5-Aza-Cytidine Enhances Terminal Polyadenylation Site Usage for Full-Length Transcripts in Cells

Samuel Ogunsola, Ling Liu, Urmi Das, Jiuyong Xie*

[1]Department of Physiology & Pathophysiology, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, R3E 0J9, Canada

Keywords: Alternative Splicing, Alternative Polyadenylation, Anti-Cancer Drug, Gene Expression

The authors declare no potential conflicts of interest.

*: Correspondence to: xiej@umanitoba.ca, phone: +1 204 975-7774

Abstract

As an inhibitor of DNA methyltransferases (DNMT) and anti-cancer drug, the many effects of 5-aza-cytidine (5-azaC) on gene expression remains unknown. Here we show that 5-azaC treatment of cultured pituitary tumour cells increases the relative usage of the genomic terminal exons (GTE) across the transcriptome. This effect is largely achieved by switching mRNA polyadenylation (poly(A)) from proximal sites to the GTE, which have a more optimal poly(A) signal consensus motif. Consistently, 5-azaC upregulates mRNA anti-terminators Scaf4 and Scaf8 while downregulating the early termination enhancer E2f2. In MOLM-13 leukaemia cells, 5-azaC similarly promotes full-length transcript production, regulating alternative polyadenylation factors in both common and cell-specific ways. Unexpectedly, PCF11, known to promote proximal poly(A) site usage, is upregulated in both cell lines, reminiscent of a homeostatic effort by the tumour/cancer cells to maintain their shortened transcripts during the 5-azaC treatment. This study highlights a previously unknown aspect of 5-azaC's effect on gene expression as a DNMT inhibitor and anti-cancer drug: to directionally promote the terminal polyadenylation site usage to switch from the shortened to full-length transcripts of cancer cells and consequently the alternative usage of strings of 3' exons.

46 **Introduction**

47

48 5-Aza-cytidine (5-azaC), an FDA-approved chemotherapeutic agent, binds and
49 inactivates DNA methyltransferases (DNMT), inhibiting DNA methylation [1]. It also affects
50 many aspects of RNA metabolism [1], but its effect on RNA processing, particularly at the
51 exome level, is not fully understood.

52 DNA methylation and gene mRNA levels appear to have a gene-region-specific
53 relationship [2]: they are negatively correlated at promoters, first exons and first introns, but
54 positively correlated within gene-bodies [2]. Whether 5-azaC has a similar region-dependent
55 effect on exon methylation and exon usage is unknown. Our study, initiated with this question,
56 reveals both similar and surprising findings.

57 Here we analysed RNA-Seq and whole genome bisulfite sequencing (WGBS) data from
58 GH3 pituitary tumour cells treated with 5-azaC. We specifically looked at the differential usage
59 and methylation of genomic first, middle, and terminal exons (GTEs, to differentiate from those
60 in transcripts). Remarkably, 5-azaC was found to promote the widespread use of strings of 3'
61 exons and polyadenylation sites in the GTEs but without much different effects on the
62 methylation of these exons. It also regulated genes involved in transcript length and pre-mRNA
63 processing. Similar effects were observed in a leukaemia cell line. This suggests that 5-azaC
64 may reverse the shortening of mRNA transcripts in tumour cells by a unique directional effect
65 on exon usage and polyadenylation, restoring their full protein domains and functions.

66

67 **Results**

68 **4. Differential effects of 5-azaC on the relative exon usage of the genomic first, middle**
69 **and terminal exons (GTE) of genes**

70 To analyze the potential exon-rank-dependent effect of 5-azaC, a DNA methylation
71 inhibitor, we treated GH3 pituitary tumor cells with 50μM 5-azaC and compared them with
72 untreated controls. Our analysis focused on the RNA sequencing data of genomic first, middle,
73 and terminal exons. In DEXSeq, the first and terminal exons were designated as E001 for the
74 positive (+) and negative (–) chromosome strands, respectively. For middle exons, we selected
75 E004, E005, and E006 from the DEXSeq exon/bin list, specifically choosing genes with more
76 than 8 exons. To further validate exon ranks, we randomly selected 100 genes from each group
77 to verify in Integrated Genomics Viewer. This approach revealed 1710 first exons, 1790 middle
78 exons, and 1875 terminal exons with a statistically significant difference upon 5-azaC treatment
79 (Fig. 1A, padj < 0.05).

80 An intriguing pattern emerged: 95% of the first exons, 59% of the middle exons, but only
81 8% of the terminal exons showed a reduction in relative usage; the rest exhibited an increase.
82 This finding suggests a distinct exon-rank-dependent impact of 5-azaC on exon usage (Fig. 1A,
83 **** p = 0, by hypergeometric density test between any two groups).

84 Furthermore, we assessed the average methylation ratio (AMR) of CpG in these samples
85 using available Whole-Genome Bisulfite Sequencing (WGBS) data. The AMR data
86 corroborated previous findings on exon-rank dependence: first exons were almost
87 unmethylated, while middle and terminal exons were highly methylated (Fig. S1) [2]. However,
88 the AMR changes induced by 5-azaC were more varied, showing both increases and decreases
89 among the three ranked exon groups (Fig. 1B). This pattern contrasts with the more uniform
90 exon usage changes of the groups, indicating that the exon-rank effect does not globally
91 correlate with AMR changes due to 5-azaC.

92 **2. Transcriptome-wide switch from the proximal to the GTE poly(A) sites induced by 5-**
93 **azaC**

94 In our study, we focused on the exon usage of 154 genes on the (+) strand and 153 genes
95 on the (-) strand, each with a minimum of 10 exons/bins identified in DEXSeq analysis. By
96 analyzing the effect of 5-azaC on the fold changes in exon usage, we observed a notable shift
97 to increased usage of 3′ (including the terminal) exons in 102 of these genes. For instance, in
98 the non-treated samples, the RNA-Seq reads for the first 6 exons of the Zmym3 gene were
99 predominantly about 5-fold higher than those for the 3′ exons (Fig. 2A). However, after 5-azaC
100 treatment, the 3′ exon levels rose to match those of the 5′ exons. The relative usage ratios of
101 these exons changed from less than 1.0 to more than 1.0 between exons 6 and 7. Further, RT-
102 PCR assays confirmed a significant increase (Fig. 2B, approximately 10-fold, $p < 0.005$, $n = 3$)
103 in the relative usage of the poly(A) site in the GTE following 5-azaC treatment.

104 The extended isoform resulting from this shift encodes 13 extra domains including
105 DUF3504, which is linked to DNA damage repair [3]. This suggests that the alternative
106 polyadenylation (APA) switch induced by 5-azaC might lead to longer protein lengths with a
107 complete set of functional domains. Overall, the 102 genes with single switches showed
108 significant clustering ($p<0.01$) in functions such as ATP binding, microtubule cytoskeleton
109 organization, ATPase activity, and microRNAs in cancer, as per DAVID functional clustering
110 analysis(Table S1).

111 Additionally, genes exhibiting multiple switches displayed alternative splicing events,
112 like exon skipping and intron retention at the switch sites. For example, intron 2 retention was
113 observed in the Gln1 gene post 5-azaC treatment (Fig. S2A). Moreover, the regions where
114 switches occurred showed significant changes in exon AMR (Fig. S2B, line graph in red). For
115 instance, in the Plekha5 and Pafah1b1 genes, notable changes in AMR were observed at the
116 switch regions (Fig. S2B). Specifically, 82 out of the 102 genes (80%) with single switches

117 exhibited AMR changes at the switch regions, indicating a possible link between methylation
118 changes and the poly(A) site switch.

119 **3. Differential enrichment of poly(A) signal motifs at the proximal and terminal exon's**
120 **poly(A) sites**

121 We then investigated whether the two poly(A) sites in the 102 genes exhibit distinct
122 poly(A) signal sequences. To this end, we analyzed the consensus poly(A) signal motifs at
123 these sites using MEME motif analysis. This analysis was based on 200 nucleotides flanking
124 the endpoint of read counts as visualized in the IGV, focusing on the proximal and GTE sites
125 of these genes (Fig. 3A). Our findings revealed that the consensus motif at the proximal site
126 was suboptimal compared to that at the GTE site. Additionally, we noted the presence of
127 slightly varied G-rich motifs enriched at both sites.

128 Furthermore, we manually inspected in IGV the RNA-Seq read ends at the GTE of 300
129 genes randomly chosen from the list of 1875 identified GTEs. Our analysis showed that for
130 these genes, 52% had their ending reads align with the annotated gene end, 38% extended
131 beyond this end, and only 10% stopped before the annotated end. This pattern indicates that 5-
132 azaC may lead to an extension of transcript lengths. It does this predominantly by shifting from
133 the proximal to the poly(A) sites located at the annotated transcript ends, or even to novel sites
134 beyond the known GTEs.

135 **4. 5-azaC regulates the expression of genes that promote distal poly(A) usage**

136 To assess whether the poly(A) site switch induced by 5-azaC is accompanied by
137 simultaneous changes in relevant genes, we conducted a DAVID functional clustering analysis.
138 This analysis focused on genes that were significantly upregulated (fold change > 2) or
139 downregulated (fold change < 0.5) by 5-azaC, as determined by edgeR analysis of the same

140 RNA-Seq data. Notably, genes involved in transcription regulation emerged as a major
141 functional cluster among those upregulated by 5-azaC (Fig. S3), while genes associated with
142 oxidative phosphorylation were prominent in the downregulated cluster.

143 Significantly, 5-azaC treatment was observed to increase the expression levels of RNA
144 polymerase II regulation factors (Fig. 3B and Table S2), such as Scaf4 and Scaf8, which are
145 known to promote the usage of distal poly(A) sites[4]. Conversely, the expression of E2f2,
146 known to promote proximal poly(A) site usage[5], was reduced. Additionally, we observed
147 that 5-azaC modulates the expression of several factors involved in alternative polyadenylation
148 and splicing. For instance, factors like Pcf11, Ssu72, Cpeb3, and Hnrnph1/h3 that show a
149 preference for G-rich sequences, were upregulated. In contrast, Esrp2, Srsf9, Rbms2, and the
150 A-rich sequence- preferring Pabpc1 were downregulated (Fig. 3C and Table S3).

151 This pattern of gene regulation suggests a consistent influence of 5-azaC on a group of
152 transcription and pre-mRNA processing factors, which aligns with the observed shift from
153 proximal to distal poly(A) sites. However, an interesting anomaly in this pattern is the
154 upregulation of Pcf11 by 5-azaC. Pcf11 is typically associated with promoting transcript
155 shortening and the downregulation of distal polyadenylation sites.

156 **5. Similar promoting effects on terminal pol(A) sites and full length transcripts by 5-azaC**
157 **are also observed in leukaemia cells upon 5-azaC treatment**

158 To explore whether the effect of 5-azaC in promoting full-length transcripts is specific to
159 GH3 cells or also observable in other cell types, particularly cancer cells, we analyzed raw
160 RNA-Seq data from a study on MOLM-13 human leukemia cells treated with 5-azaC. Focusing
161 on genes with at least 10 exons/bins as indicated in the DEXSeq output, we examined the
162 relative exon usage changes of the first, middle, and last exons. The results revealed similar

163 patterns of fold changes in these cells (Fig. 4A), although the effects were somewhat less
164 pronounced, likely due to the lower concentration of 5-azaC used in this experiment (7 μM)
165 compared to our GH3 cell study (50 μM). Notably, the fold changes in the first and the genomic
166 last exons were again in opposite directions, with a significant shift towards a higher percentage
167 of exons having fold changes greater than 1 (p = 7.62E-123, 6.05E-53, and 0 for the
168 comparisons between the first and last, first and middle, and middle and last exons, respectively,
169 as per the hypergeometric density test).

170 For instance (Fig. 4B-C), in the MOLM-13 cells, the tumor suppressor gene GADD45B
171 was predominantly expressed as a truncated isoform, evidenced by the high read peaks for only
172 the first two exons. This shorter isoform lacks the p38-interacting domain necessary for
173 mediating TGF-beta-induced apoptosis [6, 7]. However, following treatment with 5-azaC, the
174 GADD45B transcript showed nearly equal levels of exons 3 and 4 compared to the first two,
175 resulting in the full-length transcript necessary for apoptosis mediation [7].

176 **6. Common and cell-type specific changes of polyA factors by 5-azaC**

177 To determine if the 5-azaC effects on the transcript levels of transcription,
178 polyadenylation or splicing factors are consistent across different cell types, we also analyzed
179 the RNA-Seq data from leukemia cells, using a similar approach as with the GH3 cell samples.
180 Our analysis revealed that several transcription, polyadenylation, and splicing factors exhibited
181 similar regulatory patterns in both cell lines, with some being upregulated (e.g., PCF11 and
182 CSTF3) and others downregulated (e.g., CPSF1) (Table I).

183 However, our study also identified factors with divergent regulation or whose regulations
184 were specific to certain cell lines. For instance, CLP1, a gene with a mouse homologue (Clp1)
185 known to promote proximal polyadenylation site usage [8], was upregulated in GH3 cells but

186 downregulated in MOLM-13 cells. These findings suggest that while 5-azaC exerts common
187 effects on certain factors influencing transcript lengths, its impact on other factors can vary
188 depending on the cell type.

189

190 **Discussion**

191 In this study, we analyzed RNA-Seq and whole-genome bisulfite sequencing (WGBS)
192 data to investigate the effect of 5-azaC on exon usage and methylation. Our findings reveal a
193 novel impact of 5-azaC: it extends transcript lengths by selectively switching poly(A) sites
194 from proximal to GTE sites, enhancing the use of multiple 3′ exons.

195 While the influence of 5-azaC on alternative splicing is known [9], the mechanism behind
196 this effect remains elusive. Our data suggest an exon rank-dependent impact linked to
197 alternative polyadenylation and a string of 3′ exons: a decrease in the relative usage of the first
198 exons and an increase in 3′ and terminal exons (Fig. 1A). The relationship between this 5-azaC
199 effect on exon usage and differential methylation of exons is not straightforward, except at
200 extreme levels (Fig. 1B). However, methylation changes around most switch sites (Fig. S2B)
201 hint at a potential role of exon methylation in poly(A) site switching, meriting further
202 exploration.

203 The control of alternative polyadenylation involves cis-acting pre-mRNA motifs and
204 trans-acting factors [10]. Besides exon methylation changes (Fig. S2B), the distinct G-rich and
205 consensus motifs at proximal and GTE sites (Fig. 3A), and the alterations in poly(A) site
206 selection factors in the transcription machinery are noteworthy. The upregulation of Scaf4 and
207 Scaf8 and downregulation of E2f2 by 5-azaC (Fig. 3B and Table S2) align with their roles in
208 promoting distal poly(A) usage [4]. Interestingly, Pcf11, typically associated with transcript

209 shortening, is upregulated by 5-azaC in both GH3 and MOLM-13 cells (Figure 3C, Table I and
210 Table S3) [11]. One explanation could be that the Pcf11 is involved in additional pathways
211 beyond APA. For example, lower Pcf11 levels in oral squamous cell carcinoma are linked to
212 increased oxidative phosphorylation and ribosomal pathways [12], whose clusters of genes are
213 also downregulated by 5-azaC (Fig. S3). More interestingly, it could also be a homeostatic
214 effort of the tumor/cancer cells to keep their shortened transcripts in the presence of 5-azaC.

215 The unidirectional effect of 5-azaC on APA, shifting from proximal to terminal sites in
216 the GTE, contrasts with transcript shortening observed in cancer cells [13] [14] [15]. This
217 difference could be a crucial aspect of 5-azaC's therapeutic action, potentially restoring
218 regulatory targets or protein functions in cancer cells. Moreover, this novel effect offers a
219 method to directionally extend transcript lengths for both basic and therapeutic research.

220

221 **Materials and Methods**

222 **Cell culture**: The rat GH3 pituitary cells were cultured as described in [16].

223 5-azaC treatment and DNA/RNA extraction: Cell cultures were treated with 5-azaCytidine
224 (50µm) or DMSO for 18hrs before cell collection. The cytoplasmic RNA was extracted for
225 RNA sequencing and the corresponding nuclear DNA was extracted for Whole Genome
226 Bisulfite Sequencing (WGBS) analysis as described in [16]. The RNA-Seq reads of 5-azaC
227 (7µM)-treated MOLM-13 cells analysed here were obtained from PRJEB61289 deposited in
228 the SRA database by the Pina group[17].

229 **RT-PCR**: RT-PCR was based on our previous procedure [16]. Briefly, for reverse transcription
230 in a 10µl reaction, 300ng of cytoplasmic RNA was added, incubated at 45°C for 50mins and
231 90°C for 5mins. For PCR, 1.5µl of RT product was amplified in a 24µl reaction for 30 cycles.
232 For Zmym3, proximal poly(A) site, we used upstream primer 5'-
233 GATTCAGTTGTGGTACAGACTGGTTCAG-3' and downstream primer 5'-
234 CTCGACCCTTCAGGCACCTTTATTTATC-3', while for the GTE Poly(A) site, we used
235 upstream primer 5'-GGTCCTGGGAAACGAAAGAGAGAAGATG-3' and downstream
236 primer 5'-GCTACCTGGTTCACGTCCATCACTGG-3'. The PCR products were resolved in
237 agarose gels and photographed under UV lights using a digital camera.

238 **Bioinformatics analysis**: We used IGV to visualize mapped reads [18], DAVID for functional
239 annotation and clustering analysis of genes [19], Interpro protein family database for protein
240 domain search [20], and MEME for motif analysis [21]. DEXSeq and edgeR analysis were
241 carried out as in the previous studies[16, 22-24].

242 **Statistical tests**: All samples were in triplicate. Two-tailed Student's t-test was used for the
243 RT-PCR products.

244

Acknowledgements: The work is supported by a discovery grant (RGPIN-2022-05023) by the Natural Sciences and Engineering Council of Canada (NSERC) to J.X.

248 **References**

249 1. Amaral, M.S., et al., Long non-coding RNA levels can be modulated by 5-azacytidine
250 in Schistosoma mansoni. Scientific Reports, 2020. 10(1).
251 2. Anastasiadi, D., A. Esteve-Codina, and F. Piferrer, Consistent inverse correlation
252 between DNA methylation of the first intron and gene expression across tissues and
253 species. Epigenetics & Chromatin, 2018. 11.
254 3. Leung, J.W.C., et al., ZMYM3 regulates BRCA1 localization at damaged chromatin
255 to promote DNA repair. Genes & Development, 2017. 31(3): p. 260-274.
256 4. Gregersen, L.H., et al., SCAF4 and SCAF8, mRNA Anti-Terminator Proteins. Cell,
257 2019. 177(7): p. 1797-1813 e18.
258 5. Elkon, R., et al., E2F mediates enhanced alternative polyadenylation in proliferation.
259 Genome Biol, 2012. 13(7): p. R59.
260 6. Tamura, R.E., et al., GADD45 proteins: central players in tumorigenesis. Curr Mol
261 Med, 2012. 12(5): p. 634-51.
262 7. Yoo, J., et al., Transforming growth factor-beta-induced apoptosis is mediated by
263 Smad-dependent expression of GADD45b through p38 activation. J Biol Chem, 2003.
264 278(44): p. 43001-7.
265 8. Monaghan, C.E., et al., The Clp1 R140H mutation alters tRNA metabolism and
266 mRNA 3' processing in mouse models of pontocerebellar hypoplasia. Proc Natl Acad
267 Sci U S A, 2021. 118(39).
268 9. Ishfaq, T., et al., DNA demethylation and double-strand breaks affect levels of the
269 BRCA2 Delta 3 alternate splicing isoform. Cancer Research, 2020. 80(16).
270 10. Newnham, C.M., et al., Alternative polyadenylation of MeCP2: Influence of cis-
271 acting elements and trans-acting factors. RNA Biol, 2010. 7(3): p. 361-72.
272 11. Kunisky, A.K., et al., Shift in MSL1 alternative polyadenylation in response to DNA
273 damage protects cancer cells from chemotherapeutic agent-induced apoptosis. Cell
274 Rep, 2021. 37(2): p. 109815.
275 12. Chen, G., et al., RNA Binding Protein-related bioinformatics analysis reveals that
276 PCF11 is a prognostic biomarker in oral squamous cell carcinoma. 2023, Research
277 Square.
278 13. Masamha, C.P., The emerging roles of CFIm25 (NUDT21/CPSF5) in human biology
279 and disease. Wiley Interdiscip Rev RNA, 2022: p. e1757.
280 14. Mohanan, N.K., et al., Alternative polyadenylation: An enigma of transcript length
281 variation in health and disease. Wiley Interdiscip Rev RNA, 2022. 13(1): p. e1692.
282 15. Ye, C., et al., Role of alternative polyadenylation dynamics in acute myeloid
283 leukaemia at single-cell resolution. RNA Biol, 2019. 16(6): p. 785-797.
284 16. Lei, L., et al., Multilevel Differential Control of Hormone Gene Expression Programs
285 by hnRNP L and LL in Pituitary Cells. Molecular and Cellular Biology, 2018. 38(12).
286 17. Keti Zeka, et al., MAT2A inhibition in AML unveils therapeutic potential of
287 combining DNA demethylating agents with UPR targeting. bioRxiv 2023: p. doi:
288 https://doi.org/10.1101/2023.06.05.543499.
289 18. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, Integrative Genomics Viewer
290 (IGV): high-performance genomics data visualization and exploration. Brief
291 Bioinform, 2013. 14(2): p. 178-92.
292 19. Huang, D.W., B.T. Sherman, and R.A. Lempicki, Systematic and integrative analysis
293 of large gene lists using DAVID bioinformatics resources. Nature Protocols, 2009.
294 4(1): p. 44-57.
295 20. Blum, M., et al., The InterPro protein families and domains database: 20 years on.
296 Nucleic Acids Research, 2021. 49(D1): p. D344-D354.

297 21. Bailey, T.L., et al., MEME SUITE: tools for motif discovery and searching. Nucleic
298 Acids Res, 2009. 37(Web Server issue): p. W202-8.
299 22. Liu, L., et al., Transcriptome-Wide Detection of Intron/Exon Definition in the
300 Endogenous Pre-mRNA Transcripts of Mammalian Cells and Its Regulation by
301 Depolarization. Int J Mol Sci, 2022. 23(17).
302 23. Anders, S., A. Reyes, and W. Huber, Detecting differential usage of exons from
303 RNA-seq data. Genome Res, 2012. 22(10): p. 2008-17.
304 24. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, edgeR: a Bioconductor package for
305 differential expression analysis of digital gene expression data. Bioinformatics, 2010.
306 26(1): p. 139-40.

307

Figure Legend

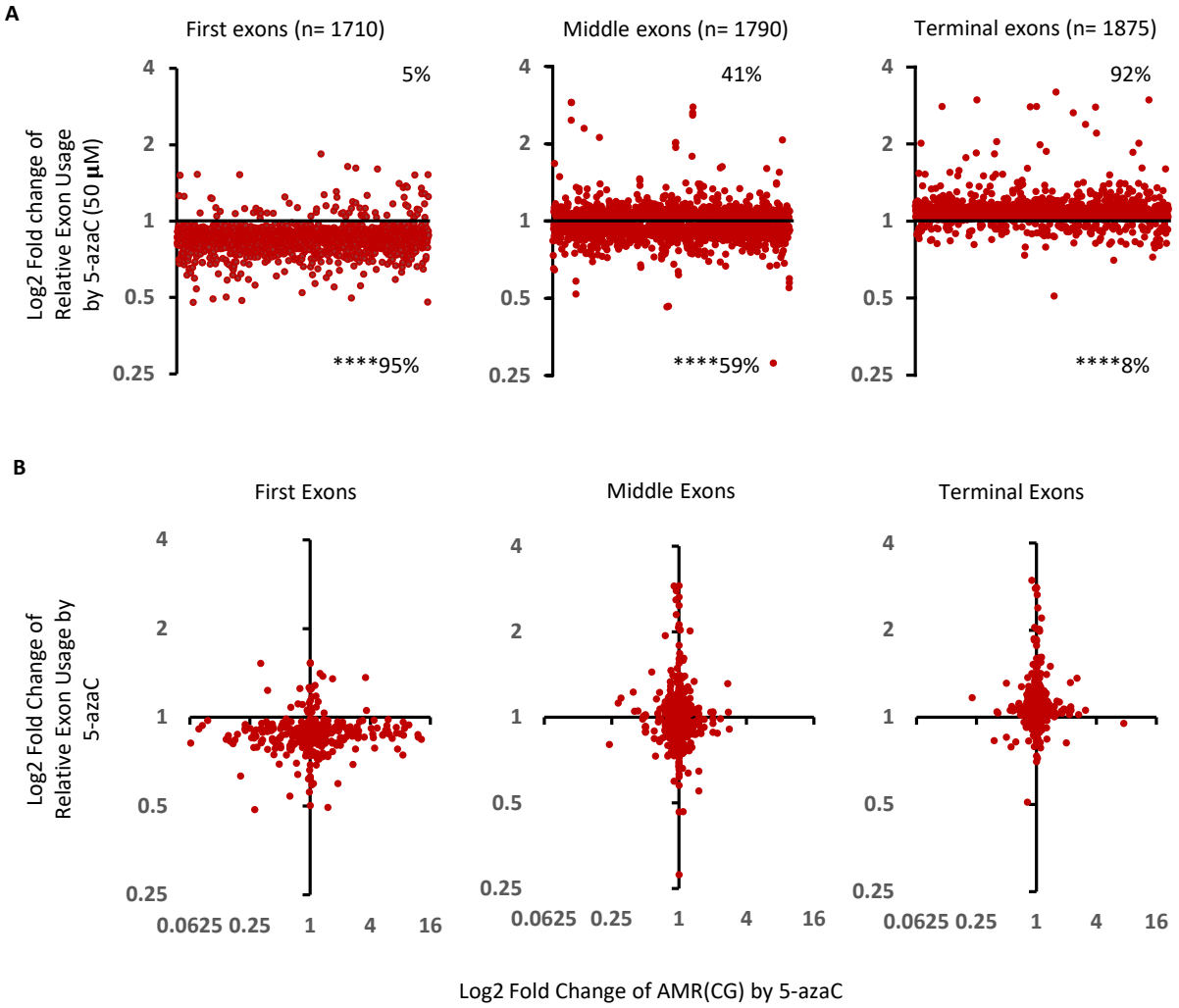
Fig. 1. Exon usage of the first, middle and terminal exons after 5-azaC treatment. Scatter plots of the log2 fold change of exon usage (**A**), and relative to the log2 fold change of average methylation ratio (**B**).

Fig. 2. The exon rank-dependent effect of 5-azaC on exon usage across a gene, representative of 102 exons. **A**. IGV view and corresponding bar plot of the 5-azaC effect on exon usage of the *Zmym3* gene. The bar graph was obtained from DEXSeq analysis data. **B**. RT-PCR validation of the reduced relative usage of the proximal poly(A) site and increased GTE poly(A) sites of the *Zmym3* gene after 5-azaC treatment (n = 3), with the PCR primers (arrowheads) diagrammed above the gel. **C**. Protein domains of the resulting short (496 amino acids, transcript ID - ENSRNOT00000076008.3) and long isoform (1371 amino acid, transcript ID - ENSRNOT00000076004.3) from the Interpro protein family database.

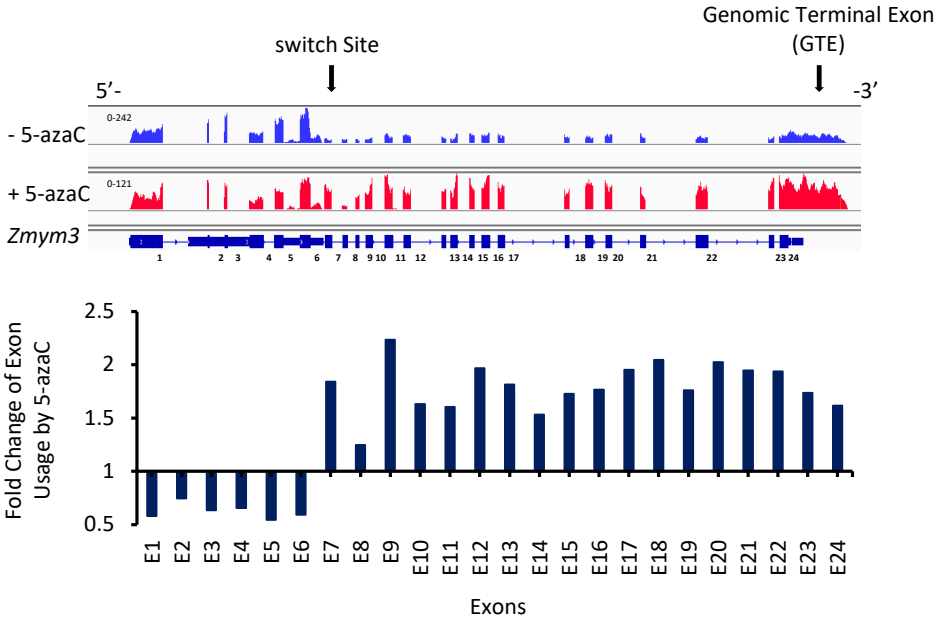
Fig. 3. Consensus motifs of different reads termination sites and mRNA transcripts changed by 5-azaC. A. Flow chart showing the selection of 200 nucleotides around the reads′ termination sites of 102 genes with single switches for MEME motif analysis (n - sequence sites contributing to the motif construction). Bar plot of average normalized counts of transcription regulation factors (*Scaf4, Scaf8 and E2f2,* **B**) and in **C**, mRNA polyadenylation and splicing factors (*Pcf11* and *Hnrnph, and Pabpc1*) that were significantly (FDR<0.01, 0.5<FC>2, ARC > 200) changed by 5-azaC treatment, from edgeR analysis of RNA-Seq data

Fig. 4. Similar 5-azaC effect promoting usage of the terminal polyadenylation sites for full length transcript production in a leukaemia cell line. A. Scatter plots of exon usage fold changes of the first, middle and GTE terminal exons, from DEXSeq analysis of the RNA-Seq

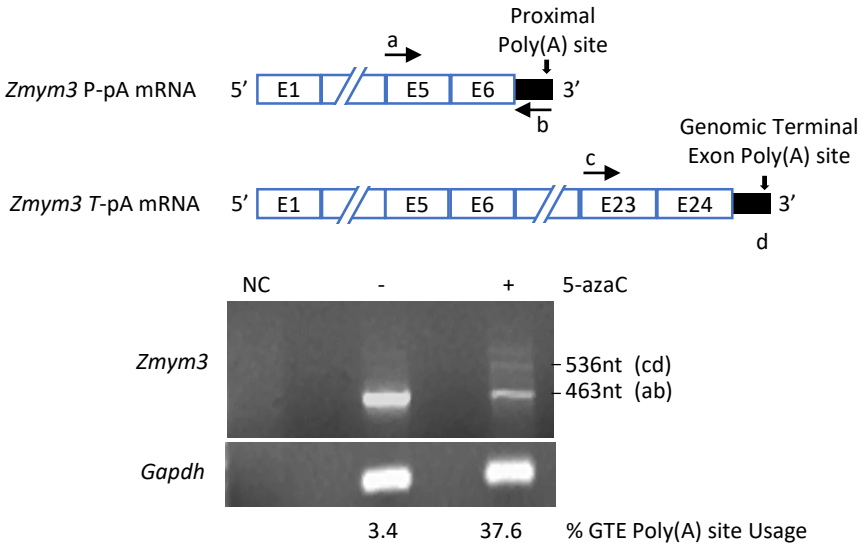
data of human MOLM-13 cells treated with 5-azaC (7μM) (Zeka, et al., 2023). The percentages of exons with fold changes above or below 1.0 are shown in each group. ****: $p < 1E-50$, by hypergeometric density test between any two groups. **B**. An example of the poly(A) site switch to the GTE of the *GADD45B* tumour suppressor gene. Shown in the IGV view of the reads peaks of each exon before and after 5-azaC treatment, with a bar graph of the fold changes of each exon below the view. Note that here the exons ranks are for each full exon instead of the bins in DEXSeq (e.g. E001). **C**. GADD45B protein domains added upon 5-azaC-enhanced terminal poly(A) site usage and full length transcript. The p38 MAP kinase interaction domain added is essential for TGF-beta-induced apoptosis through GADD45B.


A
First exons (n= 1710)
Middle exons (n= 1790)
Terminal exons (n= 1875)
Log2 Fold change of Relative Exon Usage by 5-azaC (50 µM)
5%
****95%
41%
****59%
92%
****8%
4
2
1
0.5
0.25
B
First Exons
Middle Exons
Terminal Exons
Log2 Fold Change of Relative Exon Usage by 5-azaC
0.0625 0.25 1 4 16
Log2 Fold Change of AMR(CG) by 5-azaC

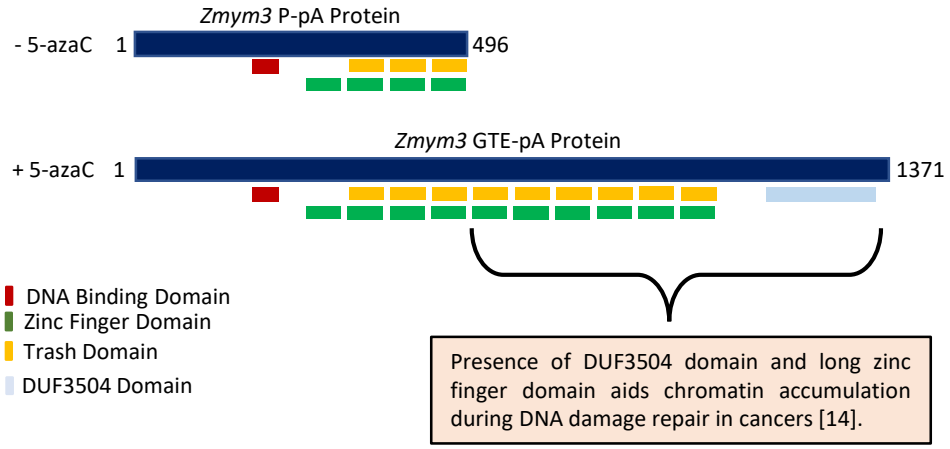
A


Genomic Terminal Exon
(GTE)
switch Site
5'-
-3'
- 5-azaC
+ 5-azaC
Zmym3
Fold Change of Exon
Usage by 5-azaC
2.5
2
1.5
1
0.5
E1 E2 E3 E4 E5 E6 E7 E8 E9 E10 E11 E12 E13 E14 E15 E16 E17 E18 E19 E20 E21 E22 E23 E24
Exons

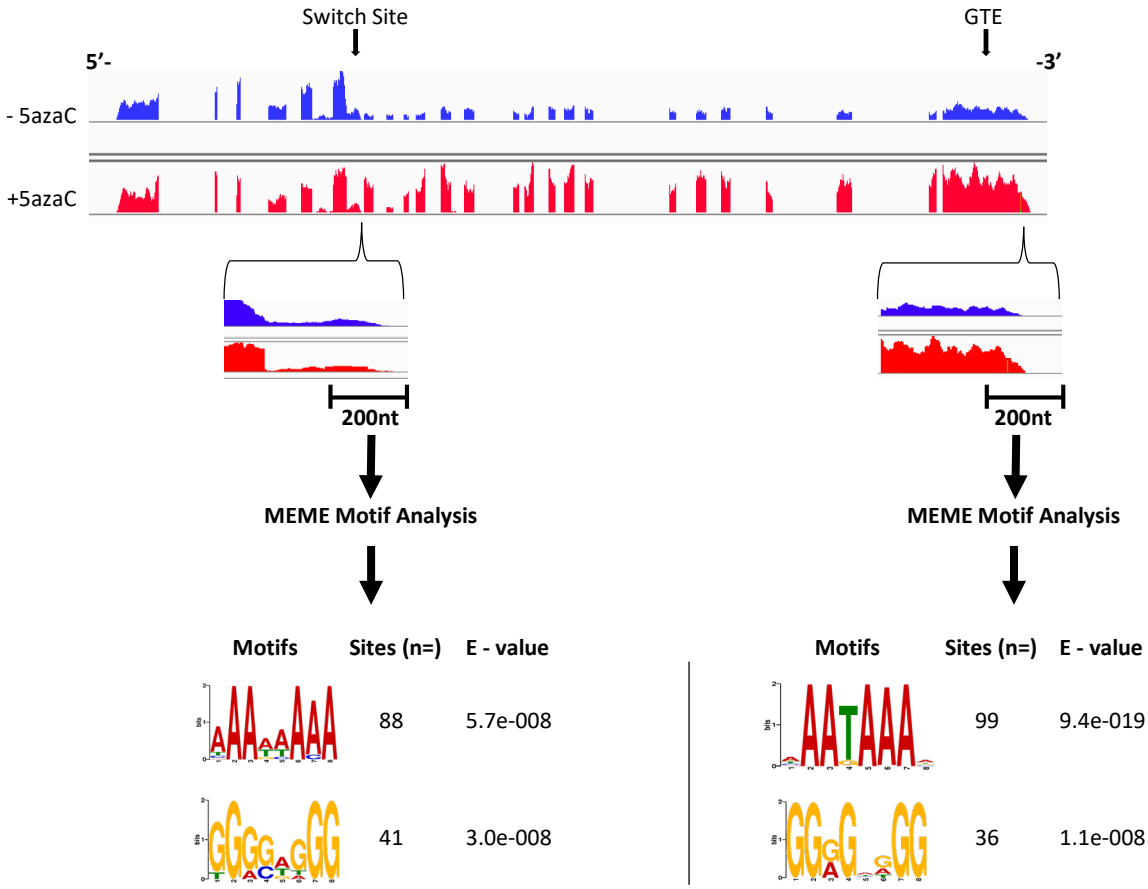
B


Proximal
Poly(A) site
a
Zmym3 P-pA mRNA
5'
E1
E5
E6
3'
b
Genomic Terminal
Exon Poly(A) site
c
Zmym3 T-pA mRNA
5'
E1
E5
E6
E23
E24
3'
d
NC
-
+
5-azaC
Zmym3
536nt (cd)
463nt (ab)
Gapdh
3.4
37.6
% GTE Poly(A) site Usage

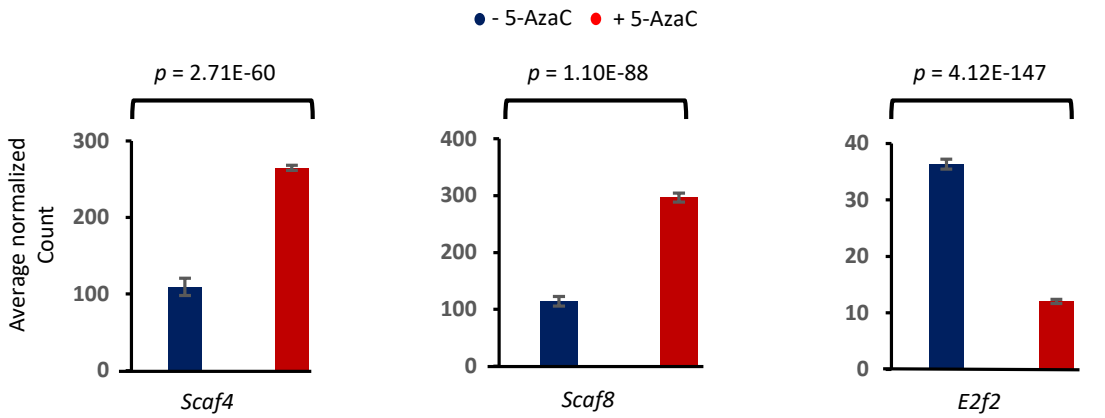
C


Zmym3 P-pA Protein
- 5-azaC
1
496
Zmym3 GTE-pA Protein
+ 5-azaC
1
1371
DNA Binding Domain
Zinc Finger Domain
Trash Domain
DUF3504 Domain
Presence of DUF3504 domain and long zinc finger domain aids chromatin accumulation during DNA damage repair in cancers [14].

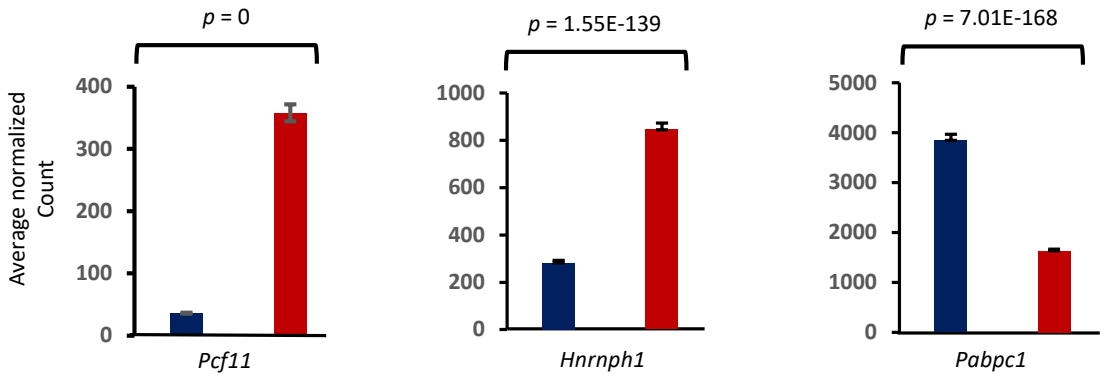
A

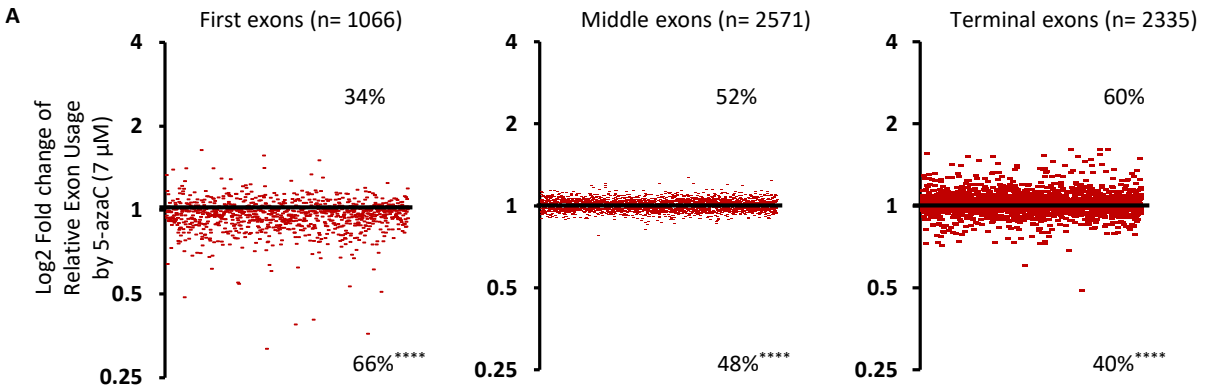

Switch Site
GTE
5'-
-3'
- 5azaC
+5azaC
200nt
200nt
MEME Motif Analysis
MEME Motif Analysis
Motifs
Sites (n=)
E - value
88
5.7e-008
41
3.0e-008
Motifs
Sites (n=)
E - value
99
9.4e-019
36
1.1e-008

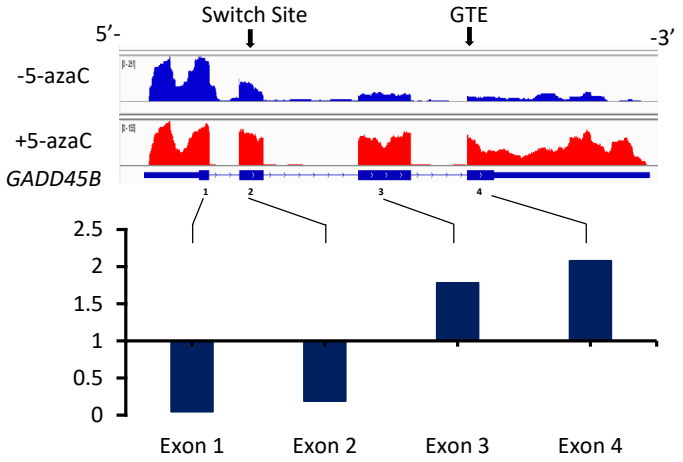
B


- 5-AzaC
+ 5-AzaC
p = 2.71E-60
p = 1.10E-88
p = 4.12E-147
Average normalized Count
Scaf4
Scaf8
E2f2

C


p = 0
p = 1.55E-139
p = 7.01E-168
Average normalized Count
Pcf11
Hnrnph1
Pabpc1


A
First exons (n= 1066)
Middle exons (n= 2571)
Terminal exons (n= 2335)
Log2 Fold change of Relative Exon Usage by 5-azaC (7 µM)
4
2
1
0.5
0.25
34%
66%****
52%
48%****
60%
40%****


B
Switch Site
GTE
5'-
-3'
-5-azaC
+5-azaC
GADD45B
1
2
3
4
2.5
2
1.5
1
0.5
0
Exon 1
Exon 2
Exon 3
Exon 4

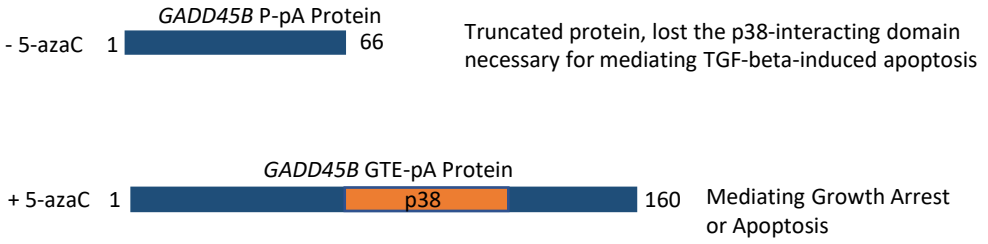

C
GADD45B P-pA Protein
- 5-azaC
1
66
Truncated protein, lost the p38-interacting domain necessary for mediating TGF-beta-induced apoptosis
GADD45B GTE-pA Protein
+ 5-azaC
1
p38
160
Mediating Growth Arrest or Apoptosis

Table 1. Common and cell type-specific changes of factors in transcription, polyadenylation and alternative splicing by 5-azaC in GH3 and/or MOLM-13 cells. Showing are lists genes from edgeR analysis of RNA-Seq data (FDR<0.1, 0.75>FC>1.25) upon 5-azaC treatment. FC: fold change, FDR: false discovery rate.

Transcription Factors				
	GH3		MOLM-13	
Genes	FDR	FC	FDR	FC
Ell	2.39E-184	3.53	4.99E-05	2.37
Atf5	3.27E-39	3.05	9.81E-09	0.39
Arid5a	3.43E-25	2.58	8.69E-05	2.20
Arid4a	6.93E-92	2.43	3.38E-07	3.09
Atf3	1.52E-22	2.36	1.18E-06	0.12
Scaf8	3.53E-87	2.32	5.65E-02	0.73
Scaf4	4.84E-59	2.16	8.02E-02	0.75
E2f4	3.42E-145	0.84	1.08E-13	0.29
Atf6b	2.67E-59	0.45	5.10E-02	1.35
Crebzf	3.45E-98	0.45	2.12E-05	0.51
Rbl2	1.41E-69	0.45	1.71E-02	1.45
Cdk2	1.03E-41	0.42	3.09E-02	0.59

Polyadenylation Factors				
	GH3		MOLM-13	
Genes	FDR	FC	FDR	FC
PCF11	0.00E+00	8.80	2.87E-02	1.45
SSU72	6.59E-213	3.83	1.80E-02	0.69
CSTF3	8.50E-149	3.65	1.93E-02	1.52
CLP1	2.75E-144	3.23	2.62E-04	0.32
CPSF7	5.20E-44	1.83	9.40E-06	0.48
WDR33	3.50E-39	1.67	7.78E-02	1.33
TFIP11	4.42E-26	1.59	7.08E-02	0.73
CPSF4	5.46E-02	0.91	1.57E-03	0.5
CPSF1	2.91E-25	0.58	7.94E-02	0.74

Splicing Factors				
	GH3		MOLM-13	
Genes	FDR	FC	FDR	FC
Prpf39	2.89E-202	4.81	1.37E-02	0.67
Prpf3	3.48E-180	3.25	9.84E-02	0.74
Sf3b3	2.16E-43	1.78	9.07E-03	0.66
Tfip11	4.42E-26	1.59	7.08E-02	0.73
Prpf6	1.11E-25	1.45	1.17E-02	1.48
Hnrnpll	1.05E-19	1.44	2.32E-04	1.77
U2af1	7.56E-04	1.19	8.08E-03	0.64
U2af2	5.13E-03	1.11	4.66E-04	0.58
Ptbp3	4.00E-15	0.70	3.95E-04	1.68
Prpf31	2.49E-30	0.55	1.77E-06	0.47
Pabpc1	7.76E-166	0.38	1.27E-05	1.74