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

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

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

Conflict of Interest: *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 (DNMTs) and an anti-cancer drug, 5-aza-cytidine (5-azaC)'s many effects on gene expression remains unclear. Here, we show that 5-azaC treatment of cultured GH_3 pituitary tumour cells increases relative usage of genomic terminal exons (GTEs) across the transcriptome. This effect is largely achieved by shifting mRNA polyadenylation from proximal poly(A) sites to GTEs, which harbour a more optimal consensus motif of poly(A) signals. Consistent with this shift, 5-azaC upregulates the mRNA anti-termination factors Scaf4 and Scaf8 while downregulating the early termination enhancer E2f2. In MOLM-13 leukaemia cells, 5-azaC similarly promotes the production of full-length transcripts and regulates alternative polyadenylation factors, some of which in the same direction as observed in GH_3 cells. Moreover, PCF11, a factor known to promote proximal poly(A) site usage, is upregulated in both cell lines, suggesting a homeostatic response by these cells to counteract transcript lengthening during 5-azaC treatment. Together, these findings uncover a previously unrecognized effect of 5-azaC on gene expression: directional promotion of terminal polyadenylation site usage, driving a transcriptome-wide switch from shortened to full-length mRNAs in tumour or cancer cells and consequently altering the alternative usage of multiple 3′ exons.

Introduction

5-aza-cytidine (5-azaC), an FDA-approved chemotherapeutic agent that binds and inactivates DNA methyltransferases (DNMTs), inhibits DNA methylation [1]. In addition to its well-established effects on DNA, 5-azaC influences multiple aspects of RNA metabolism [1]. However, its impact on RNA processing, particularly at the exon level, remains poorly understood.

DNA methylation and gene expression exhibit a gene-region-specific relationship: methylation level of promoters, first exons and first introns, but not the rest of gene body, is negatively correlated with gene expression [2, 3]. Whether 5-azaC exerts similarly region-dependent effects on exonic DNA methylation and exon usage has not been explored. This question motivated the present study and led to both expected and unexpected findings.

Here, we analysed RNA-Seq and whole-genome bisulfite sequencing (WGBS) data from GH_3 pituitary tumour cells treated with 5-azaC. We focused on differential methylation and usage of genomic first, internal, and terminal exons (GTEs, to distinguish them from transcript-defined alternative terminal exons). Remarkably, 5-azaC promoted widespread usage of multiple 3′ exons and terminal polyadenylation sites, accompanied by more frequent changes of DNA methylation of GTEs. In parallel, 5-azaC altered the expression of genes involved in transcript length regulation and pre-mRNA processing. Importantly, similar effects were observed in a leukaemia cell line. Together, these findings suggest that 5-azaC may counteract transcript shortening in cancer cells through a unique, directional effect on 3′ exon usage, polyadenylation, and exon-specific DNA methylation, thereby restoring full-length mRNAs, their encoded protein domains, and associated functions.

Results

1. Differential effects of 5-azaC on the relative exon usage of the genomic first, internal and terminal exons (GTE) of genes

To analyze potential exon-rank-dependent effects by the DNA methylation inhibitor 5-azaC, we treated GH_3 pituitary tumor cells with 50μM 5-azaC and compared them with untreated controls [4]. Our analysis focused on the RNA sequencing data of genomic first, internal, and terminal exons by using DEXSeq [5], where first and terminal exons were designated as E001 for the positive (+) and negative (-) chromosome strands, respectively. For internal exons, we selected E004, E005, and E006 from the DEXSeq exon/bin list of genes with more than 8 exons. To further validate exon ranks, we randomly selected 100 genes from each group to verify in the Integrated Genomics Viewer [6]. This approach revealed 1709 first exons, 1790 internal exons, and 1874 terminal exons with a statistically significant difference upon 5-azaC treatment (Fig. 1A, $padj < 0.05$). Among the three groups of exons emerged an intriguing pattern: 5% of the first, 41% of the internal, and 92% of the terminal exons showed increased relative usage. This increasing percentages suggest a distinct exon-rank-dependent impact of 5-azaC on exon usage (Fig. 1A).

Moreover, we assessed the average CG methylation ratio (AMR) of exons using the samples' available WGBS data [4]. AMR levels corroborated previous findings on exon-rank dependent methylation of expressed genes [2, 3]: most first exons were nearly unmethylated, whereas most internal and terminal exons were highly methylated in both the 5-azaC-untreated or treated cells (Fig. S1). AMR increases or decreases were distributed near evenly among exons with increased usages in each exon group; the same was observed among ones with reduced usages (Fig. 1B). However, the highest proportion of exons showing AMR changes occurred in the terminal exon group (84.6%, $p < 2.8E-120$, compared to the first exons),

particularly in the first quadrant, suggesting a preferential association between AMR changes and increased terminal exon usage.

2. Transcriptome-wide switch from proximal to distal poly(A) sites induced by 5-azaC

We next examined exon usage in 154 and 153 genes on the (+) and (-) strands, respectively, each containing at least 10 exons/bins identified by DEXSeq analysis. Analysis of 5-azaC-induced fold changes in exon-usage revealed a shift toward increased relative usage of 3′ exons, including terminal exons, in 102 genes. For example, in the *Nfx1* gene upon 5-azaC treatment, the read ratios of the first 16 exons relative to untreated controls were mostly below 1 (Fig. 2A). In contrast, the read ratio for all downstream 3′ exons shifted to values greater than 1, indicating increased expression of a longer *Nfx1* isoform upon 5-azaC treatment. This isoform switch was independently confirmed by quantitative real-time PCR (qPCR) of RT products (Fig. 2B; $p < 0.01$, $n = 3$). The resulting longer isoform encodes an NFX1 protein containing an additional R3H domain (Fig. 2C), which binds single-stranded DNA or RNA [7], including hTERT mRNA in HPV16 E6-mediated enhancement of hTERT expression [8].

Similarly, the *Zmym3* gene exhibited a poly(A) site switch following 5-azaC treatment (Fig. S2A). In untreated samples, RNA-seq reads corresponding to the first six exons were approximately fivefold higher than those of the downstream 3′ exons. Upon 5-azaC treatment, expression of the 3′ exons increased to levels comparable to those of the 5′ exons, with relative usage ratios shifting from below 1.0 to above 1.0 at the junction between exons 6 and 7, indicative of a switch to the distal poly(A) site. The resulting long isoform encodes 13 additional protein domains, including DUF3504 implicated in DNA damage repair [9] (Fig. S2C).

Together, these findings suggest that 5-azaC-induced switches of alternative polyadenylation (APA) sites preferentially generate longer protein isoforms containing complete functional domain sets. These domains are among proteins with such functions as ATP binding, microtubule cytoskeleton organization, ATPase activity, and microRNAs in cancer (Table S1), as shown in the DAVID functional clustering analysis of the 102 genes.

In addition to single APA switches, a subset of genes displayed multiple switching events accompanied by alternative splicing, including exon skipping and intron retention at the switch regions. For example, intron 2 retention was detected in the *Gln1* gene following 5-azaC treatment (Fig. S3A). Notably, regions undergoing poly(A) site switching also showed pronounced changes in exon AMR. In particular, *Plekha5* and *Pafah1b1* exhibited marked AMR alterations coinciding with their switch regions (Fig. S3B, red lines in graph). Moreover, 82 of the 102 genes (80%) with single APA switches displayed significant AMR changes at the corresponding switch sites, supporting a potential mechanistic link between DNA methylation dynamics and poly(A) site selection.

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

We next asked whether the two poly(A) sites in the 102 genes possess distinct poly(A) signal features. To address this, we analyzed consensus poly(A) signal motifs using MEME motif analysis[10], based on 200-nucleotide regions flanking the endpoints of RNA-seq read-counts visualized in IGV (Fig. 3A). We found that the consensus motif at the proximal site was relatively suboptimal compared with that at the GTE site. In addition, both sites were enriched for related but slightly divergent G-rich motifs.

To further assess transcript termination patterns, we manually inspected RNA-seq read ends at the GTEs of 300 genes randomly selected from the 1,874 identified GTEs using IGV. Among these genes, 52% of read ends aligned precisely with the annotated gene end, 38% extended beyond, and only 10% terminated upstream the annotated end. Together, these results indicate that 5-azaC preferentially promotes full-length transcript production, primarily by shifting poly(A) site usage from proximal sites to annotated gene-end sites, and in many cases also to previously unannotated downstream poly(A) sites.

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

To determine whether 5-azaC-induced poly(A) site switching is accompanied by coordinated changes in relevant regulatory genes, we performed DAVID functional clustering analysis on genes significantly upregulated (fold change > 2) or downregulated (fold change < 0.5) by 5-azaC, as identified by edgeR analysis of the same RNA-seq dataset. Genes involved in transcriptional regulation emerged as a dominant functional cluster among the upregulated genes (Fig. S4), whereas genes associated with oxidative phosphorylation were prominently enriched among the downregulated genes.

Particularly, 5-azaC treatment increased the expression of multiple RNA polymerase II regulatory factors (Fig. 3B and Table S2), including Scaf4 and Scaf8, which promote distal poly(A) site usage [11]. In contrast, expression of E2f2, a factor known to favor proximal poly(A) site selection [12], was reduced. In addition, 5-azaC altered the expression of several factors implicated in alternative polyadenylation and/or splicing. Specifically, factors with reported preferences for G-rich sequences, such as Pcf11, Ssu72, Cpeb3, and Hnrnph1/h3, were upregulated, whereas Esrp2, Srsf9, Rbms2, and the A-rich–sequence–binding factor Pabpc1 were downregulated (Fig. 3C and Table S3).

Collectively, this pattern of gene regulation suggests that 5-azaC exerts a coordinated influence on transcriptional and pre-mRNA processing machinery, with the changes of at least some factors consistent with the observed global shift from proximal to distal poly(A) site usage. An apparent exception to this trend is the upregulation of Pcf11, which has been typically associated with transcript shortening and repression of distal polyadenylation sites [13], suggesting potential context-dependent or compensatory roles for this factor under 5-azaC treatment.

5. Similar promoting effects on terminal poly(A) sites and full-length transcripts in leukaemia cells by 5-azaC

To determine whether the ability of 5-azaC to promote full-length transcript production is specific to GH_3 cells or also occurs in other cell types, particularly cancer cells, we analyzed raw RNA-Seq data from a study of 5-azaC-treated MOLM-13 human leukemia cells [14], deposited in the NCBI SRA (Table S4). Although a lower concentration of 5-azaC (7 µM) was used in this study [14], presence of a similar poly(A) site-switching trend would provide strong independent support for our observations in GH_3 cells.

By using the DEXSeq analysis, we focused on genes with at least 10 exons/bins and examined exon-usage changes across first, internal, and terminal exons. Despite the reduced magnitude of the effect, likely reflecting the lower 5-azaC concentration and/or cell-type differences compared with our GH_3 experiments (50 µM), a highly similar pattern of exon-usage shifts was observed (Fig. 4A). Specifically, fold changes for first and genomic terminal exons again shifted in opposite directions, with a highly significant enrichment of terminal exons exhibiting fold changes greater than 1.0 (p = 6.0E-42).

A representative example is the tumor suppressor gene *GADD45B* (Fig. 4B-C). In untreated MOLM-13 cells, *GADD45B* was predominantly expressed as a truncated isoform, as indicated by strong RNA-seq read coverage limited to the first two exons. This shorter isoform lacks the p38-interacting domain required for mediating TGF-β-induced apoptosis [15, 16]. Following 5-azaC treatment, read coverage across exons 3 and 4 increased to levels comparable to those of the first two exons, consistent with restoration of the full-length GADD45B transcript necessary for apoptotic signaling [16].

6. Common changes in poly(A) factors between 5-azaC-treated cell lines

To determine whether the effects of 5-azaC on transcript levels of transcription-, polyadenylation-, and splicing-related factors are conserved between the two experimental systems, we analyzed the RNA-Seq data from MOLM-13 leukemia cells using the same edgeR analytical parameter applied to the GH_3 cell samples. This comparison revealed that multiple transcription, polyadenylation, and splicing factors exhibited concordant regulatory trends in both cell lines, with several factors being consistently upregulated (e.g., PCF11 and CSTF3) and others downregulated (e.g., CPSF1) (Table 1). We also identified factors showing divergent regulation between the two systems. For example, CLP1, whose mouse homolog (Clp1) has been reported to promote proximal poly(A) site usage [17], was upregulated in GH_3 but downregulated in MOLM-13 cells. Despite the differences, these results indicate that 5-azaC induces a shared regulatory group of transcriptional and RNA-processing factors that are known to influence transcript length and poly(A) site selection.

Discussion

In this study, we analyzed RNA-seq and whole-genome bisulfite sequencing (WGBS) data to investigate the effects of 5-azaC on exon usage and exon methylation. Our results reveal a previously unrecognized impact of 5-azaC: it extends transcript length by selectively switching poly(A) site usage from proximal sites to genomic terminal exon (GTE) sites, thereby enhancing the coordinated usage of multiple downstream 3′ exons.

Although 5-azaC is known to influence alternative splicing [4], the underlying mechanisms remain poorly understood. Our data indicate a pronounced exon rank-dependent effect associated with alternative polyadenylation and a string of 3′ exons, characterized by reduced relative usage of first exons and increased usage of downstream 3′ and terminal exons (Fig. 1A). Interestingly, changes of exonic DNA methylation upon 5-azaC treatment also occurred on most of the terminal exons, and as reported [4], inversely correlate with exon usage changes at extreme levels (Fig. 1B). Together with consistent methylation changes surrounding most poly(A) switch sites (Fig. S2B), these observations imply a potential role for exon DNA methylation in poly(A) site selection.

Alternative polyadenylation is regulated by both *cis*-acting pre-mRNA motifs and *trans*-acting factors [18]. In addition to exon methylation changes (Fig. S2B), we observed distinct G-rich and consensus poly(A) motifs at proximal versus GTE sites (Fig. 3A), as well as significant alterations in poly(A) site selection factors within the transcriptional machinery. Particularly, the upregulation of Scaf4 and Scaf8 and downregulation of E2f2 following 5-azaC treatment (Fig. 3B; Table S2) are consistent with their known roles in promoting distal poly(A) usage [11]. In contrast, Pcf11, a factor typically associated with transcript shortening, is upregulated by 5-azaC in both GH_3 and MOLM-13 cells (Fig. 3C; Table 1; Table S3)) [13].

This apparent contradiction may reflect a homeostatic response by tumor cells attempting to preserve shortened transcripts in the presence of the anti-cancer drug 5-azaC.

The unidirectional effect of 5-azaC on alternative polyadenylation, shifting poly(A) usage from proximal to terminal GTE sites, stands in sharp contrast to the transcript shortening commonly observed in cancer cells [19-21]. This distinction may represent a key component of 5-azaC's therapeutic action, potentially restoring regulatory elements or protein functions that are lost through aberrant transcript shortening in cancer. More broadly, this newly identified effect provides a strategy to directionally extend transcript length, offering both mechanistic insight into RNA processing and a potential experimental or therapeutic avenue applicable to other existing or candidate anti-cancer drugs.

Materials and Methods

Cell culture: The rat GH_3 pituitary cells were cultured as described in [4].

5-azaC treatment and DNA/RNA extraction: Cell cultures were treated with 5-aza-cytidine (50µM) or DMSO for 18hrs before cell collection. The cytoplasmic RNA was extracted for RNA sequencing and the corresponding nuclear DNA was extracted for WGBS analysis as described in [4, 22]. The RNA-Seq reads of 5-azaC (7µM)-treated MOLM-13 cells analysed here were obtained from PRJEB61289 deposited in the NCBI SRA database by the Pina group [14].

Quantitative real-time PCR (qPCR): For cDNA preparation by reverse transcription, 300ng of cytoplasmic RNA was added to a 10µl reaction, incubated at 45ºC for 50 minutes and 90ºC for 5 minutes. The resulting cDNA was then used to quantitate the levels of *Nfx1* proximal and GTE poly(A) site for the 5-azaC-treated and non-treated samples, using primers *Nfx1* E15F (5'-GTCCATCTGTCCTCCTACCACG-3') and *Nfx1* E16R (5'-GTTCCAGACAGCTTACTGACG-3') for the proximal poly(A) site and *Nfx1* E23F (5'-CATTCTTGTCATAGTGAGGAGAAG-3') and *Nfx1* E24R (5'-GTCAACTTAACAAACAAATATGTC-3') for the GTE poly(A) site. The samples were pre-confirmed by carrying out RT-PCR on the *Prl* gene using *Prl* -specific primers for reduced Prl expression by 5-azaC, as described previously [4]. qPCR was conducted with PowerUp™ SYBR™ Green (Applied Biosystems) on a QuantStudio 3 instrument and analysed with QuantStudio™ Design and Analysis software. Samples were amplified in triplicate and normalized to *Gh1* (growth hormone gene) using sequence-specific primers as described in [4]. Error bars represent 95% confidence intervals of the quotient of *Nfx1* GTE/proximal poly(A) site normalized to *Gh1* results.

Bioinformatics analysis: We used IGV to visualize mapped reads [6], DAVID for functional annotation and clustering analysis of genes [23], Interpro protein family database for protein

domain search [24], and MEME for motif analysis [10]. DEXSeq and edgeR analysis were carried out as in the previous studies[5, 22, 25, 26].

Statistical tests: All samples were in triplicate. Two-tailed Student's t-test was used for qPCR.

Acknowledgements: This work was supported by a Discovery Grant (RGPIN-2022-05023) from the Natural Sciences and Engineering Research Council of Canada (NSERC) to J.X. S.O. is supported by a Vanier Canada Graduate Scholarship (FRN - CGV_186989). We thank Peisan Lew for assistance with qPCR.

References

1. Amaral, M.S., et al., *Long non-coding RNA levels can be modulated by 5-azacytidine in Schistosoma mansoni.* Scientific Reports, 2020. **10**(1): p. e21565.
2. Anastasiadi, D., A. Esteve-Codina, and F. Piferrer, *Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species.* Epigenetics & Chromatin, 2018. **11**: p. e37.
3. Brenet, F., et al., *DNA methylation of the first exon is tightly linked to transcriptional silencing.* PLoS One, 2011. **6**(1): p. e14524.
4. Liu, L., et al., *Epigenetic control of adaptive or homeostatic splicing during interval-training activities.* Nucleic Acids Res, 2024. **52**(12): p. 7211-7224.
5. Anders, S., A. Reyes, and W. Huber, *Detecting differential usage of exons from RNA-seq data.* Genome Res, 2012. **22**(10): p. 2008-17.
6. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.* Brief Bioinform, 2013. **14**(2): p. 178-92.
7. Grishin, N.V., *The R3H motif: A domain that binds single-stranded nucleic acids.* Trends in Biochemical Sciences, 1998. **23**(9): p. 329-330.
8. Katzenellenbogen, R.A., et al., *NFX1-123 increases hTERT expression and telomerase activity posttranscriptionally in human papillomavirus type 16 E6 keratinocytes.* J Virol, 2009. **83**(13): p. 6446-56.
9. Leung, J.W.C., et al., *ZMYM3 regulates BRCA1 localization at damaged chromatin to promote DNA repair.* Genes & Development, 2017. **31**(3): p. 260-274.
10. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W202-8.
11. Gregersen, L.H., et al., *SCAF4 and SCAF8, mRNA Anti-Terminator Proteins.* Cell, 2019. **177**(7): p. 1797-1813 e18.
12. Elkon, R., et al., *E2F mediates enhanced alternative polyadenylation in proliferation.* Genome Biol, 2012. **13**(7): p. R59.
13. Kunisky, A.K., et al., *Shift in MSL1 alternative polyadenylation in response to DNA damage protects cancer cells from chemotherapeutic agent-induced apoptosis.* Cell Rep, 2021. **37**(2): p. 109815.
14. Keti Zeka, et al., *MAT2A inhibition in AML unveils therapeutic potential of combining DNA demethylating agents with UPR targeting.* bioRxiv 2023: p. doi: https://doi.org/10.1101/2023.06.05.543499.
15. Tamura, R.E., et al., *GADD45 proteins: central players in tumorigenesis.* Curr Mol Med, 2012. **12**(5): p. 634-51.
16. Yoo, J., et al., *Transforming growth factor-beta-induced apoptosis is mediated by Smad-dependent expression of GADD45b through p38 activation.* J Biol Chem, 2003. **278**(44): p. 43001-7.
17. Monaghan, C.E., et al., *The Clp1 R140H mutation alters tRNA metabolism and mRNA 3' processing in mouse models of pontocerebellar hypoplasia.* Proc Natl Acad Sci U S A, 2021. **118**(39): p. e2110730118.
18. Newnham, C.M., et al., *Alternative polyadenylation of MeCP2: Influence of cis-acting elements and trans-acting factors.* RNA Biol, 2010. **7**(3): p. 361-72.
19. Masamha, C.P., *The emerging roles of CFIm25 (NUDT21/CPSF5) in human biology and disease.* Wiley Interdiscip Rev RNA, 2022: p. e1757.
20. Mohanan, N.K., et al., *Alternative polyadenylation: An enigma of transcript length variation in health and disease.* Wiley Interdiscip Rev RNA, 2022. **13**(1): p. e1692.
21. Ye, C., et al., *Role of alternative polyadenylation dynamics in acute myeloid leukaemia at single-cell resolution.* RNA Biol, 2019. **16**(6): p. 785-797.

22. Liu, L., et al., *Transcriptome-Wide Detection of Intron/Exon Definition in the Endogenous Pre-mRNA Transcripts of Mammalian Cells and Its Regulation by Depolarization.* Int J Mol Sci, 2022. **23**(17): p. e10157.
23. Huang, D.W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nature Protocols, 2009. **4**(1): p. 44-57.
24. Blum, M., et al., *The InterPro protein families and domains database: 20 years on.* Nucleic Acids Research, 2021. **49**(D1): p. D344-D354.
25. Lei, L., et al., *Multilevel Differential Control of Hormone Gene Expression Programs by hnRNP L and LL in Pituitary Cells.* Molecular and Cellular Biology, 2018. **38**(12): p. e00651.
26. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.

Figure Legend

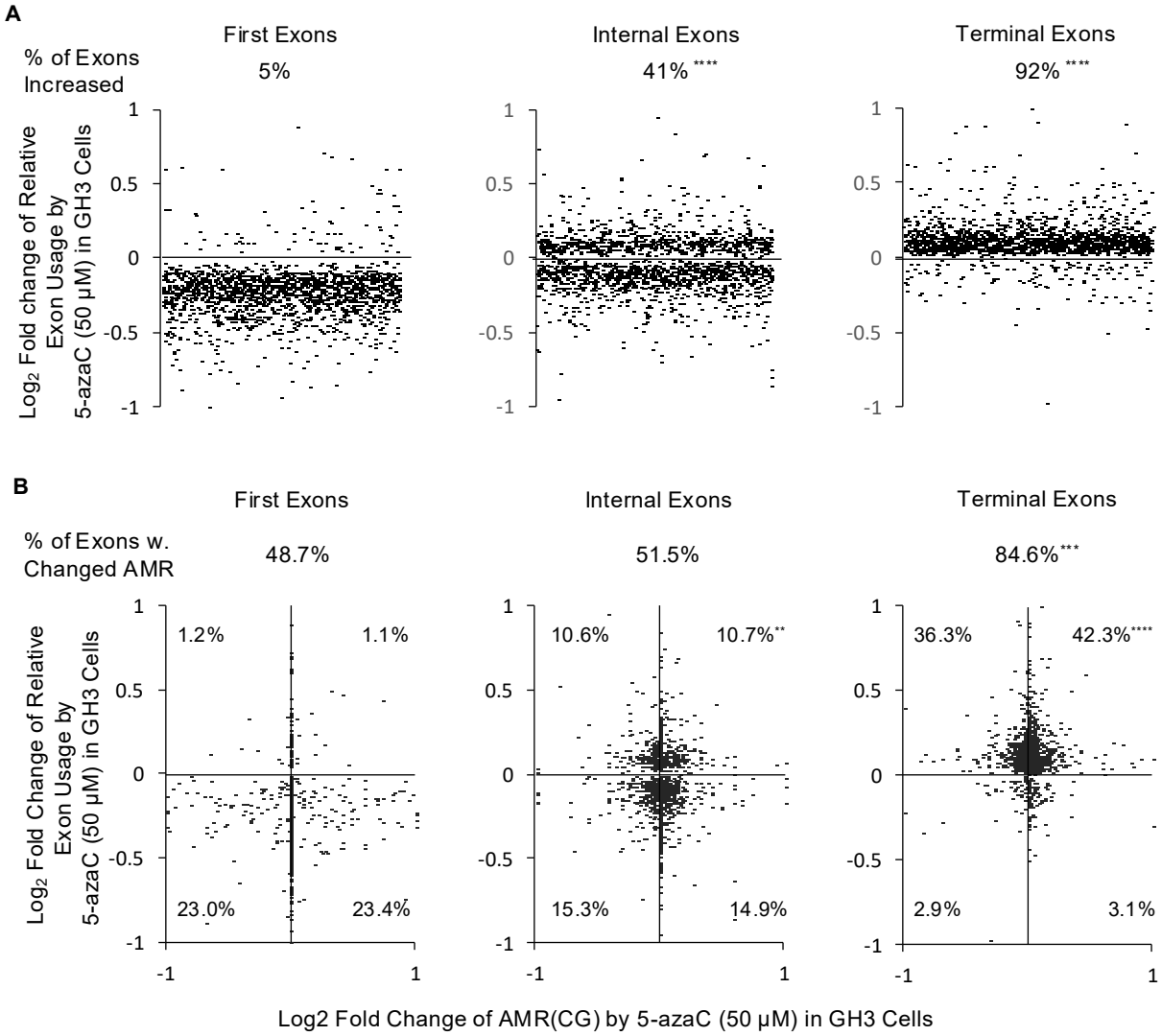
Fig. 1. Exon usage changes of first, internal, and terminal exons following 5-azaC (50 μM) treatment in GH_3 cells. A. Scatter plots showing $\log_2$ fold changes in exon usage. B. Relationship between $\log_2$ fold changes in exon usage and average methylation ratio (AMR). Percentages indicate exons exhibiting AMR changes upon 5-azaC treatment in each quadrant. Enrichment was assessed by hypergeometric density tests using first exons as the reference group ($^{**}p < 1.1E-37$; $^{***}p < 2.8E-120$, $^{****}p < 9.6E-162$). Sample sizes were n = 1,709 first exons, 1,790 internal exons, and 1,874 terminal exons.

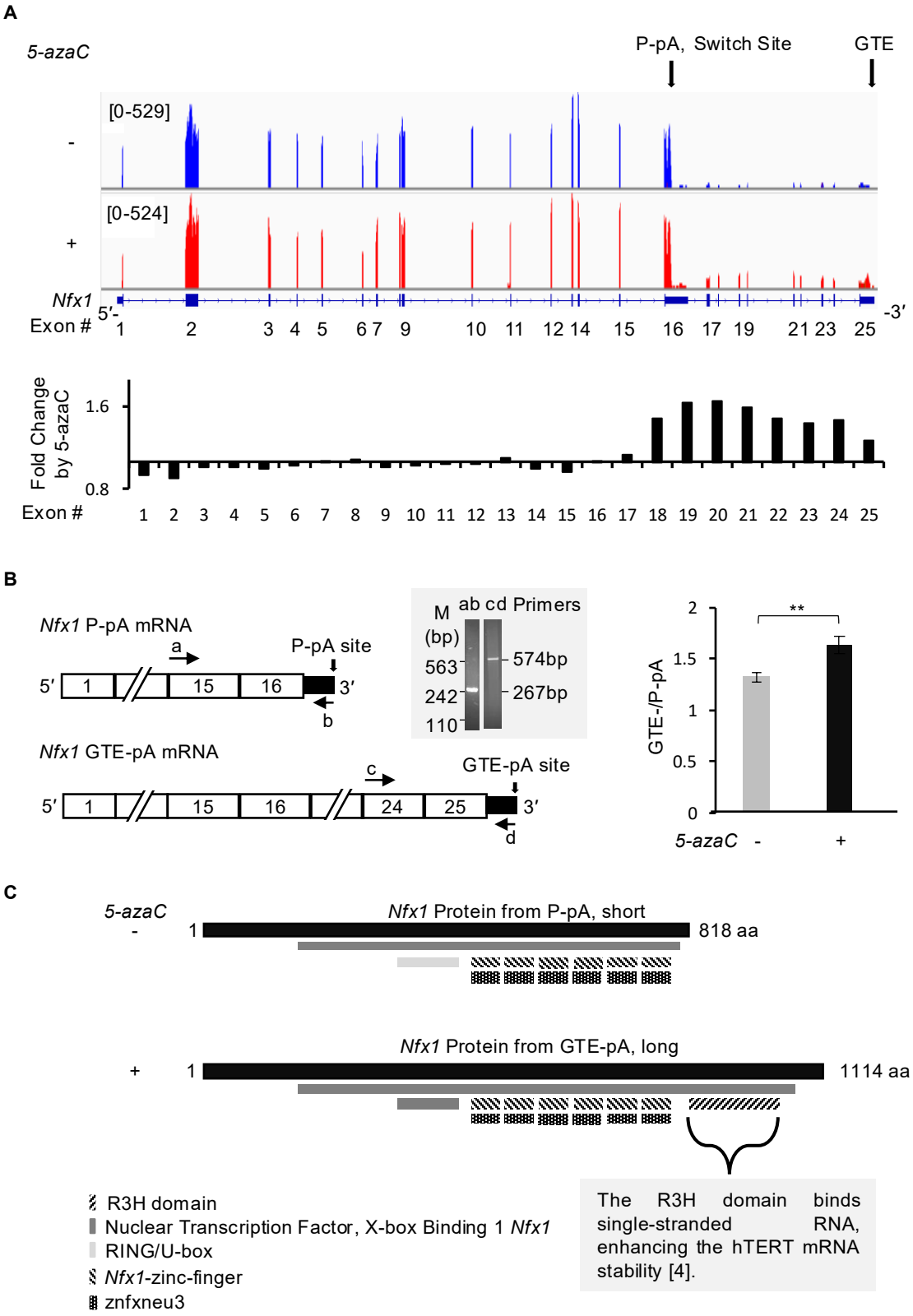
Fig. 2. Exon rank-dependent effects of 5-azaC (50 μM) on exon usage in GH_3 cells A. IGV view and corresponding bar plot showing the effect of 5-azaC on exon usage across the *Nfx1* gene. Bar plots of fold changes were derived from DEXSeq analysis. B. Quantitative real-time PCR (qPCR) validation of reverse-transcribed cDNA demonstrating reduced relative usage of the proximal poly(A) site and increased usage of genomic terminal exon (GTE) poly(A) sites in *Nfx1* following 5-azaC treatment (n = 3). PCR primer positions are indicated by arrowheads, and single specific products from the primers were confirmed by agarose gel electrophoresis. C. Protein domain architecture of the resulting short and long NFX1 protein isoforms, translated from transcripts ENSRNOT00000105959.2 (short) and ENSRNOT00000060692.6 (long), respectively, as annotated in the InterPro protein family database.

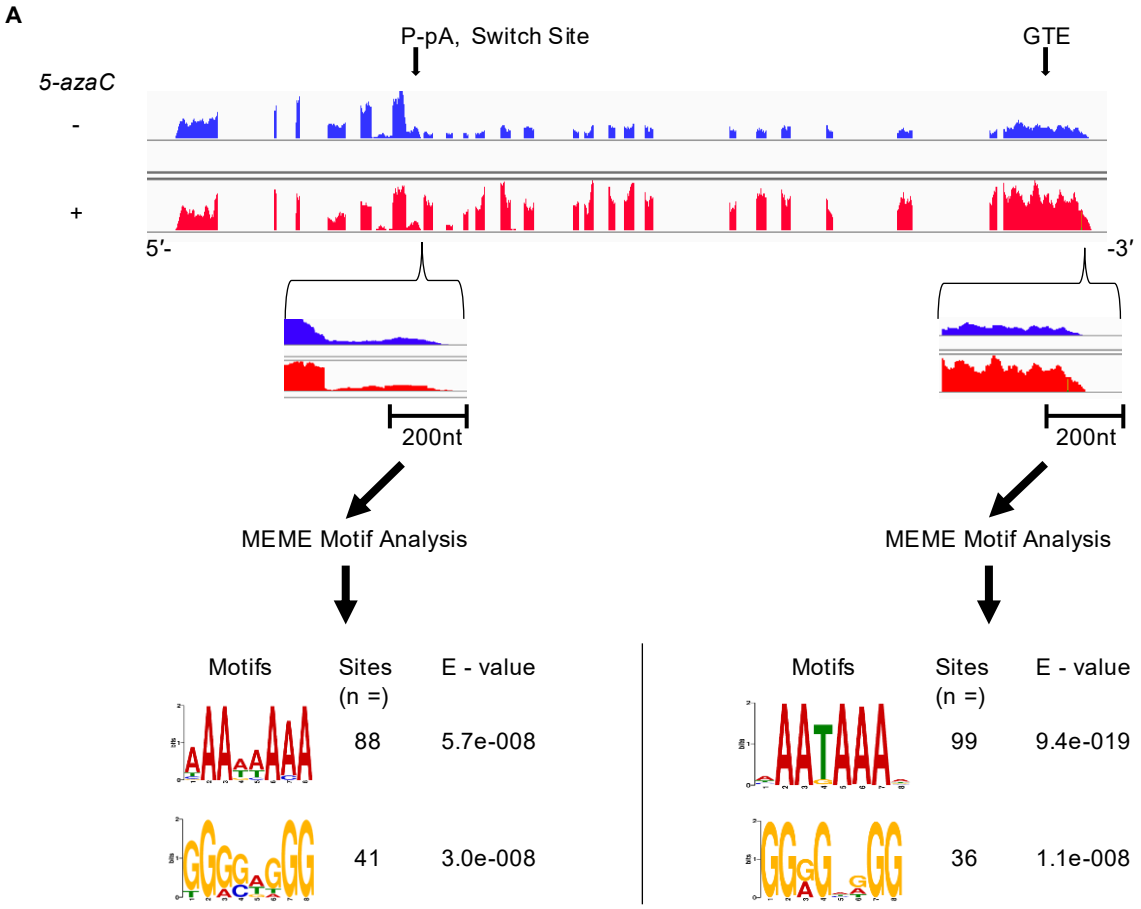
Fig. 3. Consensus motifs of alternative termination sites and mRNA transcripts altered by 5-azaC (50 μM) in GH_3 cells. A. Flow chart illustrating the selection of 200-nt sequences flanking read termination sites from 102 genes exhibiting a single termination-site switch for MEME motif analysis (n denotes the number of sequence sites contributing to motif construction). B. Bar plot showing average normalized counts of transcription regulatory factors (Scaf4, Scaf8, and E2f2) significantly altered by 5-azaC treatment.

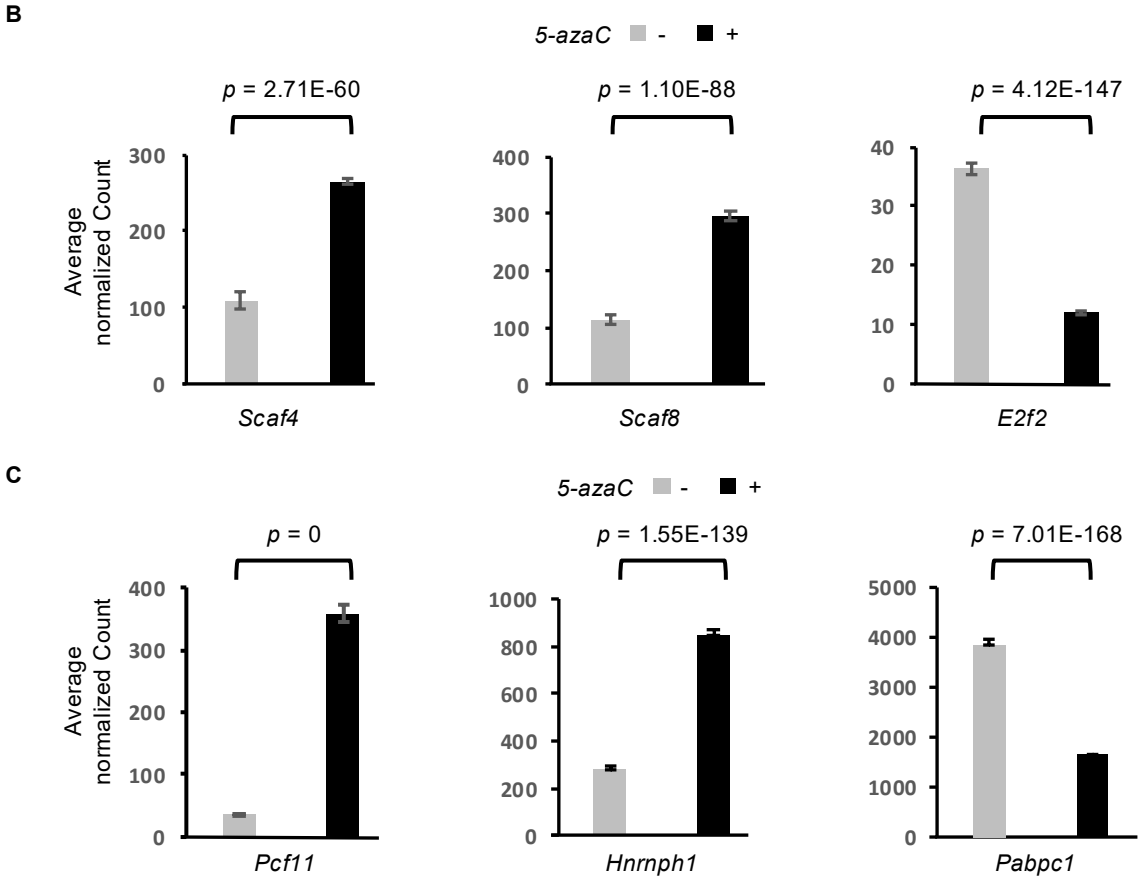
C. Bar plot of average normalized counts of mRNA polyadenylation and splicing factors (Pcf11, Hnrnph, and Pabpc1) significantly regulated by 5-azaC. Differential expression was determined by edgeR analysis of RNA-seq data (FDR < 0.01, 0.5 < fold change < 2, ARC > 200).

Fig. 4. Similar 5-azaC-induced promotion of terminal polyadenylation site usage and full-length transcript production in a leukaemia cell line. A. Scatter plots showing log2 fold changes of exon usage of first, internal, and genomic terminal exons (GTEs), derived from DEXSeq analysis of RNA-Seq data from human MOLM-13 cells treated with 5-azaC (7 µM) (Zeka et al., 2023). Percentages indicate exons with fold changes > 1.0 ($p < 0.05$) in each group. Enrichment was assessed by hypergeometric density testing relative to the first-exon group ($^{***}p$ = 4.4E-27, $^{****}p$ = 6.0E-42). Sample sizes were n = 227 first exons, 704 internal exons, and 478 terminal exons. B. Example of a poly(A) site switch to the GTE of the GADD45B tumour suppressor gene. IGV views show exon-read coverage before and after 5-azaC treatment, with corresponding bar plots of exon-usage fold changes shown below. C. Protein domains gained upon 5-azaC–enhanced terminal poly(A) site usage and full-length GADD45B transcript production. The added p38 MAP kinase-interaction domain is essential for TGF-β-induced apoptosis mediated by GADD45B.


A
First Exons
Internal Exons
Terminal Exons
% of Exons Increased
5%
41% ****
92% ****
Log2 Fold change of Relative Exon Usage by 5-azaC (50 µM) in GH3 Cells
1
0.5
0
-0.5
-1
B
First Exons
Internal Exons
Terminal Exons
% of Exons w. Changed AMR
48.7%
51.5%
84.6%***
1.2%
1.1%
23.0%
23.4%
10.6%
10.7%**
15.3%
14.9%
36.3%
42.3%****
2.9%
3.1%
Log2 Fold Change of Relative Exon Usage by 5-azaC (50 µM) in GH3 Cells
Log2 Fold Change of AMR(CG) by 5-azaC (50 µM) in GH3 Cells


A
5-azaC
P-pA, Switch Site
GTE
[0-529]
-
[0-524]
+
Nfx1
5'-
-3'
Exon #
1 2 3 4 5 6 7 9 10 11 12 14 15 16 17 19 21 23 25
Fold Change by 5-azaC
1.6
0.8
Exon #
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
B
Nfx1 P-pA mRNA
a
P-pA site
5'
1
15
16
3'
b
Nfx1 GTE-pA mRNA
c
GTE-pA site
5'
1
15
16
24
25
3'
d
M
ab cd Primers
(bp)
563
242
110
574bp
267bp
GTE-/P-pA
2
1.5
1
0.5
0
**
5-azaC
-
+
C
5-azaC
-
Nfx1 Protein from P-pA, short
1
818 aa
Nfx1 Protein from GTE-pA, long
+
1
1114 aa
R3H domain
Nuclear Transcription Factor, X-box Binding 1 Nfx1
RING/U-box
Nfx1-zinc-finger
znfxneu3
The R3H domain binds single-stranded RNA, enhancing the hTERT mRNA stability [4].


A
P-pA, Switch Site
GTE
5-azaC
-
+
5′-
-3′
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 - +
p = 2.71E-60
p = 1.10E-88
p = 4.12E-147
Average normalized Count
Scaf4
Scaf8
E2f2
C
5-azaC - +
p = 0
p = 1.55E-139
p = 7.01E-168
Average normalized Count
Pcf11
Hnrnph1
Pabpc1

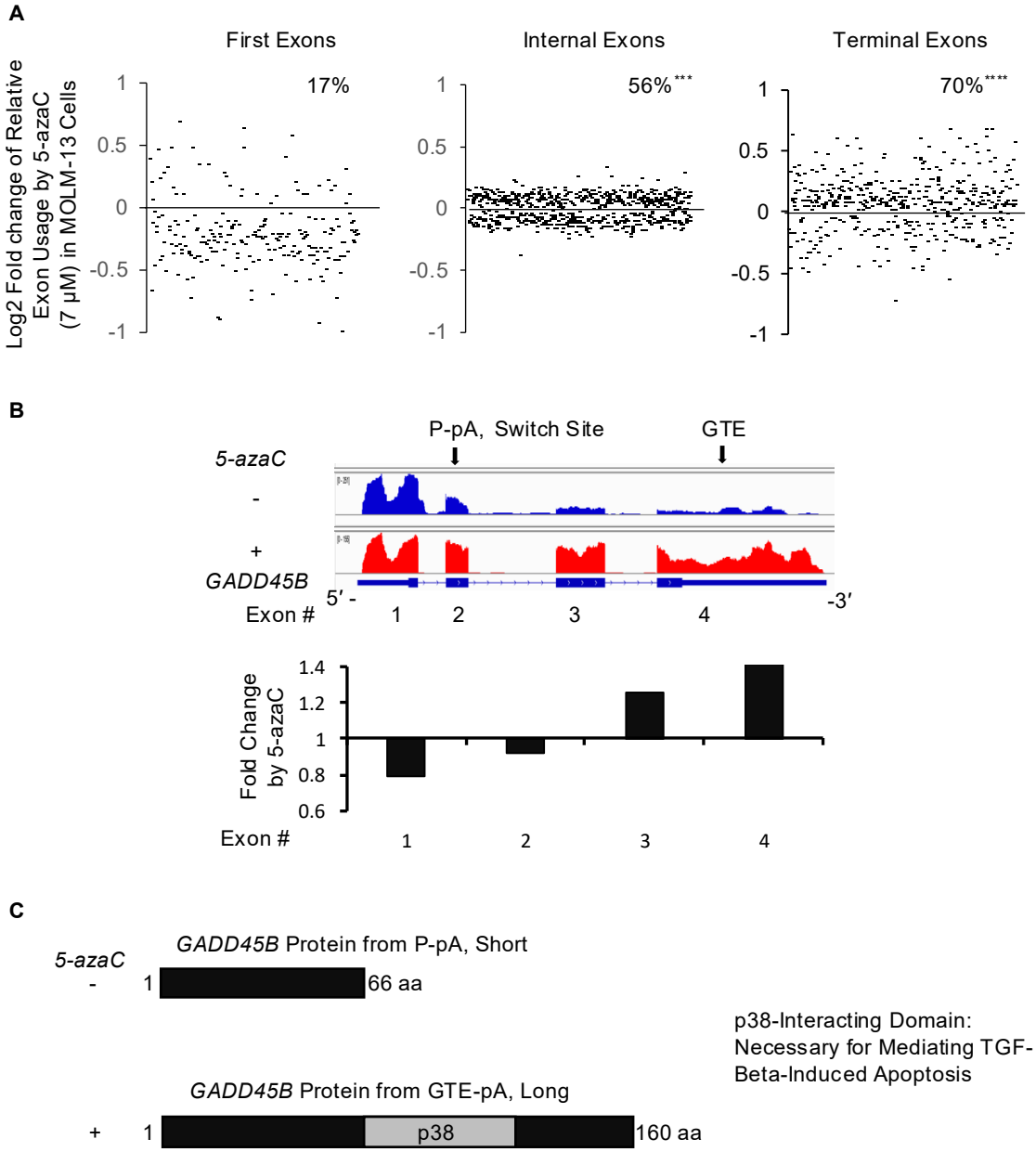

A
First Exons
Internal Exons
Terminal Exons
Log2 Fold change of Relative Exon Usage by 5-azaC (7 μM) in MOLM-13 Cells
17%
56%***
70%****
1
0.5
0
-0.5
-1
B
P-pA, Switch Site
GTE
5-azaC
-
+
GADD45B
5′ -
-3′
Exon #
1
2
3
4
Fold Change by 5-azaC
1.4
1.2
1
0.8
0.6
C
GADD45B Protein from P-pA, Short
66 aa
p38-Interacting Domain: Necessary for Mediating TGF-Beta-Induced Apoptosis
GADD45B Protein from GTE-pA, Long
p38
160 aa

Table 1. Common factors that showed changes by 5-azaC in GH_3 (treated with 50 µM 5-azaC) and/or MOLM-13 cells (treated with 7 µM 5-azaC). Listed are changed 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. Genes with names in shades: RNA-Seq reads increased (pink) or reduced (light blue) in both cell lines.

Transcription Factors				
	GH_3		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				
	GH_3		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				
	GH_3		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