








# The polyA tail facilitates splicing of last introns with weak 3' splice sites via PABPN1

Li Huang<sup>1,†</sup> , Guangnan Li<sup>1,†</sup> , Chen Du<sup>1,†</sup> , Yu Jia<sup>1</sup>, Jiayi Yang<sup>1</sup> , Weiliang Fan<sup>1</sup> , Yong-Zhen Xu<sup>1</sup>, Hong Cheng<sup>2</sup>  & Yu Zhou<sup>1,3,4,5,\*</sup> 

## Abstract

The polyA tail of mRNAs is important for many aspects of RNA metabolism. However, whether and how it regulates pre-mRNA splicing is still unknown. Here, we report that the polyA tail acts as a splicing enhancer for the last intron via the nuclear polyA binding protein PABPN1 in HeLa cells. PABPN1-depletion induces the retention of a group of introns with a weaker 3' splice site, and they show a strong 3'-end bias and mainly locate in nuclear speckles. The polyA tail is essential for PABPN1-enhanced last intron splicing and functions in a length-dependent manner. Tethering PABPN1 to nonpolyadenylated transcripts also promotes splicing, suggesting a direct role for PABPN1 in splicing regulation. Using TurboID-MS, we construct the PABPN1 interactome, including many spliceosomal and RNA-binding proteins. Specifically, PABPN1 can recruit RBM26&27 to promote splicing by interacting with the coiled-coil and RRM domain of RBM27. PABPN1-regulated terminal intron splicing is conserved in mice. Together, our study establishes a novel mode of post-transcriptional splicing regulation via the polyA tail and PABPN1.

**Keywords** last intron splicing; PABPN1; polyA tail; RBM26&27; weak 3' splice site

**Subject Categories** RNA Biology

**DOI** 10.15252/embr.202357128 | Received 6 March 2023 | Revised 4 August 2023 | Accepted 16 August 2023 | Published online 4 September 2023

**EMBO Reports (2023) 24: e57128**

## Introduction

The 3'-end cleavage and polyadenylation of eukaryotic pre-mRNAs is indispensable for mRNA maturation. Most nascent pre-mRNAs are cleaved by a macromolecular complex consisting of the CPSF, CstF, CF I, and CF II subcomplexes, and then polyadenylated by the polyA polymerase (PAP) resulting in a 3'-end polyA tail (Colgan &

Manley, 1997; Shi *et al*, 2009). The poly(A) tail (polyA) is ~ 90 nt in yeast and ~ 250 nt in mammals (Wahle, 1995; Mangus *et al*, 2003). In the nucleus, it is reported that once polyA is synthesized to more than 10 adenosines, the nuclear polyA binding protein PABPN1 immediately binds to the short A stretches and stimulates its continuing synthesis, and PABPN1 would cover the whole polyA sequence (Wahle, 1991). In the cytosol, another group of cytoplasmic polyA binding proteins (PABPCs) occupies the polyA tails, which are considered to stabilize mRNAs and promote translation (Otero *et al*, 1999; Gray *et al*, 2000; Eliseeva *et al*, 2013).

As the primary polyA binding protein in the nucleus, PABPN1 plays an important role in many aspects of RNA metabolism. First, PABPN1 is well characterized for its function in stimulating the activity of PAP in polyA synthesis (Wahle, 1991; Kerwitz *et al*, 2003; Kuhn *et al*, 2009). Second, PABPN1 can regulate the alternative polyadenylation (APA) by inhibiting the recognition of the proximal polyA site (PAS), as depletion of PABPN1 results in more usage of proximal PAS of specific genes (de Klerk *et al*, 2012; Jenal *et al*, 2012). Recently, PABPN1 has also been found to facilitate transporting mRNAs to cytosol synergistically with Aly/REF (Shi *et al*, 2017), and to involve in PAXT (polyA tail exosome targeting) complex-mediated nuclear RNA degradation of specific promoter upstream transcripts (PROMPTs) and long noncoding RNAs (lncRNAs) (Beaulieu *et al*, 2012; Bresson & Conrad, 2013; Bresson *et al*, 2015a; Meola *et al*, 2016a). PABPN1 can act as a zygotic factor mediating maternal mRNA decay during mouse early embryonic development (Zhao *et al*, 2020, 2022). Moreover, 5' short GCG expansion in the *PABPN1* coding region is found to be associated with oculopharyngeal muscular dystrophy (OPMD), a late-onset genetic disease in humans. The mutated PABPN1 with an extended polyalanine tract in the N terminal is found to form nuclear or intercellular aggregation granules in patients, a pathologic characteristic of the disease (Jenal *et al*, 2012; Anvar *et al*, 2013; Vest *et al*, 2017).

The pre-mRNA splicing is regulated at multiple layers of RNA biogenesis and processing, such as transcription, capping, 3'-end

<sup>1</sup> College of Life Sciences, TaiKang Center for Life and Medical Sciences, RNA Institute, Wuhan University, Wuhan, China

<sup>2</sup> Key Laboratory of RNA Science and Engineering, Chinese Academy of Sciences, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai, China

<sup>3</sup> Frontier Science Center for Immunology and Metabolism, Wuhan University, Wuhan, China

<sup>4</sup> Institute of Advanced Studies, Wuhan University, Wuhan, China

<sup>5</sup> State Key Laboratory of Virology, Wuhan University, Wuhan, China

\*Corresponding author. Tel: +86 15827526352; E-mail: yu.zhou@whu.edu.cn

<sup>†</sup>These authors contributed equally to this work

cleavage and polyadenylation (CPA) (Cooke & Alwine, 1996; Fong & Bentley, 2001; Maniatis & Reed, 2002). The coupling between transcription and splicing has been well studied, during which RNA polymerase II and some transcriptional elongation factors function in coordinating the co-transcription splicing (McCracken et al, 1997; Bentley, 2005). Similarly, early experiments have shown that the CPA process and splicing are not independent but highly coordinated (Zhao et al, 1999; Rigo & Martinson, 2008; Kaida, 2016). Based on *in vitro* experiments, on the one hand, splicing of the last introns can stimulate CPA, and mutation in the 3' splice site (ss) of last introns inhibits splicing as well as polyadenylation (Niwa et al, 1990; Antoniou et al, 1998; Cooke et al, 1999; Dye & Proudfoot, 2001); On the other hand, CPA can also affect splicing, and attenuated mutant of 'AAUAAA' motif significantly reduces the splicing efficiency of introns, especially of final introns (Niwa & Berget, 1991). Specifically, it has been showed that PAP can directly interact with U2AF to promote splicing (Vagner et al, 2000). However, whether the product of the CPA process, polyA tail regulates pre-mRNA splicing in the nucleus is still unknown.

In this study, we perturb the nuclear polyA binding protein PABPN1 in HeLa cells and find that PABPN1 depletion can induce the last intron retention of a considerable subset of genes *in vivo*. Using multiple types of omics data and biochemical assays, we show that PABPN1 can act as a splicing enhancer, together with RBM26&27 to promote splicing of last introns with weaker 3' ss. Importantly, the polyA tail is essential in PABPN1-enhanced last intron splicing, and the effect is polyA tail length-dependent, uncovering a novel mode of polyA tail-coupled splicing regulation.

## Results

### PABPN1 knockdown preferentially induces retention of the last introns

To investigate the function of polyA tail in the nucleus, we started to focus on the primary nuclear polyA binding protein PABPN1. We performed genome-wide RNA profiling in HeLa cells by knocking down PABPN1 (PABPN1-KD) using anti-PABPN1 siRNAs, in comparison with scrambled siRNAs. All siRNAs of PABPN1 used in this study showed strong interference in PABPN1's expression, whereas scrambled siRNAs (siNCs) had no effect (Fig EV1A). Our Ribominus RNA-seq data show good reproducibility between independent replicates from two different siRNAs (Fig EV1B). Interestingly, we noticed that loss of PABPN1 induces retention of many introns, and we quantified the intron retention ratio (IR) using IRFinder (Middleton et al, 2017). Requiring a  $\Delta$ IR (siPABPN1-siNC) cutoff of 0.15 and a *P*-value cutoff at 0.01, we identified 287 IR-upregulated introns and 61 IR-downregulated introns in siPABPN1-treated cells (Fig 1A). Thus, PABPN1 shows a more functional role in enhancing rather than repressing intron splicing (about 4.7 folds difference). As shown by the representative gene example, *COPSS* (Fig 1B), the IRs of the last introns in *COPSS* increased by about 0.25 upon PABPN1 knockdown.

When examining the RNA-seq signals, we noticed several retained introns, such as the example above, are located at the 3'-end of genes. To globally investigate the potential positional bias of regulated introns, we analyzed the distribution of those introns in

genes by calculating the position of the regulated intron within gene using the annotation of its most abundant isoform quantified by the Salmon program (Patro et al, 2017). For example, the position or the rank of the regulated intron, being the 9<sup>th</sup> intron in a major isoform with 10 introns, is calculated as 0.9 (9/10). Collectively, as shown in Fig 1C, PABPN1-KD induced up-regulated retained introns (red line) show a stronger 3'-end bias than uniformly distributed introns (black dashed line), and also than those down-regulated introns (blue line), about 62.45% (143/229) introns being the "gene last introns". Meanwhile, we identified 235 retained introns in control samples with an IR ratio  $\geq$  0.35 and found that they show little 3'-end bias (Fig 1C, green line). Notably, only 13 of those retained introns are changed (5 up-regulated and 8 down-regulated) under PABPN1 knockdown (Fig EV1C), suggesting that PABPN1's effect is independent of the intrinsic splicing efficiency of the introns.

To further validate the regulated last introns, we quantified the spliced and unspliced isoforms with designed exon-exon junctions (EEJs) and exon-intron junction (EIJ) primers, respectively, for six representative genes: *LARP7*, *DUSP12*, *SNRNP48*, *COPSS*, *BRCA1*, and *NSUN5*. PABPN1-KD induced a significant increase in EIJ signal and a decrease in EEJ signal in all these last introns leading to increased intron retention levels without affecting the expression of ACTB control (Fig 1D and E). Furthermore, using RT-qPCR, we also found that the retention of last introns is significantly increased upon PABPN1 knockdown, as shown for the three representative genes *LARP7*, *DUSP12*, and *BRCA1*, respectively (Fig 1F). Together, these results indicate that loss of PABPN1 results in an increase of transcripts with last retained introns at the steady-state level, which may be due to either loss of splicing or increased RNA half-life upon PABPN1 knockdown.

To further investigate whether the last intron retention upon PABPN1 knockdown is a direct result of PABPN1, we performed rapid depletion of PABPN1 protein by dTAG system (Nabet et al, 2018), for which we knocked in a novel degrader FKBP12<sup>F36V</sup> at the start codon of endogenous PABPN1 by CRISPR (Fig EV1D). Treating the cells with the drug dTAG13 can induce quick decay of the target protein, and we found that 3- or 6-h treatment induced almost complete degradation of PABPN1 (Fig 1G). We performed high-throughput RNA-seq for samples (dTAG treatment at 0, 3, and 6 h) from two independent dTAG cell lines. The normalized RNA-seq signals of many last retained introns showed gradual accumulation at 3 and 6 h in comparison to 0 h in both cell lines, as exemplified by *COPSS* (Fig 1H), *LARP7* (Fig EV1E), and *DUSP12* (Fig EV1F). These effects were validated by RT-PCR (Fig 1I). Further analysis revealed that 93% PABPN1 regulated last introns (133 in 143) identified from PABPN1 knockdown assay show increased IR following induction of acute degradation of PABPN1 (Fig 1J). We note that the extents of increased  $\Delta$ IR are not as evident as those from the KD assay. This is likely due to the difference between short-term (3 or 6 h) and long-term (48 h) treatment. Together, these data demonstrate that the loss of PABPN1 causes the retention of many last introns.

### PABPN1 could promote last intron splicing without the CPA process

Because PABPN1 has been previously reported to regulate the selection of alternative polyA sites (Jenal et al, 2012), to investigate

whether PABPN1 regulated last intron splicing is associated with APA, we extracted all possible polyA sites (PASs) in HeLa cells from our recently published data (Data ref: Tang et al, 2021) and calculated the PUDI (Percentage of Distal polyA site Usage Index) of all genes using our RNA-seq data from PABPN1 knockdown and control samples. Under the cutoffs of PUDI difference  $|\Delta PUDI| > 0.15$  and  $FDR < 0.05$ , we identified 572 genes switching to use the proximal PAS and 132 genes switching to use the distal PAS upon PABPN1 knockdown (Fig EV1G). This is consistent with the reported pattern that loss of PABPN1 induces more proximal PAS usage (de Klerk et al, 2012; Jenal et al, 2012). However, those genes have little overlap with the IR-upregulated or -downregulated genes

induced by PABPN1 knockdown (Fig EV1H), suggesting PABPN1-mediated splicing regulation is independent of its role in APA regulation.

Since PABPN1 is also considered part of the CPA complex (Shi et al, 2009), we next asked whether CPA complex-mediated 3' end processing can affect the splicing of terminal introns regulated by PABPN1. To confirm our suspicion, we constructed a series of reporters by replacing endogenous PAS with mouse histone H2A 3' end, which is recognized by U7 snRNP and then cleaved by histone pre-mRNA cleavage complex (HCC) to produce a conservative stem loop (SL) structure, or replacing the PAS with MALAT1 3' end, both being distinct from the CPA process and independent of PABPN1

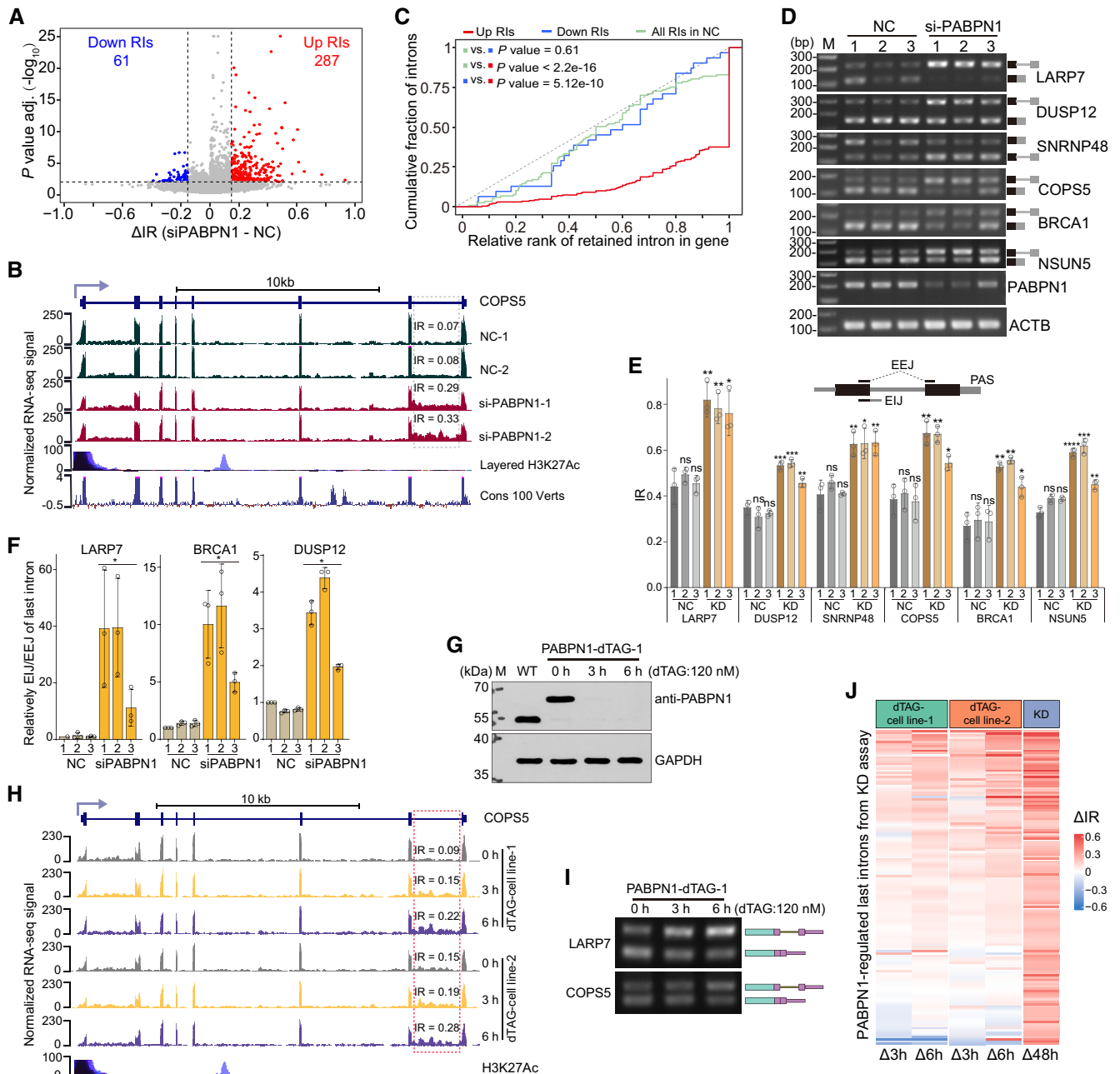


Figure 1.

**Figure 1. PABPN1-loss induced preferential retention of last introns.**

- A Differential intron retention (IR) analysis of RNA-seq data from samples under PABPN1 knockdown (siPABPN1) and control (NC) conditions. Significantly repressed and induced intron retention events (retained introns, RIs) are indicated by blue and red dots, respectively. The threshold of IR difference ( $\Delta$ IR, siPABPN1-siNC) is 0.15, and the FDR cutoff is 0.01.
- B UCSC genome browser view showing the normalized RNA-seq signals of *COP55* in HeLa cells under PABPN1 knockdown (si-PABPN-1 and si-PABPN-2) and control (NC-1 and NC-2) conditions. The quantified intron retention levels (IR score) of the last introns (dotted box) are labeled for all samples, respectively.
- C Cumulative frequency distributions of retained introns (RIs), up-regulated (red), and down-regulated (blue) upon PABPN1 knockdown, and of all retained introns in control (NC, green) samples by their relative intronic ranks in genes.
- D, E Validation of last intron retention of six representative genes in response to PABPN1 knockdown by semi-quantitative PCR in HeLa cells treated with three different NC and PABPN1 siRNAs (D). The bottom box diagram shows a schematic view of the primers for spliced (EE), exon–exon junction) and unspliced (EI), exon–intron junction) isoforms. The quantified intron retention (IR) scores of the last introns are shown on the right for all samples, respectively (E).
- F Relative ratios of the unspliced (EI) to spliced (EE) isoforms of the last introns for *LARP7*, *BRCA1*, and *DUSP12* genes in HeLa cells treated in the same manner as in (D).
- G Western blot validating the degradation effect of PABPN1 under the treatment of dTAG13 for 0, 3, and 6 h.
- H UCSC genome browser view showing the normalized RNA-seq signals of *COP55* in PABPN1-dTAG cells under PABPN1 rapid degradation for 0, 3, and 6 h. The quantified intron retention levels (IR score) of the last introns (dotted box) are labeled for all samples, respectively.
- I The last intron splicing analysis of *LARP7* and *COP55* in response to PABPN1 rapid degradation by dTAG13 for 0, 3, and 6 h by RT–PCR in PABPN1 dTAG cell lines.
- J Heatmap showing the IR ratio change of PABPN1 knockdown-induced retained last introns under rapid degradation of PABPN1 for 3 and 6 h compared to 0 h. The  $\Delta$ IR values of 143 PABPN1-regulated last introns from the knockdown assay are shown on the right for comparison.

Data information: (E, F) Data are from  $n = 3$  biologically independent experiments. The error bar represents SD. The  $P$ -values are calculated by unpaired  $t$ -test (\*\*\*\* $P < 0.0001$ ; \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; \* $P < 0.05$ ; ns, nonsignificant). See also Fig EV1.

Source data are available online for this figure.

(Marzluff & Koreski, 2017; Sun *et al*, 2020). We selected two IR-upregulated genes in response to PABPN1-KD (*LARP7*, Fig EV2A and *COP55*, 1B), and three control genes, not regulated by PABPN1 but with similar expression levels and similar sizes of terminal introns and exons (*TDRD9*, *CENPV*, and *SMTN*, Fig EV2B–D) to construct reporters. As depicted in Fig 2A, for each gene, we constructed 4 GFP reporters: C1 and C2 are the spliced and unspliced controls, respectively; C3 and C4 contain the last intron along with its flanking exons ending with the endogenous PAS (C3) or H2A/MALAT1 3' end (C4). The unspliced control C2 was designed using the intronic stop codon from the IR isoform. We transfected these reporters into HeLa cells and examined the splicing of the last intron by the GFP fusion proteins from the four reporters. We found that replacing the endogenous PAS with histone H2A or MALAT1 3' end strongly repressed the splicing of PABPN1-regulated last introns of *LARP7* and *COP55*, but had little impact on PABPN1 nonregulated last introns of *TDRD9*, *CENPV*, and *SMTN* (Fig 2B). Furthermore, we confirmed the PAS effect on PABPN1-regulated last introns on mRNA level using RT-PCR (Fig 2C). These results showed that the accumulation of transcripts with the last retained intron was gene-specific and dependent on PAS with polyA tail, which is consistent with RNA-seq results that loss of PABPN1 results in an increase of those transcripts.

It has been shown that once the CPA produces a 3' end, PABPN1 binds to the polyA tail immediately (Wahle, 1991). To distinguish whether the PABPN1-loss induced last intron splicing is due to the CPA complex or PABPN1 protein directly, we designed an MS2-PABPN1 tethering system by introducing 5 copies of MS2 sequence into the above C4 reporter downstream of the 3' UTR and upstream of the histone H2A/MALAT1 3' end (Fig 2D). After transfection of the MS2 reporter with MBP, PABPN1, or MBP–PABPN1, we performed RT–PCR and western blot to examine the splicing of the last intron in the reporter. For both *LARP7* and *COP55*, though in the absence of the endogenous PAS, tethering PABPN1 to the MS2 SL enhances the splicing of the last introns, whereas MBP or PABPN1 alone has little effect, both at the mRNA level (Fig 2E) and the protein level (Fig EV2E). These data demonstrated that the

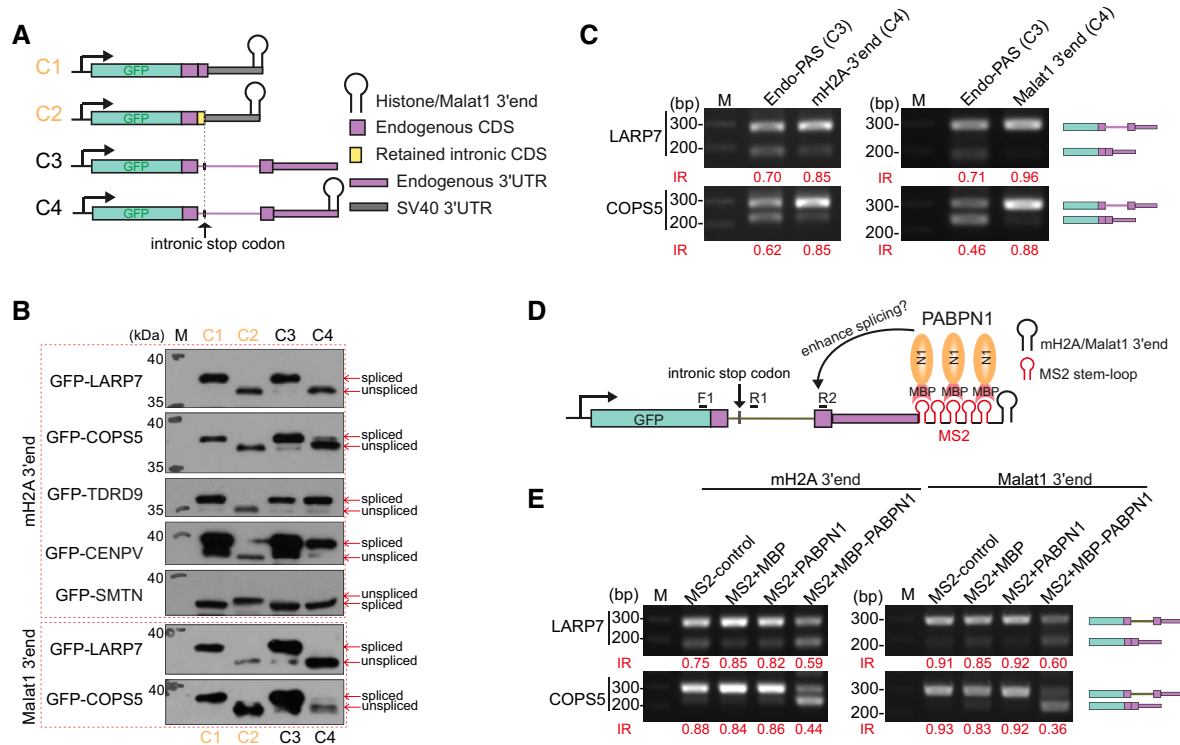
PABPN1 binding is sufficient to promote the terminal intron splicing of specific mRNAs.

Interestingly, by inspecting our RNA-seq data, we found a special gene, *FBXO4*, in which part of the last intron is retained upon PABPN1 knockdown (Fig EV2F). We reasoned that this is due to the inactivation of the last 3' ss caused activation of an upstream cryptic 3' ss (A3SS), which has the largest 3' ss score in the intron based on spliceAI program (Fig EV2F). We further constructed four reporters for this gene as in Fig 2A, with an additional control reporter for the A3SS isoform. We found that the last intron of *FBXO4*, is largely retained when replacing endogenous PAS with histone H2A 3' end at the protein level (Fig EV2G) and also at the RNA level (Fig EV2H). Together, these data demonstrate that PABPN1 can directly promote the last intron splicing without the CPA process.

**PABPN1 binds to the polyA tails of IR-upregulated genes *in vivo***

To investigate whether PABPN1 binds to the IR-upregulated genes, we re-analyzed the eCLIP-seq data of PABPN1 generated by the ENCODE consortium (data ref: Van Nostrand *et al*, 2016b). We first examined the enriched 8-mers in the sequencing reads from PABPN1 IP and input samples, and found that the 'AAAAAAA' motif is the most enriched motif for PABPN1, significantly higher than the input (Fig 3A). Furthermore, we counted the number of reads ending with a polyA tail (at least 6 or 30 continuous As) and found that PABPN1 IP samples have significantly higher fractions of those reads than the input sample (Fig 3B). Those reads ending with a polyA tail are mapped to the genomic region close to the annotated PASs (Fig EV3A). Together, these results demonstrate that PABPN1 can bind to the polyA tails of mRNAs, consistent with its role of acting as a polyA-binding protein.

We next inspected whether the IR-upregulated last introns are directly regulated by PABPN1. By requiring a gene to have PABPN1 eCLIP peak(s) in its 3' UTR or have uniquely mapped read(s) ending with polyA tail, we found that about 73% (105 in 143) of the last intron retained genes upon PABPN1 knockdown are directly bound by PABPN1 (Fig EV3B), among which there are 517 CLIP sites



**Figure 2. Replacing endogenous PAS with histone or Malat1 3' end without CPA-mediated cleavage and polyadenylation mimics the phenomenon of PABPN1-loss-induced last intron retention.**

- A** Schematic diagrams of the GFP reporter constructs for downstream splicing assay. The associated elements are annotated on the right. For one endogenous gene, four reporters are constructed: C1 & C2 are the spliced and unspliced CDS controls with SV40 3' UTR and histone 3' end, respectively; C3 & C4 contain the last intron and its flanking exons of the endogenous gene, with the endogenous 3' end (C3), histone or Malat1 3' end (C4).
- B** Western blots of the C1–C4 reporters for two endogenous regulated genes (*LARP7* and *COPS5*) and three endogenous nonregulated control genes (*TDRD9*, *CENPV*, and *SMTN*).
- C** RT–PCR analysis of the spliced and unspliced isoforms from the C3 and C4 reporters for two endogenous regulated genes *LARP7* and *COPS5*.
- D** Schematic diagram of MS2 tethering reporter for PABPN1. 5× MS2 stem loops are inserted to the C4 reporters constructed in Fig 2A which are behind the 3' UTR and in front of histone/Malat1 3' end.
- E** RT–PCR quantification of MS2 reporters of histone 3' end (left) and Malat1 3' end (right) for two PABPN1-regulated genes *LARP7* and *COPS5*.

Data information: See also Fig EV2.

Source data are available online for this figure.

(72.2%) at the polyA tails, 168 CLIP sites (23.5%) at the A-stretches within the 3' UTR, and only 31 CLIP sites (4.3%) are within the body of 3' UTR without A-stretch (Fig EV3C). As shown in a representative gene example (Fig EV3D), *AMDHD2* has broad PABPN1 binding signals in its 3' UTR and its last intron is more retained upon loss of PABPN1. In considering the eCLIP data are from HepG2 cells, we conclude that PABPN1-enhanced last introns identified in HeLa cells are mainly regulated by PABPN1 binding to the 3'-end of their RNAs.

### Long polyA tail promotes the splicing of PABPN1-enhanced last introns

Based on the fact that PABPN1 binds to the polyA tail *in vivo*, we next asked whether the polyA tail regulates last intron splicing. To this end, we designed another reporter system using polyA sequences with different lengths to replace the 5 copies of MS2 SLs (Fig 3C). For three PABPN1-regulated endogenous genes, we

constructed five reporters with different lengths of polyA sequences (35, 68, 100, 135, or 168 nt), confirmed via double digestion assay (Fig EV3E). Interestingly, we found that the spliced isoforms of all the reporters gradually increase along with the growth of polyA length, and the corresponding IR levels gradually decreased (Fig 3D). The spliced-to-unspliced ratios (EEJ/EIJ) of *LARP7* and *COPS5* also validate this polyA length-dependent enhancement (Fig 3E). Consistently, we observed the protein levels of the spliced isoforms gradually increase for *LARP7*, *COPS5*, and *NSUN5* with H2A 3' end (Fig EV3F) or with MALAT1 3' end (Fig EV3G–I). Based on these data, we speculate that PABPN1-regulated genes may possess a longer polyA tail *in vivo*. We thus analyzed the polyA tail lengths of the transcripts with unspliced last introns using the Nanopore long-reads sequencing data published recently (data ref: Tang *et al*, 2021). We observed that the polyA tail lengths of PABPN1-regulated last intron-retained genes are longer than those of nonregulated genes (Fig 3F). Together, these results demonstrate that PABPN1 enhances the 3' terminal introns splicing in a polyA tail

length-dependent manner, as a longer polyA tail may recruit more PABPN1 proteins.

To further investigate whether the polyA reporters are still dependent on PABPN1, we knocked down PABPN1 and transfected the polyA(168) and control reporters into the cells. The results show that for three case genes (*LARP7*, *COP55*, and *NSUN5*), the polyA(168) reporters have decreased splicing efficiency upon PABPN1 knockdown similar to endo-PAS reporters, while the H2A reporters always show little splicing and are not affected by PABPN1

knockdown (Fig 3G), validating that the splicing enhancement is PABPN1-dependent. Collectively, these results suggest that the long polyA tail promotes the last intron splicing through PABPN1.

**PolyA polymerase contributes to PABPN1-enhanced last intron splicing**

To further investigate the functional roles of polyA tails on terminal intron splicing, we knocked down PolyA polymerase (PAP), the

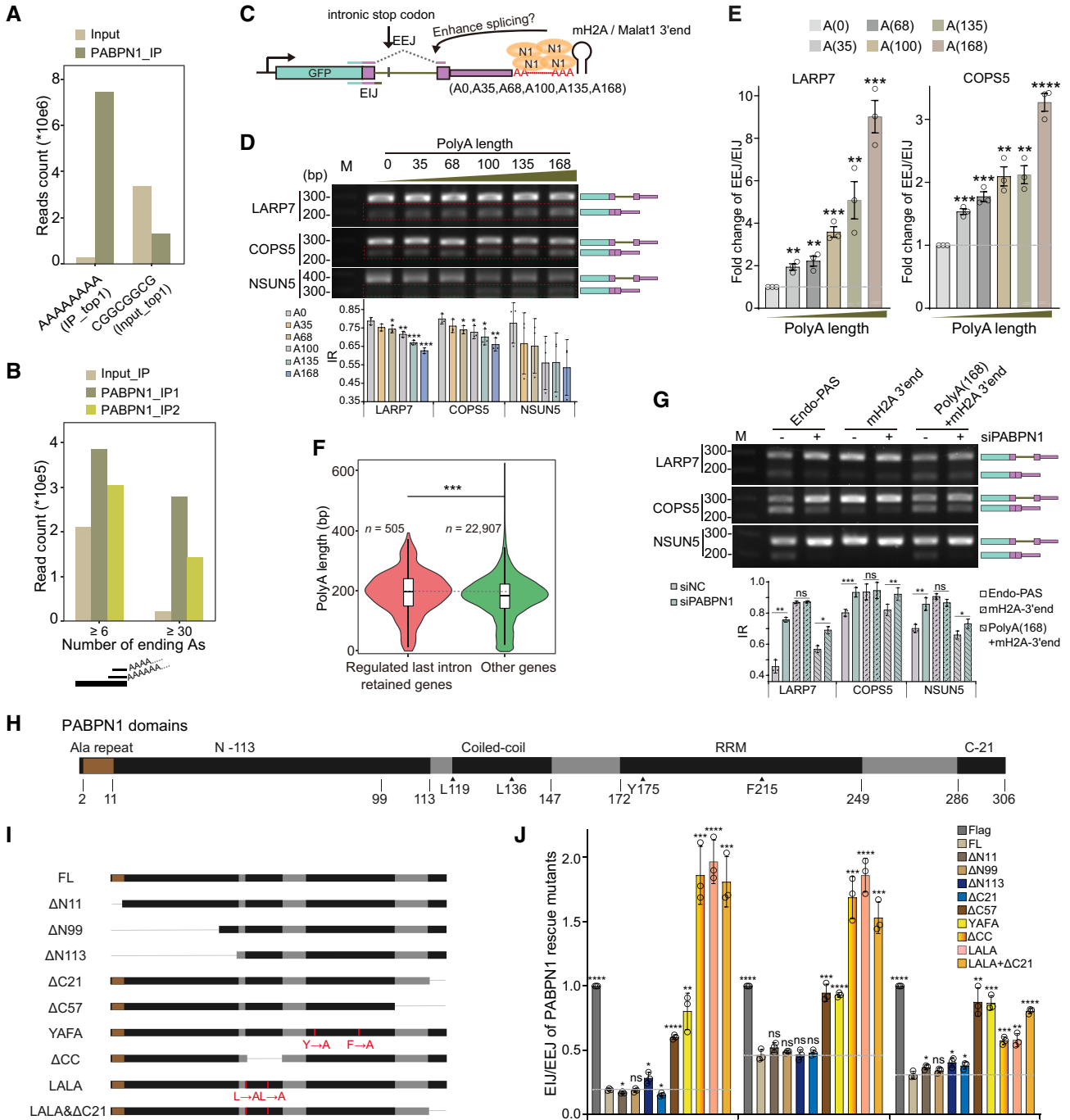


Figure 3.

**Figure 3. Long polyA tail promotes the splicing of PABPN1-enhanced last introns.**

- A, B Bar plot showing the enrichment of polyA sequences in PABPN1 immunoprecipitated RNAs using 8-mer motif analysis (A) or examining the reads ending with 6 or 30 more As.
- B In (D), only the top motif enriched in PABPN1 IP or input control is shown. The PABPN1 eCLIP data was from ENCODE project (ENCSR820UYE).
- C Schematic diagram of the splicing reporters by replacing MS2 stem loops in Fig 3A with polyA sequences of different lengths.
- D RNA quantification by RT-PCR for the spliced and unspliced isoforms (top) and calculated IR scores (bottom) of the reporters in Fig 3F.
- E Quantification of the spliced (EE) over unspliced isoforms (EI) of the reporters in Fig 3F.
- F The polyA tail length distribution of regulated last intron-retained genes and other genes by nanopore sequencing. The *P*-values are calculated by the Wilcoxon test ( $****P < 0.0001$ ).
- G Splicing analysis of the C3, C4, and polyA reporters in Fig 3F for endogenous genes *LARP7*, *COP55*, and *NSUN5* by RT-PCR upon PABPN1 knockdown (top) and the quantified intron retention levels (IR score) of these reporters (bottom).
- H Protein domains of PABPN1 with known important mutation sites marked by triangles.
- I Schematic diagrams of PABPN1 mutants for downstream rescue assay.
- J Relative ratios of the unspliced isoforms (EI) to the spliced isoforms (EE) of the last introns for *LARP7*, *DUSP12*, and *BRCA1* genes in HeLa cells treated with si-PABPN1 and then rescuing PABPN1 mutants constructed in G.

Data information: (D, E, G, and J) Data are from  $n = 3$  biologically independent experiments. The error bars represent SD. The *P*-values are calculated by unpaired t-test ( $****P < 0.0001$ ;  $***P < 0.001$ ;  $**P < 0.01$ ;  $*P < 0.05$ ; ns, nonsignificant). See also Fig EV3.

Source data are available online for this figure.

dominant adenosine polymerase to generate polyA tails for mRNAs. As PABPN1 has been reported to regulate RNA decay (Bresson & Conrad, 2013; Bresson *et al*, 2015a), we also knocked down some exosome component proteins (RRP40, ZFC3H1, and ZCCHC8). We found that the unspliced-to-spliced ratios (EI/EEJ) of the last introns from *LARP7* and *DUSP12* increase upon PAP knockdown, although the ratio of the last intron from *BRCA1* shows little change (Fig EV3J). Loss of exosome components neither affects the splicing of the last introns nor the RNA levels of PAP and PABPN1 (Fig EV3J–K). These results indicate that the terminal intron retention upon PABPN1 knockdown is not caused by the suppression of RNA degradation. To further assess PAP contribution in siPABPN1-induced last intron retention, we analyzed previously published RNA-seq data under PAP knockdown in 293T cells (data ref: Bresson *et al*, 2015b). We found that knocking down PAP alters the splicing of 24 introns under  $|\Delta IR|$  (NC - siPAP) cutoff at 0.1 and FDR cutoff at 0.01 (Fig EV3L). We observed that one-third (8 out of 24) of PAP-upregulated introns are also upregulated by PABPN1 (Fig EV3M), as shown in the common target gene *LARP7* (Fig EV3N). In consideration of only a few IR events induced by PAP knockdown, we reasoned that the RNAs with shortened polyA tails caused by PAP-KD might be missed during polyA RNA-seq library construction, or the PAP-KD effects are compensated by some noncanonical polyA polymerases in human cells (Schmidt & Norbury, 2010).

Next, to explore which domain of PABPN1 is responsible for its function in splicing, in referring to the functional domains of PABPN1 (Fig 3H), we constructed PABPN1 mutants with domain deletion or point mutation of important functional amino acids for specific domains (Fig 3I; Kerwitz *et al*, 2003). Under the condition of knocking down endogenous PABPN1, we rescued its expression with these mutants. In all cases (*LARP7*, *DUSP12*, and *BRCA1*), we found that the EI/EEJ ratio of the last intron is significantly reduced when rescuing the full-length (wildtype PABPN1) and N-terminal depleted mutants (including  $\Delta N11$ ,  $\Delta N99$ , and  $\Delta N113$ ), suggesting that the N-terminal domain of PABPN1 has no significant effect on promoting last intron splicing. In contrast, the YAFA (RNA binding domain, RRM mutated) and C-terminal depleted mutants ( $\Delta C57$ ) cannot rescue the intron splicing. Interestingly, the  $\Delta CC$  (coiled-coil domain), LALA (LALA mutation within CC domain in stimulating

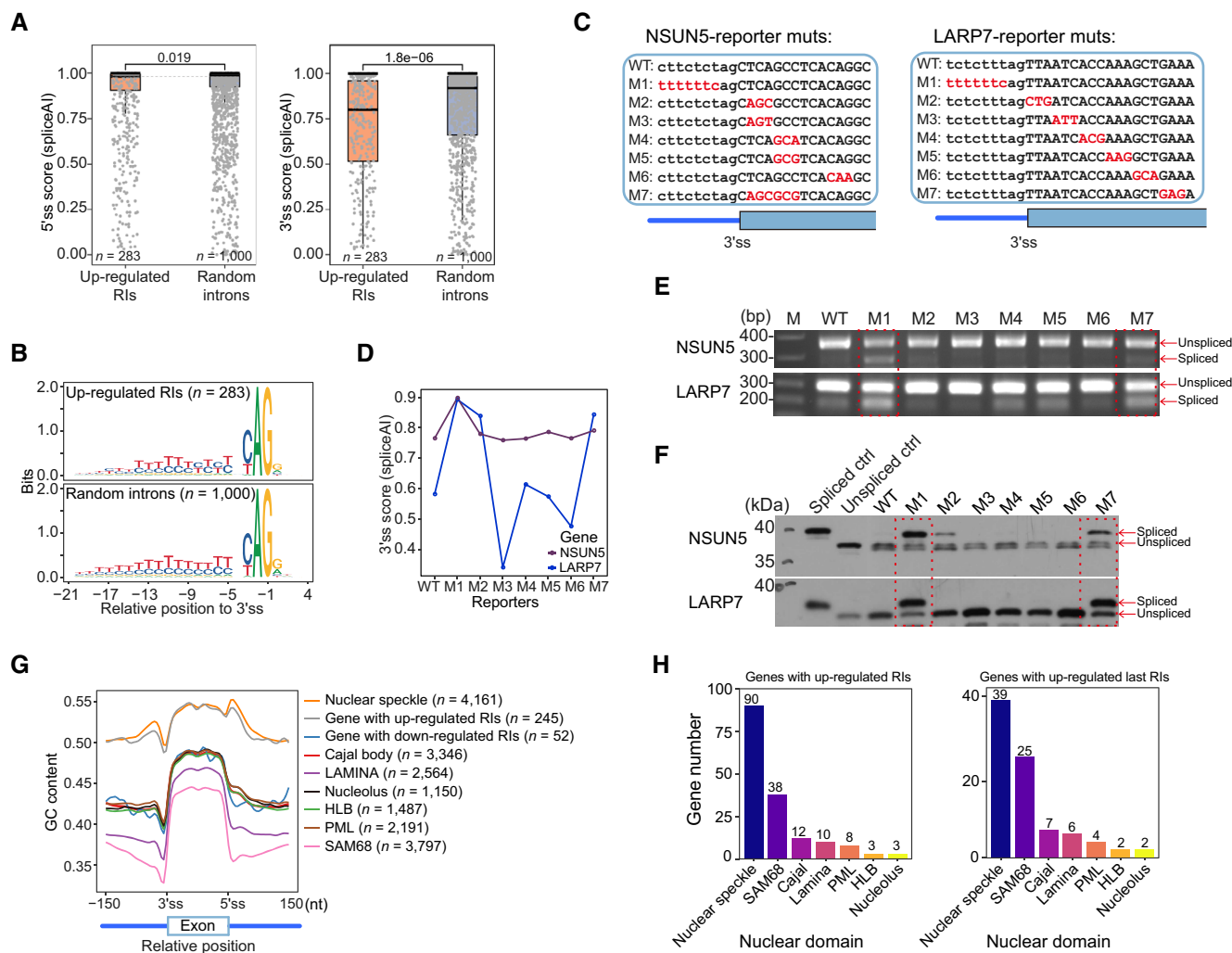
PAP), and  $\Delta CC + \Delta C21$  (combination of  $\Delta CC$  and  $\Delta C21$ ) have stronger rescue effects (Fig 3J). In sum, only the mutants that disrupt the interaction with PAP and the activity of PAP cannot rescue the splicing efficiency, indicating that PAP contributes to PABPN1-enhanced last intron splicing.

**PABPN1 loss-induced up-regulated retained introns have weaker 3' ss**

To understand how PABPN1 enhances splicing, we first tried to identify potential cis-features of PABPN1-regulated genes. Using spliceAI (Jaganathan *et al*, 2019), we analyzed the 5' and 3' splice site scores of IR-upregulated introns induced by PABPN1 knockdown and found that IR-upregulated introns have weaker 5' ss and 3' ss, especially the 3' ss than those of control introns randomly sampled based on the similar length and expression level (Fig 4A). Motif analysis of IR-upregulated introns and controls revealed a lower U but higher C content in the polypyrimidine tract region for IR-upregulated introns, in contrast to the canonical pattern for control introns (Figs 4B and EV4A). This abnormal C/U content might be one of the reasons explaining why the 3' ss of IR-upregulated introns are weaker. Furthermore, we divided these IR-upregulated introns into low ( $0.15 \leq IR \leq 0.3$ ) and high ( $IR > 0.3$ ) groups, and as expected, we found that introns with higher IR levels possess weaker 3' splice sites (Fig EV4B).

To verify whether weaker 3' ss could inhibit the splicing of PABPN1-enhanced last introns, we performed a series of base mutations (M1–M7) around the 3' ss of *NSUN5* (Fig 4C left) and *LARP7* (Fig 4C right) C4 reporters made in Fig 2A. Of note, the open reading frames were not changed by using synonymous mutations in the coding regions. The 3' ss scores of the mutants are predicted by spliceAI, and the mutant M1 shows the highest score (Fig 4D). Expectedly, the splicing of the M1 mutants replacing the original sequences from *LARP7* and *NSUN5* with the canonical 3' ss sequences is apparently enhanced at both the RNA level (Fig 4E) and the protein level (Fig 4F). Interestingly, we noticed that the 3' ss score of the mutant M7 from *LARP7* increases although the mutation site is 15 nt away, and consistently, the splicing of the intron is still enhanced.

Together, our data demonstrate that unilateral enhancing the 3' ss could compensate for the splicing defect caused by PABPN1



**Figure 4. Genomic features of PABPN1-depletion induced up-regulated retained introns (RIs).**

- A** Box plot showing the splicing score analysis of the 5' ss (left) and 3' ss (right) of PABPN1 loss-induced up-regulated retained introns (orange) and random last introns (gray) calculated by spliceAI.
- B** Consensus sequences of the PABPN1 loss-induced up-regulated retained introns (top) and randomly control introns (bottom) aligned at 3' ss (20 nt intron sequence and 3 nt exon sequence).
- C** Mutants construct near the 3' splice sites for *NSUN5* (left) and *LARP7* (right). All the mutants are based on the C4 reporters in Fig 2A.
- D** 3' splice site score of *NSUN5* (purple) and *LARP7* (blue) mutated reporters predicted by SpliceAI.
- E, F** Splicing analysis of *NSUN5* and *LARP7* reporter mutants by RT-PCR (E) and Western blot (F).
- G** The exon GC contents of PABPN1-regulated genes with up- or down-regulated RIs and other groups of genes located in different nuclear substructures including the nucleolus, nuclear speckle, nuclear LAMINA, Cajal body, HLB, PML, and SAM68.
- H** Distribution of PABPN1-regulated genes with up-regulated RIs (left) and those genes whose last intron is up-regulated RI (right) in nuclear substructures.

Data information: (A) The central line of the box plot represents the median value, and the lower and upper whiskers of the box represent the first and third quartiles, respectively. The upper whisker extends from the hinge to the largest value no further than  $1.5 \times \text{IQR}$  from the hinge (where IQR is the inter-quartile range). The lower whisker extends from the hinge to the smallest value at most  $1.5 \times \text{IQR}$  of the hinge. There are 283 up-regulated RIs and 1,000 random control RIs for splice score calculation by 5' ss and 3' ss, respectively. The P-values are calculated by the unpaired two-sided Wilcoxon test. See also Fig EV4.

Source data are available online for this figure.

depletion, implying that PABPN1 might contribute to the splicing of weak 3' ss.

### PABPN1-targeted genes are associated with nuclear speckle

The nuclear speckle is an important splicing hub and possesses a large number of splicing factors. Post-transcriptional splicing was

reported to occur in nuclear speckles (Girard *et al*, 2012), and recent studies suggest that proximity to nuclear speckles has a positive correlation with both co- and post-transcriptional splicing (Gordon *et al*, 2021). Recent studies by Barutcu *et al* have shown that nuclear speckle tends to enrich intron-retaining genes with high-GC content by systematic mapping transcripts associated with different nuclear domains (mainly including Nucleolus, Nuclear speckle,



Cajal bodies, Histone locus bodies [HLBs], PML, and SAM68 bodies; data ref: Barutcu *et al*, 2022b). We calculated the GC content of PABPN1-regulated genes and analyzed their distributions in different nuclear domains characterized by Barutcu *et al* (2022a). The GC content of genes in different nuclear domains is consistent with the original study, and the GC content of the transcripts associated with nuclear speckle is significantly higher than that of other nuclear domains (Fig 4G). Unexpectedly, the genes with up-regulated RIs (gray line) have a similar GC content to those enriched in nuclear speckle, while the genes with down-regulated RIs (blue line) have a relatively low-GC content (Fig 4G). Furthermore, not only the genes with any up-regulated RIs (Fig 4H left) but also those with up-regulated terminal RIs (Fig 4H right) are apparently enriched in the nuclear speckle, which composes 55 and 46% of genes, respectively. These data suggested that PABPN1-regulated genes with upregulated RIs are closely associated with the nuclear speckle.

### PABPN1 interacts with a large number of spliceosome proteins and RBPs

Next, to understand how PABPN1 promotes the last intron splicing, we first focus on investigating the PABPN1 interactome. To this end, we generated inducible HeLa cell lines expressing TurboID-eGFP-PABPN1 and took advantage of proximity labeling and immunoprecipitation combined with mass spectrometry to identify the interactome of PABPN1 (Fig 5A). We constructed three independent stable inducible HeLa cell lines (eGFP-PABPN1-TurboID-1/2; TurboID-eGFP-PABPN1), treated the cells with doxycycline for 48 h, and then labeled the cells with biotin for 30 min or 2 h according to the method developed by Branon *et al* (2018). We harvested cells and captured the proteins using streptomycin magnetic beads (Fig EV5A). We performed mass spectrometry analysis for 10 labeled samples and 4 control samples without TurboID-PABPN1 and identified 207 significantly enriched proteins under a log<sub>2</sub> fold-change cutoff of 1 and an FDR cutoff of 0.05 (Fig 5B).

To characterize the PABPN1 interactome, we first compared the 207 proteins with the proteins associated with spliceosome complex (Hegele *et al*, 2012), RNA degradation complex (Schmid & Jensen, 2019), and mRNA 3' end processing complex (Shi *et al*, 2009), and found that about 32.4% (79/244), 11.9% (5/42), and 31.8% (27/85) of the three complexes are identified in our mass spectrometry data (Fig EV5B), which are consistent with the known functional roles of PABPN1. Next, we performed GO-term analysis for the PABPN1 interacting proteins and found that they are significantly associated with RNA splicing and spliceosome complex (Fig EV5C). Particularly, many PABPN1 interacting proteins are associated with multiple U2-type complexes, in addition to the mRNA 3' end processing complex (Fig EV5D).

To further examine the interaction between PABPN1 and U2 complex proteins, we performed immunoprecipitation of endogenous PABPN1 and found that U2 snRNP-related proteins (U2AF2, U2AF1, and SF3B3) interacted with PABPN1 and the interactions are RNA independent. As a positive control, we also identified that the mRNA 3' end processing factors, such as CPSF7 and NUDT21, can interact with PABPN1 (Fig EV5E). To investigate whether PABPN1 regulates last intron splicing is dependent on the U2 complex, we treated HeLa cells with pladienolide B (Pla-B), an inhibitor of SF3B (Kotake *et al*, 2007), which is a critical component of U2

snRNP (Sun, 2020). We found that both the endogenous regulated genes (*LARP7* and *COPS5*) and the reporter minigenes (GFP-LARP7/COPS5) show repressed splicing of the last intron. In contrast, the nonregulated gene (*TDRD9*) shows no significant splicing changes under Pla-B treatment (Fig EV5F). These results demonstrate that PABPN1 interacts with U2 snRNP-related proteins, and U2 snRNP contributes to the splicing of PABPN1-enhanced last introns.

### RBM26&27 knockdown induces intron retention of PABPN1-regulated last RIs

Among all PABPN1 interacting proteins, we noted that RBM26&27 is highly enriched in all 8 groups of TurboID-labeled samples (Fig 5B). Considering that RBM26&27 are both RNA-binding proteins and have a potential role in regulating splicing, we knocked down RBM26/27 or both to investigate whether they contribute to the PABPN1-regulated last intron splicing. RBM26/27 knockdown does not affect the expression of PABPN1 (Appendix Fig S1A and B). Interestingly, knocking down both RBM26 and RBM27 significantly inhibits the splicing of the last introns for 5 PABPN1-regulated genes (*LARP7*, *COPS5*, *NSUN5*, *BRCA1*, and *DUSP12*), although individual knockdown of RBM26 or RBM27 only has a moderate effect on a subset of genes (Fig 5C). We further validated their effects using RT-qPCR to quantify the relative ratios of the last exon-intron junction (EIJ) over the exon-exon junction (EEJ) (Fig 5D), indicating that RBM26&27 can enhance the last intron splicing of PABPN1-regulated genes *in vivo*. In addition, to investigate whether RBM26&27 are also functional for reporters with mH2A 3' end (without endo-PAS), we constructed the MBP-RBM26/27 fusion proteins (Appendix Fig S1C) and tethered the fusion proteins to the *LARP7* and *COPS5* MS2 reporters. We found that the tethering of RBM26, RBM27, or both, can promote the splicing of the MS2 reporters (Fig 5E), which are not affected by PABPN1 knockdown (Appendix Fig S1D), suggesting that RBM26&27 might be the regulatory factors downstream of PABPN1. The results of polyA reporter assays further demonstrate this hypothesis as inhibiting RBM26, RBM27, or both leads to similarly increased levels of intron retention as PABPN1 knockdown (Fig 5F).

### PABPN1 interacts with RBM26&27 in promoting mRNA last intron splicing

To globally explore the regulated retained introns of RBM26&27, we analyzed previously published RNA-seq data under RBM26&27 knockdown in HeLa cells (data ref: Silla *et al*, 2020b) and found that loss of RBM26&27 induces 436 up-regulated retained introns and 67 down-regulated retained introns under  $|\Delta\text{IR}|$  (siRBM26&27 - NC) cut-off at 0.15 and an FDR cutoff at 0.01 (Appendix Fig S1E), suggesting an important role of RBM26&27 in regulating RNA splicing. Surprisingly, RBM26&27-KD induced up-regulated retained introns also show a strong 3'-end bias (69.3% [149/215] RIs being last introns), similar to PABPN1-KD, whereas ZC3H3 and ZFC3H1 knockdown induced RIs display no or very weak 3'-end bias (Fig 5G and Appendix Fig S1F and G). Considering that RBM26&27, ZC3H3, and ZFC3H1 were also identified as the RNA decay factors in the polyA tail exosome targeting (PAXT) complex (data ref: Meola *et al*, 2016b; Silla *et al*, 2020b), we evaluated to what extent RNA degradation affects PABPN1-regulated intron retention. By

comparing the up-regulated retained introns between PABPN1, RBM26 & 27, ZC3H3, and ZFC3H1 knockdown, we found that 213 retained introns are uniquely regulated by PABPN1, 56 RIs are co-regulated by PABPN1 and RBM26&27, and very few are co-targets of ZC3H3 and ZFC3H1 (Appendix Fig S1H and I). These results

indicate that RNA decay has only a limited effect on intron retention induced by PABPN1-KD, and RBM26&27 can promote the last intron splicing with PABPN1 independent of or prior to RNA degradation.

Next, we aimed to investigate the molecular mechanism of how PABPN1 interacts with RBM26&27. First, we constructed myc- and

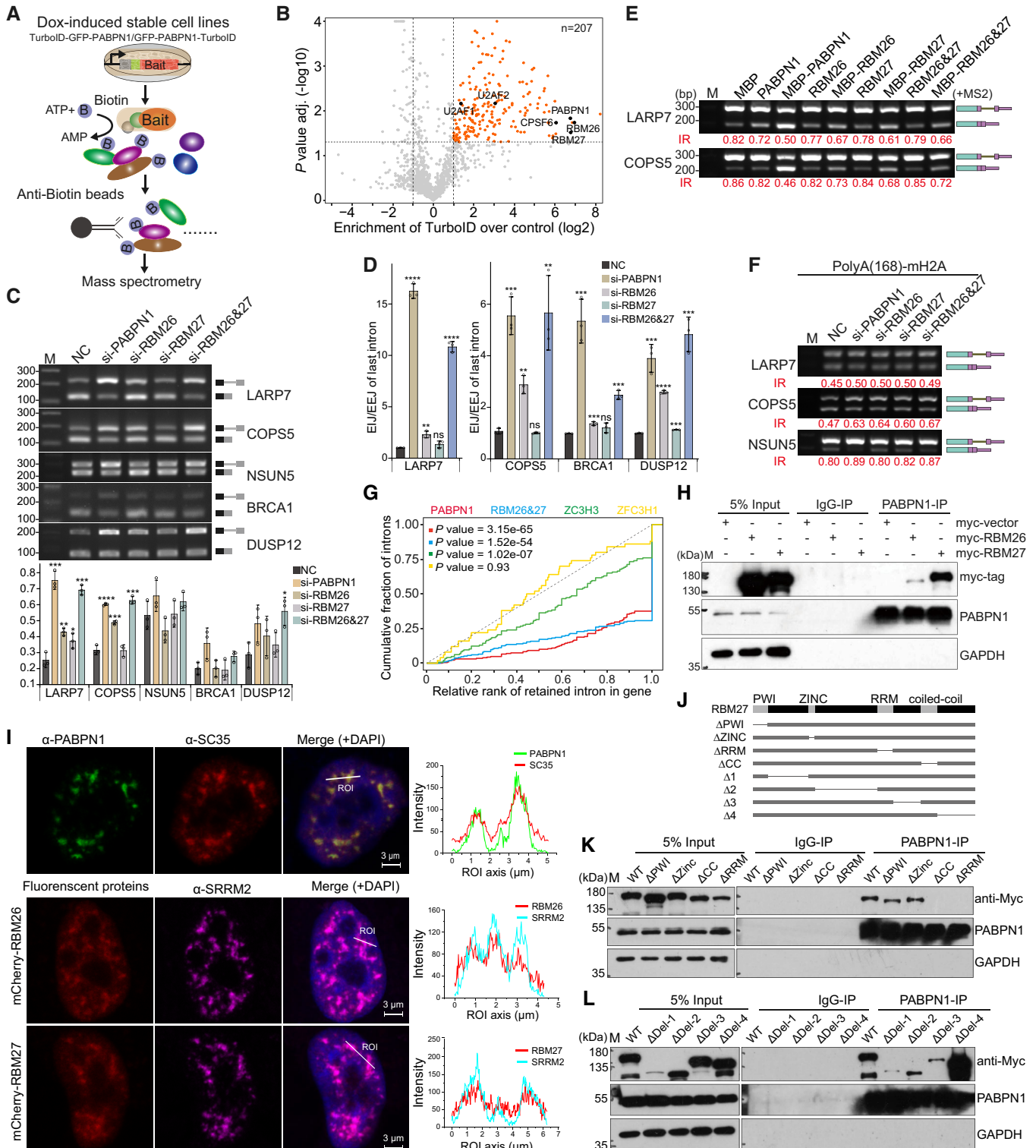


Figure 5.

**Figure 5. Identification of PABPN1 interactome and RBM26&27 knockdown induces intron retention of PABPN1-regulated last retained introns (RIs).**

- A Flowchart of TurboID-GFP-PABPN1 proximity labeling coupled mass spectrometry (MS) assay.  
 B Volcano plots of the enriched proteins identified by PABPN1 TurboID-MS. Significantly enriched proteins are colored in orange, and a few representative proteins are highlighted with labels.  
 C RT-PCR analysis of the last intron splicing of five representative PABPN1-dependent endogenous genes upon RBM26/27 knockdown with quantified IR values.  
 D Relative ratios of the unspliced (EI) to spliced (EE) isoforms of the last introns for *LARP7*, *COP55*, *BRCA1*, and *DUSP12* genes in HeLa cells treated in the same manner as in (C).  
 E Splicing analysis of MS2 reporters for two PABPN1-regulated genes *LARP7* and *COP55* when tethering RBM26/27 by RT-PCR.  
 F Splicing analysis of the polyA reporters for *LARP7*, *COP55*, and *NSUN5* under the condition of RBM26/27 knockdown.  
 G Cumulative frequency distributions of up-regulated RIs upon PABPN1 (red), RBM26&27 (blue), ZC3H3 (green), and ZFC3H1 (yellow) knockdown by their relative intronic rank in genes.  
 H Western blot validation of endogenous PABPN1 interacting with the myc-tagged RBM26 and RBM27.  
 I Representative immunofluorescence images showing the localization of endogenous PABPN1 relative to SC35, and the mCherry-tagged RBM26&27 relative to SRRM2 *in vivo*.  
 J Protein domains of RBM27 and the schematic diagrams of constructed myc-tagged RBM27 mutants.  
 K, L Western blot of endogenous PABPN1 immunoprecipitated myc-tagged RBM27 mutated proteins.

Data information: (C–D) Data are from  $n = 3$  biologically independent experiments. The error bars represent SD. The  $P$ -values are calculated by unpaired  $t$ -test (\*\*\*\* $P < 0.0001$ ; \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; \* $P < 0.05$ ; ns, nonsignificant). See also Fig EV5 and Appendix Fig S1. Source data are available online for this figure.

mCherry-tagged RBM26/27 fusion proteins (Appendix Fig S1J) and performed endogenous PABPN1 immunoprecipitation. We found that PABPN1 can strongly interact with RBM27 and have a relatively weak interaction with RBM26 (Fig 5H). Consistent with its association with the nuclear speckle above, endogenous PABPN1 locates in the nuclear speckle, while mCherry-RBM26 or RBM27 also shows great colocalization with the marker protein SRRM2 of the nuclear speckle (Fig 5I). These results imply that RBM26&27 might closely interact with PABPN1 in nuclear speckles.

To further dissect the interaction site in RBM27 that has strong interaction with PABPN1, we constructed a series of RBM27 deletion mutants tagged with myc (Fig 5J). By endogenous PABPN1 immunoprecipitation, we found that the coiled-coil (CC) and the RNA binding (RRM) domains of RBM27 mediate the interaction between PABPN1 and RBM27 (Fig 5K and L). We note that the level of  $\Delta$ Del-1 mutant is much lower than other mutants, probably due to its low stability. Collectively, these results demonstrated that PABPN1 might directly recruit RBM27 via its coiled-coil and RNA binding domains in enhancing the splicing of 3' terminal introns.

**The pattern of PABPN1-regulated last intron splicing is conserved in mice**

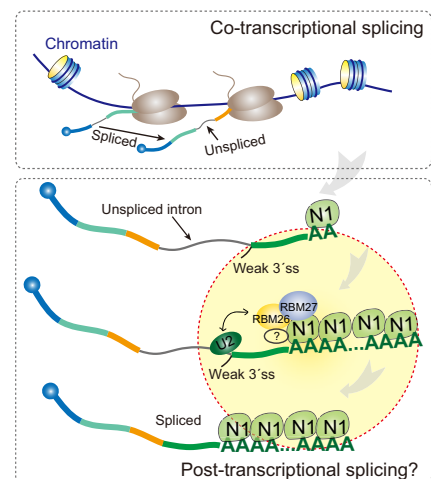
In addition to human cells, we also want to know whether the polyA tail-PABPN1 axis regulates splicing in other species. We analyzed recently published RNA-seq data of *Pabpn1*-knockout (*Pabpn1*-KO) in mouse oocytes (data ref: Zhao et al, 2021) and found that loss of *Pabpn1* also induced splicing dysregulation of many genes (Appendix Fig S2A). As the representative example, the last intron of *Hsd17b8* is retained upon *Pabpn1*-KO in mice (Appendix Fig S2B). Unexpectedly, among all *Pabpn1* up-regulated retained introns, about 52.6% (41/78) are the 3' terminal introns (Appendix Fig S2C), showing a strong 3'-end bias as observed in human cells, although the regulated genes are largely different. These data demonstrate that the regulatory pattern of PABPN1 in promoting last intron splicing is a conserved mechanism.

In summary, we find a group of last introns with weak 3' splice sites are post-transcriptionally spliced in at least subsets of transcripts with polyA tails, considering that binding of PABPN1 to

polyA tail enhances the splicing of the last introns. The unspliced RNAs generally localize in the nuclear speckle and possess long polyA tails, which would promote PABPN1 binding. PABPN1 recruits RBM26&27 or other factors to probably enhance the assembly of U2 spliceosome on the weak 3' splice site, which enables splicing of the last intron, as illustrated in the model (Fig 6). These results uncover a novel mode of post-transcriptional splicing enhanced by the polyA tail-PABPN1 axis.

**Discussion**

The mRNA polyA tail is the most prevalent process on mRNAs, as almost all mRNAs are polyadenylated. Although the function of the

**Figure 6. Graphical model for polyA tail-PABPN1 axis in enhancing last intron splicing of specific RNAs.**

Within the nucleus, many introns are spliced co-transcriptionally, while certain introns remain unspliced with 3'-end already being polyadenylated. The added polyA tail can recruit nuclear polyA binding protein PABPN1 (N1), which may further cooperate with RBM26&27 to facilitate 3' splice site (ss) recognition by U2 snRNP, and thus functions as an enhancer for the post-transcriptional splicing of last introns with weaker 3' ss.

mRNA polyA tail is well studied in the cytosol, how the polyA tail regulates gene expression in the nucleus is largely unknown. In this study, we discovered that the polyA tail acts as a splicing enhancer for the last introns via the nuclear polyA binding protein PABPN1. PABPN1 binds to the polyA tail and interacts with many spliceosome and RNA-binding proteins, specifically recruiting RBM26&27 to promote the splicing of many last introns with a weak 3' splice site.

### Functions of mRNA polyA tail

The function of the mRNA polyA tail is investigated more in the cytosol, where the polyA tail is bound by PABPCs. PABPC1 enables the rapid circulation of the ribosome and promotes translation by promoting the polyA tail and 5' cap to form a closed-loop mRNP structure (Munroe & Jacobson, 1990; Preiss & Hentze, 1998; Kapp & Lorsch, 2004). In addition, the stability of polyA RNAs is also considered to be closely related to the polyA tail. Because the mRNA 3' to 5' degradation begins with rapid deadenylation of the polyA tail (Goldstrohm & Wickens, 2008), PABPC1 binding to the polyA tail can resist deadenylation to protect mRNA from degradation. However, recent insights from polyA tail sequencing methods showed that the function of polyA tail on mRNA expression is not completely consistent with previous knowledge, where highly expressed mRNAs frequently possess shorter polyA tails (Lima et al, 2017; Tudek et al, 2021). Therefore, the full picture of the long polyA tail on RNA translation and stability is still unclear.

In contrast to the cytoplasmic functions above, the function of polyA tail in the nucleus has been rarely reported, except for one study reporting that exogenous insertion of polyA sequences can promote the transporting of short transcripts to the cytoplasm (Fuke & Ohno, 2008), and another one implying that hyperpolyadenylation (400 nt) of nuclear RNAs might link to RNA decay (Bresson & Conrad, 2013). Here, our results have linked the polyA tail to RNA splicing by using the modified reporter assay of Muniz et al (2015). Using different polyA sequence reporters, we investigated the effect of polyA tail length on mRNA splicing and found that longer polyA enabled more efficient splicing. We showed that the polyA tail preferentially enhanced the last intron splicing on reporters, and within a certain range, the longer the polyA tail, the more effective the splicing of the last introns. Of note, currently, we could only insert a polyA sequence up to 168 nt, and even longer polyA tails need further investigation. In addition to reporter assays on individual genes, our genome-wide RNA profiling data upon PABPN1 knockdown or fast degradation provide global evidence supporting that 3'-end polyA tail can promote intron splicing, which has enriched and expanded the previous findings from West lab (Muniz et al, 2015). We think it also merits further studies to understand the functional roles of polyA tails in coupling different RNA processing events in the nucleus and the cytoplasm, and even during RNA export.

Workman et al (2019) reported the phenomenon that a set of transcripts with retained introns tend to have longer polyA tails, probably due to those transcripts are restricted in the nucleus and thus avoiding deadenylation in the cytoplasm. Based on our data, we interpreted that longer polyA tails may recruit more PABPN1, which helps to enhance the splicing of the last retained introns. However, it is difficult to dissect the causal relationship between

intron retention, polyA length, and polyA tail-PABPN1 enhanced splicing, which warrants deep investigation in the future.

### PolyA tail and PABPN1 in splicing regulation

In this study, we find that polyA tail and PABPN1 can function as a splicing activator of the last introns by interacting with RBM26&27 (Fig 6). Interestingly, RBM26&27 also acts in the polyA tail exosome targeting (PAXT) complex to induce RNA decay (Silla et al, 2020a). However, we found that other PAXT factors except for RBM26&27, such as ZFC3H1 and ZC3H3, might not regulate PABPN1-targeted last introns splicing. This indicates that the “polyA tail-PABPN1-RBM26&27” axis could have multiple roles in nuclear pre-mRNA processing.

How the polyA tail facilitates last intron splicing via PABPN1 seems to be complex, and our results suggest that PABPN1 may exert different functions by recruiting different protein factors, especially for the introns not regulated by RBM26&27. Our TurboID-MS data show that PABPN1 is closely associated with spliceosome proteins, especially U2 spliceosome (Fig EV5). Interestingly, PABPN1 interacts with U2AFs and U2-snRNP representing the first step of splicing, and also interacts with PRP19 complex and U4/U6-U5 tri-snRNP complex representing further catalytic steps of splicing (Wahl et al, 2009). These data imply that PABPN1 is physically closed to the spliceosome through the splicing cycle, and there could be crosstalk between the polyA tail and spliceosome mediated by PABPN1. The detailed mechanism of PABPN1 in facilitating splicing, including how the PABPN1 or PABPN1-RBM26&27 complex communicates with U2 spliceosome, remains to be dissected by additional studies in the future.

### PABPN1, nuclear speckle, and post-transcriptional splicing

The nuclear speckle is a subcellular domain that has longtime been supposed as the place where post-transcriptional splicing occurs (Girard et al, 2012). We uncovered that the genes with the last retained introns regulated by PABPN1 showed similar features with nuclear speckle-associated mRNAs and endogenous PABPN1 perfectly overlapped with nuclear speckle (Fig 5I). Considering that the polyadenylated RNAs must be transcriptional-finished RNAs and the core layer of nuclear speckle is not fluid enough for spliceosome assembly (Fei et al, 2017), we speculated that PABPN1-enhanced splicing belongs to post-transcriptional splicing and it takes place in the periphery of nuclear speckles. What's more, the last-intron-retained genes identified in our work might be ideal cases to explore the function of nuclear speckles in mRNA post-transcriptionally splicing.

## Materials and Methods

### Cell culture, transfection, RNAi, and pladienolide B treatment

HeLa cells (free of mycoplasma contamination) were cultured in DMEM (Gibco) supplemented with 10% FBS (Hyclone) and penicillin-streptomycin (Gibco) at 37°C with 5% CO<sub>2</sub>. Lipofectamine 2000 (Invitrogen) was used for DNA transfection; siRNA transfection was carried out with Lipofectamine RNAiMax (Invitrogen)

following the manufacturer's protocol. Pladienolide B (Pla-B) was a gift from Dr. Peng Du, Peking University, and the cells were treated with 10, 25, 50, and 100 nM Pla-B for 24 h. The siRNA sequences and the targets are listed in Table EV1.

### Minigene construction

All minigene vectors were constructed based on the backbone of pEGFP-C1. For endogenous minigenes, the sequence of the retained intron and its flanking exons and the endogenous polyA signal were inserted downstream of EGFP-CDS. For mH2A 3' end and Malat1 3' end minigenes, the SV40-polyA signal region in pEGFP-C1 was replaced with mouse H2A mRNA 3' end or Malat1 3' end, and the sequence of the retained intron and its flanking exons without endogenous polyA signal was inserted in-frame downstream of EGFP-CDS. For spliced control vectors, spliced exon-exon junction sequence up to the stop codon was inserted in-frame downstream of EGFP-CDS. For intronic control vectors, intronic sequence up to the predicted stop codon was inserted in-frame downstream of EGFP-CDS. Intron-containing sequences were amplified from human genomic DNA, and exon-exon junction sequences were amplified from cDNA with KOD-Plus-Neo enzyme (TOYOBO). The primers are listed in Table EV2.

### MS2 tethering and artificial polyA reporters

Reporters for MS2 tethering assay and artificial polyA assay were constructed based on mH2A-3' end and Malat1 3' end minigenes. 5× MS2 SL or artificial polyA sequence with distinct lengths was inserted between the minigene region and mH2A/Malat1 3' end region. 5× MS2 sequence was synthesized by Tsingke and then cloned into the vector by double digestion and ligation. Distinct artificial polyA sequences were cloned by Golden Gate Assembly with Type IIS restriction endonucleases BbsI. First, the 35 A sequence containing 5' BbsI site was inserted by double digestion and ligation. Then, a second 35 A sequence was inserted by BbsI digestion and ligation to construct the reporter with 68 A. Similarly, the reporters with 100, 135, and 168 A were constructed.

### RNA isolation, reverse transcription, and PCR analysis

Total RNA was extracted with TRIzol (Invitrogen) and treated with TURBO™ DNase (Invitrogen) to remove genomic DNA. cDNA was synthesized from 1 µg of total RNA with random primer using M-MLV reverse transcriptase (Promega). According to the manufacturer's instructions, semi-quantitative PCR was carried out with KOD-Plus-Neo DNA Polymerase (TOYOBO). Quantitative PCR was performed using Hieff UNICON® qPCR SYBR Green Master Mix (Yeasen). The primers are listed in Table EV2.

### Western blotting, immunofluorescence, and antibodies

HeLa cells transfected with siRNA or plasmid were harvested 48 or 24 h after transfection, respectively. Cells were lysed with SDS-PAGE loading buffer and boiled for 10 min at 95°C. Western blotting was performed according to the standard protocol. For immunofluorescence staining, cells were fixed with 4% PFA, permeabilized with 0.1% Triton X-100 in 1× PBS, blocked with 1% BSA in 1× PBS, and

hybridized with primary and secondary antibodies sequentially supplied in 1× PBS containing 0.5% BSA in room temperature for 1 h. Antibodies used in this study include PABPN1 (rabbit, Abcam), SC35 (mouse, Lab own),  $\alpha$ -SRRM2 (rabbit, Abcam),  $\alpha$ -GFP (mouse, Proteintech), SF3B3 (rabbit, Abclon), U2AF2 (rabbit, Abclon), U2AF1 (mouse, Proteintech), PABPC1 (rabbit, Abclon), CPSF6 (rabbit, Proteintech), NUDT21 (mouse, Abclon), myc-tag (mouse, MBL), beta-actin (mouse, Proteintech), Alexa Fluor 488 Goat anti-rabbit (Invitrogen), Alexa Fluor 594 Goat anti-mouse (Invitrogen), Alexa Fluor 647 Goat anti-rabbit (Invitrogen), and mCherry (rabbit, Abclon).

### rRNA-depleted RNA-seq

Cells were harvested for total RNA extraction after 48 h siRNA transfection. The RNA integrity was evaluated by 1% agarose gel electrophoresis. The purity and quantity of RNA were evaluated by NanoDrop One. One microgram of total RNA was hybridized with 2.5 µg anti-rRNA DNA probes and then digested with RNaseH (Invitrogen). DNA probes were then degraded with TURBO™ DNase (Invitrogen), and the RNA was purified with RNAClean XP beads (Beckman). The rRNA-depleted RNA libraries were constructed following the NEBNext Ultra II Directional RNA Library workflow for NovaSeq PE150 sequencing.

### Proximity labeling and immunoprecipitation

The TurboID-EGFP-PABPN1 fusion proteins (doxycycline-inducible) were constructed based on the pLVX-Puro backbone. The TurboID tag was inserted at the C terminus or N terminus of (N)-EGFP-PABPN1 (-C) to construct EGFP-PABPN1-TurboID or TurboID-EGFP-PABPN1 fusion proteins. Cells expressing TurboID fused PABPN1 were screened by a lentiviral packaging system. Monoclonal cells were picked, cultured, and tested by Western blotting and immunofluorescence. Monoclonal cells were seeded in 15 cm dishes and proliferated until 85% confluence with doxycycline induction. Biotin solution was added into freshly supplied medium to 500 µM working concentration and the cells were cultured for 30 or 120 min for labeling. The labeled cells were scraped, and the nucleus fraction was separated and reserved. The MNase was added to degrade DNA and RNA to release all labeled proteins. The MNase-treated nucleus suspension was lysed with RIPA buffer, and the supernatant was acquired for immunoprecipitation.

For co-immunoprecipitation, cells expressing myc-tagged RBM26/27 were harvested at 48 h post-transfection. Cell pellets proliferated in one 60 mm dish were resuspended with 0.5 ml IP lysis buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, and newly added protease inhibitor) and then incubated on ice for 30 min. The suspensions were then sonicated for 5 cycles (30 s/30 s) in high-power mode (Bioruptor Plus sonication device) at 4°C, and centrifugated at 16,000 g for 30 min. The supernatants were then harvested for Co-IP by incubating with 1.5 µg anti-PABPN1 or anti-rabbit IgG (Proteintech, 10284-1-AP) for 2 h at room temperature. Then, 25 µl washed protein G magnetic beads (Thermo Scientific) were added. After 1.5 h incubation at room temperature, the beads were washed with lysis buffer 3 times for 10 min each. Finally, the beads were boiled with protein loading buffer followed by western blotting.

### Fast degradation of PABPN1 by dTAG

The sgRNAs targeting near the start codon of PABPN1 were designed through the website (<https://chopchop.cbu.uib.no/>) developed by Zhang's lab. The synthetic oligos were inserted into the PX330 CRISPR plasmid. The puromycin-T2A-FKBP12<sup>F36V</sup>-3x flag fragment was cloned from the original vector provided by Cheng lab, and the sequences of the upstream and downstream homologous arms were cloned from the HeLa genomic DNA. The donor vector was constructed by ligating the above fragments on the T vector in the order of "upstream arm, FKBP12<sup>F36V</sup>, downstream arm". Next, the sgRNA and donor vectors were transfected into HeLa cells in a 6-well plate in a ratio of 1:2, which were transferred into a 10 cm dish after 48 h. Puromycin was added to the dish with a final concentration of 1.5 µg/ml when the cells were attached to the dish. After 48 h, a fresh medium with puromycin was replaced. The single-cell clones were isolated from the remaining cells not killed by puromycin for proliferation. An appropriate number of cells were taken for Western blot verification to select the correct knock-in cell lines. The dTAG13 with a concentration of 120 nM dTAG13 was added to the cells for fast degradation of PABPN1.

### RNA-seq data analysis

The raw pair-end reads were preprocessed to remove the adapters with cutadapt (v.3.5) (Martin, 2011). The quality of the resulting reads was evaluated using FastQC (v.0.11.9) (de Sena Brandine & Smith, 2019). Reads were firstly mapped to rRNA sequences from NCBI, and those reads not mapped to rRNA sequences were aligned to the human genome (hg38) with GENCODE v32 gene annotation using STAR aligner (v.2.7.9a) (Dobin & Gingeras, 2015).

### Calculation of intron retention and definition of "gene last intron"

To estimate the intron retention level, we used IRFinder (v.1.3.1) (Middleton et al, 2017) to calculate the IR ratio and used the "analysisWithLowReplicates.pl" script to perform differential analysis under  $|\Delta\text{IR ratio}|$  cutoff of 0.15 and a *P*-value cutoff of 0.01. To call retained introns in transfecting NC siRNA samples, we required the IR ratios in two negative control samples to be over 0.35. To define the position or rank of the retained introns, the software Salmon (Patro et al, 2017) was used to quantify the expression of all annotated isoforms per gene, of which the most abundant isoform with more than half of the total expression level was defined as the major isoform. The intron ranks were calculated from the 3'- to 5'-end direction in major isoforms. The last intron in the 3'-end of the major isoform per gene was defined as the "gene last intron".

Fisher's exact test was used to evaluate the statistical significance for the overlapping of a Venn diagram on the  $2 \times 2$  contingency table, using the number of protein-coding gene unique introns not in the diagram as the double negative set.

### Alternative polyadenylation analysis

First, we identified the polyA sites (PASs) from our published SLAM-seq data as previously described (Tang et al, 2022). Then, we calculated the PDUI (Percentage of Distal polyA site Usage Index) for RNA-seq data from PABPN1 knockdown and control samples. Briefly, the

mean coverage for each region between consecutive PASs was calculated; for each PAS, we defined its mean coverages of upstream and downstream regions as  $\text{Cov}_{\text{up}}$  and  $\text{Cov}_{\text{down}}$ , respectively. The mean coverage from the 5' terminal exons to the closest pPAS was defined as  $\text{Cov}_{\text{first}}$ . The dPAS was defined as the most 3' PAS with the mean coverage of its 200-nt downstream region ( $\text{Cov}_{\text{down}}$  for dPAS)  $< 5\%$  of the coverage of the first 100 nt of the 5' terminal exon. An upstream PAS was defined as proximal PAS (pPAS) if its  $\text{Cov}_{\text{up}}/\text{Cov}_{\text{down}}$  ratio was over 1.2. For a specific pPAS, the PDUI was calculated using this formula:  $\text{PDUI} = \text{Cov}_{\text{down}}/\text{Cov}_{\text{first}}$ . If there were more than one pPAS for a gene, only the most significant one (with the minimum FDR) was used as the PDUI of this gene. Fisher's exact test was used to determine the *P*-value of PDUI difference between PABPN1 knockdown and control samples, which were further adjusted by the Benjamini-Hochberg procedure to compute the false-discovery rate (FDR). A gene with  $\text{FDR} \leq 0.01$  and  $|\Delta\text{PDUI}| \geq 0.15$  was considered as having APA between two conditions.

### Analysis of eCLIP-seq and Nanopore RNA-seq data

We processed ENCODE PABPN1 eCLIP-seq data (ENCSR820UYE) using the standard pipeline (Van Nostrand et al, 2016a). To estimate the polyA content in PABPN1 eCLIP and input reads, we counted the frequency of 8-mers in the Fastq file with jellyfish (v.1.1.12) (Marcais & Kingsford, 2011). Besides, we counted the number of reads ending with consecutive As with cutadapt (v.3.5) (Martin, 2011) and custom R script. The reads ending with 6 or more consecutive As were mapped to the human genome (hg38) using STAR aligner (v.2.7.9a) (Dobin & Gingeras, 2015) after trimming the polyAs.

We analyzed the association between last intron retention with polyA tail length using our Nanopore RNA-seq data published recently (Tang et al, 2022). Briefly, the Nanopore long reads were mapped to the human genome (hg38) using the minimap2 (v.2.17-r941) program. The polyA tail lengths of the mapped reads were calculated using Nanopolish (Workman et al, 2019), and the polyA length distributions of the last intron-containing reads from PABPN1 regulated and nonregulated genes were analyzed.

### Prediction of splice site scores

For a specific intron, we used spliceAI (Jaganathan et al, 2019) to calculate its 5' and 3' splice site scores by taking the intron sequence and its upstream 10 kb and downstream 10 kb as input.

## Data availability

The raw sequencing data from this study are deposited in the Genome Sequence Archive in BIG Data Center (CNGB-NGDC Members and Partners, 2022), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under the accession number: HRA002497 (<https://ngdc.cnbc.ac.cn/gsa-human/browse/HRA002497>). The mass spectrometry proteomics data have been deposited into the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD039864 (<https://www.ebi.ac.uk/pride/archive/projects/PXD039864>).

**Expanded View** for this article is available [online](#).

## Acknowledgments

This study was supported by grants from the National Natural Science Foundation of China [31922039 and 31871316]; Hubei Provincial Natural Science Foundation of China [2020CFA057 and 2020CFA017]; the Fundamental Research Funds for the Central Universities [2042022dx0003 and 2042020kf1069]. We thank the core facility of the College of Life Sciences at Wuhan University for providing services including mass spectrometry and confocal imaging. We thank the members of Zhou lab for insightful discussions during the process of this investigation. We acknowledge Dr. Peng Du for providing us with pladienolide B. Part of the computation in this work was done on the supercomputing system in the Supercomputing Center of Wuhan University.

## Author contributions

**Li Huang:** Conceptualization; data curation; formal analysis; investigation; visualization; methodology; writing – original draft; writing – review and editing. **Guangan Li:** Conceptualization; data curation; investigation; visualization; methodology. **Chen Du:** Data curation; formal analysis; investigation; visualization; methodology; writing – original draft. **Yu Jia:** Investigation. **Jiayi Yang:** Data curation; formal analysis. **Weiliang Fan:** Formal analysis; visualization. **Yong-Zhen Xu:** Resources; writing – review and editing. **Hong Cheng:** Validation; writing – review and editing. **Yu Zhou:** Conceptualization; resources; supervision; funding acquisition; validation; visualization; methodology; writing – original draft; project administration; writing – review and editing.

## Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

## References

- Antoniou M, Geraghty F, Hurst J, Grosveld F (1998) Efficient 3'-end formation of human beta-globin mRNA *in vivo* requires sequences within the last intron but occurs independently of the splicing reaction. *Nucleic Acids Res* 26: 721–729
- Anvar SY, Raz Y, Verway N, van der Sluijs B, Venema A, Goeman JJ, Vissing J, van der Maarel SM, t Hoen PA, van Engelen BG *et al* (2013) A decline in PABPN1 induces progressive muscle weakness in oculopharyngeal muscle dystrophy and in muscle aging. *Aging (Albany NY)* 5: 412–426
- Barutcu AR, Wu M, Braunschweig U, Dyakov BJA, Luo Z, Turner KM, Durbic T, Lin ZY, Weatheritt RJ, Maass PG *et al* (2022a) Systematic mapping of nuclear domain-associated transcripts reveals speckles and lamina as hubs of functionally distinct retained introns. *Mol Cell* 82: 1035–1052.e9
- Barutcu AR, Wu M, Braunschweig U, Dyakov BJA, Luo Z, Turner KM, Durbic T, Lin ZY, Weatheritt RJ, Maass PG *et al* (2022b) Gene Expression Omnibus GSE176439 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE176439>). [DATASET]
- Beaulieu YB, Kleinman CL, Landry-Voyer AM, Majewski J, Bachand F (2012) Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. *PLoS Genet* 8: e1003078
- Bentley DL (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol* 17: 251–256
- Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinikina T, Carr SA, Feldman JL, Perrimon N, Ting AY (2018) Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* 36: 880–887
- Bresson SM, Conrad NK (2013) The human nuclear poly(a)-binding protein promotes RNA hyperadenylation and decay. *PLoS Genet* 9: e1003893
- Bresson SM, Hunter OV, Hunter AC, Conrad NK (2015a) Canonical poly(A) polymerase activity promotes the decay of a wide variety of mammalian nuclear RNAs. *PLoS Genet* 11: e1005610
- Bresson SM, Hunter OV, Hunter AC, Conrad NK (2015b) Gene Expression Omnibus GSE73678 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73678>). [DATASET]
- CNCB-NGDC Members and Partners (2022) Database resources of the National Genomics Data Center, China National Center for Bioinformatics in 2022. *Nucleic Acids Res* 50: D27–D38
- Colgan DF, Manley JL (1997) Mechanism and regulation of mRNA polyadenylation. *Genes Dev* 11: 2755–2766
- Cooke C, Alwine JC (1996) The cap and the 3' splice site similarly affect polyadenylation efficiency. *Mol Cell Biol* 16: 2579–2584
- Cooke C, Hans H, Alwine JC (1999) Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Mol Cell Biol* 19: 4971–4979
- Dobin A, Gingeras TR (2015) Mapping RNA-seq reads with STAR. *Curr Protoc Bioinformatics* 51: 11.14.11–11.14.19
- Dye MJ, Proudfoot NJ (2001) Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell* 105: 669–681
- Eliseeva IA, Lyabin DN, Ovchinnikov LP (2013) Poly(A)-binding proteins: structure, domain organization, and activity regulation. *Biochemistry (Mosc)* 78: 1377–1391
- Fei J, Jadalaha M, Harmon TS, Li ITS, Hua B, Hao Q, Holehouse AS, Reyer M, Sun Q, Freier SM *et al* (2017) Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *J Cell Sci* 130: 4180–4192
- Fong N, Bentley DL (2001) Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev* 15: 1783–1795
- Fuke H, Ohno M (2008) Role of poly (A) tail as an identity element for mRNA nuclear export. *Nucleic Acids Res* 36: 1037–1049
- Girard C, Will CL, Peng J, Makarov EM, Kastner B, Lemm I, Urlaub H, Hartmuth K, Luhrmann R (2012) Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat Commun* 3: 994
- Goldstrohm AC, Wickens M (2008) Multifunctional deadenylase complexes diversify mRNA control. *Nat Rev Mol Cell Biol* 9: 337–344
- Gordon JM, Phizicky DV, Neugebauer KM (2021) Nuclear mechanisms of gene expression control: pre-mRNA splicing as a life or death decision. *Curr Opin Genet Dev* 67: 67–76
- Gray NK, Collier JM, Dickson KS, Wickens M (2000) Multiple portions of poly (A)-binding protein stimulate translation *in vivo*. *EMBO J* 19: 4723–4733
- Hegele A, Kamburov A, Grossmann A, Sourlis C, Wowro S, Weimann M, Will CL, Pena V, Luhrmann R, Stelzl U (2012) Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell* 45: 567–580
- Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB *et al* (2019) Predicting splicing from primary sequence with deep learning. *Cell* 176: 535–548.e24
- Jenal M, Elkon R, Loayza-Puch F, van Haften G, Kuhn U, Menzies FM, Oude Vrielink JA, Bos AJ, Drost J, Rooijers K *et al* (2012) The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* 149: 538–553
- Kaida D (2016) The reciprocal regulation between splicing and 3'-end processing. *Wiley Interdiscip Rev RNA* 7: 499–511
- Kapp LD, Lorsch JR (2004) The molecular mechanics of eukaryotic translation. *Annu Rev Biochem* 73: 657–704

- Kerwitz Y, Kuhn U, Lilie H, Knoth A, Scheuermann T, Friedrich H, Schwarz E, Wahle E (2003) Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. *EMBO J* 22: 3705–3714
- de Klerk E, Venema A, Anvar SY, Goeman JJ, Hu O, Trollet C, Dickson G, den Dunnen JT, van der Maarel SM, Raz V et al (2012) Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Res* 40: 9089–9101
- Kotake Y, Sagane K, Owa T, Mimori-Kiyosue Y, Shimizu H, Uesugi M, Ishihama Y, Iwata M, Mizui Y (2007) Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat Chem Biol* 3: 570–575
- Kuhn U, Gundel M, Knoth A, Kerwitz Y, Rudel S, Wahle E (2009) Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. *J Biol Chem* 284: 22803–22814
- Lima SA, Chipman LB, Nicholson AL, Chen YH, Yee BA, Yeo GW, Collier J, Pasquinelli AE (2017) Short poly(A) tails are a conserved feature of highly expressed genes. *Nat Struct Mol Biol* 24: 1057–1063
- Mangus DA, Evans MC, Jacobson A (2003) Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol* 4: 223
- Maniatis T, Reed R (2002) An extensive network of coupling among gene expression machines. *Nature* 416: 499–506
- Marcais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27: 764–770
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17: 3
- Marzluff WF, Koreski KP (2017) Birth and death of histone mRNAs. *Trends Genet* 33: 745–759
- McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385: 357–361
- Meola N, Domanski M, Karadoulama E, Chen Y, Gentil C, Pultz D, Vitting-Seerup K, Lykke-Andersen S, Andersen JS, Sandelin A et al (2016a) Identification of a nuclear exosome decay pathway for processed transcripts. *Mol Cell* 64: 520–533
- Meola N, Domanski M, Karadoulama E, Chen Y, Gentil C, Pultz D, Vitting-Seerup K, Lykke-Andersen S, Andersen JS, Sandelin A et al (2016b) Gene Expression Omnibus GSE84172 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84172>). [DATASET]
- Middleton R, Gao D, Thomas A, Singh B, Au A, Wong JJ, Bomane A, Cosson B, Eyraas E, Rasko JE et al (2017) IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol* 18: 51
- Muniz L, Davidson L, West S (2015) Poly(A) polymerase and the nuclear poly(A) binding protein, PABPN1, coordinate the splicing and degradation of a subset of human pre-mRNAs. *Mol Cell Biol* 35: 2218–2230
- Munroe D, Jacobson A (1990) mRNA poly(A) tail, a 3' enhancer of translational initiation. *Mol Cell Biol* 10: 3441–3455
- Nabet B, Roberts JM, Buckley DL, Paulk J, Dastjerdi S, Yang A, Leggett AL, Erb MA, Lawlor MA, Souza A et al (2018) The dTAG system for immediate and target-specific protein degradation. *Nat Chem Biol* 14: 431–441
- Niwa M, Berget SM (1991) Mutation of the AAUAAA polyadenylation signal depresses invitro splicing of proximal but not distal introns. *Gene Dev* 5: 2086–2095
- Niwa M, Rose SD, Berget SM (1990) Invitro polyadenylation is stimulated by the presence of an upstream intron. *Gene Dev* 4: 1552–1559
- Otero LJ, Ashe MP, Sachs AB (1999) The yeast poly(A)-binding protein Pab1p stimulates *in vitro* poly(A)-dependent and cap-dependent translation by distinct mechanisms. *EMBO J* 18: 3153–3163
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14: 417–419
- Preiss T, Hentze MW (1998) Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast. *Nature* 392: 516–520
- Rigo F, Martinson HG (2008) Functional coupling of last-intron splicing and 3'-end processing to transcription *in vitro*: the poly(A) signal couples to splicing before committing to cleavage. *Mol Cell Biol* 28: 849–862
- Schmid M, Jensen TH (2019) The nuclear RNA exosome and its cofactors. *Adv Exp Med Biol* 1203: 113–132
- Schmidt MJ, Norbury CJ (2010) Polyadenylation and beyond: emerging roles for noncanonical poly(A) polymerases. *Wiley Interdiscip Rev RNA* 1: 142–151
- de Sena Brandine G, Smith AD (2019) Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Res* 8: 1874
- Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR 3rd, Frank J, Manley JL (2009) Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* 33: 365–376
- Shi M, Zhang H, Wu X, He Z, Wang L, Yin S, Tian B, Li G, Cheng H (2017) ALYREF mainly binds to the 5' and the 3' regions of the mRNA *in vivo*. *Nucleic Acids Res* 45: 9640–9653
- Silla T, Schmid M, Dou Y, Garland W, Milek M, Imami K, Johnsen D, Polak P, Andersen JS, Selbach M et al (2020a) The human ZC3H3 and RBM26/27 proteins are critics for PAXT-mediated nuclear RNA decay. *Nucleic Acids Res* 48: 2518–2530
- Silla T, Schmid M, Dou Y, Garland W, Milek M, Imami K, Johnsen D, Polak P, Andersen JS, Selbach M et al (2020b) Gene Expression Omnibus GSE131255 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131255>). [DATASET]
- Sun C (2020) The SF3b complex: splicing and beyond. *Cell Mol Life Sci* 77: 3583–3595
- Sun Y, Zhang Y, Aik WS, Yang XC, Marzluff WF, Walz T, Dominski Z, Tong L (2020) Structure of an active human histone pre-mRNA 3'-end processing machinery. *Science* 367: 700–703
- Tang P, Yang Y, Li G, Huang L, Wen M, Ruan W, Guo X, Zhang C, Zuo X, Luo D et al (2021) Gene Expression Omnibus GSE165742 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165742>). [DATASET]
- Tang P, Yang Y, Li G, Huang L, Wen M, Ruan W, Guo X, Zhang C, Zuo X, Luo D et al (2022) Alternative polyadenylation by sequential activation of distal and proximal PolyA sites. *Nat Struct Mol Biol* 29: 21–31
- Tudek A, Krawczyk PS, Mroczek S, Tomecki R, Turtola M, Matylla-Kulinska K, Jensen TH, Dziembowski A (2021) Global view on the metabolism of RNA poly(A) tails in yeast *Saccharomyces cerevisiae*. *Nat Commun* 12: 4951
- Vagner S, Vagner C, Mattaj JW (2000) The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev* 14: 403–413
- Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K et al (2016a) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* 13: 508–514
- Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K et al (2016b) Gene Expression Omnibus GSE77634 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77634>). [DATASET]
- Vest KE, Phillips BL, Banerjee A, Apponi LH, Dammer EB, Xu W, Zheng D, Yu J, Tian B, Pavlath GK et al (2017) Novel mouse models of oculopharyngeal



- muscular dystrophy (OPMD) reveal early onset mitochondrial defects and suggest loss of PABPN1 may contribute to pathology. *Hum Mol Genet* 26: 3235–3252
- Wahl MC, Will CL, Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* 136: 701–718
- Wahle E (1991) A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 66: 759–768
- Wahle E (1995) Poly(A) tail length control is caused by termination of processive synthesis. *J Biol Chem* 270: 2800–2808
- Workman RE, Tang AD, Tang PS, Jain M, Tyson JR, Razaghi R, Zuzarte PC, Gilpatrick T, Payne A, Quick J et al (2019) Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods* 16: 1297–1305
- Zhao J, Hyman L, Moore C (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* 63: 405–445
- Zhao LW, Zhu YZ, Chen H, Wu YW, Pi SB, Chen L, Shen L, Fan HY (2020) PABPN1L mediates cytoplasmic mRNA decay as a placeholder during the maternal-to-zygotic transition. *EMBO Rep* 21: e49956
- Zhao LW, Zhu YZ, Wu YW, Pi SB, Shen L, Fan HY (2021) Gene Expression Omnibus GSE174032 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE174032>). [DATASET]
- Zhao LW, Zhu YZ, Wu YW, Pi SB, Shen L, Fan HY (2022) Nuclear poly(A) binding protein 1 (PABPN1) mediates zygotic genome activation-dependent maternal mRNA clearance during mouse early embryonic development. *Nucleic Acids Res* 50: 458–472