



Neuronal Cells Display Distinct Stability Controls of Alternative Polyadenylation mRNA Isoforms, Long Non-Coding RNAs, and Mitochondrial RNAs

Aysegul Guvenek^{1,2†}, Jihae Shin^{1†}, Lidia De Filippis³, Dinghai Zheng¹, Wei Wang¹, Zhiping P. Pang³ and Bin Tian^{1,4*}

OPEN ACCESS

Edited by:

Athma A. Pai,
University of Massachusetts Medical
School, United States

Reviewed by:

Thomas Gonatopoulos-Pournatzis,
National Cancer Institute at Frederick
(NIH), United States
Zhiping Zhang,
University of Nevada, Reno,
United States

*Correspondence:

Bin Tian
btian@wistar.org

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Genetics

Received: 21 December 2021

Accepted: 28 February 2022

Published: 18 May 2022

Citation:

Guvenek A, Shin J, De Filippis L,
Zheng D, Wang W, Pang ZP and
Tian B (2022) Neuronal Cells Display
Distinct Stability Controls of Alternative
Polyadenylation mRNA Isoforms, Long
Non-Coding RNAs, and
Mitochondrial RNAs.
Front. Genet. 13:840369.
doi: 10.3389/fgene.2022.840369

¹Department of Microbiology, Biochemistry and Molecular Genetics, Rutgers New Jersey Medical School, Newark, NJ, United States, ²Rutgers School of Graduate Studies, Newark, NJ, United States, ³Department of Neuroscience and Cell Biology, Child Health Institute of New Jersey, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, United States, ⁴Program in Gene Expression and Regulation, Center for Systems and Computational Biology, The Wistar Institute, Philadelphia, PA, United States

RNA stability plays an important role in gene expression. Here, using 3' end sequencing of newly made and pre-existing poly(A)+ RNAs, we compare transcript stability in multiple human cell lines, including HEK293T, HepG2, and SH-SY5Y. We show that while mRNA stability is generally conserved across the cell lines, specific transcripts having a high GC content and possibly more stable secondary RNA structures are relatively more stable in SH-SY5Y cells compared to the other 2 cell lines. These features also differentiate stability levels of alternative polyadenylation (APA) 3'UTR isoforms in a cell type-specific manner. Using differentiation of a neural stem cell line as a model, we show that mRNA stability difference could contribute to gene expression changes in neurogenesis and confirm the neuronal identity of SH-SY5Y cells at both gene expression and APA levels. In addition, compared to transcripts using 3'-most exon cleavage/polyadenylation sites (PASSs), those using intronic PASs are generally less stable, especially when the PAS-containing intron is large and has a strong 5' splice site, suggesting that intronic polyadenylation mostly plays a negative role in gene expression. Interestingly, the differential mRNA stability among APA isoforms appears to buffer PAS choice in these cell lines. Moreover, we found that several other poly(A)+ RNA species, including promoter-associated long noncoding RNAs and transcripts encoded by the mitochondrial genome, are more stable in SH-SY5Y cells than the other 2 cell lines, further highlighting distinct RNA metabolism in neuronal cells. Together, our results indicate that distinct RNA stability control in neuronal cells may contribute to the gene expression and APA programs that define their cell identity.

Keywords: alternative polyadenylation, 3' UTR, RNA stability, mitochondrial RNA, long non-coding RNA, GC content, RNA structure

1 INTRODUCTION

Regulation of RNA stability is an important step of expression control of both protein-coding and noncoding genes (Garneau et al., 2007; Schoenberg and Maquat, 2012; Kilchert et al., 2016; Schmid and Jensen, 2019). Distinct decay mechanisms exist in cytoplasm and nucleus, responsible for modulating the life spans of cytoplasm-enriched mRNAs and nucleus-enriched long non-coding RNAs (lncRNAs), respectively (Garneau et al., 2007; Schoenberg and Maquat, 2012; Kilchert et al., 2016; Schmid and Jensen, 2019). RNA decay regulation is also critical for removing aberrant RNA species in the cell (Wolin and Maquat, 2019). Many of the mRNA decay mechanisms are modulated by 3'UTR sequence motifs, such as AU-rich elements (AREs) and GU-rich elements (GREs) (Chen and Shyu, 1995; Barreau et al., 2005; Vlasova et al., 2008; Lee et al., 2010; Vlasova-St Louis et al., 2013), through interactions with a large repertoire of RNA binding proteins (RBPs) (Hentze et al., 2018). In addition, structured RNAs in 3'UTRs could regulate mRNA decay by interacting with RBPs that have direct roles in RNA stability (Park and Maquat, 2013; Fischer et al., 2020) or by influencing decay kinetics (Wu and Bartel, 2017). Moreover, 3'UTR size *per se* has also been reported as a destabilizing feature due to 3'UTR size-dependent interaction with UPF1 (Hogg and Goff, 2010).

Cleavage/polyadenylation of RNA precursors defines the 3' end of almost all mRNAs and lncRNAs. About 70% of mRNA genes in mammals have multiple cleavage/polyadenylation sites (PASs) leading to mRNA isoforms, a phenomenon called alternative polyadenylation (APA) (Tian and Manley, 2017; Gruber and Zavolan, 2019). APA is dynamic in cell growth, differentiation and development (Sandberg et al., 2008; Ji et al., 2009; Shepard et al., 2011). Different tissue or cell types display distinct global APA profiles (Zhang et al., 2005). For example, neuronal cells prefer the usage of APA sites that are distal to gene promoter compared to other cell types (Zhang et al., 2005; Miura et al., 2013; Guvenek and Tian, 2018). In contrast, immune cells and professional secretory cells globally prefer proximal APA sites (Singh et al., 2018; Cheng et al., 2020). Genes that display APA tend to be ubiquitously expressed (Lianoglou et al., 2013) and have long evolutionary history (Lee et al., 2008; Wang et al., 2018b). Notably, brain-specific genes were found to have a higher APA site conservation rate than other genes (Wang et al., 2018b), suggesting that APA isoform expression may be particularly important in neuronal cells.

APA sites in the 3'-most exon lead mostly to isoforms with different 3'UTR sizes, whereas APA sites in introns change both the coding region and 3'UTR. Because of the central role of 3'UTR in mRNA metabolism (Mayr, 2018), APA isoforms with different 3'UTR sizes are expected to have distinct fates in the cell. Indeed, a growing number of studies have shown that 3'UTR isoforms can differ substantially in stability, translational efficiency, or localization (Spies et al., 2013; Tushev et al., 2018). For example, long 3'UTR isoforms in general are less stable than short 3'UTR isoforms in mouse NIH3T3 cells (Spies et al., 2013; Zheng et al., 2018) and human HEK293T cells (Shin et al., 2021). However, to what extent

stability difference between isoforms varies across cell types remains an open question.

In this study, by comparing poly(A)+ transcript abundance in newly made and pre-existing pools, we examine how APA isoforms differ in stability in multiple human cell lines. We examine how stability difference between isoforms varies across cell types and analyze the interplay between stability control and APA site choice. We further examine stability controls of lncRNAs and poly(A)+ RNAs encoded by the mitochondrial genome. Our results reveal 3'UTR features correlated with transcript stability differences between cell types and indicate potential contributions of RNA stability to neuronal cell identity.

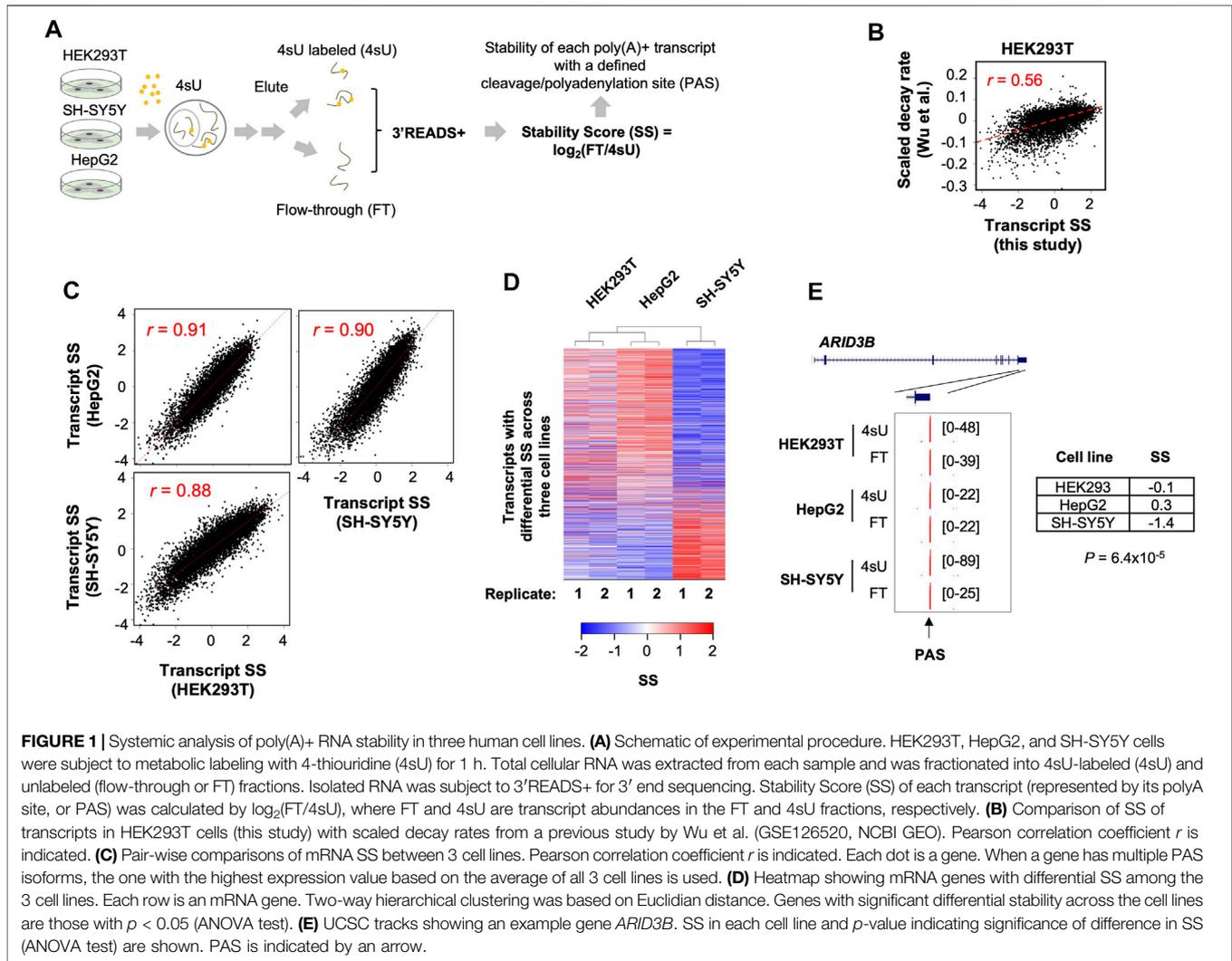
2 RESULTS

2.1 Global Stability Analysis of poly(A)+ RNAs in Three Human Cell Lines

We hypothesized that transcript stability might be distinct in different cell types. To test this, we carried out metabolic labeling of cellular RNA with 4-thiouridine (4sU) in human HEK293T, HepG2, and SH-SY5Y cells (**Figure 1A**). These cell lines, all widely used in biomedical research, have distinct characteristics: HEK293T was derived from human embryonic kidney cells (DuBridge et al., 1987); HepG2 was from a liver hepatocellular carcinoma (Knowles et al., 1980) and has been a cell model for metabolism studies; SH-SY5Y was from a neuroblastoma (Biedler et al., 1978) and has been a model to study neuronal cell functions (Kovalevich and Langford, 2013).

After 4sU labeling for 1 h, total cellular RNA was extracted from each cell line (two biological replicates each), followed by separation of 4sU-labeled and non-labeled RNAs (see Methods for details). These two RNA pools, named 4sU labeled (4sU) and flow-through (FT), respectively, represent newly made and pre-existing RNAs, respectively. We subjected 4sU and FT RNA samples, as well as the total RNA, to the 3'READS+ method for 3' end sequencing (Jin et al., 2015; Zheng et al., 2016). The abundance of each transcript is represented by reads mapped to its PAS (Zheng et al., 2016). We calculated the ratio of transcript abundance in the FT pool to that in the 4sU pool, or $\log_2(\text{FT}/4\text{sU})$, named Stability Score (SS) for simplicity, to represent RNA stability (Zheng et al., 2018). We found that SS values of transcripts in HEK293T cells were well correlated ($r = 0.56$, Pearson Correlation, **Figure 1B**) with scaled decay rates measured by using transcriptional shutdown with Actinomycin D (Wu et al., 2019), supporting the validity of using SS to reflect transcript stability.

Requiring at least five reads in at least one sample as evidence of expression (**Supplementary Figure S1A**), we identified over 100,000 PASs in the 3 cell lines (**Supplementary Figure S1A**). We then used RefSeq and GENCODE databases to assign identified PASs to genes (Frankish et al., 2019). Unassigned reads were additionally annotated with the Fantom database, which has a good coverage of lncRNAs (Lizio et al., 2015; Hon et al., 2017), including promoter-associated lncRNA (p-lncRNA) and enhancer-associated RNAs (eRNAs). Overall, we identified



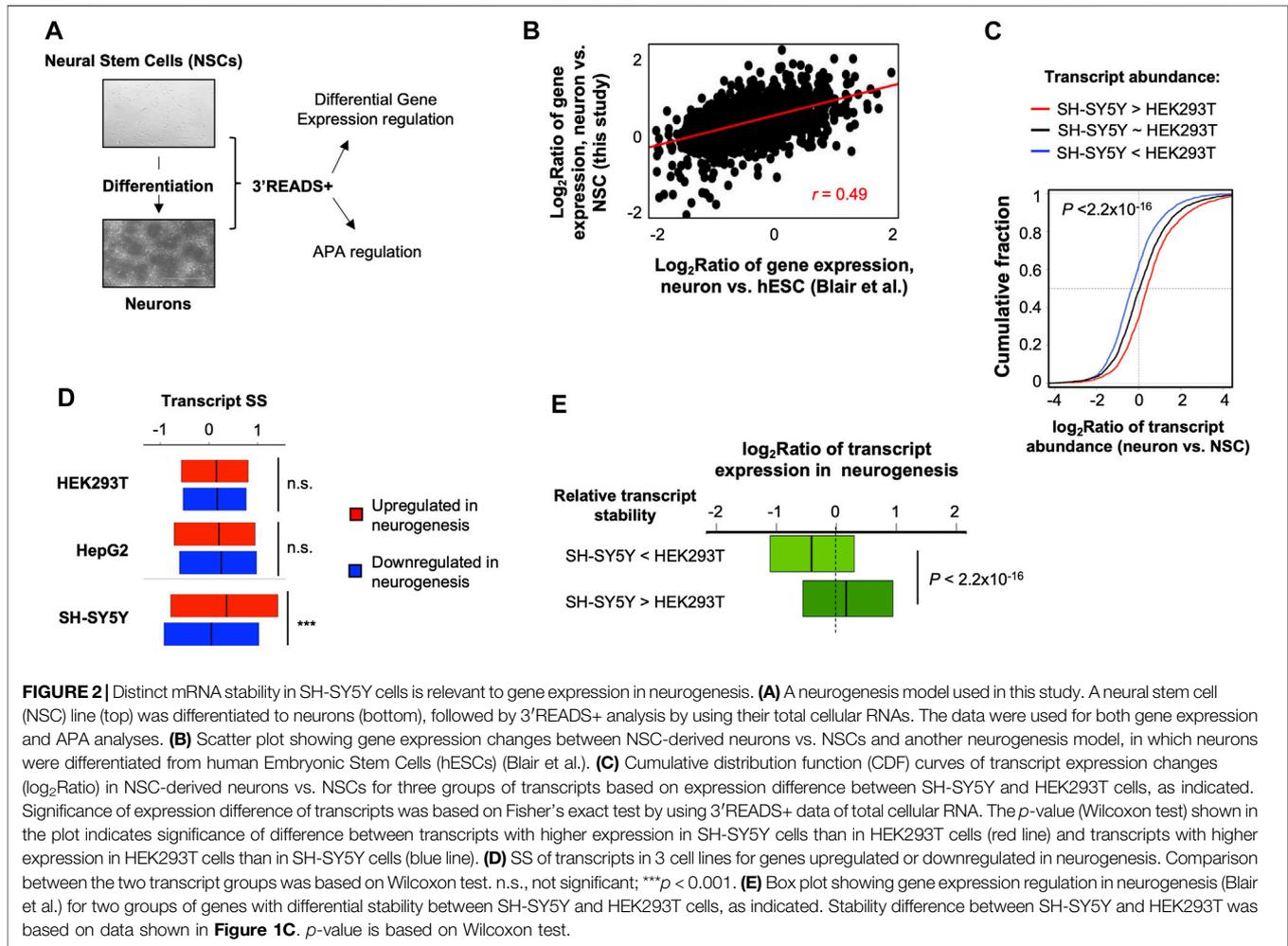
PASs to over 15,000 mRNA and lncRNA genes (Supplementary Tables S1, S2), of which 11,555 were detected in all 3 cell types. Our gene types included mRNA, lncRNA, eRNA, p-lncRNA, pseudogene, mitochondrial genome gene, and intergenic transcripts (Supplementary Figure S1B). Interestingly, 4sU and FT samples showed distinct distributions for different gene types (Supplementary Figure S1B), suggesting different RNA species are grossly different in stability.

2.2 Consistent and Distinct mRNA Stability Across Cell Lines

We first focused on mRNA genes, which accounted for the majority of poly(A)+ RNAs (Supplementary Figure S1B). Overall, SS of mRNA transcripts were well correlated among the 3 cell lines ($r = 0.88$ – 0.91 , Pearson Correlation, Figure 1C), indicating that mRNA stability is largely conserved across cell types. Gene Ontology (GO) analysis of the transcripts that were among the most stable (top 20% in SS) or the least stable (bottom 20% in SS) in all 3 cell lines (Supplementary Table S3) indicated

that stable mRNAs were related to aspects of cell metabolism, such as “small molecule metabolic process”, “ion transport”, “oxidation-reduction”, and “NADH metabolic process”, whereas unstable mRNAs were related to gene expression regulation, such as “transcription from RNA polymerase II promoter”, “negative regulation of macromolecule biosynthetic process”, and “cell differentiation”. In addition, several well-known mRNA stability-related gene features were well correlated with SS (Supplementary Figure S2), such as GC content (both gene and last exon) and intron/exon junction density as positive features for mRNA stability and U content (both gene and 3'UTR) and 3'UTR size as negative features for mRNA stability. Both GO and gene feature analysis results are in line with previous studies of mRNA stability (Lee et al., 2010; Spies et al., 2013), further supporting the suitability of using SS to examine RNA stability.

While transcripts were generally correlated in SS between the 3 cell lines, some differences were discernable (Figure 1C). We thus employed an ANOVA analysis to specifically identify transcripts that displayed significant stability difference among



the cell lines, yielding 1,607 transcripts with $p < 0.05$ (ANOVA test, **Figure 1D**). Two-way hierarchical clustering of transcript SS revealed that SH-SY5Y was distinct from HEK293T and HepG2 (**Figure 1D**). An example gene *ARID3B* is shown in **Figure 1E**, whose SS in SH-SY5Y cells was much lower than those in HEK293T and HepG2 (-1.4 vs. -0.1 and 0.3 , $p = 6.4 \times 10^{-5}$, ANOVA test, **Figure 1E**). Together, these results indicate that while transcript stability is largely similar among cell types, some transcripts display cell type-specific stability.

2.3 Distinct Stability Control in SH-SY5Y Cells is Relevant to Neuronal Cell Identity

Given the distinct mRNA stability profile in SH-SY5Y cells compared to HEK293 and HepG2 cells, we wondered whether mRNA stability might play a role in their neuronal cell identity. To this end, we carried out 3'READS+ analysis of a neural stem cell (NSC) line (Vescovi et al., 1999) and its differentiated neurons (**Figure 2A**, see Materials and Methods for detail). Consistent with the notion that differentiation of this NSC line provides a simple and meaningful analysis of neurogenesis (Vescovi et al., 1999), we found that gene expression changes in NSC versus derived neurons were

well correlated with those in differentiation of human embryonic stem cells (hESCs) to mature neurons (Blair et al., 2017) ($r = 0.49$, Pearson Correlation, **Figure 2B**).

Based on total RNA samples, we found that the transcripts with a higher abundance in SH-SY5Y cells compared to HEK293T cells tended to be upregulated in neurogenesis, whereas those with a lower abundance in SH-SY5Y cells compared to HEK293T cells tended to be downregulated in neurogenesis (**Figure 2C**). This result is in line with the neuronal identity of SH-SY5Y cells and suggests that SH-SY5Y versus HEK293T comparison could provide insights into differences between neuronal and nonneuronal cells.

Interestingly, we found that genes upregulated in neurogenesis had higher SS than those downregulated in neurogenesis in SH-SY5Y cells ($p < 0.001$, Wilcoxon test, **Figure 2D**). By contrast, the SS difference between these two gene sets was not statistically significant in either HEK293T or HepG2 cells (**Figure 2D**). This result indicates that neuron-specific genes are more stable in SH-SY5Y cells than in nonneuronal cells. In line with this result, we found that transcripts tended to be upregulated in neurogenesis if they were more stable in SH-SY5Y cells than in HEK293T cells, and tended to be downregulated in neurogenesis if they were less

TABLE 1 | Top biological processes enriched for transcripts with differential stability in SH-SY5Y cells versus HEK293T cells.

Biological process	p-value
Transcripts more stable in SH-SY5Y	
cotranslational protein targeting to membrane	1.8E-08
protein targeting to membrane	1.8E-08
translational initiation	2.5E-08
SRP-dependent cotranslational protein targeting to membrane	4.9E-08
protein localization to endoplasmic reticulum	1.6E-07
NADH regeneration	7.2E-04
canonical glycolysis	7.2E-04
monosaccharide catabolic process	1.9E-03
cofactor metabolic process	2.9E-03
hexose catabolic process	3.8E-03
Transcripts more stable in HEK293T	
RNA processing	7.2E-08
ncRNA metabolic process	5.5E-06
regulation of tolerance induction	6.7E-05
N-terminal peptidyl-methionine acetylation	6.7E-05
ncRNA processing	1.2E-04
S-adenosylmethionine metabolic process	2.6E-02
RNA polyadenylation	2.1E-02
cellular macromolecule catabolic process	2.1E-02
RNA 3'-end processing	4.2E-02
regulation of mRNA metabolic process	4.8E-02

stable in SH-SY5Y cells than in HEK293T cells ($p < 2.2 \times 10^{-16}$, K-S test, **Figure 2E**).

Gene Ontology analysis indicated that genes whose transcripts were more stable in SH-SY5Y cells than HEK293T cells tended to be associated with protein targeting to membrane and translation (**Table 1**), including “co-translational protein targeting to membrane”, “protein targeting to membrane”, “translational initiation”, etc. By contrast, genes whose transcripts were less stable in SH-SY5Y cells than in HEK293T cells tended to be associated with RNA processing and metabolism functions (**Table 1**), such as “RNA processing”, “ncRNA metabolic process”, “regulation of tolerance induction”, etc. Together, these results indicate that SH-SY5Y cells have distinct mRNA stability controls that impact specific functional gene groups and may play a role in establishing neuronal identity.

2.4 Long 3'UTR Isoforms are Generally Less Stable than Short 3'UTR Isoforms

APA in 3'UTR leads to isoforms with different 3'UTR sizes (**Figure 3A**). Previous studies have shown that 3'UTR isoforms could have different stability levels (Spies et al., 2013; Tushev et al., 2018; Zheng et al., 2018). We identified 3'UTR APA isoforms in 57% of all genes (6,568 out of 11,555) in the 3 cell lines. For simplicity, we focused on the top two 3'UTR APA isoforms based on the combined expression level of each isoform across the 3 cell lines. Based on the relative positions of their PASs to the 5' end of gene, the two isoforms were named proximal PAS (pPAS) isoform and distal PAS (dPAS) isoform, respectively (**Figure 3A**); the

3'UTR portion between the two PASs was named alternative 3'UTR or aUTR (**Figure 3A**).

We calculated the relative expression (RE, \log_2 Ratio of transcript abundance, illustrated in **Figure 3A**) of dPAS isoform versus pPAS isoform for each gene in each cell line (**Figure 3B**). Based on the median RE value of all genes, we found that, in both 4sU and FT samples, SH-SY5Y cells had much higher RE values than HepG2 and HEK293T cells (**Figure 3B**). This result is in good agreement with the notion that neuronal cells preferentially express long 3'UTRs compared to other cell types (Zhang et al., 2005; Miura et al., 2013; Guvenc and Tian, 2018; Ha et al., 2018). HepG2 cells had slightly lower RE values than HEK293T cells (**Figure 3B**), indicating that the former had the shortest 3'UTRs overall among the 3 cell lines.

We next directly compared SH-SY5Y cells with HEK293T cells for 3'UTR isoform expression levels (**Figure 3C**). We found that genes showing longer 3'UTRs in SH-SY5Y cells than HEK293T cells outnumbered those showing the opposite trend by 4.4-fold (red dots vs. blue dots, **Figure 3C**). Importantly, genes showing higher RE values (red dots in **Figure 3C**) in SH-SY5Y cells also tended to have higher RE values in neurogenesis (neurons versus NSCs, **Figure 3D** and **Supplementary Figure S3A**). Notably, both SH-SY5Y versus HEK293T and neurons versus NSCs showed aUTR size-dependent RE Difference (RED) increase (**Supplementary Figure S3B**), indicating that the larger the aUTR the more likely there is a switch from pPAS usage to dPAS usage, or 3'UTR lengthening. An example gene *ERCC1* is shown in **Supplementary Figure S3C**. These results indicate that comparison of SH-SY5Y with HEK293T recapitulates 3'UTR APA isoform changes in neurogenesis, both in terms of direction and degree.

Interestingly, we found that, in all cell lines, RE values were higher in 4sU samples than in FT samples (**Figure 3B**), indicating that 3'UTRs are longer in newly made RNAs than in pre-existing RNAs. As expected, RE values were similar between total and FT samples, indicating that pre-existing RNA (FT) is similar to steady state RNA (total). This result suggests that 3'UTR isoforms have distinct stability levels, leading to abundance differences in newly made versus pre-existing or steady state pools.

To examine 3'UTR isoform stability difference directly, we compared pPAS isoforms with dPAS isoforms for their SS in the 3 cell lines. Consistent with our previous data with mouse NIH3T3 cells (Zheng et al., 2018), we found that, in all these human cell lines, pPAS isoforms were significantly more stable than dPAS isoforms ($p < 0.05$, DEXSeq, **Figure 3E**). The number of genes whose pPAS isoforms were more stable than dPAS isoforms (blue genes in **Figure 3E**) outnumbered those with the opposite trend (red genes in **Figure 3E**) by 7.4-, 8.0-, and 16.8-fold in HEK293T, HepG2, and SH-SY5Y cells, respectively. The larger fold difference in SH-SY5Y cells than those in other 2 cell lines indicates that the extent to which long 3'UTR isoforms are less stable than short 3'UTR isoforms is the greatest in SH-SY5Y cells. We also found that as aUTR size increased, the difference in abundance between pPAS and dPAS isoforms became larger (**Figure 3F**). This trend appeared more obvious in SH-SY5Y cells than in HEK293T or HepG2 cells, further indicating that

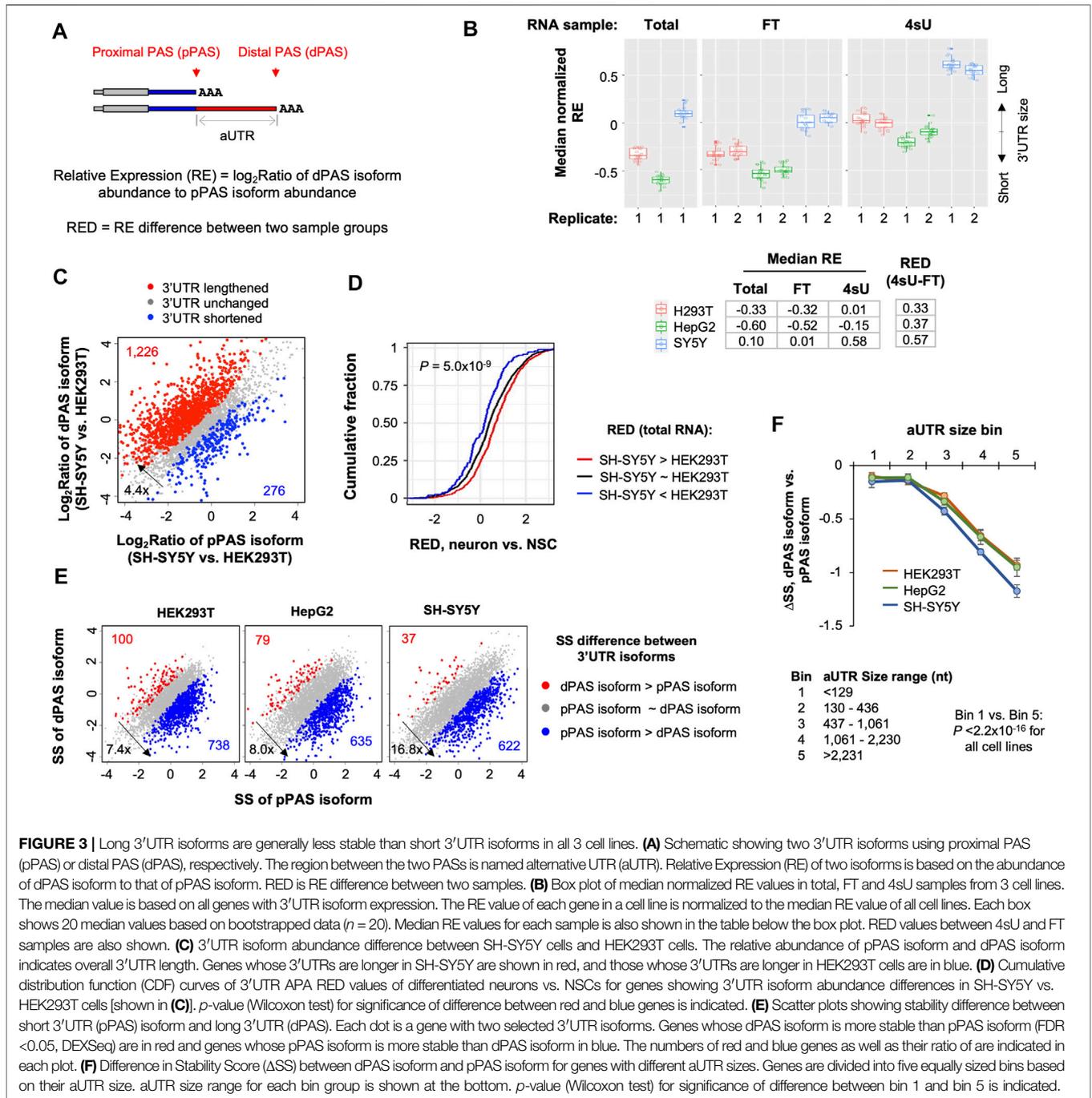


FIGURE 3 | Long 3'UTR isoforms are generally less stable than short 3'UTR isoforms in all 3 cell lines. **(A)** Schematic showing two 3'UTR isoforms using proximal PAS (pPAS) or distal PAS (dPAS), respectively. The region between the two PASs is named alternative UTR (aUTR). Relative Expression (RE) of two isoforms is based on the abundance of dPAS isoform to that of pPAS isoform. RED is RE difference between two samples. **(B)** Box plot of median normalized RE values in total, FT and 4sU samples from 3 cell lines. The median value is based on all genes with 3'UTR isoform expression. The RE value of each gene in a cell line is normalized to the median RE value of all cell lines. Each box shows 20 median values based on bootstrapped data ($n = 20$). Median RE values for each sample is also shown in the table below the box plot. RED values between 4sU and FT samples are also shown. **(C)** 3'UTR isoform abundance difference between SH-SY5Y cells and HEK293T cells. The relative abundance of pPAS isoform and dPAS isoform indicates overall 3'UTR length. Genes whose 3'UTRs are longer in SH-SY5Y are shown in red, and those whose 3'UTRs are longer in HEK293T cells are in blue. **(D)** Cumulative distribution function (CDF) curves of 3'UTR APA RED values of differentiated neurons vs. NSCs for genes showing 3'UTR isoform abundance differences in SH-SY5Y vs. HEK293T cells [shown in **(C)**]. p -value (Wilcoxon test) for significance of difference between red and blue genes is indicated. **(E)** Scatter plots showing stability difference between short 3'UTR (pPAS) isoform and long 3'UTR (dPAS). Each dot is a gene with two selected 3'UTR isoforms. Genes whose dPAS isoform is more stable than pPAS isoform (FDR < 0.05, DEXSeq) are in red and genes whose pPAS isoform is more stable than dPAS isoform are in blue. The numbers of red and blue genes as well as their ratio of are indicated in each plot. **(F)** Difference in Stability Score (Δ SS) between dPAS isoform and pPAS isoform for genes with different aUTR sizes. Genes are divided into five equally sized bins based on their aUTR size. aUTR size range for each bin group is shown at the bottom. p -value (Wilcoxon test) for significance of difference between bin 1 and bin 5 is indicated.

stability difference between 3'UTR isoforms is greater in SH-SY5Y cells compared to HEK293T and HepG2 cells.

2.5 3'UTR Motifs and Structures Contribute to Distinct mRNA Stability Controls in SH-SY5Y Cells

We next wanted to identify transcript features that made RNA stability control in SH-SY5Y cells distinct as compared to other

cells. Using a regression model, we examined various transcript features that correlated with transcript SS differences (Δ SS) between SH-SY5Y cells and HEK293T cells. The features we used included exonic and intronic sizes, splicing parameters, nucleotide contents of different regions, etc. (see Section 4 for detail). We found that GC and U contents were the top two features correlated to Δ SS (Figure 4A). Interestingly, whereas GC contents of the whole gene and of the last exon positively contributed to Δ SS,

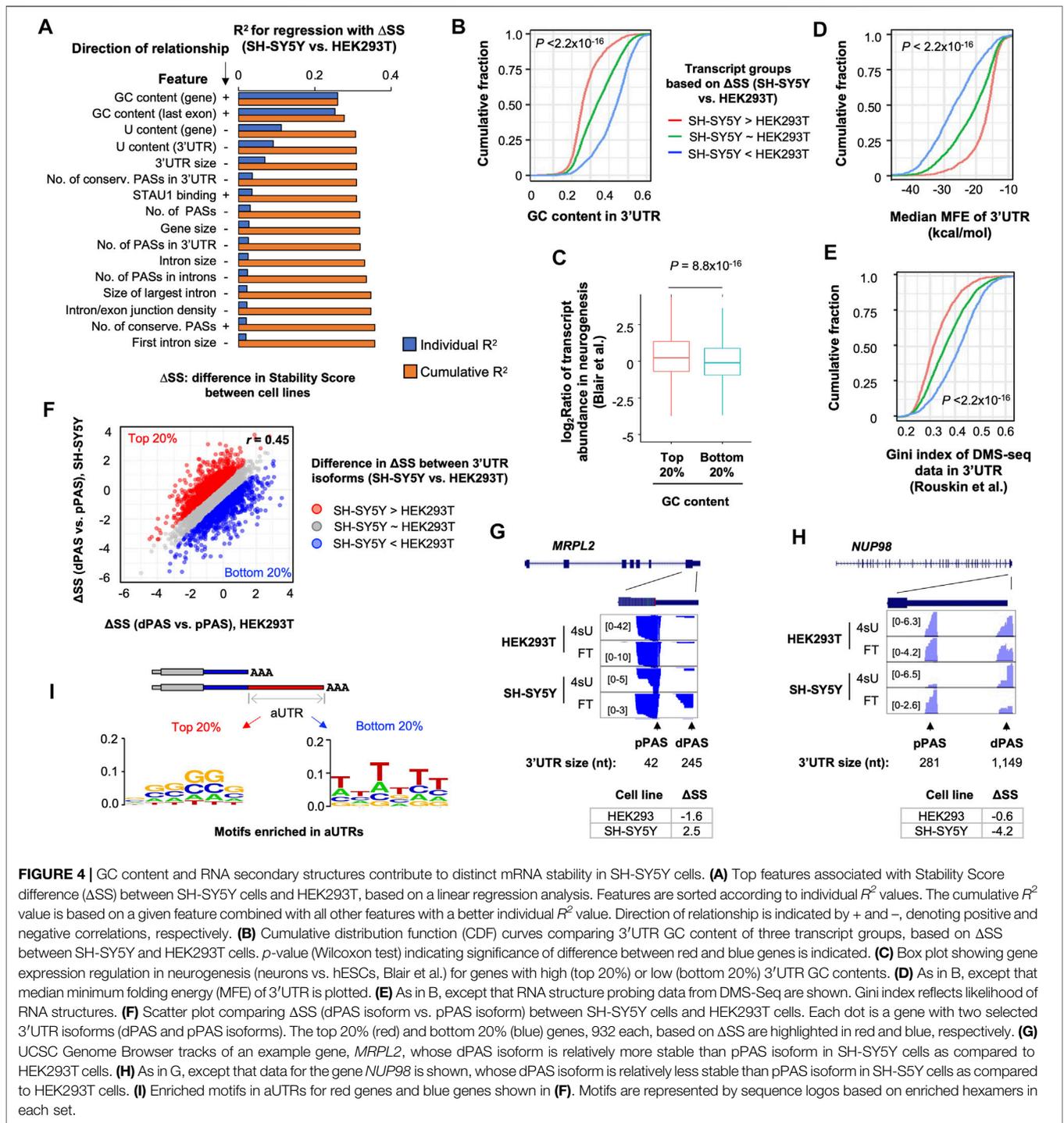


FIGURE 4 | GC content and RNA secondary structures contribute to distinct mRNA stability in SH-SY5Y cells. **(A)** Top features associated with Stability Score difference (ΔSS) between SH-SY5Y cells and HEK293T, based on a linear regression analysis. Features are sorted according to individual R² values. The cumulative R² value is based on a given feature combined with all other features with a better individual R² value. Direction of relationship is indicated by + and -, denoting positive and negative correlations, respectively. **(B)** Cumulative distribution function (CDF) curves comparing 3'UTR GC content of three transcript groups, based on ΔSS between SH-SY5Y and HEK293T cells. *p*-value (Wilcoxon test) indicating significance of difference between red and blue genes is indicated. **(C)** Box plot showing gene expression regulation in neurogenesis (neurons vs. hESCs, Blair et al.) for genes with high (top 20%) or low (bottom 20%) 3'UTR GC contents. **(D)** As in B, except that median minimum folding energy (MFE) of 3'UTR is plotted. **(E)** As in B, except that RNA structure probing data from DMS-Seq are shown. Gini index reflects likelihood of RNA structures. **(F)** Scatter plot comparing ΔSS (dPAS isoform vs. pPAS isoform) between SH-SY5Y cells and HEK293T cells. Each dot is a gene with two selected 3'UTR isoforms (dPAS and pPAS isoforms). The top 20% (red) and bottom 20% (blue) genes, 932 each, based on ΔSS are highlighted in red and blue, respectively. **(G)** UCSC Genome Browser tracks of an example gene, *MRPL2*, whose dPAS isoform is relatively more stable than pPAS isoform in SH-SY5Y cells as compared to HEK293T cells. **(H)** As in G, except that data for the gene *NUP98* is shown, whose dPAS isoform is relatively less stable than pPAS isoform in SH-SY5Y cells as compared to HEK293T cells. **(I)** Enriched motifs in aUTRs for red genes and blue genes shown in **(F)**. Motifs are represented by sequence logos based on enriched hexamers in each set.

U contents of these regions negatively contributed to ΔSS (Figure 4A). In addition, 3'UTR size and number of PASs in 3'UTRs also negatively impacted ΔSS, albeit to a much lesser extent (Figure 4A). The effect of 3'UTR GC content on transcript stability difference between SH-SY5Y and HEK293 cells could also be demonstrated by the significant difference in GC content between the transcripts that were more stable in SH-SY5Y (top 20% of ΔSS) and those that were

more stable in HEK293T (bottom 20% of ΔSS) ($p < 2.2 \times 10^{-16}$, K-S test, Figure 4B). Importantly, transcripts with high GC contents (top 20%) were more likely to be upregulated in neurogenesis than transcripts with low GC contents (bottom 20%, Figure 4C), highlighting the potential functional relevance of GC content for gene expression in neuronal cells.

Because RNAs with higher GC contents could adopt more stable secondary structures than those with lower GC contents

(Zheng et al., 2020), we next set out to examine how RNA structures might be related to the RNA stability difference between SH-SY5Y and HEK293T cells. We first calculated the medium minimum folding energy of all sub-sequences of a given 3'UTR (100 nt each, see Materials and Methods for detail). We found that transcripts that were more stable in SH-SY5Y cells (top 20% Δ SS, SH-SY5Y vs. HEK293T) had significantly lower MFE values than those that were more stable in HEK293T cells (bottom 20% Δ SS, SH-SY5Y vs. HEK293T) ($p < 2.2 \times 10^{-16}$, K-S test, **Figure 4D**), supporting the notion that RNA structures positively contribute to Δ SS (SH-SY5Y vs. HEK293T).

To further corroborate RNA structure prediction results, we analyzed a dataset previously generated by Rouskin et al., which probed RNA structures *in vivo* using dimethyl sulfate treatment followed by sequencing (DMS-seq) (Rouskin et al., 2014). With DMS-Seq data, RNA structures are represented by Gini indices. A high Gini index indicates a high possibility of RNA secondary structures. Consistent with the MFE-based RNA structure prediction result, transcripts that were more stable in SH-SY5Y cells (top 20% Δ SS, SH-SY5Y vs. HEK293T) showed significantly higher DMS-seq Gini indices than those that were more stable in HEK293T cells (bottom 20% Δ SS, SH-SY5Y vs. HEK293T) ($p < 2.2 \times 10^{-16}$, K-S test, **Figure 4E**). Taken together, these results indicate that high GC contents and likely more stable RNA secondary structures may make transcripts more stable in SH-SY5Y cells compared to HEK293T cells.

We next asked whether the GC content-related RNA stability difference between SH-SY5Y and HEK293T cells could impact 3'UTR isoform stability differences in different cells. To this end, we first calculated 3'UTR isoform stability difference, or Δ SS, between dPAS isoform and pPAS isoform per gene in both SH-SY5Y and HEK293T cells (**Figure 4F**). While Δ SS values were generally correlated between the 2 cell types ($r = 0.45$, Pearson Correlation, **Figure 4F**), some differences were discernable. Based on Δ SS differences, we identified genes that had higher Δ SS values in SH-SY5Y cells (top 20%, red genes in **Figure 4F**) and genes that had higher Δ SS values in HEK293T cells (bottom 20%, blue genes in **Figure 4F**). Two example genes *MPPL2* and *NUP98* are shown in **Figures 4G,H**, respectively. Whereas *MPPL2* had a larger Δ SS value in SH-SY5Y cells than HEK293T cells (2.5 vs. -1.6), *NUP98* showed the opposite trend (-0.6 vs. -4.2). Interestingly, we found that GC-rich motifs were highly enriched in aUTRs of red genes (**Figure 4I**, left), whereas U-rich motifs were enriched in aUTRs of blue genes (**Figure 4I**, right, U shown as T). This result indicates that GC and U contents in aUTRs contribute to 3'UTR isoform stability variations between SH-SY5Y and HEK293T cells, corroborating our results based on transcript comparison across genes (**Figures 4A–E**). Interestingly, we also found that red and blue genes were enriched with GO terms related to mitochondrial functions (**Supplementary Figure S4A**) and cytosolic metabolic functions (**Supplementary Figure S4B**), respectively, suggesting that 3'UTR isoform difference in stability may have distinct functional consequences in the 2 cell types.

Taken together, our results indicate that GC-rich motifs and likely RNA structures could make transcripts and isoforms more stable in SH-SY5Y cells than HEK293T cells.

2.6 Intronic Polyadenylation Isoforms are Generally Unstable

A sizable fraction of APA isoforms use PASs in introns (illustrated in **Figure 5A**) (Tian et al., 2007; Hoque et al., 2013; Singh et al., 2018). With our data, we identified 36,368 intronic polyadenylation (IPA) sites in 10,191 genes (**Supplementary Figure S5A**). Similar numbers of genes in the 3 cell lines expressed IPA isoforms (**Supplementary Figure S5**). By comparing transcript abundances of IPA isoforms versus 3'-most exon APA isoforms (called TPA isoforms for simplicity, illustrated in **Figure 5A**), we found that genes showing IPA suppression in SH-SY5Y versus HEK293T (red genes, **Figure 5B**) outnumbered those showing IPA activation in SH-SY5Y versus HEK293T (blue genes, **Figure 5B**) by 4.6-fold. This result indicates global IPA suppression in SH-SY5Y compared to HEK293T. Notably, IPA was also suppressed in NSC differentiation to neurons (**Supplementary Figure S5B**). These results are in good agreement with the notion that neuronal cells prefer to use PASs in the last exon (Zhang et al., 2005; Taliaferro et al., 2016).

We next compared transcript stability between IPA isoforms and TPA isoforms. We found that median Δ SS (IPA isoform vs. TPA isoforms) in all 3 cell lines were negative (**Figure 5C**), indicating that IPA isoforms in general were less stable than TPA isoforms. Interestingly, based on median Δ SS, IPA isoforms were significantly less unstable in SH-SY5Y cells compared to the two other cell lines ($p < 0.05$, *t*-test, **Figure 5C**). Therefore, for both 3'UTR APA isoforms (**Figure 3**) and IPA isoforms, stability difference between isoforms varies across cell lines, and, strikingly, the more stable isoforms in a cell also appeared to be less abundant at the steady state level as compared to other isoforms (see **Section 3**).

We next wanted to examine gene features related to IPA isoform stability as we did with 3'UTR isoforms. However, because our 3'READS+ reads for IPA transcripts were largely located in the middle of an intron, we could not accurately derive sequence features, such as CDS size, 3'UTR size, etc., for most IPA isoforms. We therefore focused on the introns in which IPA sites were identified, including intron size, 5' splice site (5'SS) strength, and 3' splice site (3'SS) strength. Splice site strengths were based on the maximum entropy (MaxEnt) method (Yeo and Burge, 2004). We divided IPA site-containing introns into five equally sized bins based on each of these three features, and examined IPA isoform SSs across the bins. We found that IPA isoform stability decreased as intron size (**Figure 5D**) and 5'SS strength (**Figure 5E**) increased in all 3 cell lines. This trend, however, was not discernible with 3'SS strength (**Figure 5F**). Consistently, transcripts in the top 20% (bin 5) and bottom 20% (bin 1) groups based on intron size and 5'SS strength were significantly different in stability ($p < 0.05$, Wilcoxon test, **Figures 5D,E**) in all 3 cell types. In contrast, the difference was

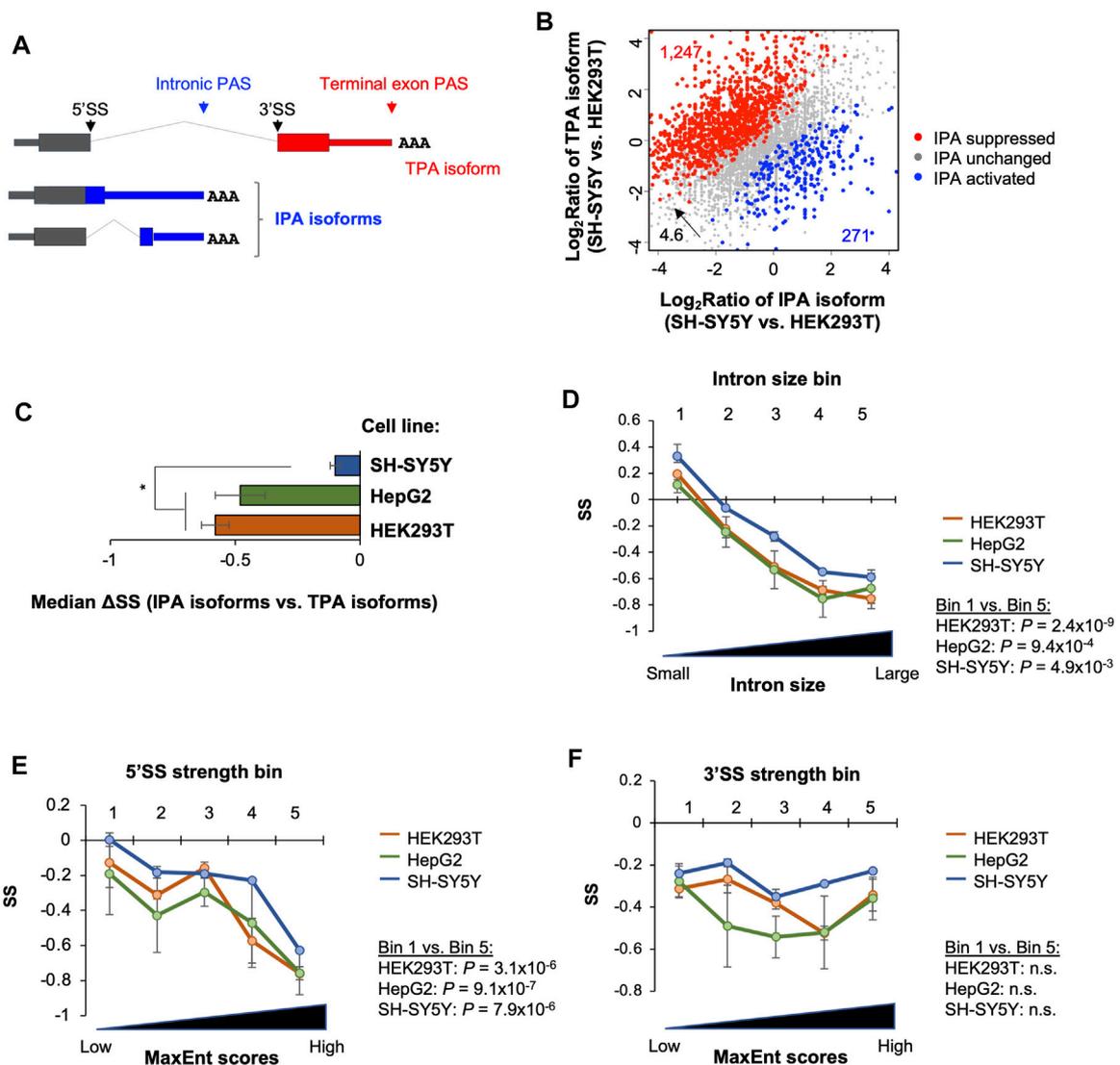


FIGURE 5 | Stability analysis of intronic polyadenylation isoforms. **(A)** Schematic of intronic polyadenylation (IPA) isoforms and 3'-most terminal exon polyadenylation (TPA) isoform. **(B)** Comparison of IPA isoform stability with TPA isoform stability in three cell lines. **(C)** Median Δ SS between IPA and TPA isoforms in three cell lines. Significance is based on the Wilcoxon test comparing the cell lines. **(D)** Impact of intron size on IPA isoform stability. IPA isoforms are divided into five equally sized bins based on the size of intron containing the IPA site. P-value (Wilcoxon test) for significance of difference between bin 1 and bin 5 in each cell line is indicated. **(E)** As in C, except that IPA isoforms are divided into five equally sized bins based on the 5'SS strength (MaxEnt score) of intron containing the IPA site. **(F)** As in C, except that IPA isoforms are divided into five equally sized bins based on the 3'SS strength (MaxEnt score) of intron containing the IPA site. n.s., not significant ($p > 0.05$, Wilcoxon test).

not significant based on 3'SS strength (Figure 5F). Together, these data indicate that IPA isoforms are generally unstable, and intron size and 5'SS strength are related to their stability control.

2.7 Distinct Stability of lncRNAs and Mitochondrial RNAs in SH-SY5Y Cells

In addition to protein-coding transcripts, our 3'READS + data identified many lncRNAs, including canonical lncRNAs, p-lncRNAs and eRNAs, RNAs encoded by the mitochondrial

genome (named mtRNAs), pseudogene transcripts, and unannotated intergenic RNAs (Supplementary Figure S1C). We calculated SS of these RNAs and asked whether their stability controls were different among the three cell types (Figure 6A). Interestingly, p-lncRNAs and mtRNAs both showed discernable stability variations across the cell lines. They both were significantly more stable in SH-SY5Y cells than in HepG2 or HEK293T cells ($p < 0.05$, Wilcoxon test, Figure 6A). In contrast, mRNAs overall did not show significant differences between the cell lines ($p > 0.05$, Wilcoxon test, Figure 6A). In addition, pseudogene and

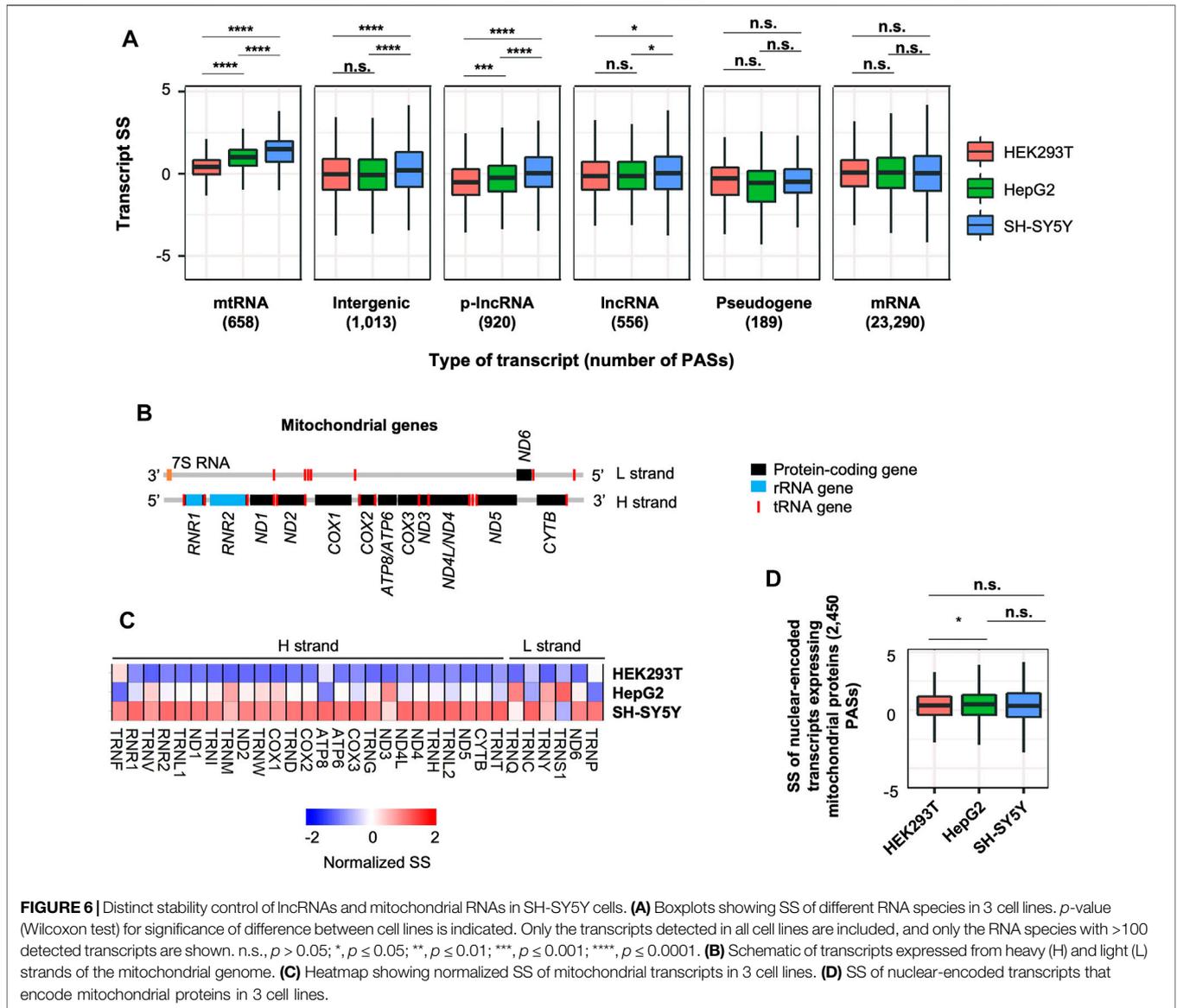


FIGURE 6 | Distinct stability control of lncRNAs and mitochondrial RNAs in SH-SY5Y cells. **(A)** Boxplots showing SS of different RNA species in 3 cell lines. *p*-value (Wilcoxon test) for significance of difference between cell lines is indicated. Only the transcripts detected in all cell lines are included, and only the RNA species with >100 detected transcripts are shown. n.s., *p* > 0.05; *, *p* ≤ 0.05; **, *p* ≤ 0.01; ***, *p* ≤ 0.001; ****, *p* ≤ 0.0001. **(B)** Schematic of transcripts expressed from heavy (H) and light (L) strands of the mitochondrial genome. **(C)** Heatmap showing normalized SS of mitochondrial transcripts in 3 cell lines. **(D)** SS of nuclear-encoded transcripts that encode mitochondrial proteins in 3 cell lines.

intergenic transcripts overall showed comparable stability levels between the cell lines (Figure 6A).

We next further examined mtRNA genes, which included 13 protein-coding genes, 2 rRNA genes, and 22 tRNA genes (Figure 6B and Supplementary Figure S6A). We found that most of them were indeed more stable in SH-SY5Y cells compared to HEK293T and HepG2 cells (Figure 6C). Importantly, this trend was not detected in nuclear-encoded RNAs that encoded for mitochondrial proteins (Figure 6D), indicating that mitochondria in SH-SY5Y cells have a distinct RNA stability control than other 2 cell types. Notably, mtRNAs are highly expressed in brain tissues based on the Genotype-Tissue Expression (GTEx) data (GTEx Consortium et al., 2017) (Supplementary Figure S6B, see Section 4 for detail), suggesting that stability control may play a role in expression regulation of mitochondrial transcripts in neurons.

3 DISCUSSION

In this study, we systematically studied poly(A)+ RNA stability in three human cell lines, namely, HEK293T, HepG2, and SH-SY5Y. We found that while mRNA stability controls are generally similar in these cells, certain transcripts display differential stability levels across the cell lines, especially in SH-SY5Y cells. We show that GC content and RNA secondary structures of transcripts correlate with distinct mRNA stability values in SH-SY5Y cells, which also impact 3'UTR isoform stability differences in these cells. Given the relevance of our SH-SY5Y versus HEK293T comparison to neurogenesis in gene expression and APA regulation, we conclude that modulation of mRNA stability could contribute to establishment of neuronal cell identity.

Our RNA stability analysis is based on comparison of transcript abundances in newly made versus pre-existing RNA pools. We show that the Stability Score (SS) derived from this

comparison are well correlated with RNA decay rates that were measured by using transcriptional shutdown. As such, our method offers a simple and meaningful approach to examine RNA stability. On the other hand, it is worth noting that the accuracy of SS could be compromised if there are alternations of transcription during the course of metabolic labeling. In other words, when newly made and pre-existing RNAs are produced in distinct transcriptional contexts, our method would be affected by large noises. However, we do not expect this to be an issue when cells are under normal growth conditions, as in this study.

We show that GC content and RNA structures could impact stability differences between 3'UTR isoforms, corroborating the gene-to-gene analysis result. It is worth noting that APA isoforms that differ in 3'UTRs offer an accurate and efficient approach to examine the impact of 3'UTR on mRNA metabolism. This is because the 3'UTR isoforms presumably have identical sequences except for their aUTRs; their comparison, therefore, addresses potential, confounding features in CDS and 5'UTR. The underlying mechanism(s) for RNA stability control via GC content and RNA secondary structures, however, is unclear. There are a group of RNA-binding proteins that specifically interact with sequence motifs with high GC contents (Dominguez et al., 2018; Van Nostrand et al., 2020) and/or RNA structures (Tian et al., 2004; Li et al., 2010; Pan et al., 2018). For example, recent studies indicated G3BP1 and UPF1 in binding highly structured 3'UTRs, which leads to structure-mediated RNA decay (Imamachi et al., 2017; Fischer et al., 2020). Intriguingly, we recently found that GC content and RNA structures in 3'UTRs help mRNA associate with the endoplasmic reticulum (ER) (Cheng et al., 2021). How RNA stability interfaces with RBP interactions and ER association needs to be studied in the future.

Our finding that long 3'UTR isoforms are generally less stable than short 3'UTR isoforms is largely in line with our previous studies (Zheng et al., 2018; Cheng et al., 2020; Shin et al., 2021). Notably, an earlier study using transcriptional shutdown by Actinomycin D reported a similar but less prominent trend in NIH3T3 cells (Spies et al., 2013). The discrepancy could stem from differences in sequencing and analysis methods. For example, in the study by Spies et al. mRNA half-life was calculated over a period of time after transcriptional shutdown. In contrast, we generated SS based on newly made and pre-existing RNAs. While SS is sensitive in detecting stability differences, it does not measure half-life *per se*. Therefore, subtle differences in half-life values could be significant in this work but not in the Spies et al. study. Nevertheless, our comparative approach using multiple cell lines enabled us to identify cell type-specific differences in RNA stability, mitigating any intrinsic biases in using SS to measure stability.

We found that IPA isoforms are generally less stable than isoforms using 3'-most exon PASSs. This result suggests that while some IPA isoforms could diversify the protein-coding potential of genes, most IPA events may lead to unstable RNAs that do not have substantial impacts on cell functions. In other words, IPA may be employed mainly to downregulate gene expression. Interestingly, we found that both intron size and 5'SS strength play negative roles in IPA isoform stability. 5'SS strength may be related to nuclear retention of IPA transcripts through U1 snRNP, as shown recently for lncRNAs (Yin et al., 2020). The

negative impact of intron size on IPA isoform stability may be due to the possibility that large introns tend to give rise to IPA isoforms with long 3'UTRs, a negative feature for transcript stability. However, because our short read data do not cover full-length sequence, we could only infer these potential mechanisms at this point. Future work using long read sequencing would provide more definitive conclusions.

Our data indicate that isoform stability variation counters APA site choice, i.e., the isoforms preferentially expressed in 1 cell type are also less stable in this cell type, as compared to other cell types. Interestingly, this trend applies to both 3'UTR APA isoforms and IPA isoforms, indicating that stability regulation of a specific transcript is connected to its 3' end processing activity. Notably, a recent study by the Struhl lab reported a similar compensatory link between transcript decay and its 3' end processing in yeast (Moqtaderi et al., 2022). This phenomenon is reminiscent of mRNA degradation-mediated transcriptional buffering that was previously reported in multiple species (Haimovich et al., 2013; Sun et al., 2013). Whether XRN1, the protein implicated in the decay-transcription buffering (Sun et al., 2013), is also involved in the link between APA isoform decay and APA site choice awaits further experimentation.

Several types of lncRNAs, including regular lncRNAs and p-lncRNAs, were found to be generally more stable in SH-SY5Y cells compared to HEK293T or HepG2 cells. Regular lncRNAs are stand-alone genes, similar to mRNAs. p-lncRNAs are associated with gene promoters, also known as upstream antisense RNAs or PROMPTs (Preker et al., 2008; Hon et al., 2017). Both RNA species are enriched, if not exclusively localized, in nucleus. As such, their stability is under the control of nuclear exosome (Kilchert et al., 2016; Schmid and Jensen, 2019). Our result, therefore, appears to indicate that nuclear degradation machinery is less active in SH-SY5Y cells than the other 2 cell lines. Whether the increased stability of lncRNAs in SH-SY5Y cells would impact gene regulation in neurogenesis would be an interesting direction to explore (Liu et al., 2016).

Our finding that mitochondrial RNAs are more stable in SH-SY5Y cells than other 2 cell lines may have important implications for neurobiology. Mitochondrial dysfunctions are known to be involved in the pathogenesis of several neurodegenerative disorders (Federico et al., 2012; Bader and Winklhofer, 2020), highlighting the role of mitochondria in a high energy demanding environment like brain. Consistently, mitochondrial genes are highly expressed in neuronal tissues compared to non-neuronal tissues (**Supplementary Figure S6B**). How mitochondrial RNA stability regulation contributes to neuronal cell functions is an open question that needs to be addressed in the future.

4 MATERIALS AND METHODS

4.1 Cell Culture and Neural Stem Cell Differentiation

Human HEK293T and HepG2 cells were cultured in high glucose DMEM with 10% fetal bovine serum (FBS). SH-SY5Y cells were cultured in DMEM/F12 with 10% FBS, 1%

sodium pyruvate, and 1% glutamate. All media contained 1% Penicillin/Streptomycin solution (Sigma). Human neural stem cells (NSCs) were obtained from Vescovi et al. (Vescovi et al., 1999). NSCs were cultured in NS-A Basal medium. Differentiation of NSCs to neurons was carried out by following the protocol described in Vescovi et al. (Vescovi et al., 1999). Deidentified primary human NSC line was previously developed (Vescovi et al., 1999). For neuronal induction, individual spheres were mechanically dissociated and transferred at a density of 2.5×10^4 cells/cm² onto matrigel-coated chamber-slides in the presence of 20 ng/ml FGF2. After 72 h, cultures were shifted to NS-A Basal medium containing 2% FBS and were grown for another 2 weeks. Neurons were harvested by using accutane to dissociate cells, and RNA was extracted by using Trizol.

4.2 Metabolic Labeling of RNA and RNA Isolation

Cells at ~70% confluency were used for metabolic labeling of RNA as previously described (Zheng et al., 2018). Briefly, cell media were supplemented with 50 μ M of 4-thiouridine (4sU, Sigma) for 1 h before cell harvest. Total RNA was extracted by using TRIzol (Thermo Fisher Scientific). Newly made (4sU-labeled) and pre-existing RNA pools were fractionated by using a previously described protocol (Radle et al., 2013). Briefly, 100 μ g of total RNA was biotinylated with 200 μ g of biotin-HPDP (Thermo Fisher Scientific), and the biotinylated RNA was captured by Dynabeads MyOne Streptavidin C1 (Thermo Fisher Scientific). The unbound, flow-through (FT) RNA was collected after extensive washing of the beads. Biotinylated RNA (4sU) was eluted from the beads by using DTT.

4.3 3'READS+

The 3'READS+ procedure was carried out as previously described (Zheng et al., 2016). Briefly, poly(A)+ RNA was captured by using oligo (dT)₂₅ magnetic beads (NEB) and was fragmented on-bead by RNase III (NEB). After washing away free RNA fragments, poly(A)+ RNA fragments were eluted from the beads and precipitated with ethanol, followed by ligation to heat-denatured 5' adapter (5'-CCUUGGCACCCGAGAAUCCANNNN) with T4 RNA ligase 1 (NEB). The ligated products were captured by biotin-T₁₅-(+TT)₅ (Exiqon, and +T indicates locked nucleic acid) bound to Dynabeads MyOne Streptavidin C1 (Thermo Fisher Scientific). After washing, RNA fragments on the beads were digested with RNase H and then eluted from the beads. After precipitation with ethanol, RNA fragments were ligated to a 5' adenylated 3' adapter (5'-rApp/NNNGATCGTCCGACTGTAGAACTCTGAAC/3ddC (Bio Scientific) with T4 RNA ligase 2 (truncated KQ version, NEB). The ligation products were then reverse transcribed by using M-MLV reverse transcriptase (Promega), followed by PCR amplification with Phusion high-fidelity DNA polymerase (NEB) and bar-coded PCR primers for 12–18 cycles. PCR products were size selected twice with AMPure

XP beads (Beckman Coulter). The size and quantity of the cDNA libraries were examined on an Agilent Bioanalyzer and sequenced on an Illumina HiSeq machine (1 \times 150 bases).

4.4 3'READS+ Data Analysis

3'READS+ data were analyzed as previously described (Hoque et al., 2013; Zheng and Tian, 2017; Zheng et al., 2018). The 5' adapter was removed using Cutadapt (Martin, 2011) and reads with <23 nt were discarded. The retained reads were mapped to the human genome (hg19) using bowtie2 local mode (Langmead and Salzberg, 2012). The six random nucleotides at the 5' end (from the 3' adapter) were removed using the setting “-5 6” in bowtie2. Reads with a mapping quality score ≥ 10 were kept for further analysis. Reads with two or more non-genomic 5'Ts after alignment were called PAS reads. Cleavage sites were clustered within 24 nt (Hoque et al., 2013) and were assigned to genes based on RefSeq (release 83) annotations (Pruitt et al., 2007). RNA-seq data-supported 3' end extension was applied to improve 3' end region annotation as previously described (Wang et al., 2018b). Genic PASs were annotated by using the RefSeq with the largest genomic span. Human lncRNA annotations were based on data from the FANTOM5 database (Hon et al., 2017). PAS reads mapped to genes were normalized by the median ratio method in DESeq (Anders and Huber, 2010). Only the isoforms with read count greater than five in at least one of the samples were used. PASs on chromosome M were annotated to mitochondrial genes by using MitoCarta2.0 (Calvo et al., 2016).

4.5 Stability Analysis of Transcripts

For each transcript with a defined PAS, its abundances (reads per million mapped, or RPM) in flow-through (FT) sample and 4sU sample were calculated and normalized by using the DESeq method. The log₂ (RPM of FT sample/RPM of 4sU sample) value is called Stability Score (SS). Two biological replicates were averaged. Scaled mRNA decay rate data was obtained from a previous study (Wu et al., 2019), which involved decay rate calculation after transcriptional shutdown by Actinomycin D. Differential stability analysis across cell lines was carried out by using the ANOVA test.

4.6 APA Isoform Analysis

For 3'UTR isoform analysis, the two APA isoforms containing 3'-most exon PASs with the highest expression levels were selected. Differential expression of proximal PAS and distal PAS isoforms was carried out by using DEXSeq (Anders et al., 2012). Significant events were those with $p < 0.05$ (Fisher's exact test or DEXSeq analysis) and relative abundance difference >5%. Relative expression (RE) of the two isoforms was calculated by log₂ (distal PAS isoform RPM/proximal PAS isoform RPM). aUTR size was the distance between the proximal and distal PASs in the 3'UTR.

4.7 Gene Ontology Analysis

Gene ontology analysis was carried out by using the GOstats package in R (Falcon and Gentleman, 2007). The Fisher's exact test was used to calculate p -values to indicate significance of association between a gene set and a GO term. GO terms

associated with more than 1,000 genes were considered too generic and were discarded. To reduce redundancy, any GO term that overlapped with a more significant term by >90% was removed.

4.8 Gene Feature Analysis

Gene features were based on RefSeq annotations. The Pearson correlation r value and individual and cumulative R^2 -value were calculated using the “cor” and “lm” functions in R. The PAS count of a gene and conservation of PAS were obtained from PolyA_DB v3 (Wang et al., 2018a). For intronic features, intron size was based on the RefSeq database, considering all RefSeq-supported splicing isoforms. The strengths of 5' and 3' splicing sites were calculated by using the MaxEntScan program (Yeo and Burge, 2004).

4.9 Sequence Motif Analysis

Nucleotide content in a specific region was calculated by using the BSgenome package in R. Hexamer frequencies in 3'UTRs were calculated by using the Biostrings package in R and were compared between gene sets. p -values indicating significance of hexamer enrichment or depletion were based on the Fisher's exact test. Sequence motifs were generated by using the Weblogo program (<https://weblogo.berkeley.edu/>).

4.10 RNA Structure Analysis

The RNAfold function of the ViennaRNA package (Hofacker and Stadler, 2006) was used to calculate Minimum Folding Energy (MFE) of folded RNA sequences. Each 3'UTR sequence was divided into a series of 100-nt sub-sequences with a 50-nt overlap between adjacent ones. The median MFE value all sub-sequences was used to represent the whole sequence. DMS-Seq data based on K562 cells (Rouskin et al., 2014) were downloaded from NCBI GEO (GSE4580). Gini indices were calculated as previously described (Rouskin et al., 2014). The median Gini Index of each 3'UTR was used for analysis.

REFERENCES

- Anders, S., and Huber, W. (2010). Differential Expression Analysis for Sequence Count Data. *Genome Biol.* 11, R106. doi:10.1186/gb-2010-11-10-r106
- Anders, S., Reyes, A., and Huber, W. (2012). Detecting Differential Usage of Exons from RNA-Seq Data. *Genome Res.* 22, 2008–2017. doi:10.1101/gr.133744.111
- Bader, V., and Winkhofer, K. F. (2020). Mitochondria at the Interface between Neurodegeneration and Neuroinflammation. *Semin. Cel Dev. Biol.* 99, 163–171. doi:10.1016/j.semcdb.2019.05.028
- Barreau, C., Paillard, L., and Osborne, H. B. (2005). AU-rich Elements and Associated Factors: Are There Unifying Principles? *Nucleic Acids Res.* 33, 7138–7150. doi:10.1093/nar/gki1012
- Biedler, J. L., Roffler-Tarlov, S., Schachner, M., and Freedman, L. S. (1978). Multiple Neurotransmitter Synthesis by Human Neuroblastoma Cell Lines and Clones. *Cancer Res.* 38, 3751–3757.
- Blair, J. D., Hockemeyer, D., Doudna, J. A., Bateup, H. S., and Floor, S. N. (2017). Widespread Translational Remodeling during Human Neuronal Differentiation. *Cel Rep.* 21, 2005–2016. doi:10.1016/j.celrep.2017.10.095
- Calvo, S. E., Clauser, K. R., and Mootha, V. K. (2016). MitoCarta2.0: an Updated Inventory of Mammalian Mitochondrial Proteins. *Nucleic Acids Res.* 44, D1251–D1257. doi:10.1093/nar/gkv1003

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in NCBI GEO database (GSE189899).

AUTHOR CONTRIBUTIONS

AG, JS, and BT conceived of and designed the experiments. JS, LF, and DZ performed the experiments. AG, JS, and WW analyzed the data. ZPP contributed critical reagents and discussed results. AG, JS, and BT wrote the paper.

FUNDING

This work was supported by National Institutes of Health grants GM129069 to BT, AA023797 and MH126420 to ZPP, NS097992 to BT and ZPP, and a grant from New Jersey Health Foundation to JS. The Pang lab at the Child Health Institute of New Jersey is supported in part by Robert Wood Johnson Foundation (grant #74260). BT is also supported by a recruitment grant to The Wistar Institute from The Pew Charitable Trusts.

ACKNOWLEDGMENTS

We thank members of BT and ZP labs for helpful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.840369/full#supplementary-material>

- Chen, C.-Y. A., and Shyu, A.-B. (1995). AU-rich Elements: Characterization and Importance in mRNA Degradation. *Trends Biochemical Sciences* 20, 465–470. doi:10.1016/s0968-0004(00)89102-1
- Cheng, L. C., Zheng, D., Baljinnayam, E., Sun, F., Ogami, K., Yeung, P. L., et al. (2020). Widespread Transcript Shortening through Alternative Polyadenylation in Secretory Cell Differentiation. *Nat. Commun.* 11, 3182. doi:10.1038/s41467-020-16959-2
- Cheng, L. C., Zheng, D., Zhang, Q., Guvenek, A., Cheng, H., and Tian, B. (2021). Alternative 3' UTRs Play a Widespread Role in Translation-independent mRNA Association with the Endoplasmic Reticulum. *Cel Rep.* 36, 109407. doi:10.1016/j.celrep.2021.109407
- Derti, A., Garrett-Engle, P., Macisaac, K. D., Stevens, R. C., Sriram, S., Chen, R., et al. (2012). A Quantitative Atlas of Polyadenylation in Five Mammals. *Genome Res.* 22, 1173–1183. doi:10.1101/gr.132563.111
- Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., et al. (2018). Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol. Cel.* 70, 854–867. doi:10.1016/j.molcel.2018.05.001
- DuBridge, R. B., Tang, P., Hsia, H. C., Leong, P. M., Miller, J. H., and Calos, M. P. (1987). Analysis of Mutation in Human Cells by Using an Epstein-Barr Virus Shuttle System. *Mol. Cel Biol* 7, 379–387. doi:10.1128/mcb.7.1.379-387.1987
- Falcon, S., and Gentleman, R. (2007). Using GOstats to Test Gene Lists for GO Term Association. *Bioinformatics* 23, 257–258. doi:10.1093/bioinformatics/btl567

- Federico, A., Cardaioli, E., Da Pozzo, P., Formichi, P., Gallus, G. N., and Radi, E. (2012). Mitochondria, Oxidative Stress and Neurodegeneration. *J. Neurol. Sci.* 322, 254–262. doi:10.1016/j.jns.2012.05.030
- Fischer, J. W., Busa, V. F., Shao, Y., and Leung, A. K. L. (2020). Structure-Mediated RNA Decay by UPF1 and G3BP1. *Mol. Cell.* 78, 70–84. doi:10.1016/j.molcel.2020.01.021
- Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., et al. (2019). GENCODE Reference Annotation for the Human and Mouse Genomes. *Nucleic Acids Res.* 47, D766–D773. doi:10.1093/nar/gky955
- Garneau, N. L., Wilusz, J., and Wilusz, C. J. (2007). The Highways and Byways of mRNA Decay. *Nat. Rev. Mol. Cell Biol.* 8, 113–126. doi:10.1038/nrm2104
- Gruber, A. J., and Zavolan, M. (2019). Alternative Cleavage and Polyadenylation in Health and Disease. *Nat. Rev. Genet.* 20, 599–614. doi:10.1038/s41576-019-0145-z
- Guvenek, A., and Tian, B. (2018). Analysis of Alternative Cleavage and Polyadenylation in Mature and Differentiating Neurons Using RNA-Seq Data. *Quant Biol.* 6, 253–266. doi:10.1007/s40484-018-0148-3
- GTEX Consortium Battle, A., Brown, C. D., Engelhardt, B. E., and Montgomery, S. B. (2017). Genetic Effects on Gene Expression across Human Tissues. *Nature* 550, 204–213. doi:10.1038/nature24277
- Ha, K. C. H., Blencowe, B. J., and Morris, Q. (2018). QAPA: a New Method for the Systematic Analysis of Alternative Polyadenylation from RNA-Seq Data. *Genome Biol.* 19, 45. doi:10.1186/s13059-018-1414-4
- Haimovich, G., Medina, D. A., Causse, S. Z., Garber, M., Millán-Zambrano, G., Barkai, O., et al. (2013). Gene Expression Is Circular: Factors for mRNA Degradation Also Foster mRNA Synthesis. *Cell* 153, 1000–1011. doi:10.1016/j.cell.2013.05.012
- Hentze, M. W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A Brave New World of RNA-Binding Proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341. doi:10.1038/nrm.2017.130
- Hofacker, I. L., and Stadler, P. F. (2006). Memory Efficient Folding Algorithms for Circular RNA Secondary Structures. *Bioinformatics* 22, 1172–1176. doi:10.1093/bioinformatics/btl023
- Hogg, J. R., and Goff, S. P. (2010). Upf1 Senses 3'UTR Length to Potentiate mRNA Decay. *Cell* 143, 379–389. doi:10.1016/j.cell.2010.10.005
- Hon, C.-C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J. L., Gough, J., et al. (2017). An Atlas of Human Long Non-coding RNAs with Accurate 5' Ends. *Nature* 543, 199–204. doi:10.1038/nature21374
- Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., et al. (2013). Analysis of Alternative Cleavage and Polyadenylation by 3' Region Extraction and Deep Sequencing. *Nat. Methods* 10, 133–139. doi:10.1038/nmeth.2288
- Imamachi, N., Salam, K. A., Suzuki, Y., and Akimitsu, N. (2017). A GC-Rich Sequence Feature in the 3' UTR Directs UPF1-dependent mRNA Decay in Mammalian Cells. *Genome Res.* 27, 407–418. doi:10.1101/gr.206060.116
- Ji, Z., Lee, J. Y., Pan, Z., Jiang, B., and Tian, B. (2009). Progressive Lengthening of 3' Untranslated Regions of mRNAs by Alternative Polyadenylation during Mouse Embryonic Development. *Pnas* 106, 7028–7033. doi:10.1073/pnas.0900028106
- Jin, Y., Geisberg, J. V., Moqtaderi, Z., Ji, Z., Hoque, M., Tian, B., et al. (2015). Mapping 3' mRNA Isoforms on a Genomic Scale. *Curr. Protoc. Mol. Biol.* 110, 4–17. doi:10.1002/0471142727.mb0423s110
- Kilchert, C., Wittmann, S., and Vasiljeva, L. (2016). The Regulation and Functions of the Nuclear RNA Exosome Complex. *Nat. Rev. Mol. Cell Biol.* 17, 227–239. doi:10.1038/nrm.2015.15
- Knowles, B. B., Howe, C. C., and Aden, D. P. (1980). Human Hepatocellular Carcinoma Cell Lines Secrete the Major Plasma Proteins and Hepatitis B Surface Antigen. *Science* 209, 497–499. doi:10.1126/science.6248960
- Kovalevich, J., and Langford, D. (2013). Considerations for the Use of SH-Sy5y Neuroblastoma Cells in Neurobiology. *Methods Mol. Biol. (Clifton, N.J.)* 1078, 9–21. doi:10.1007/978-1-62703-640-5_2
- Langmead, B., and Salzberg, S. L. (2012). Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923
- Lee, J. E., Lee, J. Y., Wilusz, J., Tian, B., and Wilusz, C. J. (2010). Systematic Analysis of Cis-Elements in Unstable mRNAs Demonstrates that CUGBP1 Is a Key Regulator of mRNA Decay in Muscle Cells. *PLoS one* 5, e11201. doi:10.1371/journal.pone.0011201
- Lee, J. Y., Ji, Z., and Tian, B. (2008). Phylogenetic Analysis of mRNA Polyadenylation Sites Reveals a Role of Transposable Elements in Evolution of the 3'-end of Genes. *Nucleic Acids Res.* 36, 5581–5590. doi:10.1093/nar/gkn540
- Li, X., Quon, G., Lipshitz, H. D., and Morris, Q. (2010). Predicting *In Vivo* Binding Sites of RNA-Binding Proteins Using mRNA Secondary Structure. *Rna* 16, 1096–1107. doi:10.1261/rna.2017210
- Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S., and Mayr, C. (2013). Ubiquitously Transcribed Genes Use Alternative Polyadenylation to Achieve Tissue-specific Expression. *Genes Dev.* 27, 2380–2396. doi:10.1101/gad.229328.113
- Liu, S. J., Nowakowski, T. J., Pollen, A. A., Lui, J. H., Horlbeck, M. A., Attenello, F. J., et al. (2016). Single-cell Analysis of Long Non-coding RNAs in the Developing Human Neocortex. *Genome Biol.* 17, 67. doi:10.1186/s13059-016-0932-1
- Lizio, M., Harshbarger, J., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., et al. (2015). Gateways to the FANTOM5 Promoter Level Mammalian Expression Atlas. *Genome Biol.* 16, 22. doi:10.1186/s13059-014-0560-6
- Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet j.* 17, 10–12. doi:10.14806/ej.17.1.200
- Mayr, C. (2018). What Are 3' UTRs Doing? *Cold Spring Harbor Perspect. Biol.* 11, a034728. doi:10.1101/cshperspect.a034728
- Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J. O., and Lai, E. C. (2013). Widespread and Extensive Lengthening of 3' UTRs in the Mammalian Brain. *Genome Res.* 23, 812–825. doi:10.1101/gr.146886.112
- Moqtaderi, Z., Geisberg, J. V., and Struhl, K. (2022). A Compensatory Link between Cleavage/polyadenylation and mRNA Turnover Regulates Steady-State mRNA Levels in Yeast. *Cell Biol.* 119, e2121488119. doi:10.1073/pnas.2121488119
- Pan, X., Rijnbeek, P., Yan, J., and Shen, H.-B. (2018). Prediction of RNA-Protein Sequence and Structure Binding Preferences Using Deep Convolutional and Recurrent Neural Networks. *BMC genomics* 19, 511. doi:10.1186/s12864-018-4889-1
- Park, E., and Maquat, L. E. (2013). Staufen-mediated mRNA Decay. *Wiley Interdiscip. Rev. RNA* 4, 423. doi:10.1002/wrna.1168
- Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M. S., Mapendano, C. K., et al. (2008). RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. *Science* 322, 1851–1854. doi:10.1126/science.1164096
- Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007). NCBI Reference Sequences (RefSeq): a Curated Non-redundant Sequence Database of Genomes, Transcripts and Proteins. *Nucleic Acids Res.* 35, D61–D65. doi:10.1093/nar/gkl842
- Rädle, B., Rutkowski, A. J., Ruzsics, Z., Friedel, C. C., Koszinowski, U. H., and Dölken, L. (2013). Metabolic Labeling of Newly Transcribed RNA for High Resolution Gene Expression Profiling of RNA Synthesis, Processing and Decay in Cell Culture. *JoVE* 78, e50195. doi:10.3791/50195
- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J. S. (2014). Genome-wide Probing of RNA Structure Reveals Active Unfolding of mRNA Structures *In Vivo*. *Nature* 505, 701–705. doi:10.1038/nature12894
- Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., and Burge, C. B. (2008). Proliferating Cells Express mRNAs with Shortened 3' Untranslated Regions and Fewer microRNA Target Sites. *Science* 320, 1643–1647. doi:10.1126/science.1155390
- Schmid, M., and Jensen, T. H. (2019). The Nuclear RNA Exosome and its Cofactors. *Adv. Exp. Med. Biol.* 1203, 113–132. doi:10.1007/978-3-030-31434-7_4
- Schoenberg, D. R., and Maquat, L. E. (2012). Regulation of Cytoplasmic mRNA Decay. *Nat. Rev. Genet.* 13, 246–259. doi:10.1038/nrg3160
- Shepard, P. J., Choi, E.-A., Lu, J., Flanagan, L. A., Hertel, K. J., and Shi, Y. (2011). Complex and Dynamic Landscape of RNA Polyadenylation Revealed by PAS-Seq. *Rna* 17, 761–772. doi:10.1261/rna.2581711
- Shin, J., Ding, Q., Wang, L., Cui, Y., Baljinnam, E., Guvenek, A., et al. (2021). CRISPRas: Programmable Regulation of Alternative Polyadenylation by dCas9. *Nucleic Acids Res.* gkab519. doi:10.1093/nar/gkab519
- Singh, I., Lee, S.-H., Sperling, A. S., Samur, M. K., Tai, Y.-T., Fulciniti, M., et al. (2018). Widespread Intronic Polyadenylation Diversifies Immune Cell Transcriptomes. *Nat. Commun.* 9, 1716. doi:10.1038/s41467-018-04112-z
- Spies, N., Burge, C. B., and Bartel, D. P. (2013). 3' UTR-Isoform Choice Has Limited Influence on the Stability and Translational Efficiency of Most mRNAs in Mouse Fibroblasts. *Genome Res.* 23, 2078–2090. doi:10.1101/gr.156919.113

- Sun, M., Schwab, B., Pirkel, N., Maier, K. C., Schenk, A., Failmezger, H., et al. (2013). Global Analysis of Eukaryotic mRNA Degradation Reveals Xrn1-dependent Buffering of Transcript Levels. *Mol. Cell* 52, 52–62. doi:10.1016/j.molcel.2013.09.010
- Taliaferro, J. M., Vidaki, M., Oliveira, R., Olson, S., Zhan, L., Saxena, T., et al. (2016). Distal Alternative Last Exons Localize mRNAs to Neural Projections. *Mol. Cell* 61, 821–833. doi:10.1016/j.molcel.2016.01.020
- Tian, B., Bevilacqua, P. C., Diegelman-Parente, A., and Mathews, M. B. (2004). The Double-Stranded-RNA-Binding Motif: Interference and Much More. *Nat. Rev. Mol. Cell Biol* 5, 1013–1023. doi:10.1038/nrm1528
- Tian, B., and Manley, J. L. (2017). Alternative Polyadenylation of mRNA Precursors. *Nat. Rev. Mol. Cell Biol* 18, 18–30. doi:10.1038/nrm.2016.116
- Tian, B., Pan, Z., and Lee, J. Y. (2007). Widespread mRNA Polyadenylation Events in Introns Indicate Dynamic Interplay between Polyadenylation and Splicing. *Genome Res.* 17, 156–165. doi:10.1101/gr.5532707
- Tushev, G., Glock, C., Heumüller, M., Biever, A., Jovanovic, M., and Schuman, E. M. (2018). Alternative 3' UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments. *Neuron* 98, 495–511. doi:10.1016/j.neuron.2018.03.030
- Van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., et al. (2020). A Large-Scale Binding and Functional Map of Human RNA-Binding Proteins. *Nature* 583, 711–719. doi:10.1038/s41586-020-2077-3
- Vescovi, A. L., Parati, E. A., Gritti, A., Poulin, P., Ferrario, M., Wanke, E., et al. (1999). Isolation and Cloning of Multipotential Stem Cells from the Embryonic Human CNS and Establishment of Transplantable Human Neural Stem Cell Lines by Epigenetic Stimulation. *Exp. Neurol.* 156, 71–83. doi:10.1006/exnr.1998.6998
- Vlasova, I. A., Tahoe, N. M., Fan, D., Larsson, O., Rattenbacher, B., Sternjohn, J. R., et al. (2008). Conserved GU-rich Elements Mediate mRNA Decay by Binding to CUG-Binding Protein 1. *Mol. Cell* 29, 263–270. doi:10.1016/j.molcel.2007.11.024
- Vlasova-St. Louis, I., Dickson, A. M., Bohjanen, P. R., and Wilusz, C. J. (2013). CELFish Ways to Modulate mRNA Decay. *Biochim. Biophys. Acta (Bba) - Gene Regul. Mech.* 1829, 695–707. doi:10.1016/j.bbaggm.2013.01.001
- Wang, R., Nambiar, R., Zheng, D., and Tian, B. (2018a). PolyA_DB 3 Catalogs Cleavage and Polyadenylation Sites Identified by Deep Sequencing in Multiple Genomes. *Nucleic Acids Res.* 46, D315–d319. doi:10.1093/nar/gkx1000
- Wang, R., Zheng, D., Yehia, G., and Tian, B. (2018b). A Compendium of Conserved Cleavage and Polyadenylation Events in Mammalian Genes. *Genome Res.* 28, 1427–1441. doi:10.1101/gr.237826.118
- Wolin, S. L., and Maquat, L. E. (2019). Cellular RNA Surveillance in Health and Disease. *Science* 366, 822–827. doi:10.1126/science.aax2957
- Wu, Q., Medina, S. G., Kushawah, G., DeVore, M. L., Castellano, L. A., Hand, J. M., et al. (2019). Translation Affects mRNA Stability in a Codon-dependent Manner in Human Cells. *eLife* 8, e45396. doi:10.7554/eLife.45396
- Wu, X., and Bartel, D. P. (2017). Widespread Influence of 3'-End Structures on Mammalian mRNA Processing and Stability. *Cell* 169, 905–917. doi:10.1016/j.cell.2017.04.036
- Yeo, G., and Burge, C. B. (2004). Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *J. Comput. Biol.* 11, 377–394. doi:10.1089/1066527041410418
- Yin, Y., Lu, J. Y., Zhang, X., Shao, W., Xu, Y., Li, P., et al. (2020). U1 snRNP Regulates Chromatin Retention of Noncoding RNAs. *Nature* 580, 147–150. doi:10.1038/s41586-020-2105-3
- Zhang, H., Lee, J., and Tian, B. (2005). Biased Alternative Polyadenylation in Human Tissues. *Genome Biol.* 6, R100. doi:10.1186/gb-2005-6-12-r100
- Zheng, D., Cho, H., Wang, W., Rambout, X., Tian, B., and Maquat, L. E. (2020). 3'READS + RIP Defines Differential Staufen1 Binding to Alternative 3'UTR Isoforms and Reveals Structures and Sequence Motifs Influencing Binding and Polysome Association. *Rna* 26, 1621–1636. doi:10.1261/rna.076133.120
- Zheng, D., Liu, X., and Tian, B. (2016). 3'READS+, a Sensitive and Accurate Method for 3' End Sequencing of Polyadenylated RNA. *Rna* 22, 1631–1639. doi:10.1261/rna.057075.116
- Zheng, D., and Tian, B. (2017). Polyadenylation Site-Based Analysis of Transcript Expression by 3'READS+. *Methods Mol. Biol. (Clifton, N.J.)* 1648, 65–77. doi:10.1007/978-1-4939-7204-3_6
- Zheng, D., Wang, R., Ding, Q., Wang, T., Xie, B., Wei, L., et al. (2018). Cellular Stress Alters 3'UTR Landscape through Alternative Polyadenylation and Isoform-specific Degradation. *Nat. Commun.* 9, 2268. doi:10.1038/s41467-018-04730-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Guvenek, Shin, De Filippis, Zheng, Wang, Pang and Tian. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.