archer KJ, Feng L, et al. Absence of the wild-type allele predicts poor prognosis in adult de novo acute myeloid leukemia with normal cytogenetics and the internal tandem duplication of *FLT3*: a Cancer and Leukemia Group B study. *Cancer Res*. 2001;61(19):7233-7239.

54. Langer C, Radmacher MD, Ruppert AS, et al. High *BAALC* expression associates with other molecular prognostic markers, poor outcome, and a distinct gene-expression signature in cytogenetically normal patients younger than 60 years with acute myeloid leukemia: a Cancer and Leukemia Group B (CALGB) study. *Blood*. 2008;111(11):5371-5379.

55. Langer C, Marcucci G, Holland KB, et al. Prognostic importance of *MN1* transcript levels, and biologic insights from *MN1*-associated gene and microRNA expression signatures in cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2009;27(19): 3198-3204.

56. Marcucci G, Baldus CD, Ruppert AS, et al. Overexpression of the ETS-related gene, *ERG*, predicts a worse outcome in acute myeloid leukemia with normal karyotype: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2005;23(36):9234-9242.

57. Schwind S, Maharry K, Radmacher MD, et al. Prognostic significance of expression of a single microRNA, *miR-181a*, in cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2010;28(36):5257-5264.

58. Marcucci G, Maharry KS, Metzeler KH, et al. Clinical role of microRNAs in cytogenetically normal acute myeloid leukemia: *miR-155* upregulation independently identifies high-risk patients. *J Clin Oncol*. 2013;31(17):2086-2093.

59. Eisfeld A-K, Marcucci G, Maharry K, et al. *miR-3151* interplays with its host gene *BAALC* and independently affects outcome of patients with cytogenetically normal acute myeloid leukemia. *Blood*. 2012;120(2):249-258.

60. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression Methods in Biostatistics: Linear, Logistic, Survival and Repeated Measures Models. New York, NY: Springer; 2005.

61. Maass PG, Glažar P, Memczak S, et al. A map of human circular RNAs in clinically relevant tissues. *J Mol Med (Berl)*. 2017;95(11):1179-1189.

62. Bachmayr-Heyda A, Reiner AT, Auer K, et al. Correlation of circular RNA abundance with proliferation--exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci Rep*. 2015;5:8057.

63. Suzuki H, Zuo Y, Wang J, Zhang MQ, Malhotra A, Mayeda A. Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res*. 2006;34(8):e63.

64. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res*. 2019;47(D1):D155-D162.

65. Agarwal V, Bell GW, Nam J-W, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife*. 2015;4:e5005.