Noncoding RNAs and Their Response Predictive Value in Azacitidine-treated Patients With Myelodysplastic Syndrome and Acute Myeloid Leukemia With Myelodysplasia-related Changes

MICHAELA DOSTALOVA MERKEROVA[1], JIRI KLEMA[2], DAVID KUNDRAT[1],
KATARINA SZIKSZAI[1], ZDENEK KREJCIK[1], ANDREA HRUSTINCOVA[1], IVA TRSOVA[1],
ANH VU LE[2], JAROSLAV CERMAK[3], ANNA JONASOVA[4] and MONIKA BELICKOVA[1]

[1]Department of Genomics, Institute of Hematology and Blood Transfusion, Prague, Czech Republic;
[2]Department of Computer Sciences, Czech Technical University, Prague, Czech Republic;
[3]Laboratory of Anemias, Institute of Hematology and Blood Transfusion, Prague, Czech Republic;
[4]First Department of Medicine, General University Hospital, Prague, Czech Republic

Abstract. *Background/Aim: Prediction of response to azacitidine (AZA) treatment is an important challenge in hematooncology. In addition to protein coding genes (PCGs), AZA efficiency is influenced by various noncoding RNAs (ncRNAs), including long ncRNAs (lncRNAs), circular RNAs (circRNAs), and transposable elements (TEs). Materials and Methods: RNA sequencing was performed in patients with myelodysplastic syndromes or acute myeloid leukemia before AZA treatment to assess contribution of ncRNAs to AZA mechanisms and propose novel disease prediction biomarkers. Results: Our analyses showed that lncRNAs had the strongest predictive potential. The combined set of the best predictors included 14 lncRNAs, and only four PCGs, one circRNA, and no TEs. Epigenetic regulation and recombinational repair were suggested as crucial for AZA response, and network modeling defined three deregulated lncRNAs (CTC-482H14.5, RP11-419K12.2, and RP11-736I24.4) associated with these processes. Conclusion: The expression of various ncRNAs can influence the effect of AZA and new ncRNA-based predictive biomarkers can be defined.*

This article is freely accessible online.

Correspondence to: Michaela Dostalova Merkerova, Ph.D., Institute of Hematology and Blood Transfusion, U Nemocnice 1, 128 20 Prague 2, Czech Republic. Tel: +42 0221977231, e-mail: michaela.merkerova@uhkt.cz

Myelodysplastic syndromes (MDS) are a heterogenous group of clonal malignancies characterized by ineffective hematopoiesis resulting in peripheral cytopenia and hypercellular bone marrow dysplasia. MDS patients have an increased tendency to transform to acute myeloid leukemia (AML). When the disease progresses to leukemia, prognosis is clearly unfavorable, with an estimated survival of less than one year (1). The prognosis of MDS patients is assessed based on the Revised International Prognostic Scoring System (IPSS-R) and depends on the number and severity of cytopenia, percentage of bone marrow (BM) blasts and cytogenetics (2).

Azacitidine (AZA) is a hypomethylating agent (HMA) that is currently used as a standard care for patients with higher-risk MDS or AML with 20-30% BM blasts when not eligible for intensive chemotherapy and allogeneic transplantation. AZA is a cytidine analogue that, at low doses, inhibits DNA methyltransferases, resulting in DNA hypomethylation. At high doses, AZA is directly cytotoxic to abnormal BM hematopoietic cells because of its incorporation into DNA and RNA (3). Although AZA is extensively used, 30-40% of patients fail to respond or relapse after treatment (4). The International Working Group (IWG) proposed standardized response criteria for evaluating clinically significant responses in MDS (5). Stratification of patients who will benefit from those who will not respond to AZA is essential. In addition to the high costs associated with the treatment, an important issue lies in prolonged exposure to AZA as current recommendations are for a minimum of 6 months of treatment before deeming it a failure. Therefore, proper prediction biomarkers that could potentially prevent prolonged unnecessary adverse effects in AZA-nonresponsive patients are needed.

There have been many efforts to characterize clinical, biological, or molecular parameters predictive of AZA treatment outcome (6-12). Because DNA hypermethylation can contribute to tumorigenesis by silencing tumor suppressor genes, it was supposed that the methylation pattern may be predictive of AZA response. Although both global and gene-specific hypermethylation have been performed in MDS (13, 14), little relation between the degree of demethylation following hypomethylating treatment and hematologic response has been evidenced (15). Furthermore, due to the number of somatic mutations found in regulators of DNA methylation, it has been hypothesized that such mutations may impact the response to AZA. Some studies have suggested the utility of *DNMT3A* (7, 8) and *TET2* (9, 10) mutations as predictive markers for AZA response, yet these datasets have not been largely confirmed. To date, no clinical, cytogenetic or molecular markers of AZA treatment outcome to support clinical decisions have been validated (16).

Gene expression profiling has repeatedly been performed in MDS and has identified many protein coding genes (PCGs) and signaling pathways associated with disease pathophysiology (17) and treatment (11). To provide biological insights into the contribution of noncoding RNAs (ncRNAs) to the AZA response or resistance and to propose novel appropriate molecular markers able to predict the response, whole transcriptome RNA sequencing (RNA-seq) was performed in CD34+ BM cells in patients with higher-risk MDS and AML with myelodysplasia-related changes (AML-MRC) before AZA therapy. A more detailed analysis of PCG data was not the major aim of this study, as they have already been addressed in a previous study (18). This study focused on the noncoding transcriptome and paid special attention to the examination of various classes of ncRNAs as predictive markers of the AZA response and compared their power to predict the response with the predictive power of PCGs. In addition to regular analysis of RNA-seq data that standardly examines the expression of PCGs and several classes of noncoding transcripts (mainly long noncoding RNAs, lncRNAs), the NGS data with regard to circular RNAs (circRNAs) and transposable elements (TEs) was also analysed. To the authors' knowledge, this is the first study addressing and integrating transcriptional data of these molecules in MDS.

Materials and Methods

Patients. The study included 26 patients with a diagnosis of MDS or AML-MRC and 9 healthy controls. BM samples were obtained from the patients before the first administration of AZA during routine clinical assessment at the Institute of Hematology and Blood Transfusion and the First Department of Internal Medicine of the General Faculty Hospital in Prague. The study included only patients with no known history of previous malignancy, chemotherapy or radiation therapy, and none of the patients received hematopoietic stem cell (HSC) transplantation. The patient's diagnoses were assessed based on the standard WHO 2016 classification criteria (19), and all the patients were classified according to the IPSS-R categories (2) at the time of sample collection. AZA was administered at 75 mg/m^2 subcutaneously for 7 days and repeated every 4 weeks. The response status of the patients was evaluated according to IWG criteria (5). Patients with complete remission (CR), partial remission (PR), marrow CR with hematological improvement (mCR with HI), or stable disease with HI (SD with HI) were considered responders (RESPs, N=11). Patients with SD without HI and with progressive disease (PD) accounted for nonresponders (nRESPs, N=15). The control group contained nine hematologically healthy donors (four males and five females; mean age 41 years; range=27-69 years). Written informed consent was obtained from all tested subjects and the study was approved by the Scientific Board and the Ethics Committee of the Institute of Hematology and Blood Transfusion and the General Faculty Hospital, and performed in accordance with the ethical standards of the Declaration of Helsinki. The detailed clinical and laboratory characteristics of the cohort, including the patient classification into subgroups, IPSS-R categories, AZA response, blood counts, cytogenetics, and mutational status, are summarized in Table I.

Cell separation and nucleic acid extraction. Mononuclear cells (MNCs) were separated from BM aspirates by Ficoll-Paque density gradient centrifugation (GE Health care, Munich, Germany). CD34+ cells were isolated from MNCs using magnetic cell separation system from Miltenyi Biotec (Bergisch Gladbach, Germany). The acid-guanidine-phenol-chloroform method was used to extract total RNA and DNase I (Qiagen, Hilden, Germany) treatment was applied to prevent genomic DNA contamination. Each RNA was quantified using Qubit 3 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), and RNA integrity was assessed using the TapeStation 4200 HS RNA kit (Agilent Technologies, Santa Clara, CA, USA).

RNA sequencing. For RNA-Seq, a rRNA depletion-based approach (RiboCop rRNA Depletion Kit v1.2, Lexogen, Vienna, Austria) was used followed by library construction with NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA). Each library was marked with unique indexed primers. Library quality and size were assessed using a TapeStation 4200 HS D1000 kit and quantified with a Qubit 3 Fluorimeter. Libraries were subsequently equimolarly pooled and 2 x 125 bp paired-end sequenced with a HiSeq SBS V4 kit on a HiSeq 2500 instrument (Illumina, San Diego, CA, USA).

Bioinformatical data processing. FASTQ files were subjected to initial quality control by FastQC (20), and adaptor trimming and low-quality sequence removal were performed by Trimmomatic (21). For identification of "regular" transcripts (*i.e.*, PCGs and lncRNAs), the reads were aligned to the human genome version hg38 using STAR (22). Genomic features in mapped data were then counted using the StringTie tool (23). For circRNA identification, the reads were mapped to the human genome version hg19 using BWA (24) because genomic coordinates of circRNAs in the circBase database are in hg19. The mapped data were further processed by CIRI2 software (25). The SalmonTE tool (26) was used for the detection of TEs from raw FASTQ files and for further

Table I. *Patient characteristics*.

Number of patients	26
Age (years)	69 (44-82)
Gender (male/female)	14/12
Diagnosis	
MDS-MLD	1
MDS-EB1	1
MDS-EB2	18
AML-MRC	6
IPSS-R category of MDS*	
intermediate/high/very high/NA	5/9/6/6
Medullary blasts (%)	16.7 (6.6-27.4)
Hemoglobin (g/l)	98 (66-127)
Neutrophils (×10⁹/l)	2.1 (0.1-11.8)
Platelets (×10⁹/l)	90 (20-578)
Cytogenetic features	
Normal karyotype	8
Isolated del(5q)	5
Complex	4
Other	9
IPSS-R karyotype	
Good/intermediate/poor/very poor	15/7/1/3
Somatic mutations	
No. of mutations per patient: 0/1/2/3/4/5/6	4/6/6/6/1/1/2
Most frequent mutations	
TP53	9
SF3B1	6
RUNX1	5
TET2	5
DNMT3A	4
Follow-up	
Mean follow-up from sample collection (months)	18 (4-47)
Number of deceased patients	19
AZA treatment	
Months from sample collection to AZA administration	1.3 (0-3.3)
Number of RESPs/nRESPs	11/6/9
Cycles administered	11 (3-38)

*The IPSS-R category was not assessed in AML-MRC patients. Average numbers are calculated as the mean of the cohort. MDS: Myelodysplastic syndrome; MLD: multilineage dysplasia; EB: excess blasts; AML-MRC: acute myeloid leukemia with myelodysplasia-related changes; IPSS-R: Revised International Prognostic Scoring System; AZA: azacytidine; RESPs: responders; nRESPs: nonresponders; NA: not available.

processing of TE data. All other subsequent statistical analyses were performed in the R statistical environment with appropriate Bioconductor packages. The pcaMethods (27) was applied for sample clustering. Differential expression analysis (DEA) was performed with edgeR (28) and DESeq2 (29). For DEA, only those transcripts that had at least 5 counts-per-million (CPM) in at least this number of libraries, which corresponded to half of the smallest sample group, were retained. In standard settings, only the transcripts with |logFC|>1 (absolute value of binary logarithm of fold change) and FDR<0.05 (false discovery rate; *p*-value adjusted for multiple testing) were considered significantly deregulated.

Three different supervised machine learning methods were used for the identification of a subset of transcripts applicable as the best prognostic classifiers of patient response. The methods that can cope with a large number of features and are not prone to overfitting were

selected. All the methods had been applied to RNA-seq data classification. In particular, the following methods were employed: support-vector machines with recursive feature elimination (SVM-RFE) (30), lasso and elastic-net regularized generalized linear models (GLMNET) (31), and variance modeling at the observational level (voom) extensions of the nearest shrunken centroids (voomNSC) (32). Considering unbalanced classes (the number of RESP and nRESP patients differs) the area under the ROC curve (AUC) was used for the evaluation of classifier performance. To assess the generalization ability of the classifiers, the evaluation was performed with the aid of 10 times repeated 10-fold cross-validation.

To define a formula calculating the prediction score of a patient based on the weight of the individual selected transcripts, GLMNET was applied to fit generalized linear models via penalized maximum likelihood. When the response was binary (RESP from nRESP patients in our case), it fit a logistic regression model:

$$\text{Prediction score} = \log \frac{Pr(RESP|X=x)}{Pr(nRESP|X=x)} = \beta_0 + \beta^T x$$

where β_0 stands for intercept, β for regression coefficients of individual transcripts (T), and x for an expression vector of a patient. The prediction score is equal to a logarithm of patient response odds; thus, a score >0 predicts future AZA response, and a score <0 suggests AZA failure.

Pathway analysis was conducted using the Gene Set Enrichment Analysis (GSEA) method (33). As a reference, the c5 (c5.all.v7.0.symbols.gmt [Gene Ontology, GO]) gene set from the Molecular Signatures Database was utilized. Computational modeling was applied for the construction of coexpressional networks as previously described (34). Briefly, ncRNA network analysis was implemented as follows. The GO terms significantly enriched in deregulated PCGs were identified. Then, ncRNAs (from all three classes of noncoding transcripts: lncRNAs, circRNAs, and TEs) were found whose expression profiles were most correlated with the PCGs associated with the given GO term. Specifically, ncRNAs were scored with preranked GSEA, and the enrichment score statistic was calculated based on absolute Spearman correlation between the ncRNA under consideration and the set of PCGs associated with the GO term. Eventually, a network was displayed that included nodes corresponding to the highest scoring ncRNAs and the core PCGs representing the given GO term. The core PCGs were those that most often contributed to the enrichment signal for the highest scoring noncoding transcripts. The edges between the network nodes represented the upper quartile correlations (bold line) and the above median correlations (regular line).

Results

Characterization of transcriptomic data. Deep RNA-seq was performed on 35 samples obtained from CD34+ BM cells isolated from 26 patients (clinical details are summarized in Table I) and 9 healthy controls and, on average, 75,8 M (range=56.8-99.9 M) of raw pair-end reads per sample were detected. The output data were aligned to hg38 and, on average, 87% reads per sample uniquely mapped to the genome. Overall, 19,873 PCGs and 38,343 ncRNAs were detected with at least one read across all the samples. Using the CIRI2 software, the data were further realigned to hg19
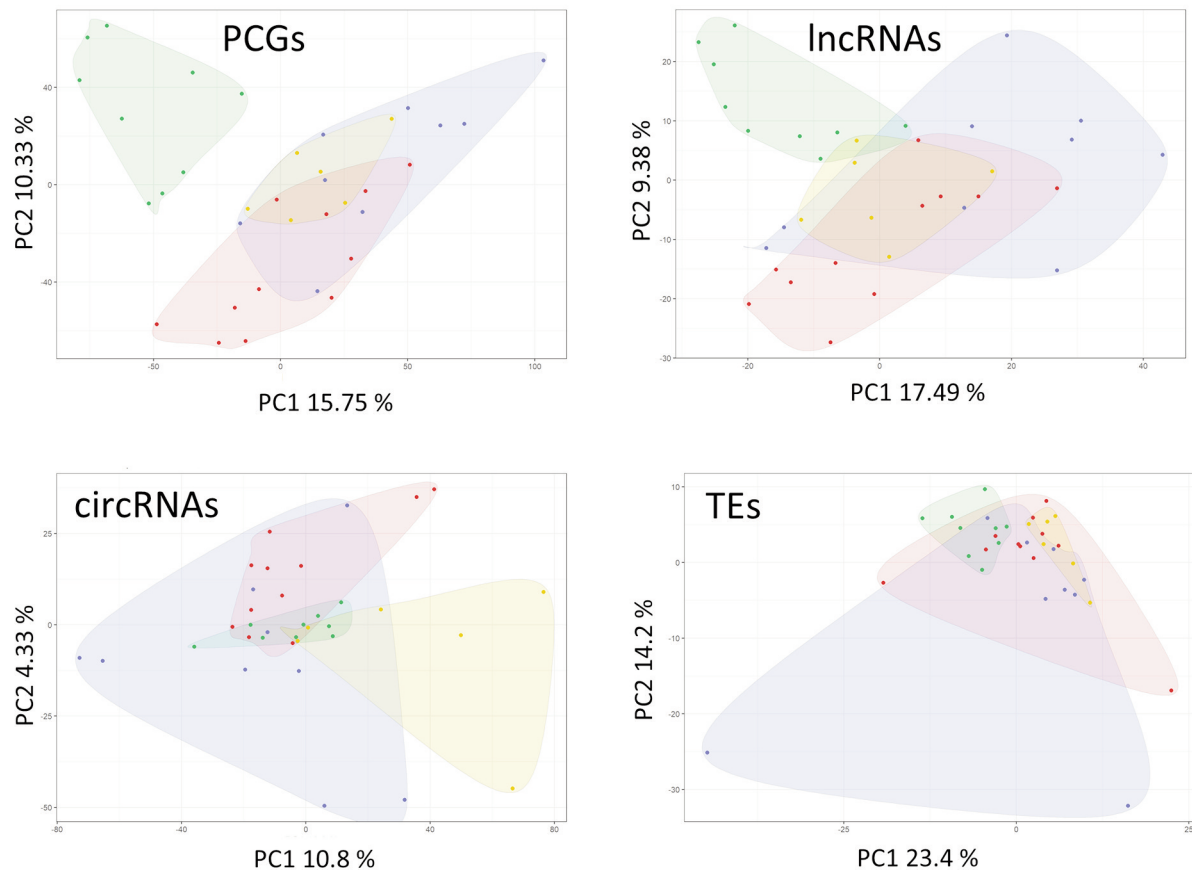
Figure 1. *Principal component analysis (PCA). Clustering of samples was performed separately with four different types of RNA-seq output data. PCGs: Protein coding genes, lncRNAs: long noncoding RNAs, circRNAs: circular RNAs; TEs: transposable elements; green – healthy controls; red – responders, yellow – patients with stable disease without hematological improvement, blue – patients with progressive disease.*

and mapped the sequences to the circBase database containing 58,216 circRNAs. On average, 3.7% of reads per sample mapped to circRNAs and 49,280 circRNAs were identified across all the samples. Finally, data analysis with the SalmonTE tool showed that 5.7% of raw reads mapped to TE sequences. Of the 687 elements utilized by the tool, 627 TEs were detected across all samples.

Only the transcripts that had ≥5 CPMs at least in any 5 samples were considered as expressed (12,330 PCGs, 2,059 ncRNAs, 6,015 circRNAs, and 554 TEs) and were therefore used for subsequent data analyses. The majority of the ncRNAs were composed of antisense RNAs (20%, N=418) or long intergenic noncoding RNAs (lincRNAs, 15%, N=375). Therefore, for the purpose of simplification, the "lncRNA" term was used for this category of transcripts within the paper. Regarding circRNAs, the majority of them were of exonic origin (94%, N=5,628) whereas only a minority of them were intronic (4%, N=233) or intergenic (2.5%, N=145). Within TE data analysis, the expression of 410 LTR (68%; *e.g.*, 235 ERV1, 41 ERV2, and 133 ERV3)

and 196 non-LTR (32%; *e.g.*, 166 L1 and 70 SINE) retrotransposons was detected.

Disease-specific transcriptome. In the initial step, the patient-specific transcriptome must be characterized. Although this was not a major aim of this study, it was necessary to explore the general molecular background in disease pathogenesis to better understand AZA-related changes. Additionally, there was almost no information about circRNA and TE transcription in MDS. This data was therefore of particular interest.

Four PCA clustering analyses (*i.e.*, separate analyses for PCGs, lncRNAs, circRNAs, and TEs) were performed to compare the contribution of different transcript groups to the disease pathogenesis. Based on PCA, the degree of variability was evaluated in gene expression among the samples. All four PCA analyses showed that healthy controls grouped into a compact homogenous sample cluster whereas expression in patient samples was more heterogeneous and future response status cannot be predicted based on any type

Table II. *The top deregulated transcripts between patients with myelodysplastic syndromes/acute myeloid leukemia with myelodysplasia-related changes and healthy controls from each RNA category defined by differential expression analysis.*

	PCGs		lncRNAs		circRNAs		TEs	
	Up	Down	Up	Down	Up	Down	Up	Down
1	COL4A5	ARPP21	MEG3	RP13-685P2.5	chr2:116066814-116101488 (DPP10)	chr3:35721278-35732497 (hsa_circ_0001281; ARPP21)	X5A_LINE	LTR87*
2	DSC2	PAX5	MIRLET7BHG	RP11-424C20.2	chr10:31661946-31676195 (hsa_circ_0000228; ZEB1)	chr17:81042813-81043199 (METRNL)	MLT2F	LTR80B*
3	CASS4	EBF1	RP6-159A1.4	AOX2P	chr3:171969049-172028671 (hsa_circ_0003692; FNDC3B)	chr5:157744372-157786654 (hsa_circ_0001548)	LOR1b_LTR	LTR18A*
4	SLITRK5	AKAP12	MT-RNR2	RP11-374F3.4	chr7:37136223-37172839 (hsa_circ_0079884; ELMO1)*	chr5:158368701-158368987 (EBF1)	MIR3	L1M3C_5*
5	LGALSL	POU2AF1	MT-RNR1	SCARNA12	chr5:50073902-50093067 (hsa_circ_0072437; PARP8)*	chr5:143042459-143052262 (AC008696.2)	MIRb*	LTR68*
6	CD5L	RAG1	TRG-AS1	AOX3P	chr20:3888572-3893281 (hsa_circ_0006873; PANK2)*	chr5:158500403-158526471 (EBF1)	PABL_BI**	MLT1I*
7	TPSB2	MME	RP11-398F12.1	RP11-124N14.4	chr1:54506428-54509198 hsa_circ_0012634 (TMEM59)*	chr5:157744372-157745153	X21_LINE*	L1ME4*
8	LOXL4	AREG	RP11-262H14.1	IGHD	chr2:230701563-230744844 (hsa_circ_0007381; TRIP12)*	chr10:125798030-125806240 (hsa_circ_0000264; CHST15)	LTR06*	LTR26E*
9	HBA1	CD79A	RP11-541N10.3	RP11-861A13.4	chr8:86193458-86193868*	chr1:2234416-2236024 (hsa_circ_0007120; SKI)	LTR75*	MER57E3*
10	MFAP3L	NEIL1	AC093627.10	AC002454.1	chr21:36206706-36259393*	chr16:16101672-16196574 (ABCC1)	HERV-K14CI*	L2*

The top 10 up-regulated (UP) and 10 down-regulated (DOWN) transcripts with the highest logFC values (FDR<0.05) in patients compared to controls are listed in each row. *Transcripts that reached only raw *p*<0.05.

of genome-wide transcriptomic data. Interestingly, patients were most clearly differentiated from healthy controls based on the levels of PCGs and lncRNAs, yet information about circRNA or TE expression was not sufficient for stratification of patients and controls (Figure 1).

In the next step, DEA was performed between samples obtained from patients and those from normal individuals. All patient samples were included in these analyses, irrespective of their future AZA response status. Table II summarizes the most deregulated transcripts of each RNA category in MDS/AML-MRC patients compared to healthy controls.

Regarding PCGs, DEA revealed 457 significantly deregulated transcripts (|logFC|>1, FDR <0.05). In patients, 331 PCGs were up-regulated (based on logFC, the 12 most up-regulated genes were *COL4A5, DSC2, CASS4, SLITRK5, LGALSL, CD5L, TPSB2, LOXL4, HBA1, MFAP3L, HBA2,* and *CANCB4*) and 126 PCGs were down-regulated (the 12

most down-regulated genes were *ARPP21, PAX5, EBF1, AKAP12, POU2AF1, RAG1, MME, AREG, CD79A, NEIL1, CD24,* and *LEF1*). These differences were comparable to those published in earlier gene expression profiling studies (34-37). For example, down-regulation of *LEF1* significantly correlated with clinical outcome and might be promising for assessing prognosis in MDS (37). Importantly, the similarity of the PCG data presented in this study with the previous findings proved the validity of our RNA-seq outputs. GSEA further revealed that the PCG deregulation in the patients was predominantly associated with epigenetic regulation of gene expression (normalized enrichment score (NES)=2.02, *p*=0.002) and chromatin silencing (NES=1.97, *p*=0.002). Core enrichment genes for these two pathways included multiple histone proteins, especially those from the *HIST1H* and *HIST2H* families. Other significantly deregulated pathways included DNA ligation involved in DNA repair (NES=1.82; *p*<0.001; core enrichment genes: *HMGB1/2, LIG1/3,*
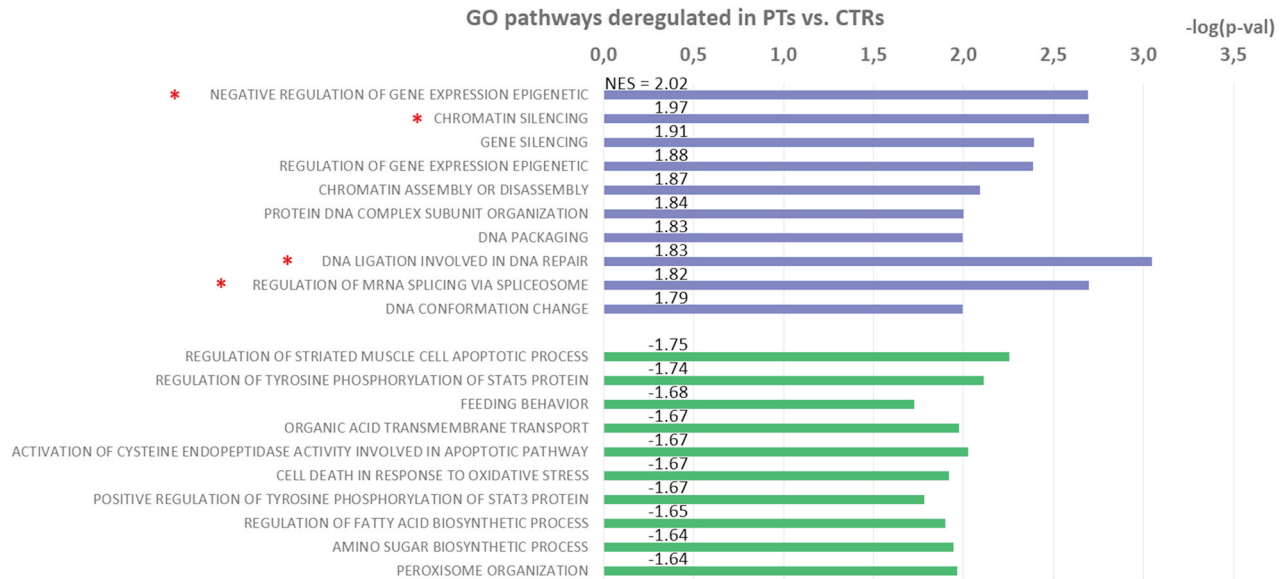
Figure 2. *Continued*

PARP1/2, XRCC1, etc.) and regulation of alternative mRNA slicing via spliceosomes (NES=1.82; *p*=0.002; core enrichment genes: splicing factors such as *SRSF2-4/6/9-10* and *SF3A1*, *etc*.). All these pathways were suppressed in patient samples compared to healthy controls, suggesting that transcriptional regulation and reparatory processes were disrupted at multiple levels in myelodysplasia. GSEA output is summarized in Figure 2. Because PCG profiling has repeatedly been performed in MDS and many signaling pathways associated with disease pathophysiology have been described (34-37), the expression of various classes of noncoding transcripts was a focus of this study.

Among the lncRNAs identified by standard RNA-seq data analysis, 80 differentially expressed transcripts (|logFC|>1, FDR<0.05; 64 up-regulated/ 16 down-regulated, Figure 3A) were found. *MEG3* and *MIRLET7BHG* were the most up-regulated lncRNAs and *RP13-685P2.5* and *RP11-424C20.2* were the most down-regulated lncRNAs (Figure 3B). Interestingly, increased expression of *MEG3* has already been associated with poor outcome in high risk MDS (38). *MIRLET7BHG* is a precursor gene for let7b miRNA which is also up-regulated in MDS (39).

Furthermore, circRNA data were analyzed by DEA. Of the 6,015 circularized transcripts used for the analysis, a significant change (|logFC|>1, FDR<0.05) was observed in the level of 23 circRNAs (3 up-regulated/20 down-regulated). Up-regulation was detected of circRNAs spliced from *DPP10*, *ZEB1*, and *FNDC3B* genes and down-regulation of circRNAs processed from *ARPP21*, *METRNL*, *EBF1*, *SKI*, *ABCC1*, and *AKAP12* genes. More detailed information about the deregulated circRNAs, including their sequences and localizations in the genome, is shown in Table III.

DEA performed on TE data showed that the most significantly deregulated clades of TEs between patient and healthy control samples were ERV1 (mean logFC=0.14, *p*=9.34E-09), ERV3 (mean logFC=0.15, *p*=5.74E-04), and SINE (mean logFC=–0.12, *p*=1.27E-03) (Figure 4A). At the level of individual elements, only four of them were identified with significant deregulation (|logFC|>1 and FDR<0.05): up-regulation of *LOR1b_LTR* (ERV1 clade), *MLT2F* (ERV3 clade), *X5A_LINE* (CR1 clade), and *MIR3* (SINE clade (Figure 4B). However, 80 TEs were significant at the cutoff level FDR <0.05, and the illustrative heatmap in Figure 4C shows their levels in the patient and control samples.

Pre-treatment transcriptional differences related to future AZA response. Following the collection of BM samples, all tested patients started to receive AZA therapy. To assess expression differences in various groups of transcripts with relation to their predictive power, patients were divided into two groups of samples, RESP (N=11) and nRESP (SD and PD, N=16) and DEA was performed. Table IV shows lists of the most deregulated transcripts from each category. Overall, DEA between RESPs and nRESPs identified significant (|logFC|>1, FDR <0.05) deregulation of 202 PCGs (111 up-regulated/91 down-regulated in RESP), 34 lncRNAs (34 down-regulated), 21 circRNAs (4 up-regulated/17 down-regulated), and down-regulation of one TE. The three PCGs most up-regulated in RESPs were *SEC14L5*, *TUBB1*, and *TREML1*, and the most down-regulated PCGs were
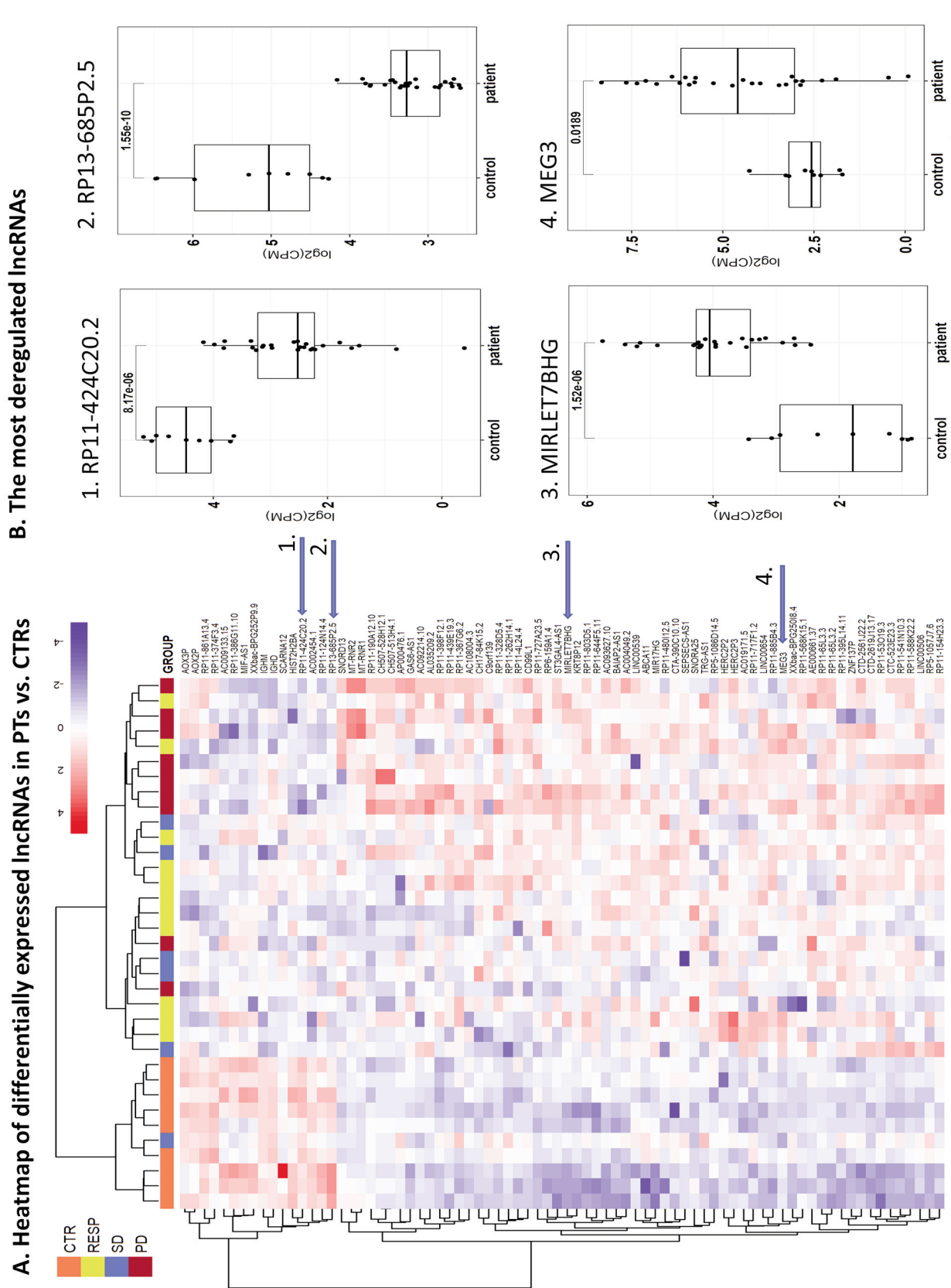
Figure 2. *Pathway analysis of differentially expressed genes in patients with myelodysplastic syndromes/acute myeloid leukemia with myelodysplasia-related changes compared to healthy controls. A plot of gene set enrichment analysis results shows p-values (logarithmically scaled) and normalized enrichment scores (NES) of the top 10 most enriched pathways in each group of samples sorted according to descended NES values. The four most significant gene sets (marked by asterisks; negative regulation of gene expression - epigenetic; chromatin silencing; DNA ligation involved in DNA repair; and regulation of alternative mRNA slicing via spliceosome) were selected for detailed visualization of the data. Enrichment plots and expression heatmaps of core enrichment genes are shown for these four gene sets.*

Figure 3. *Differentially expressed lncRNAs in patients with myelodysplastic syndromes/acute myeloid leukemia with myelodysplasia-related changes compared to healthy controls. (A) Heatmap of differentially expressed transcripts (|logFC|>1, FDR<0.05, N=80). (B) Boxplots of the four most deregulated lncRNAs.*

Table III. *Differentially expressed circRNAs between patients with myelodysplastic syndromes/acute myeloid leukemia with myelodysplasia-related changes and healthy controls (|logFC|>1, FDR<0.05).*

Genomic locus	circBase ID	Gene	Type	Strand	logFC	CPM (CTR)	CPM (PTS)	FDR
circRNAs up-regulated in patients								
chr2:116066814-116101488	-	*DPP10*	Exon	+	5.1	0.0	185.1	1.14E-03
chr10:31661946-31676195	hsa_circ_0000228	*ZEB1*	Exon	+	2.7	120.8	847.8	2.45E-04
chr3:171969049-172028671	hsa_circ_0003692	*FNDC3B*	Exon	+	2.6	98.2	614.2	1.95E-04
CircRNAs down-regulated in patients								
chr3:35721278-35732497	hsa_circ_0001281	*ARPP21*	Exon	+	–4.7	231.0	3.8	1.59E-05
chr17:81042813-81043199	-	*METRNL*	Exon	+	–4.7	789.5	16.7	1.89E-17
chr5:157744372-157786654	hsa_circ_0001548	-	Intergenic	+	–4.5	279.3	7.6	1.09E-05
chr5:158368701-158368987	-	*EBF1*	Intron	-	-4.3	294.2	9.5	9.69E-06
chr5:143042459-143052262	-	AC008696.2	Exon	+	–4.2	305.0	11.3	9.69E-06
chr5:158500403-158526471	-	*EBF1*	Exon	-	–4.2	484.8	27.7	2.24E-08
chr5:157744372-157745153	-	-	Intergenic	+	–3.9	240.9	11.0	3.57E-04
chr10:125798030-125806240	hsa_circ_0000264	*CHST15*	Exon	-	–3.7	219.0	11.6	1.31E-03
chr1:2234416-2236024	hsa_circ_0007120	*SKI*	Exon	+	–3.6	292.9	18.7	1.52E-04
chr16:16101672-16196574	-	*ABCC1*	Exon	+	–3.6	310.5	19.7	1.40E-04
chr6:151669845-151674887	hsa_circ_0078299	*AKAP12*	Exon	+	–3.6	715.9	65.4	6.58E-09
chr2:55252221-55254621	hsa_circ_0054597	*RTN4*	Exon	-	–3.5	132.7	4.7	2.49E-02
chr6:90916287-90981660	hsa_circ_0001626	*BACH2*	Exon	-	–3.0	248.6	24.9	6.93E-03
chr1:230313963-230339036	hsa_circ_0000192	*GALNT2*	Exon	+	–2.9	186.8	15.9	2.97E-02
chr5:142779220-142780417	hsa_circ_0001543	*NR3C1*	Exon	-	–2.4	701.5	121.6	1.09E-05
chr17:2297336-2298748	-	*MNT*	Exon	-	–2.4	280.0	47.7	2.92E-02
chr18:77170402-77171501	hsa_circ_0048023	*NFATC1*	Exon	+	–2.3	342.7	62.6	1.36E-02
chr2:242644067-242651486	hsa_circ_0001124	*ING5*	Exon	+	–2.0	588.4	140.1	1.15E-03
chr3:59997096-59999990	-	*FHIT*	Intron	-	–1.9	1047.6	276.2	1.03E-04
chr18:77455224-77464917	hsa_circ_0006209	*CTDP1*	Exon	+	–1.8	550.7	139.5	7.99E-03

CPM: Counts per million.

ADGRB3, *LRAT*, and *ARHGAP20*. Interestingly, several of the most up-regulated PCGs were involved in megakaryocyte differentiation (*e.g.*, *TUBB1*, *TREML1*, *PPBT*, *ITGB3*, and *THBS1*). Detailed pathway analysis of PCG results is described below in a separate chapter. Among the lncRNAs with significantly altered levels between RESPs and nRESPs (details included in Table V), four lincRNAs were found from the 21p11.2 chromosomal locus (*CH507-513H4.3-6*) and several other lncRNAs from different categories (lincRNAs, antisense RNAs, rRNAs, miRNA precursors, *etc.*), the majority of which are completely unannotated (their details are included in Table VI). Regarding deregulated circRNAs, significant deregulation was observed in a circRNA processed from *KDM1A* histone demethylase gene, a circRNA from *MED21* (transcriptional regulator of RNA polymerase II), two circRNAs processed from zing-finger transcription factors (*ZNF208* and *ZNF608*), and circRNAs backspliced from transmembrane proteins (*SLC8A1/3*, *PLXNC1)*, etc. Among TEs, only L1 element *L1PBB_5* showed significant down-regulation between RESPs and nRESPs, but some other TEs were moderately altered at raw *p*<0.05 (Table IV).

Molecular markers for the prediction of patient response to AZA. Based on differences in gene expression levels, this study aimed to predict future responses to AZA treatment, which can help us individualize patient care in routine clinical practice. Although DEA was effective in the identification of individual transcripts stratifying RESPs from nRESPs, machine learning methods also take into consideration relations among molecules and might thus eliminate similarly expressed genes from the final set of predictive markers. A classifier combining several independent molecular markers of diverse characteristics should provide more reliable classification of patients compared to simple information on an individual gene level. In this study, a machine learning approach was used and three different classification algorithms (SVM-RFE, voomNSC, and GLMNET) were applied to identify the most powerful set of molecular markers. Using all datasets together (*i.e.*, PCGs, lncRNAs, circRNAs, and TEs), the highest classification accuracy was clearly reached by GLMNET (Figure 5) and this algorithm was therefore chosen for further investigations. The GLMNET performance graph suggests that the classifier was able to

DEA: Transposable elements in MDS/AML patients vs. healhy donors

A. Differential expression of TEs sorted by clades

C. Boxplots of the most significantly deregulated TEs
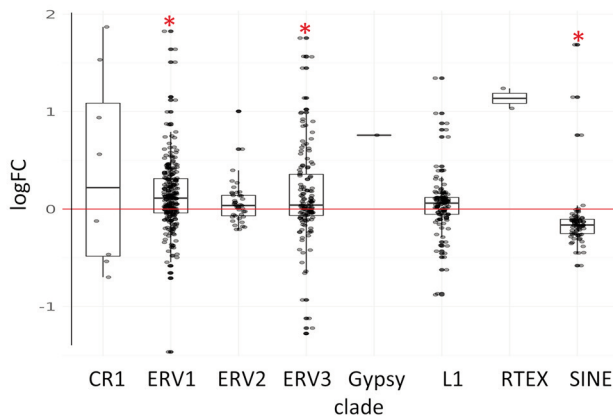
B. The most significantly deregulated TEs
(|logFC|>1, FDR≤0.05)

name	class	clade	logFC	FDR
LOR1b_LTR	LTR Retrotransposon	ERV1	1.15	6.97E-04
MLT2F	LTR Retrotransposon	ERV3	1.75	6.97E-04
X5A_LINE	Non-LTR Retrotransposon	CR1	1.87	1.29E-03
MIR3	Non-LTR Retrotransposon	SINE	1.15	4.80E-02
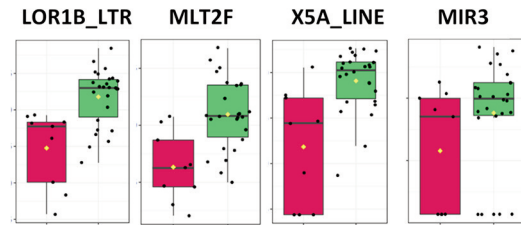
D. Heatmap of the TEs deregulated at FDR<0.05
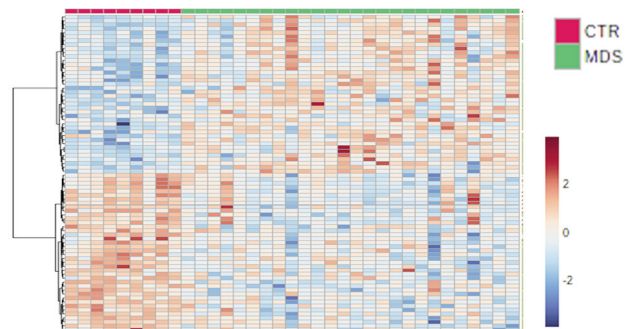
Figure 4. *Deregulation of transposable elements in patients with myelodysplastic syndromes/acute myeloid leukemia with myelodysplasia-related changes and controls. (A) Boxplot showing deregulation (logFC) of different clades of TEs. (B) List of the four most significantly (|logFC|>1, FDR<0.05) deregulated transposable elements and (C) their expression boxplots. (D) Illustrative heatmap of 80 transposable elements deregulated with the cutoff FDR<0.05.*

reasonably distinguish between RESPs and nRESPs because the observed AUC areas were significantly higher than the value of 0.5 (corresponding to a random classification), and a significant difference could be observed for a broad scale of signature sizes. The GLMNET classifier also reached the optimum performance for a limited signature size of 100 transcripts (AUC=0.85) and its performance did not further grow with its extension. This observation led us to the conclusion that it was possible to distinguish between the patient classes with a relatively small set of transcripts.

In addition to the choice of an appropriate data-classification algorithm, contribution of different transcript categories to the response prediction was tested by evaluating each transcript category separately. Surprisingly, lncRNAs overperformed the other types of molecules, as they provided the best classification power between RESPs from nRESPs (AUC=0.83 defined by GLMNET) (Figure 6A).

After optimization of the methodology, GLMNET was employed to identify a particular small set of the most powerful markers. A compromise was reached between the signature size (the number of used transcripts) and its

performance, and the size was kept as small as possible while approaching the best possible AUC of 0.85 observed earlier. The final set contained 19 transcripts, of which 14 were lncRNAs, only 4 PCGs and 1 circRNA, and none TE (Figure 6A and B, Table VII). This distribution reinforces the earlier observation that lncRNA molecules represented the core of the RESP/nRESP classification signature. Furthermore, a formula calculating the prediction score of a patient was defined based on the weights of the 19 transcripts selected by GLMNET. The regression score was calculated as follows:

Prediction score=−12.43+βTx

The score uses expression values (x) and regression coefficients (β) for these 19 transcripts (T) (β coefficients can be found in Table VII). When simplified and binarized, a score >0 predicts that the tested patient will respond to AZA, whereas a score <0 suggests future AZA failure.

Pathway analysis of transcripts associated with the AZA response. To explore cellular processes associated with patient

Table IV. *The top deregulated transcripts between responders and nonresponders from each RNA category defined by differential expression analysis.*

	PCGs		lncRNAs		circRNAs		TEs	
	Up	Down	Up	Down	Up	Down	Up	Down
1	SEC14L5	ADGRB3	RP11-520F9.2*	C20orf200	chr2:40655612-40657441 (hsa_circ_0005232; SLC8A1)	chr1:23376879-23385660 (hsa_circ_0003889, KDM1A)	MLT1H2*	L1PBB_5
2	TUBB1	LRAT	RP11-520F9.1*	MT-RNR2	chr7:149191294-149318263 (ZNF746)	chr12:94603365-94613939 (PLXNC1)	LTR24*	X21_LINE*
3	TREML1	ARHGAP20	IL8RBP*	CH507-513H4.6	chr4:105439733-105440611 (hsa_circ_0124919)	chr22:17117929-17119630 (TPTEP1)	LTR2*	MER66B*
4	PPBP	ABI3BP	CTB-158D10.3*	CH507-513H4.4	chr2:40655612-40657444 (hsa_circ_0000994; SLC8A1)	chr21:9827155-9827307	LTR41*	MLT2F*
5	LGALSL	LRCH2	AC115617.2*	CH507-513H4.3	chr1:12638745-12639440 (hsa_circ_0010023; DHRS3)*	chr5:124036706-124036962 (hsa_circ_0001523; ZNF608)	LTR40B*	LTR77*
6	ITGB3	RP5-900K19.2	MEG8*	CH507-513H4.5	chr6:17671287-17672025 (NUP153)*	chrY:22669237-22683186	MER9B*	LTR38A1*
7	THBS1	PYGO1	RP11-1143G9.4*	RNA5-8SN4	chr3:59908071-59999990 (FHIT)*	chr19:22157530-22171711 (ZNF208)	AluYd8*	LTR2B*
8	PRLR	CNTN1	RP11-354E11.2*	RNA5-8SN3	chr2:47177501-47206046 (hsa_circ_0008101; TTC7A)*	chr6:77981056-77983024	L1MB6_5*	L2B*
9	HPSE	RASSF6	CTD-2540B15.13*	RNA5-8S5	chr4:178272533-178274882 (NEIL3)*	chr4:183812555-183836728 (hsa_circ_0002212; DCTD)	MLT1D*	HARLEQUIN*
10	HBG2	NLGN1	FAM45B*	RP11-259P6.2	chr1:118003110-118045592 (hsa_circ_0002059; MAN1A2)*	chr12:27183090-27185614 (MED21)	-	MER74A*

The top 10 up-regulated (UP) and 10 down-regulated (DOWN) transcripts with the highest logFC values (FDR<0.05) in responders compared to nonresponders are listed in each row. *Transcripts that reached only raw *p*<0.05.

Table V. *Differentially expressed lncRNAs between azacitidine responders and nonresponders (|logFC|>1, FDR<0.05, all of them were down-regulated in RESP patients).*

lncRNA type	lncRNAs significantly down-regulated in responders
lincRNAs	AC007092.1, AP000476.1, AP001171.1, CH507-513H4.3, CH507-513H4.4, CH507-513H4.5, CH507-513H4.6, RP11-1195F20.2, RP11-266K4.14, RP11-77P16.4, RP1-90L6.2
Antisense RNAs	C20orf200, CTD-2516F10.2, RP11-190A12.10
miRNA precursors	MIR3687-1, MIR6724-1/2/3
rRNAs/mitochondrial-rRNAs	MT-RNR1/2, MT-TI, RNA5-8S5, RNA5-8SN3/4
Miscellaneous lncRNAs	AC103965.1, AC116366.4, HCG4P7, RP11-259P6.2, RP11-353N4.5, RP11-553L6.5, RP11-573G6.4, SCARNA6, SNORA47

capability to benefit from AZA treatment, GSEA was performed on PCGs (Figure 7). Many of the processes associated with the AZA response were also generally deregulated in myelodysplasia (described above and summarized in Figure 2). In particular, epigenetic regulation of gene expression and DNA repair, which were generally suppressed in patients (compared to healthy donors), were suppressed specifically in AZA nRESPs and activated in AZA RESPs. Based on these observations, two GO gene sets were selected, recombinational repair (NES=1.95; *p*=0.006; core enrichment genes included *BRCA1/2, H2AFX, POLQ, RAD51, MCM8,* and *XRCC2*; Figure 8A and B) and negative epigenetic regulation of gene expression (NES=1.91; *p*=0.014;

core enrichment genes included multiple *HIST1H* transcripts and other epigenetics-related factors, e.g., *DNMT1/3B* and *EZH2*; Figure 9A and B), for further investigation.

This study further aimed to explore specific noncoding transcripts related to these two processes (epigenetic regulation and DNA repair) in the context of future AZA responses. Computational modeling was applied and coexpressional networks were constructed based on correlation rates among all transcripts (PCGs, lncRNAs, circRNAs, and TEs) deregulated between AZA RESPs and nRESPs. Interaction graphs for both GO terms displaying 20 PCGs the most representative given GO term (the most deregulated leading-edge genes) and 12 noncoding

Table VI. *Differentially expressed circRNAs between azacitidine responders and nonresponders (|logFC|>1, FDR<0.05).*

Genomic locus	circBase ID	Gene	Type	Strand	logFC	CPM RESP	CPM nRESP	FDR
circRNAs up-regulated in RESP								
chr2:40655612:40657441	hsa_circ_0005232	*SLC8A1*	Exon	-	2.6	479.1	86.0	2.48E-03
chr7:149191294:149318263	-	-	Intergenic	-	2.4	595.5	111.1	2.42E-03
chr4:105439733:105440611	-	-	Intron	-	2.0	727.3	173.4	2.42E-03
chr2:40655612:40657444	hsa_circ_0000994	*SLC8A1*	Exon	-	1.9	3043.2	848.2	2.23E-04
circRNAs down-regulated in RESP								
chr1:23376879:23385660	hsa_circ_0003889	*KDM1A*	Exon	+	−4.6	0.0	122.4	3.40E-02
chr12:94603365:94613939	-	*PLXNC1*	Exon	+	−4.5	0.0	111.4	3.77E-02
chr22:17117929:17119630	-	*TPTEP1*	Exon	+	−3.8	24.2	365.7	2.23E-04
chr21:9827155:9827307	-	-	Intergenic	-	−3.8	11.4	312.7	1.74E-03
chr5:124036706:124036962	hsa_circ_0001523	*ZNF608*	Exon	-	−3.5	10.4	165.2	3.40E-02
chrY:22669237:22683186	-	-	Intergenic	-	−3.0	29.2	261.9	3.40E-02
chr19:22157530:22171711	-	*ZNF208*	Exon	-	−2.6	82.0	554.3	9.44E-04
chr6:77981056:77983024	-	-	Intergenic	-	−2.5	59.1	314.4	3.82E-02
chr4:183812555:183836728	hsa_circ_0002212	*DCTD*	Exon	-	−2.4	115.9	632.8	9.44E-04
chr12:27183090:27185614	-	*MED21*	Intron	+	−2.4	46.7	270.0	4.57E-02
chrX:79544404:79565732	-	*CHMP1B2P*	Exon	-	−2.2	119.0	566.8	2.48E-03
chr6:147527106:147560406	hsa_circ_0078150	*STXBP5*	Exon	+	−2.2	160.9	751.4	9.44E-04
chr4:9642694:9643074	-	-	Intergenic	+	−2.2	187.4	872.3	7.08E-04
chr14:70633355:70635201	hsa_circ_0000551	*SLC8A3*	Exon	-	−1.9	148.8	574.7	1.71E-02
chr18:6237962:6312055	hsa_circ_0046760	*L3MBTL4*	Exon	-	−1.9	130.0	475.2	3.40E-02
chr4:82025523:82065484	-	*PRKG2*	Exon	-	−1.9	148.0	533.1	3.40E-02
chr21:16363521:16415895	hsa_circ_0002711	*NRIP1*	Exon	-	−1.5	555.6	1587.1	2.33E-02

CPM: Counts per million.

transcripts (the most correlated with these PCGs) are shown in Figure 8C and Figure 9C. Although no TE was identified as related to any of the two processes, four circRNAs (hsa_circ_0001580 from *KDM1B* gene, hsa_circ_0008737 from *CAMTA1* gene, chr2:136,620,176-136,622,733 from *MCM6* gene, and chr17:59,853,761-59,878,835 from *BRIP6* gene) were associated with recombinational repair pathway, and three circRNAs (chr2:61,722,589-61,727,029 from *XPO1* gene, chr5: 134,681,712-134,688,690 from *H2AFY* gene, and chr14:97,299,803-97,342,457 from *VRK1* gene) with epigenetic regulation processes. Regarding lncRNAs, several transcripts with unknown functions were included in the networks, associating them with a cellular process for the first time (*e.g.*, epigenetic regulation: *CTC-298B17.1, CTB-83J4.2, RP11-449P15.1*; DNA repair: *AC004159.2, RP5894D12.5, C20orf166*; others can be found in Figure 8C and Figure 9C). Interestingly, three of the lncRNAs defined by GLMNET as the strongest predictors of AZA response were defined by computational modeling as significantly associated with recombinational repair (*CTC-482H14.5, RP11-419K12.2,* and *RP11-736I24.4*) and/or negative epigenetic regulation of gene expression (*CTC-482H15*.5). *CTC-482H14.5* (up-regulated in nRESPs) is an antisense RNA of the *KDM4B* gene, *RP11-419K12.2* (down-regulated in nRESPs) belongs to the lincRNA category, and *RP11-*

736I24.4 (down-regulated in nRESPs) is a transcribed unprocessed pseudogene of the *OTUD7A* gene. Additional descriptions of these lncRNAs are included in Table VIII.

Discussion

Over the last several years, the understanding of pathogenesis of MDS and AML-MRC has expanded. Multiple publications have described alterations of PCGs involved in chromatin modification, DNA methylation, RNA splicing, cohesin complex, transcription factors, cell signaling, and DNA damage (40). In this study, suppression of a multitude of PCGs associated with epigenetic mechanisms, DNA repair, and RNA splicing was documented, confirming the previous findings and proving validity for our current data. However, the expression of various classes of noncoding transcripts and their involvement in the abovementioned MDS/AML-associated pathways were focused in relation to their contribution to the response to AZA treatment. Classic types of lncRNAs ordinarily identified by classic RNA-seq data analyses were examined but special attention was paid to two additional classes of noncoding transcripts: circRNAs and TEs. The deregulation of lncRNAs in MDS has previously been addressed (34), but circRNAs and TEs have only marginally been studied in the context of this disease.
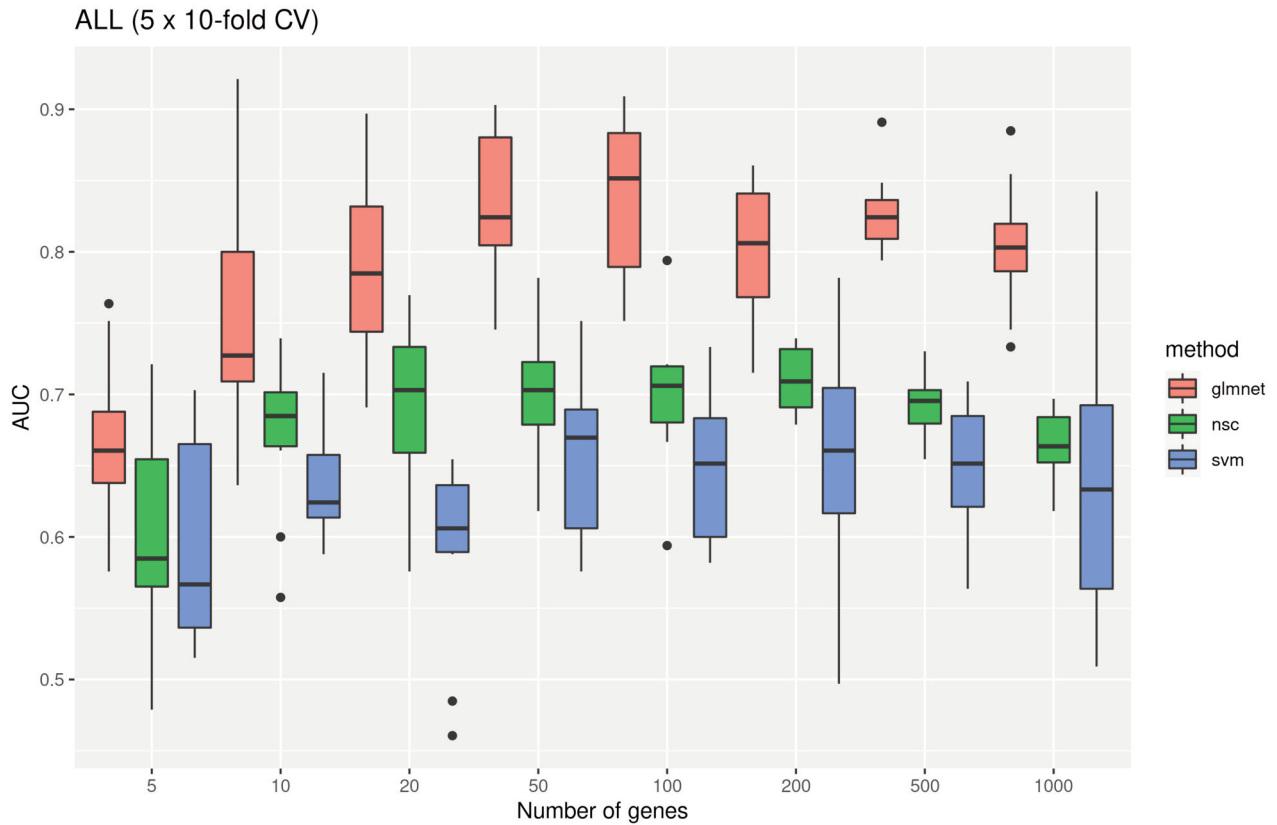
ALL (5 x 10-fold CV)

Figure 5. *Application of three different machine learning classification algorithms (SVM-RFE, voomNSC, and GLMNET) for classification of response to the azacitidine treatment. The graph suggests that GLMNET outperforms the other methods and that a limited set of transcripts is sufficient for the best split. The larger signatures only lead to overfitting.*

The expression profiles of the four types of transcripts (PCGs, lncRNAs, circRNAs, and TEs) were compared, and it was noticed that although PCGs and lncRNAs were able to stratify MDS/AML-MRC patients from healthy controls, the transcription of circRNAs and TEs was more heterogeneous, with substantial interindividual variability among patients. Therefore, PCGs and lncRNAs seem to be more relevant molecules in disease pathogenesis. However, circRNAs and TEs are much less numerous groups of transcripts than PCGs and lncRNAs, which might worsen the clustering output. Additionally, some bioinformatical bias might affect the results. For example, low numbers of reads covering circRNA junctions in comparison to the detection of linear transcripts or repetitiveness in TE sequences affecting read alignment might influence data processing. As information on circRNAs and TEs has only recently started being uncovered, the possibility of false discoveries and some differences in the number of detected molecules must be expected. For data annotation, only one bioinformatical tool for each type of transcripts was chosen (*i.e.*, CIRI2 for circRNAs and SalmonTE for TEs). Although these tools are widely

recommended, multiple annotation tools with different algorithm strategies should ideally be combined in more detailed, ongoing studies to achieve more reliable results. However, the comparison of multiple bioinformatics tools was out of the scope of this early study. The major focus of this study was an early characterization of expression changes of circRNAs and TEs and their comparison with standard gene expression data in relation to AZA treatment prediction.

CircRNAs are circularized transcripts produced by the so-called back-splicing process. They can act as regulators of gene expression but only limited information about the precise functions of individual circRNAs is available thus far. Due to their unusual stability and expression specificity, they appear to be important candidates for clinical biomarker research (41). In MDS/AML-MRC, several circRNAs processed from genes encoding known regulators of hematopoiesis and/or oncogenesis were deregulated, *e.g.*, down-regulation of circRNAs derived from *METRNL* (42), *EBF1* (43), *SKI* (44), *ABCC1* (45), and *AKAP12* (46), and up-regulation of circRNAs derived from *ZEB1* (47, 48) and *FNDC3B* (49, 50). For example, the *SKI* protooncogene is an important regulator
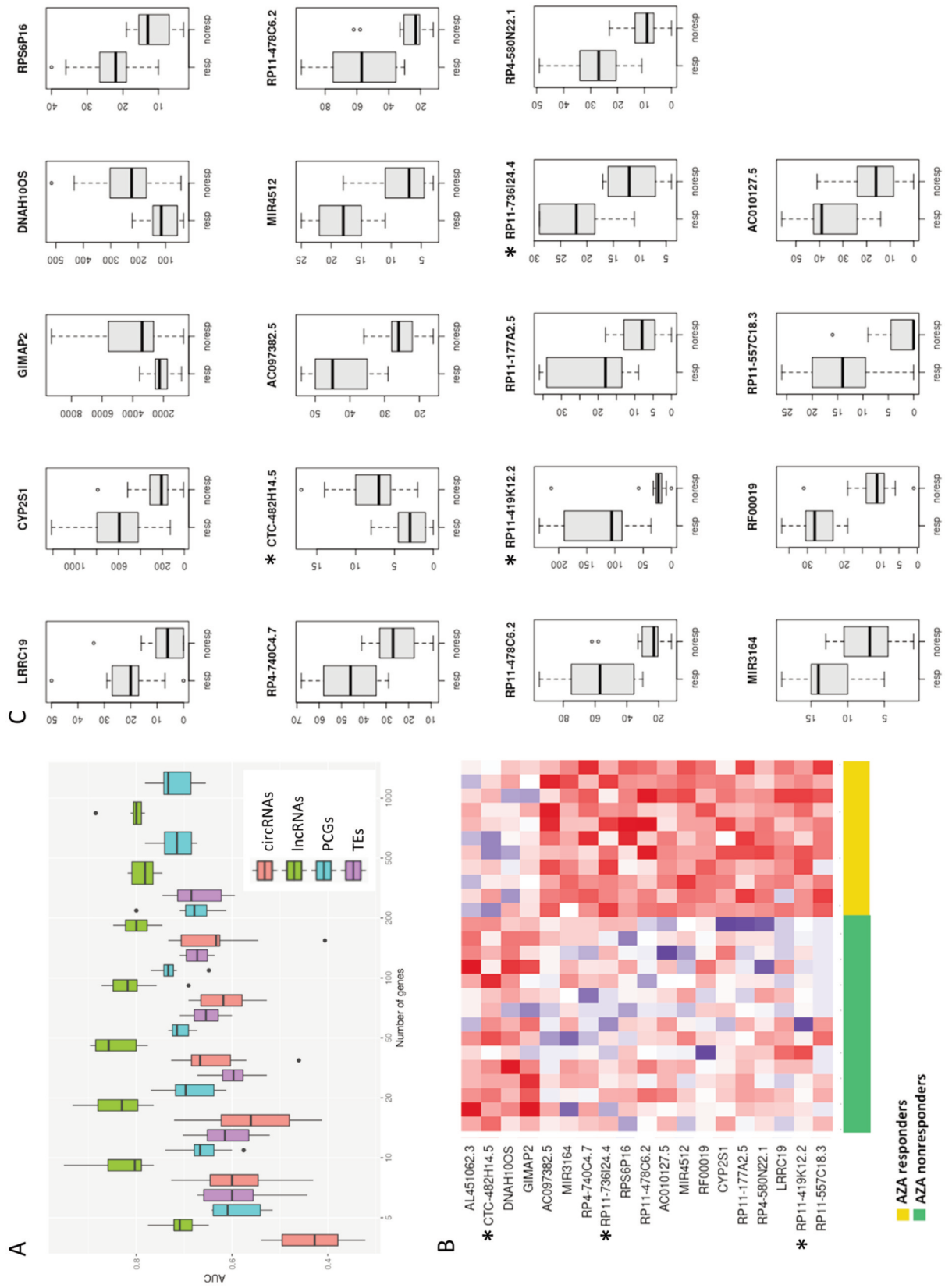
Figure 6. *The most powerful markers of azacitidine response defined by machine learning (GLMNET). (A) The four classes of transcripts had substantially different stratification power to discriminate azacitidine responders and nonresponders. (B) Heatmap showing the expression of the final set of response markers. (C) Boxplot of expression values of individual transcripts discriminating azacitidine responders and nonresponders. The lncRNAs marked with asterisks are discussed in the text in detail.*

Table VII. *Final set of the most powerful markers of future patient response to azacitidine defined by GLMNET. The lncRNAs marked with asterisks are discussed in the text in detail.*

Gene name	Gene id	Description	Mean expr.	logFC	FDR	DEA rank	Coefficient β
CYP2S1	ENSG00000167600.13	PCG: cytochrome P450 superfamily member	407.2	−1.09	0.349	3,745	0.00008
GIMAP2	ENSG00000106560.10	PCG: immune-associated GTPase	3087.1	1.01	0.003	3	−0.00007
LRRC19	ENSG00000184434.7	PCG: pathogen-recognition receptor	14.1	−1.28	0.669	12,212	0.00082
DNAH10OS	ENSG00000250091.2	PCG: uncharacterized protein	184.8	1.23	0.005	5	−0.00088
AC010127.5	ENSG00000224490.5	LNC: antisense RNA to TTC21B	22.9	−0.97	0.231	2,079	0.01516
*CTC-482H14.5	ENSG00000267550.1	LNC: antisense RNA to KDM4B	6.4	1.49	0.061	279	−0.00724
RP11-557C18.3	ENSG00000253607.1	LNC: antisense RNA to DERL1	6.6	−2.36	0.194	1,447	0.05217
RP4-580N22.1	ENSG00000229463.2	LNC: lincRNA	17.0	−1.21	0.205	1,643	0.06677
*RP11-419K12.2	ENSG00000285108.1	LNC: lincRNA	62.7	−2.07	<0.001	1	0.00823
MIR4512	ENSG00000266589.1	LNC: miRNA precursor	12.1	−0.99	0.124	809	0.08748
MIR3164	ENSG00000265539.1	LNC: miRNA precursor	9.6	−0.65	0.603	9,933	0.01013
RF00019	ENSG00000199875.1	LNC: miscellaneous RNA	18.1	−1.02	0.111	684	0.13092
RPS6P16	ENSG00000243250.1	LNC: ribosomal protein pseudogene	16.2	−0.85	0.344	3,649	0.00485
RP11-478C6.2	ENSG00000248564.1	LNC: pseudogene similar to ZNF74	39.0	−0.90	0.137	937	0.00403
RP11-177A2.5	ENSG00000232750.2	LNC: ribosomal protein pseudogene	13.9	−1.16	0.191	1,417	0.00039
RP4-740C4.7	ENSG00000272420.1	LNC: sense intronic RNA to MORN1	34.9	−0.73	0.343	3,632	0.02876
AC097382.5	ENSG00000187904.3	LNC: sense overlapping TBC1D14	32.7	−0.59	0.388	4,527	0.13947
*RP11-736I24.4	ENSG00000284783.1	LNC: transcribed unprocessed OTUD7A pseudogene	15.2	−0.84	0.217	1,842	0.00167
hsa_circ_0006595	circRNA processed from AL451062.3 gene (ENSG00000284989)	CIRC: exonic circRNA from AL451062.3	5.4	1.90	0.048	208	−0.02487

Mean expression values indicate average normalized expression of a given transcript in all samples, logFC and FDR values are computed by DEA between RESPs and nRESPs, and DEA rank indicates the order of a given transcript in DEA (when sorted according to FDR). Coefficient values β express the weight of the individual selected transcripts in the logistic regression model and are applicable to the regression formula calculating the prediction score (score=−12.43+β^{T}x).

of hematopoietic stem cell activity and its overexpression leads to myeloproliferative disease (44). In this study, the expression of *SKI*-derived hsa_circ_0007120 was down-regulated in MDS/AML-MRC. This circRNA was found to be strongly down-regulated during the differentiation of mesenchymal stem cells into cardiomyocyte-like cells, suggesting its involvement in cell proliferation and differentiation-related processes (51). Furthermore, the protein encoded by the *AKAP12* gene acts as a tumor suppressor regulating cell cycle progression and inhibiting Src-mediated oncogenic signaling (52). The *AKAP12* promoter is unmethylated and expressed in normal HSPCs but exhibits profound methylation and transcriptional silencing in leukemic myeloblasts (46). In MDS/AML-MRC, down-regulation of both, linear and circular (hsa_circ_0078299) forms of *AKAP12* was detected, suggesting common regulation within the locus. Interestingly, hsa_circ_0078299 was identified as a candidate biomarker of human stroke with a potential role in gene expression control and inflammation, and *SRSF2* splicing factor was predicted as its target gene (53). Another potentially relevant circRNA deregulated in MDS/AML-MRC patients is hsa_circ_0000228 transcribed from the *ZEB1* gene. The product of the *ZEB1* gene, a zinc finger transcription factor,

acts as a transcriptional regulator in hematopoiesis, coordinating HSC self-renewal, apoptosis, and multilineage differentiation fates. Additionally, it is required to suppress leukemic potential in AML (48). Several *ZEB1*-derived circRNAs have been shown to function as oncogenes in breast cancer (54), gastric cancer (55, 56), and hepatocellular carcinoma (57) by sponging various complementary miRNAs. Although the expression of the linear form of *ZEB1* was not changed in the MDS/AML-MRC patient samples, an increased rate of circularization of *ZEB1* transcripts (*i.e.*, an increase in hsa_circ_0000228) might exert specific roles in regulatory processes in CD34+ cells.

TEs represent detrimental units of DNA whose harmfulness lies in their inherent mobile nature. Their expression can lead to insertional mutagenesis, chromosomal rearrangements, and genomic instability (58). The significance of TEs has previously been addressed in various subtypes of primary AML (59-62), and it has been shown that TE expression was able to predict prognosis independent of mutation-based and coding gene expression-based risk stratification (59). In addition to AML, Colombo *et al.* (60) also compared the activity of TEs in high-risk and low-risk MDS on a limited number of samples (six high-risk and six
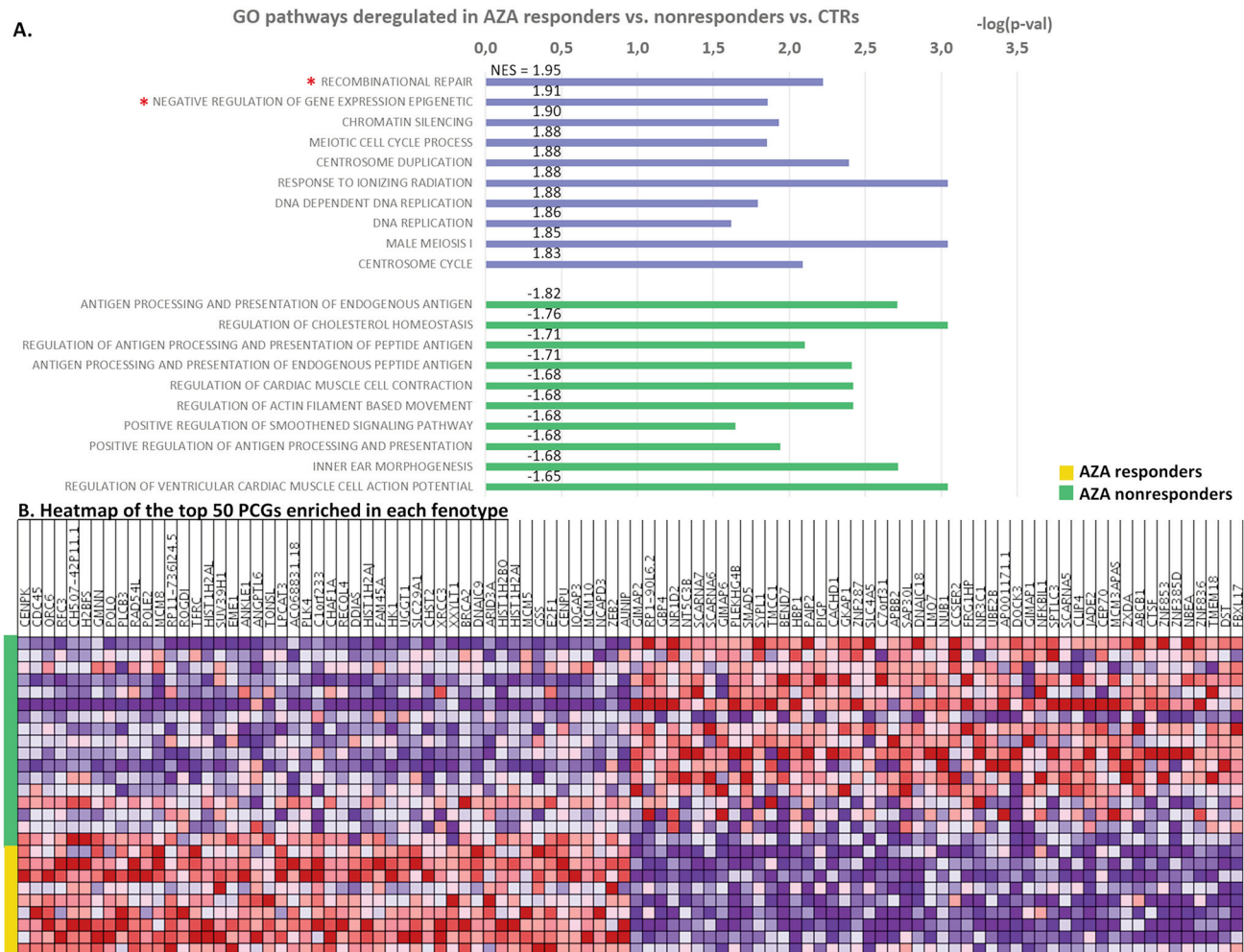
Figure 7. *Pathway analysis of differentially expressed genes in azacitidine responders and nonresponders. Plot of gene set enrichment analysis results shows p-values (logarithmically scaled) and normalized enrichment scores (NES) of the top 10 most enriched pathways in each group of samples sorted according to descended NES values. Two significant gene sets (marked by asterisks; recombinational repair and negative regulation of gene expression - epigenetic) were selected for further evaluation. Heatmap shows the top 50 protein coding genes differentially regulated between the two groups of samples.*

low-risk MDS patients) and noticed that TE expression is significantly down-regulated in high-risk disease. They proposed that the induction of TEs in low-risk cases may be a potential mechanism for immune-mediated clearance of cancer cells *via* the viral recognition pathway, whereas TE suppression in high-risk cases may enable immune escape of cells (60). In this study, CD34+ BM cells from only high-risk MDS patients were examined. Therefore, it was not possible to compare TE activity in low-risk disease. However, the rate of TE deregulation in our high-risk MDS samples was much lower than that of PCGs and other types of noncoding transcripts, negating the fundamental roles of TEs in high-risk disease pathogenesis.

Prediction of response to AZA therapy is an important challenge of applied research in MDS/AML. Current research activities are usually focused on the identification of an ideal, single biomarker whose alterations (such as a mutational change or transcription deregulation) would be predictive of treatment response with high sensitivity and specificity. The patient capability to respond to AZA and the identification of prediction markers have been repeatedly addressed (11, 12, 18). However, given the genomic heterogeneity of the disease, no universal biomarker has been identified thus far. This failure is likely triggered by complex interactions among multiple genomic alterations, affecting the resulting phenotype. Therefore, a combined approach studying different
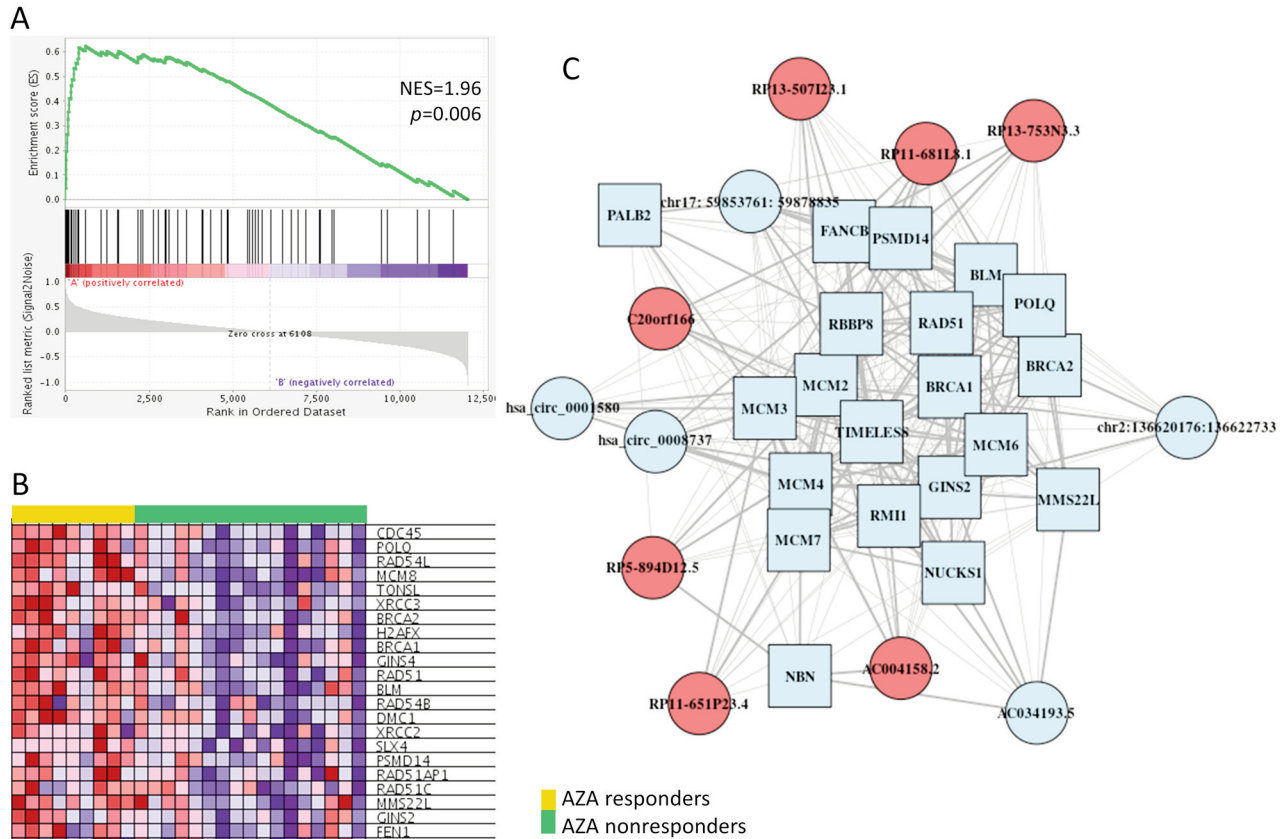
RECOMBINATIONAL REPAIR

Figure 8. *Association of the recombinational repair pathway with the azacitidine response. (A) gene set enrichment analysis enrichment plot, (B) expression heatmap of core enriched genes, and (C) coexpression network of transcripts associated with this process. Blue – down-regulation in nonresponders, red – up-regulation in nonresponders, circle-noncoding transcript, square - protein coding gene.*

types of molecules and their relations is required to yield an understanding of how these alterations define responsiveness to therapy. In addition to alterations in PCGs, the efficiency of AZA therapy can be influenced by the expression of various noncoding transcripts, contributing together to the clinical effects of the drug. At the level of individual genes, several noncoding RNAs significantly deregulated between AZA RESPs and nRESPs were identified. However, the pathological importance of most of them was somewhat questionable given that they are completely unannotated in current databases. This was especially the case for the most deregulated lncRNAs (*e.g.*, down-regulation of four lincRNAs *CH507-513H4.3-6* in AZA RESPs) because the only information available about them is often their genomic location. CircRNAs appeared to be a more relevant category of noncoding transcripts, as altered circularization was observed of several genes potentially relevant for AZA responsiveness (*e.g.*, *KDM1A* histone demethylase or *MED21* regulating transcription with RNA polymerase II). Regarding

TEs, only one TE was significantly deregulated between AZA RESPs and nRESPs (down-regulation L1 element *L1PBB_5* in AZA RESPs), with merely slight changes in the expression profiles of other TEs. The absence of altered TE transcription doubts the hypothesis suggesting that the clinical effects of AZA might be caused, at least partly, by demethylation of TE regions leading to increased production of TE transcripts, which might in turn activate the innate immune system through the viral defense pathway (60, 63).

Considering the above-discussed data on alterations of individual transcripts predisposing to AZA response or failure, a multibiomarker approach combining several features of different natures seems more appropriate to predict the response with high robustness. By machine learning methods, we compared the four categories of transcripts, and unexpectedly, lncRNAs substantially overperformed the prediction potential of PCGs. However, neither circRNAs nor TEs provided significant potential to stratify AZA RESPs and nRESPs. The combined set of the
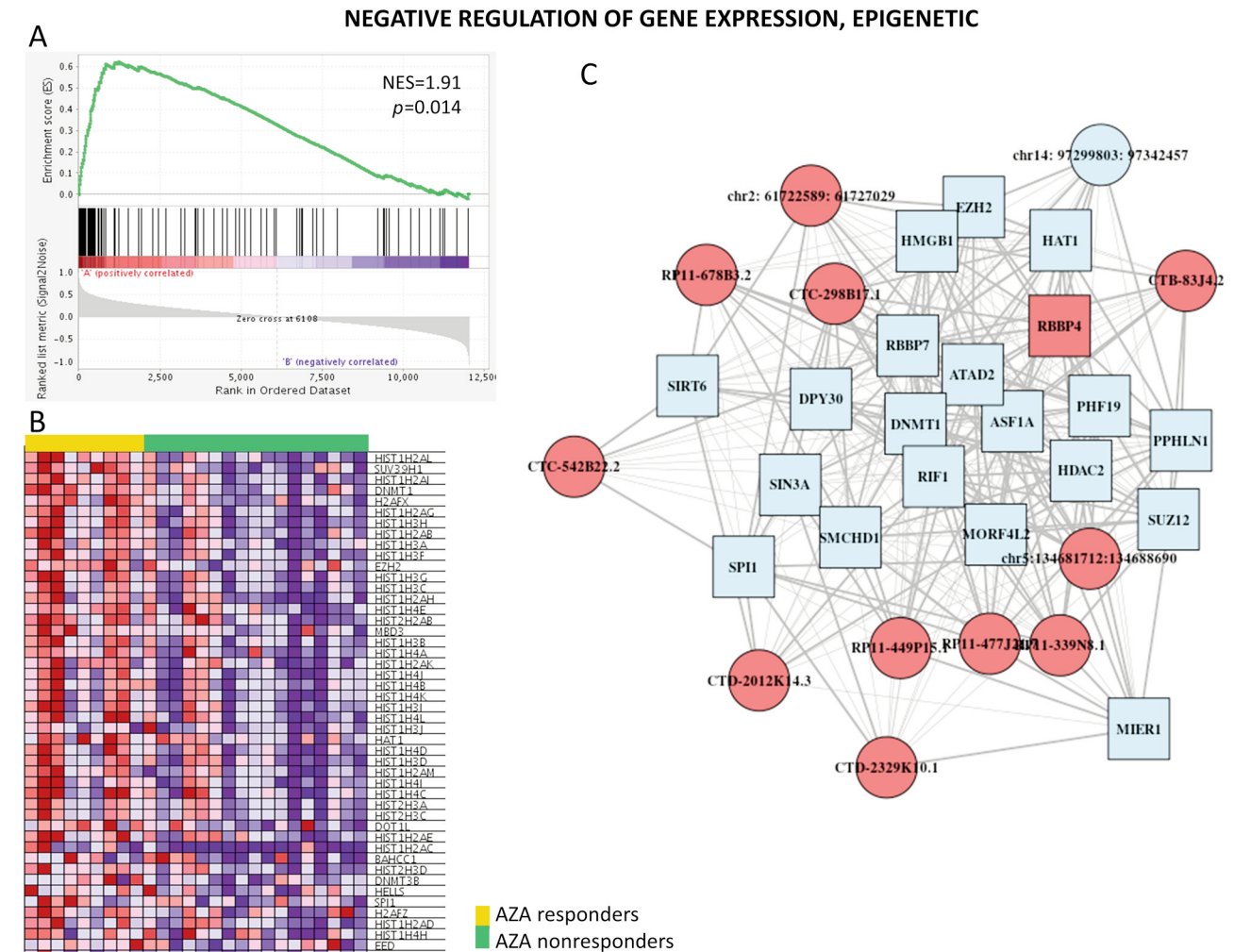
NEGATIVE REGULATION OF GENE EXPRESSION, EPIGENETIC

Figure 9. *Association of negative epigenetic regulation of gene expression with azacitidine response. (A) gene set enrichment analysis enrichment plot, (B) expression heatmap of core enriched genes, and (C) coexpressional network of transcripts associated with this process. Blue – down-regulation in nonresponders, red – up-regulation in nonresponders, circle - noncoding transcript, square - protein coding gene.*

best predictors of AZA response finally included 19 genes, and surprisingly, 14 of them were of lncRNA nature, whereas only four PCGs, one circRNA, and no TE were added to this final list of the most powerful stratifying molecules.

To better understand the contribution of deregulated noncoding transcripts to the molecular biology of MDS/AML treatment, important cellular pathways associated with AZA were first defined. Previously, the quiescent phenotype of HSCs was proposed as a key factor in the prediction of treatment outcome in MDS (64). In a preceding publication, it was identified that the BM of AZA RESPs contains actively cycling cells poised for erythromyeloid differentiation, whereas in nRESPs, BM HSPCs display a signature of a quiescent state (repression of genes related to active cell cycling, replication, and DNA damage response) (18). These

metabolic pathways seem to be closely interconnected with epigenetic modifications; the noncycling status has been associated with decreased expression of the *HIST1* cluster and therefore with the transcription-nonpermissive chromatin configuration maintaining cell quiescence (18). In this study, it was confirmed that AZA nRESPs display suppressed expression of epigenetic regulators and many chromatin related genes (especially those from the *HIST1H* and *HIST2H* families) in addition to genes involved in recombinational repair. In the process of epigenetic modulation, *DNMT1* (a maintenance DNA methyltransferase) was defined as one of the core PCGs related to the AZA response. Previously, it was demonstrated that *DNMT1* serves as a pharmacologic target of HMAs (65), and down-regulation of *DNMT1* expression in AZA nRESPs has been attributed to noncycling status and

Table VIII. *LncRNAs predicting future azacitidine response (assessed by GLMNET as the best predictors) and significantly (FDR<0.05) associated with recombinational repair and/or negative epigenetic regulation of gene expression GO terms.*

Name	Gene ID	Aliases	Description	Localization (hg38)	FDR	NES	Leading Edge PCGs
CTC-482H14.5	ENSG00000267550.1	*AC022517.1, Lnc-KDM4B-2*	antisense RNA to *KDM4B*	chr19:5,178, 119-5,178,464	GO RECOMBINATIONAL REPAIR:		
					6,88E-10	1,70	*MCM7, MCM3, FANCM, MCM8, MCM5, DMC1, RPA1, TIMELESS, BLM, RAD54L*
					GO NEGATIVE REGULATION OF GENE EXPRESSION, EPIGENETIC:		
					3,40E-02	1,29	*KMT2D, EZH2, CDC45, AEBP2, DNMT1, RBBP4, SIRT1, ATAD2, DPY30, PHF1*
RP11-419K12.2	ENSG00000285108.1	*AC103718.2, Lnc-GSDMC-35*	lincRNA	chr8:129,664, 195-129, 724,677	GO RECOMBINATIONAL REPAIR:		
					3,55E-02	1,25	*SLX4, RAD51AP1, CDC45, RAD51C, RAD54L, AUNIP, BRCA2, XRCC3, RECQL4, GINS4*
RP11-736I24.4	ENSG00000284783.1	*AC091057.4*	transcribed unprocessed OTUD7A pseudogene	chr15:30,750, 406-30, 759,859	GO RECOMBINATIONAL REPAIR:		
					4,07E-03	1,32	*POLL, GEN1, HELQ, RAD51AP1, NIPBL, XRCC2, FANCB,* RAD54L, RAD52, GINS4

NES: Normalized enrichment score.

reduced replication of HSPCs in these patients (18). Another core PCG of epigenetic mechanisms involved in the response to AZA was the *EZH2* gene, which was also specifically down-regulated in our cohort of AZA nRESPs. *EZH2*, a histone methyltransferase gene, is recurrently mutated in MDS/AML patients, and its mutations are associated with a poor prognosis of these patients (66). *EZH2* is significantly suppressed in MDS (67), and its CpG methylation level is reduced by AZA (68). Regarding the DNA damage response, another pathway described in the context of the AZA response (18), we identified *RAD51*, *BRCA1/2*, and several members of the *MCM* gene family as core PCGs down-regulated in the patients with future AZA failure. Alterations in *RAD51* (36, 69) and *BRCA* (70) genes have previously been reported in MDS. The expression of *MCM* genes, which are involved in the initiation and elongation phases of DNA replication, was reduced following epigenetic therapy (71). Similar to downmodulation of epigenetic machinery, attenuation of DNA repair and response pathways has also

been described as a characteristic feature of HSC quiescence, which underlies DNA damage accumulation in these cells (72). In summary, suppression of epigenetic regulation and recombinational repair seem to be crucial for the cell behavior of MDS/AML-MRC HSPCs with respect to the AZA response. Therefore, ncRNAs related to these two processes were explored to identify noncoding transcripts necessary for AZA-triggered mechanisms.

Although many deregulated noncoding transcripts are being described in cancer, their functional characterization is difficult. Construction of co-expression networks of PCGs with other RNAs, which are regulated similarly to PCGs, enables functional analysis of transcripts with unknown functions. This "guilt-by-association" strategy works with the principle that genes with related functions tend to have similar expression profiles. PCGs in a coexpressional module are associated with signaling pathways, attributing the same functions to unknown RNAs in this network. In this study, coexpressional networks were constructed for genes associated with epigenetic regulation

and recombinational repair pathways. Unsurprisingly, no TE was related to these processes, repeatedly suggesting little significance of these elements in high-risk MDS. In contrast, several lncRNAs and circRNAs co-regulated with core PCGs of the two processes were defined. Interestingly, the majority of the circRNAs found by co-expressional networks were processed from genes already known to be involved in MDS-related processes (recombinational repair - *MCM6* and *BRIP6;* epigenetic regulation - *KDM1B* and *H2AFY;* transcription, cell cycle and proliferation regulators - *CAMTA1, XPO1* and *VRK1).* In particular, *KDM* genes seem to be closely related to responsiveness to AZA. *KDM*s encode demethylases of histone lysine residues (*KDM1* mediates demethylation of H3K4 and H3K9, *KDM4* can act on H3K9, H3K36, and H1K26), yet they can also demethylate a number of nonhistone proteins such as p53, E2F1, DNMT1, or STAT3. *KDMs* are often differentially expressed in leukemia, and cooperate with transcription factors to activate or repress gene expression (73). In this study, deregulation of several noncoding transcripts was observed related to *KDM*s – *CTC-482H14.5* (asRNA to *KDM4B* gene, down-regulated in AZA RESPs), hsa_circ_0003889 (circRNA processed from *KDM1A*, down-regulated in AZA RESPs), and hsa_circ_0001580 (circRNA processed from *KDM1B*, up-regulated in AZA RESPs). Unlike these noncoding transcripts, the levels of *KDM* PCGs (namely *KDM1A/1B/4B*) were not altered between AZA RESPs and nRESPs, suggesting that these ncRNAs play specific roles independent of their host genes.

Finally, the best predictor genes defined by GLMNET were linked to cellular processes related to MDS. Among these predictors, three ncRNAs potentially associated with epigenetic regulation and/or recombinational repair were identified. First, *CTC-482H14.5* is an antisense RNA of the *KDM4B* gene and is coexpressed with several MDS-related PCGs, *e.g.*, *MCMs* (recombination repair) and *DNMT1* (epigenetic regulation). Second, *RP11-419K12.2* and *RP11-736I24.4* link to recombinational repair through coregulation with several *RAD* genes, *XRCC2/3*, *BRCA2*, or *FANCB*. *RP11-419K12.2* is a lincRNA with an unknown function located in proximity to *GSDMC* (gasdermin C, melanoma-derived leucine zipper-containing extranuclear factor), and *RP11-736I24.4* is a transcribed unprocessed pseudogene of *OTUD7A (*OTU deubiquitinase 7A). *OTUD7A* is a possible tumor suppressor gene controlling NF-κB expression through TRAF6 (74), a critical process implicated in the pathogenesis of MDS (75). Therefore, novel functions in epigenetic regulation and recombinational repair processes were predicted for these three unexplored lncRNAs, pointing to possible mechanisms by which their deregulation can be associated with AZA responsiveness.

In conclusion, this study addressed the expression and roles of various types of noncoding transcripts in AZA treatment of high-risk MDS and AML-MRC patients, suggesting particular ncRNAs associated with AZA responsiveness. Importantly, it was shown that lncRNAs can be utilized as novel predictive markers and that the combination of several markers of different natures is a more robust approach than monitoring the expression of a single gene. To the authors' knowledge, this is the first study identifying particular members of circRNA and TE transcript families deregulated in high-risk MDS and AML-MRC. In particular, deregulation of several circRNAs produced from hematopoiesis- and oncology-relevant genes was identified, whereas altered transcription of TEs seemed of less significance in this disease. However, the data in this pilot study have to be validated in independent cohorts of patients, different bioinformatical approaches analyzing circRNA and TE levels should be used, and functional studies proving the predicted roles in cellular processes have to be performed before rendering final conclusions about the roles of these noncoding transcripts in myelodysplasia.

Data Availability

Conflicts of Interest

The Authors declare that there are no competing interests.

Authors' Contributions

Conceptualization, Michaela Dostalova Merkerova; Funding acquisition, Michaela Dostalova Merkerova; Investigation, Katarina Szikszai, Andrea Hrustincova, Iva Trsova; Methodology, Zdenek Krejcik; Bioinformatics, Jiri Klema, David Kundrat, Anh Vu Le; Resources, Jaroslav Cermak, Anna Jonasova; Supervision, Monika Belickova, Jaroslav Cermak, Michaela Dostalova Merkerova; Writing – original draft, Michaela Dostalova Merkerova; Writing – review & editing, Zdenek Krejcik, Monika Belickova and Jaroslav Cermak.

Acknowledgements

References

1 Montalban-Bravo G and Garcia-Manero G: Myelodysplastic syndromes: 2018 update on diagnosis, risk-stratification and management. Am J Hematol *93(1)*: 129-147, 2018. PMID: 29214694. DOI: 10.1002/ajh.24930

2   Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Solé F, Bennett JM, Bowen D, Fenaux P, Dreyfus F, Kantarjian H, Kuendgen A, Levis A, Malcovati L, Cazzola M, Cermak J, Fonatsch C, Le Beau MM, Slovak ML, Krieger O, Luebbert M, Maciejewski J, Magalhaes SM, Miyazaki Y, Pfeilstöcker M, Sekeres M, Sperr WR, Stauder R, Tauro S, Valent P, Vallespi T, van de Loosdrecht AA, Germing U and Haase D: Revised international prognostic scoring system for myelodysplastic syndromes. Blood *120(12)*: 2454-2465, 2012. PMID: 22740453. DOI: 10.1182/blood-2012-03-420489

3   Loiseau C, Ali A and Itzykson R: New therapeutic approaches in myelodysplastic syndromes: Hypomethylating agents and lenalidomide. Exp Hematol *43(8)*: 661-672, 2015. PMID: 26123365. DOI: 10.1016/j.exphem.2015.05.014

4   Silverman LR, Demakos EP, Peterson BL, Kornblith AB, Holland JC, Odchimar-Reissig R, Stone RM, Nelson D, Powell BL, DeCastro CM, Ellerton J, Larson RA, Schiffer CA and Holland JF: Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. J Clin Oncol *20(10)*: 2429-2440, 2002. PMID: 12011120. DOI: 10.1200/JCO.2002.04.117

5   Cheson BD, Greenberg PL, Bennett JM, Lowenberg B, Wijermans PW, Nimer SD, Pinto A, Beran M, de Witte TM, Stone RM, Mittelman M, Sanz GF, Gore SD, Schiffer CA and Kantarjian H: Clinical application and proposal for modification of the International Working Group (IWG) response criteria in myelodysplasia. Blood *108(2)*: 419-425, 2006. PMID: 16609072. DOI: 10.1182/blood-2005-10-4149

6   Itzykson R, Thépot S, Quesnel B, Dreyfus F, Beyne-Rauzy O, Turlure P, Vey N, Recher C, Dartigeas C, Legros L, Delaunay J, Salanoubat C, Visanica S, Stamatoullas A, Isnard F, Marfaing-Koka A, de Botton S, Chelghoum Y, Taksin AL, Plantier I, Ame S, Boehrer S, Gardin C, Beach CL, Adès L, Fenaux P and Groupe Francophone des Myelodysplasies(GFM): Prognostic factors for response and overall survival in 282 patients with higher-risk myelodysplastic syndromes treated with azacitidine. Blood *117(2)*: 403-411, 2011. PMID: 20940414. DOI: 10.1182/blood-2010-06-289280

7   Traina F, Visconte V, Elson P, Tabarroki A, Jankowska AM, Hasrouni E, Sugimoto Y, Szpurka H, Makishima H, O'Keefe CL, Sekeres MA, Advani AS, Kalaycio M, Copelan EA, Saunthararajah Y, Olalla Saad ST, Maciejewski JP and Tiu RV: Impact of molecular mutations on treatment response to DNMT inhibitors in myelodysplasia and related neoplasms. Leukemia *28(1)*: 78-87, 2014. PMID: 24045501. DOI: 10.1038/leu.2013.269

8   Jung SH, Kim YJ, Yim SH, Kim HJ, Kwon YR, Hur EH, Goo BK, Choi YS, Lee SH, Chung YJ and Lee JH: Somatic mutations predict outcomes of hypomethylating therapy in patients with myelodysplastic syndrome. Oncotarget *7(34)*: 55264-55275, 2016. PMID: 27419369. DOI: 10.18632/oncotarget.10526

9   Bejar R, Lord A, Stevenson K, Bar-Natan M, Pérez-Ladaga A, Zaneveld J, Wang H, Caughey B, Stojanov P, Getz G, Garcia-Manero G, Kantarjian H, Chen R, Stone RM, Neuberg D, Steensma DP and Ebert BL: TET2 mutations predict response to hypomethylating agents in myelodysplastic syndrome patients. Blood *124(17)*: 2705-2712, 2014. PMID: 25224413. DOI: 10.1182/blood-2014-06-582809

10   Itzykson R, Kosmider O, Cluzeau T, Mansat-De Mas V, Dreyfus F, Beyne-Rauzy O, Quesnel B, Vey N, Gelsi-Boyer V, Raynaud S, Preudhomme C, Adès L, Fenaux P, Fontenay M and Groupe

Francophone des Myelodysplasies (GFM): Impact of TET2 mutations on response rate to azacitidine in myelodysplastic syndromes and low blast count acute myeloid leukemias. Leukemia *25(7)*: 1147-1152, 2011. PMID: 21494260. DOI:10.1038/leu.2011.71

11   Belickova M, Merkerova MD, Votavova H, Valka J, Vesela J, Pejsova B, Hajkova H, Klema J, Cermak J and Jonasova A: Up-regulation of ribosomal genes is associated with a poor response to azacitidine in myelodysplasia and related neoplasms. Int J Hematol *104(5)*: 566-573, 2016. PMID: 27416819. DOI: 10.1007/s12185-016-2058-3

12   Krejcik Z, Belickova M, Hrustincova A, Votavova H, Jonasova A, Cermak J, Dyr JE and Merkerova MD: MicroRNA profiles as predictive markers of response to azacitidine therapy in myelodysplastic syndromes and acute myeloid leukemia. Cancer Biomark *22(1)*: 101-110, 2018. PMID: 29630523. DOI: 10.3233/CBM-171029

13   Figueroa ME, Skrabanek L, Li Y, Jiemjit A, Fandy TE, Paietta E, Fernandez H, Tallman MS, Greally JM, Carraway H, Licht JD, Gore SD and Melnick A: MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. Blood *114(16)*: 3448-3458, 2009. PMID: 19652201. DOI: 10.1182/blood-2009-01-200519

14   Jiang Y, Dunbar A, Gondek LP, Mohan S, Rataul M, O'Keefe C, Sekeres M, Saunthararajah Y and Maciejewski JP: Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. Blood *113(6)*: 1315-1325, 2009. PMID: 18832655. DOI: 10.1182/blood-2008-06-163246

15   Voso MT, Santini V, Fabiani E, Fianchi L, Criscuolo M, Falconi G, Guidi F, Hohaus S and Leone G: Why methylation is not a marker predictive of response to hypomethylating agents. Haematologica *99(4)*: 613-619, 2014. PMID: 24688109. DOI: 10.3324/haematol.2013.099549

16   Kuendgen A, Müller-Thomas C, Lauseker M, Haferlach T, Urbaniak P, Schroeder T, Brings C, Wulfert M, Meggendorfer M, Hildebrandt B, Betz B, Royer-Pokora B, Gattermann N, Haas R, Germing U and Götze KS: Efficacy of azacitidine is independent of molecular and clinical characteristics - an analysis of 128 patients with myelodysplastic syndromes or acute myeloid leukemia and a review of the literature. Oncotarget *9(45)*: 27882-27894, 2018. PMID: 29963245. DOI: 10.18632/oncotarget.25328

17   Theilgaard-Mönch K, Boultwood J, Ferrari S, Giannopoulos K, Hernandez-Rivas JM, Kohlmann A, Morgan M, Porse B, Tagliafico E, Zwaan CM, Wainscoat J, Van den Heuvel-Eibrink MM, Mills K and Bullinger L: Gene expression profiling in MDS and AML: potential and future avenues. Leukemia *25(6)*: 909-920, 2011. PMID: 21445077. DOI: 10.1038/leu.2011.48

18   Koralkova P, Belickova M, Kundrat D, Dostalova Merkerova M, Krejcik Z, Szikszai K, Kaisrlikova M, Vesela J, Vyhlidalova P, Stetka J, Hlavackova A, Suttnar J, Flodr P, Stritesky J, Jonasova A, Cermak J and Divoky V: Low plasma citrate levels and specific transcriptional signatures associated with quiescence of CD34[+] progenitors predict azacitidine therapy failure in MDS/AML patients. Cancers (Basel) *13(9)*: 2161, 2021. PMID: 33946220. DOI: 10.3390/cancers13092161

19   Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M and Vardiman JW: The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood *127(20)*: 2391-2405, 2016. PMID: 27069254. DOI: 10.1182/blood-2016-03-643544

20 Andrews S: FastQC - A quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Last accessed on January 14, 2022]

21 Bolger AM, Lohse M and Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15): 2114-2120, 2014. PMID: 24695404. DOI: 10.1093/bioinformatics/btu170

22 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR: STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1): 15-21, 2013. PMID: 23104886. DOI: 10.1093/bioinformatics/bts635

23 Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL and Pertea M: Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol 20(1): 278, 2019. PMID: 31842956. DOI: 10.1186/s13059-019-1910-1

24 Li H and Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14): 1754-1760, 2009. PMID: 19451168. DOI: 10.1093/bioinformatics/btp324

25 Gao Y, Zhang J and Zhao F: Circular RNA identification based on multiple seed matching. Brief Bioinform 19(5): 803-810, 2018. PMID: 28334140. DOI: 10.1093/bib/bbx014

26 Jeong HH, Yalamanchili HK, Guo C, Shulman JM and Liu Z: An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. Pac Symp Biocomput 23: 168-179, 2018. PMID: 29218879.

27 Stacklies W, Redestig H, Scholz M, Walther D and Selbig J: pcaMethods—a bioconductor package providing PCA methods for incomplete data. Bioinformatics 23(9): 1164-1167, 2007. PMID: 17344241. DOI: 10.1093/bioinformatics/btm069

28 Robinson MD, McCarthy DJ and Smyth GK: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1): 139-140, 2010. PMID: 19910308. DOI: 10.1093/bioinformatics/btp616

29 Love MI, Huber W and Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15(12): 550, 2014. PMID: 25516281. DOI: 10.1186/s13059-014-0550-8

30 Guyon I, Weston J, Barnhill S and Vapnik V: Gene selection for cancer classification using support vector machines. Mach Learn 46: 389-422, 2002. DOI: 10.1023/A:1012487302797

31 Friedman J, Hastie T and Tibshirani R: Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33(1): 1-22, 2010. PMID: 20808728.

32 Zararsiz G, Goksuluk D, Klaus B, Korkmaz S, Eldem V, Karabulut E and Ozturk A: voomDDA: discovery of diagnostic biomarkers and classification of RNA-seq data. PeerJ 5: e3890, 2017. PMID: 29018623. DOI: 10.7717/peerj.3890

33 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102(43): 15545-15550, 2005. PMID: 16199517. DOI: 10.1073/pnas.0506580102

34 Szikszai K, Krejcik Z, Klema J, Loudova N, Hrustincova A, Belickova M, Hruba M, Vesela J, Stranecky V, Kundrat D, Pecherkova P, Cermak J, Jonasova A and Dostalova Merkerova M: LncRNA profiling reveals that the deregulation of H19, WT1-AS, TCL6, and LEF1-AS1 is associated with higher-risk myelodysplastic syndrome. Cancers (Basel) 12(10): 2726, 2020. PMID: 32977510. DOI: 10.3390/cancers12102726

35 Zhu Y and Wu L: Identification of latent core genes and pathways associated with myelodysplastic syndromes based on integrated bioinformatics analysis. Hematology 25(1): 299-308, 2020. PMID: 32772642. DOI: 10.1080/16078454.2020.1802917

36 Valka J, Vesela J, Votavova H, Dostalova-Merkerova M, Horakova Z, Campr V, Brezinova J, Zemanova Z, Jonasova A, Cermak J and Belickova M: Differential expression of homologous recombination DNA repair genes in the early and advanced stages of myelodysplastic syndrome. Eur J Haematol 99(4): 323-331, 2017. PMID: 28681469. DOI: 10.1111/ejh.12920

37 Pellagatti A, Benner A, Mills KI, Cazzola M, Giagounidis A, Perry J, Malcovati L, Della Porta MG, Jädersten M, Verma A, McDonald EJ, Killick S, Hellström-Lindberg E, Bullinger L, Wainscoat JS and Boultwood J: Identification of gene expression-based prognostic markers in the hematopoietic stem cells of patients with myelodysplastic syndromes. J Clin Oncol 31(28): 3557-3564, 2013. PMID: 24002510. DOI: 10.1200/JCO.2012.45.5626

38 Merkerova MD, Remesova H, Krejcik Z, Loudova N, Hrustincova A, Sziksai K, Cermak J, Jonasova A and Belickova M: Relationship between altered miRNA expression and DNA methylation of the DLK1-DIO3 region in azacitidine-treated patients with myelodysplastic syndromes and acute myeloid leukemia with myelodysplasia-related changes. Cells 7(9): 138, 2018. PMID: 30223454. DOI: 10.3390/cells7090138

39 Karoopongse E, Yeung C, Byon J, Ramakrishnan A, Holman ZJ, Jiang PY, Yu Q, Deeg HJ and Marconedes AM: The KDM2B- let-7b -EZH2 axis in myelodysplastic syndromes as a target for combined epigenetic therapy. PLoS One 9(9): e107817, 2014. PMID: 25225797. DOI: 10.1371/journal.pone.0107817

40 Menssen AJ and Walter MJ: Genetics of progression from MDS to secondary leukemia. Blood 136(1): 50-60, 2020. PMID: 32430504. DOI: 10.1182/blood.2019000942

41 Li X, Yang L and Chen LL: The biogenesis, functions, and challenges of circular RNAs. Mol Cell 71(3): 428-442, 2018. PMID: 30057200. DOI: 10.1016/j.molcel.2018.06.034

42 Ushach I, Burkhardt AM, Martinez C, Hevezi PA, Gerber PA, Buhren BA, Schrumpf H, Valle-Rios R, Vazquez MI, Homey B and Zlotnik A: METEORIN-LIKE is a cytokine associated with barrier tissues and alternatively activated macrophages. Clin Immunol 156(2): 119-127, 2015. PMID: 25486603. DOI:10.1016/j.clim.2014.11.006

43 Nechanitzky R, Akbas D, Scherer S, Györy I, Hoyler T, Ramamoorthy S, Diefenbach A and Grosschedl R: Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. Nat Immunol 14(8): 867-875, 2013. PMID: 23812095. DOI: 10.1038/ni.2641

44 Singbrant S, Wall M, Moody J, Karlsson G, Chalk AM, Liddicoat B, Russell MR, Walkley CR and Karlsson S: The SKI proto-oncogene enhances the in vivo repopulation of hematopoietic stem cells and causes myeloproliferative disease. Haematologica 99(4): 647-655, 2014. PMID: 24415629. DOI: 10.3324/haematol.2013.093971

45 Liu B, Li LJ, Gong X, Zhang W, Zhang H and Zhao L: Co-expression of ATP binding cassette transporters is associated with poor prognosis in acute myeloid leukemia. Oncol Lett 15(5): 6671-6677, 2018. PMID: 29616130. DOI: 10.3892/ol.2018.8095

46 Flotho C, Paulun A, Batz C and Niemeyer CM: AKAP12, a gene with tumour suppressor properties, is a target of promoter DNA methylation in childhood myeloid malignancies. Br J Haematol 138(5): 644-650, 2007. PMID: 17686059. DOI: 10.1111/j.1365-2141.2007.06709.x

47 Soen B, Vandamme N, Berx G, Schwaller J, Van Vlierberghe P and Goossens S: ZEB proteins in leukemia: friends, foes, or friendly foes? Hemasphere *2(3)*: e43, 2018. PMID: 31723771. DOI: 10.1097/HS9.0000000000000043

48 Almotiri A, Alzahrani H, Menendez-Gonzalez JB, Abdelfattah A, Alotaibi B, Saleh L, Greene A, Georgiou M, Gibbs A, Alsayari A, Taha S, Thomas LA, Shah D, Edkins S, Giles P, Stemmler MP, Brabletz S, Brabletz T, Boyd AS, Siebzehnrubl FA and Rodrigues NP: Zeb1 modulates hematopoietic stem cell fates required for suppressing acute myeloid leukemia. J Clin Invest *131(1)*: e129115, 2021. PMID: 33108352. DOI: 10.1172/JCI129115

49 Rajasagi M, Shukla SA, Fritsch EF, Keskin DB, DeLuca D, Carmona E, Zhang W, Sougnez C, Cibulskis K, Sidney J, Stevenson K, Ritz J, Neuberg D, Brusic V, Gabriel S, Lander ES, Getz G, Hacohen N and Wu CJ: Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. Blood *124(3)*: 453-462, 2014. PMID: 24891321. DOI: 10.1182/blood-2014-04-567933

50 Cheng CK, Wang AZ, Wong THY, Wan TSK, Cheung JS, Raghupathy R, Chan NPH and Ng MHL: FNDC3B is another novel partner fused to RARA in the t(3;17)(q26;q21) variant of acute promyelocytic leukemia. Blood *129(19)*: 2705-2709, 2017. PMID: 28314734. DOI: 10.1182/blood-2017-02-767707

51 Ruan ZB, Chen GC, Zhang R and Zhu L: Circular RNA expression profiles during the differentiation of human umbilical cord-derived mesenchymal stem cells into cardiomyocyte-like cells. J Cell Physiol, 2019. PMID: 30756390. DOI: 10.1002/jcp.28310

52 Akakura S and Gelman IH: Pivotal role of AKAP12 in the regulation of cellular adhesion dynamics: Control of cytoskeletal architecture, cell migration, and mitogenic signaling. J Signal Transduct *2012*: 529179, 2012. PMID: 22811901. DOI: 10.1155/2012/529179

53 Silva PW, M Shimon SM, de Brito LM, Reis-das-Mercês L, Magalhães L, Araújo G, Ribeiro-Dos-Santos Â and Vidal AF: Novel insights toward human stroke-related epigenetics: circular RNA and its impact in poststroke processes. Epigenomics *12(22)*: 1957-1968, 2020. PMID: 33242258. DOI: 10.2217/epi-2020-0128

54 Pei X, Zhang Y, Wang X, Xue B, Sun M and Li H: Circular RNA circ-ZEB1 acts as an oncogene in triple negative breast cancer via sponging miR-448. Int J Biochem Cell Biol *126*: 105798, 2020. PMID: 32629026. DOI: 10.1016/j.biocel.2020.105798

55 Lin S, Song S, Sun R, Zhang M, Du Y, Zhang D, Xu W and Wang H: Oncogenic circular RNA Hsa-circ-000684 interacts with microRNA-186 to upregulate ZEB1 in gastric cancer. FASEB J *34(6)*: 8187-8203, 2020. PMID: 32388910. DOI: 10.1096/fj.201903246R

56 Ren C, Zhang Z, Wang S, Zhu W, Zheng P and Wang W: Circular RNA hsa_circ_0001178 facilitates the invasion and metastasis of colorectal cancer through upregulating ZEB1 via sponging multiple miRNAs. Biol Chem *401(4)*: 487-496, 2020. PMID: 31747371. DOI: 10.1515/hsz-2019-0350

57 Gong Y, Mao J, Wu D, Wang X, Li L, Zhu L and Song R: Circ-ZEB1.33 promotes the proliferation of human HCC by sponging miR-200a-3p and upregulating CDK6. Cancer Cell Int *18*: 116, 2018. PMID: 30123094. DOI: 10.1186/s12935-018-0602-3

58 Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, Mager DL and Feschotte C: Ten things you should know about transposable elements. Genome Biol *19(1)*: 199, 2018. PMID: 30454069. DOI: 10.1186/s13059-018-1577-z

59 Colombo AR, Triche T Jr and Ramsingh G: Transposable element expression in acute myeloid leukemia transcriptome and prognosis. Sci Rep *8(1)*: 16449, 2018. PMID: 30401833. DOI: 10.1038/s41598-018-34189-x

60 Colombo AR, Zubair A, Thiagarajan D, Nuzhdin S, Triche TJ and Ramsingh G: Suppression of transposable elements in leukemic stem cells. Sci Rep *7(1)*: 7029, 2017. PMID: 28765607. DOI: 10.1038/s41598-017-07356-9

61 Zeng Y, Cao Y, Halevy RS, Nguyen P, Liu D, Zhang X, Ahituv N and Han JJ: Characterization of functional transposable element enhancers in acute myeloid leukemia. Sci China Life Sci *63(5)*: 675-687, 2020. PMID: 32170627. DOI: 10.1007/s11427-019-1574-x

62 Deniz Ö, Ahmed M, Todd CD, Rio-Machin A, Dawson MA and Branco MR: Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. Nat Commun *11(1)*: 3506, 2020. PMID: 32665538. DOI: 10.1038/s41467-020-17206-4

63 Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, Hein A, Rote NS, Cope LM, Snyder A, Makarov V, Budhu S, Slamon DJ, Wolchok JD, Pardoll DM, Beckmann MW, Zahnow CA, Merghoub T, Chan TA, Baylin SB and Strick R: Inhibiting DNA Methylation Causes an Interferon Response in Cancer *via* dsRNA Including Endogenous Retroviruses. Cell *162(5)*: 974-986, 2015. PMID: 26317466. DOI: 10.1016/j.cell.2015.07.011

64 Unnikrishnan A, Papaemmanuil E, Beck D, Deshpande NP, Verma A, Kumari A, Woll PS, Richards LA, Knezevic K, Chandrakanthan V, Thoms JAI, Tursky ML, Huang Y, Ali Z, Olivier J, Galbraith S, Kulasekararaj AG, Tobiasson M, Karimi M, Pellagatti A, Wilson SR, Lindeman R, Young B, Ramakrishna R, Arthur C, Stark R, Crispin P, Curnow J, Warburton P, Roncolato F, Boultwood J, Lynch K, Jacobsen SEW, Mufti GJ, Hellstrom-Lindberg E, Wilkins MR, MacKenzie KL, Wong JWH, Campbell PJ and Pimanda JE: Integrative genomics identifies the molecular basis of resistance to azacitidine therapy in myelodysplastic syndromes. Cell Rep *20(3)*: 572-585, 2017. PMID: 28723562. DOI: 10.1016/j.celrep.2017.06.067

65 Ghoshal K, Datta J, Majumder S, Bai S, Kutay H, Motiwala T and Jacob ST: 5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal. Mol Cell Biol *25(11)*: 4727-4741, 2005. PMID: 15899874. DOI: 10.1128/MCB.25.11.4727-4741.2005

66 Zhang Q, Han Q, Zi J, Ma J, Song H, Tian Y, McGrath M, Song C and Ge Z: Mutations in EZH2 are associated with poor prognosis for patients with myeloid neoplasms. Genes Dis *6(3)*: 276-281, 2019. PMID: 32042866. DOI: 10.1016/j.gendis.2019.05.001

67 Cabrero M, Wei Y, Yang H, Ganan-Gomez I, Bohannan Z, Colla S, Marchesini M, Bravo GM, Takahashi K, Bueso-Ramos C and Garcia-Manero G: Down-regulation of EZH2 expression in myelodysplastic syndromes. Leuk Res *44*: 1-7, 2016. PMID: 26970171. DOI: 10.1016/j.leukres.2016.02.009

68 Gawlitza AL, Speith J, Rinke J, Sajzew R, Müller EK, Schäfer V, Hochhaus A and Ernst T: 5-Azacytidine modulates CpG methylation levels of EZH2 and NOTCH1 in myelodysplastic syndromes. J Cancer Res Clin Oncol *145(11)*: 2835-2843, 2019. PMID: 31506740. DOI: 10.1007/s00432-019-03016-9

69 Valka J, Vesela J, Votavova H, Dostalova-Merkerova M, Urbanova Z, Jonasova A, Cermak J and Belickova M: Genetic variant screening of DNA repair genes in myelodysplastic syndrome identifies a novel mutation in the XRCC2 gene. Oncol Res Treat *42(5)*: 263-268, 2019. PMID: 30861523. DOI: 10.1159/000497209

70 Kaplan HG, Calip GS and Malmgren JA: Maximizing breast cancer therapy with awareness of potential treatment-related blood disorders. Oncologist *25(5)*: 391-397, 2020. PMID: 32073195. DOI: 10.1634/theoncologist.2019-0099

71 Belinsky SA, Grimes MJ, Picchi MA, Mitchell HD, Stidley CA, Tesfaigzi Y, Channell MM, Liu Y, Casero RA Jr, Baylin SB, Reed MD, Tellez CS and March TH: Combination therapy with vidaza and entinostat suppresses tumor growth and reprograms the epigenome in an orthotopic lung cancer model. Cancer Res *71(2)*: 454-462, 2011. PMID: 21224363. DOI: 10.1158/0008-5472.CAN-10-3184

72 Beerman I, Seita J, Inlay MA, Weissman IL and Rossi DJ: Quiescent hematopoietic stem cells accumulate DNA damage during aging that is repaired upon entry into cell cycle. Cell Stem Cell *15(1)*: 37-50, 2014. PMID: 24813857. DOI: 10.1016/j.stem.2014.04.016

73 Arifuzzaman S, Khatun MR and Khatun R: Emerging of lysine demethylases (KDMs): From pathophysiological insights to novel therapeutic opportunities. Biomed Pharmacother *129*: 110392, 2020. PMID: 32574968. DOI: 10.1016/j.biopha.2020.110392

74 Mevissen TE, Hospenthal MK, Geurink PP, Elliott PR, Akutsu M, Arnaudo N, Ekkebus R, Kulathu Y, Wauer T, El Oualid F, Freund SM, Ovaa H and Komander D: OTU deubiquitinases reveal mechanisms of linkage specificity and enable ubiquitin chain restriction analysis. Cell *154(1)*: 169-184, 2013. PMID: 23827681. DOI: 10.1016/j.cell.2013.05.046

75 Muto T, Walker CS, Choi K, Hueneman K, Smith MA, Gul Z, Garcia-Manero G, Ma A, Zheng Y and Starczynowski DT: Adaptive response to inflammation contributes to sustained myelopoiesis and confers a competitive advantage in myelodysplastic syndrome HSCs. Nat Immunol *21(5)*: 535-545, 2020. PMID: 32313245. DOI: 10.1038/s41590-020-0663-z