nature structural & molecular biology

ARTICLES

https://doi.org/10.1038/s41594-021-00709-z

Check for updates

Alternative polyadenylation by sequential activation of distal and proximal PolyA sites

Peng Tang[1,2,6], Yang Yang[1,6], Guangnan Li[1], Li Huang[1], Miaomiao Wen[3], Wen Ruan[1], Xiaolong Guo[1], Chen Zhang[1], Xinxin Zuo[1], Daji Luo[4], Yongzhen Xu[1], Xiang-Dong Fu[5]✉ and Yu Zhou[1,2,3]✉

Analogous to alternative splicing, alternative polyadenylation (APA) has long been thought to occur independently at proximal and distal polyA sites. Using fractionation-seq, we unexpectedly identified several hundred APA genes in human cells whose distal polyA isoforms are retained in chromatin/nuclear matrix and whose proximal polyA isoforms are released into the cytoplasm. Global metabolic PAS-seq and Nanopore long-read RNA-sequencing provide further evidence that the strong distal polyA sites are processed first and the resulting transcripts are subsequently anchored in chromatin/nuclear matrix to serve as precursors for further processing at proximal polyA sites. Inserting an autocleavable ribozyme between the proximal and distal polyA sites, coupled with a Cleave-seq approach that we describe here, confirms that the distal polyA isoform is indeed the precursor to the proximal polyA isoform. Therefore, unlike alternative splicing, APA sites are recognized independently, and in many cases, in a sequential manner. This provides a versatile strategy to regulate gene expression in mammalian cells.

Alternative polyadenylation (APA) is a widespread phenomenon[1], which has been linked to the regulation of important cellular functions[2–6]. Dysregulation of APA has also been attributed to many pathological processes[7–10]. Based on the localization of polyA sites (PASs), APA can be classified into intronic polyadenylation or tandem polyadenylation. The former tends to produce different proteins[4,6,7,10], while the latter generates RNA isoforms with different lengths of 3′ untranslated region (UTR)[2,3,5,8,9]. Transcripts with different 3′ UTR lengths are thought to contribute to regulated gene expression at the posttranscriptional level by altering RNA stability, translation efficiency or cellular localization[2,3,5,8,9,11]. However, two recent genome-wide studies suggest that 3′ UTR appears to have marginal effects on RNA stability and translation[12,13], implying that APA isoforms may not be coupled with differential microRNA actions, as widely perceived.

The mechanism for the polyadenylation reaction has been well elucidated. The specificity and efficiency of the reaction are dictated by the coordinated interactions of multiple proteins[14]. In general, these factors form three subcomplexes: cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF) and cleavage factor (CFI and CFII), which respectively recognize the polyA signal (AAUAAA or its variants), downstream U/GU-rich element, and upstream UGUA motif[15]. Different PASs are composed of different combinations of these *cis*-acting elements, thus exhibiting different processing efficiencies[15,16]. Early studies on Ad2 RNA[17] and various endogenous genes in CHO and HeLa cells[18] show that the kinetics of polyA addition is ultrafast, usually occurring soon after the PAS is transcribed, which in many cases may even take place ahead of some upstream splicing events[17]. This implies that the 3′ end formation at different PASs may mostly take place in the first come-first served basis. Additionally, it was reported that RNA polymerase II is associated with 3′ end processing factors and facilitates 3′ end formation in vitro[19,20], which in turn tightly regulates transcription termination[16,21,22]. Together, all existing evidence appears to point to a general model for APA where multiple PASs may be efficiently presented to the 3′ end formation machinery during transcription to produce RNA isoforms with different 3′ UTRs, similar to the mutually exclusive recognition of competing splice sites during alternative splicing.

However, through a series of genome-wide experiments, we now uncover an unprecedented model for APA where transcripts may be processed in two phases: in the first phase, APA occurs independently at different PASs with efficiency dictated by the strength of each site during transcription elongation; in the second phase, the longer APA isoform ended at the distal PAS serves as the substrate for further processing at the proximal site.

Results

Tight association of RNA biogenesis with the nuclear matrix. To study the fate of RNA after transcription, we sought to isolate RNA from different cellular compartments (Fig. 1a). We first separated cytoplasmic RNA (CY) from the nuclei and then separated nascent RNA from those released into the nucleoplasm (NP) by treating isolated nuclei with high salt[23–25]. It has been widely assumed that the remaining RNA correspond to chromatin-associated RNA[23]. Because chromatin could be removed by extensive DNase I treatment[26,27], we asked whether it is possible to further differentiate chromatin-associated RNA from polyA$^+$ RNA tightly anchored in the nuclear matrix (NM)[28,29]. Unexpectedly, we detected little RNA after exhausted digestion of DNA, which was sufficient to quantitatively release Histone 3 (Fig. 1b and Extended Data Fig. 1a). Almost all hyper-phosphorylated Pol II (Pol IIo), which is associated with transcription elongation, was also tethered to the NM (Extended Data Fig. 1b), suggesting that nascent RNA along with elongating Pol II are readily assembled into ribonucleoprotein complexes during transcription to become part of NM. Therefore, the NM fraction

[1]State Key Laboratory of Virology, College of Life Sciences, Wuhan University, Wuhan, China. [2]Frontier Science Center for Immunology and Metabolism, Wuhan University, Wuhan, China. [3]Institute for Advanced Studies, Wuhan University, Wuhan, China. [4]State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Science, Wuhan, China. [5]Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, University of California, San Diego, La Jolla, CA, USA. [6]These authors contributed equally: Peng Tang, Yang Yang.
✉e-mail: xdfu@health.ucsd.edu; yu.zhou@whu.edu.cn

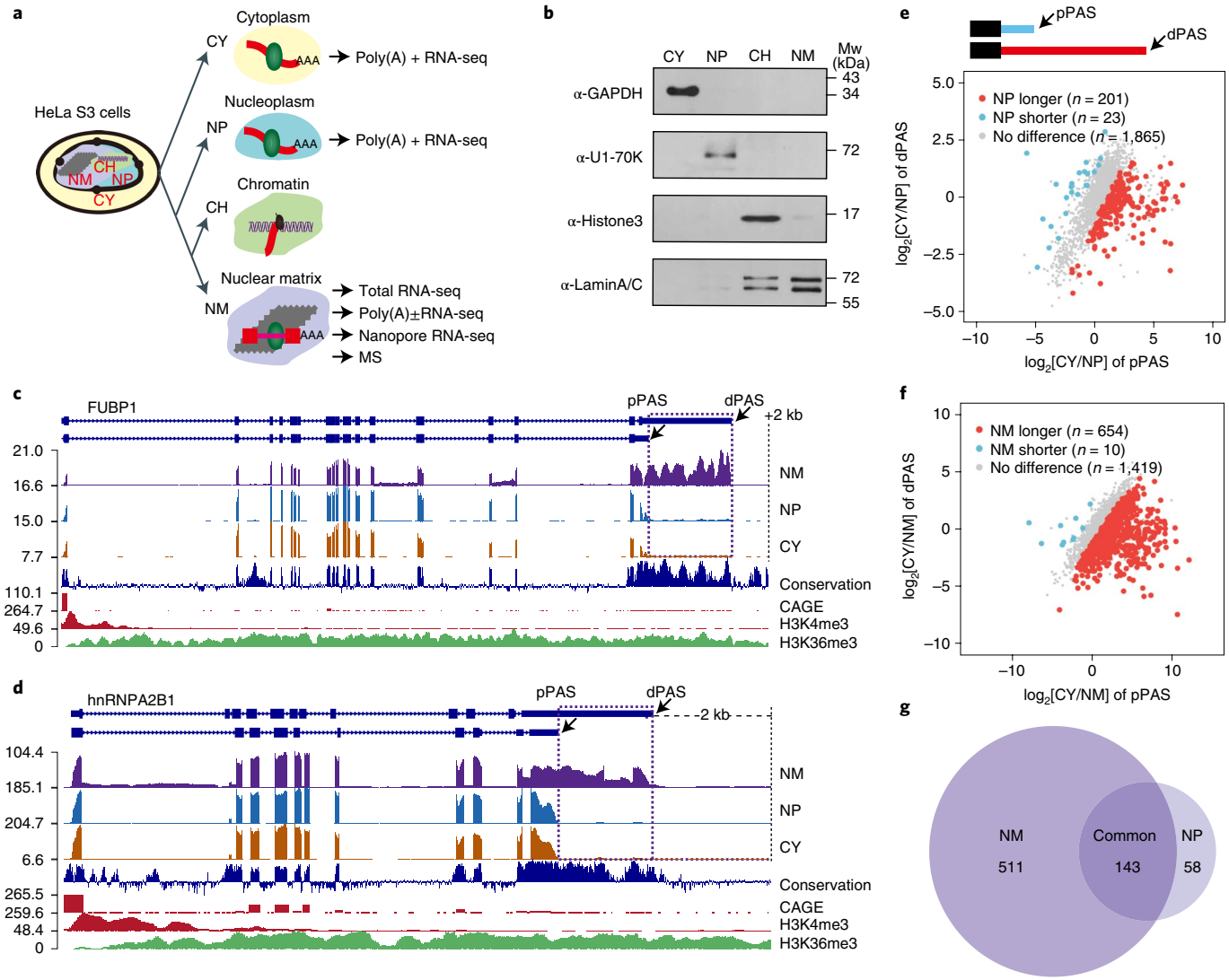

Fig. 1 | Isoforms with extended 3′ UTR are restricted in the NM. a, Flowchart of cellular fractionation and high-throughput profiling. HeLa S3 cells are first treated with hypotonic buffer to separate CY from the nuclei, then high salt buffer is used to separate the nuclear soluble component (NP) from others resistant to high salt. The obtained pellet is digested with DNase I in CSK buffer, from which the released fraction is chromatin (CH) and the remaining insoluble fraction is the NM. **b**, Immunoblot of the extracts from different fractions. Anti-GAPDH, anti-U1-70K, anti-Histone3 and anti-LaminA/C antibodies are markers for the CY, NP, CH and NM fractions, respectively. Data were from $n = 2$ independent experiments. **c**,**d**, UCSC genome browser tracks showing the enrichment of long 3′ UTR isoforms (dotted box) of FUBP1 (**c**) and hnRNPA2B1 (**d**) genes in the NM from polyA$^+$ RNA-seq. The genome conservation, CAGE-seq, H3K4me3 and H3K36me3 signal tracks from ENCODE project are included. Two arrows mark the proximal and distal polyadenylation sites, pPAS and dPAS, respectively. **e**,**f**, Global analysis of the relative enrichment of transcripts with different 3′ UTRs (dPAS versus pPAS) between different cellular fractions: CY versus NP (**e**) and CY versus NM (**f**), respectively. The genes with enriched long 3′ UTR in the NP (**e**) or NM (**f**), and CY are highlighted in red and blue, respectively. **g**, A Venn diagram showing the intersection of genes with enriched long 3′ UTR in the NP and NM fractions from **e** and **f**.

harbors both chromatin-associated RNA and those released from chromatin but still anchored in the NM.

Because little RNA was detectable in the chromatin fraction, we focused on analyzing RNA from CY, NP and NM. We first performed NM RNA-sequencing (RNA-seq) after depleting ribosomal RNA and noted obviously retained intronic signals in NM based on the completed splicing index (coSI) (Extended Data Fig. 1c). Furthermore, such NM-associated nascent RNAs display a 5′-to-3′ descending polarity in intron removal (Extended Data Fig. 1d) and a sawtooth pattern visible in many long introns (Extended Data Fig. 1e), characteristic of cotranscriptional splicing as reported early[24,25]. These observations suggest that NM-associated RNA represent a unique population of RNA that are tightly retained in the nucleus for both co- and posttranscriptional processing.

Retention of longer 3′ UTR isoforms in the NM. We were particularly interested in the critical features of already polyadenylated RNA in NM. Thus, we further performed polyA$^+$ RNA-seq for CY, NP and NM RNA (Fig. 1a). As expected, polyA$^+$ messenger RNAs with various partially spliced introns were specially enriched in NM (Fig. 1c,d), consistent with a report that polyadenylation can precede splicing and only completely spliced mRNAs are released into the NP before nuclear export[25]. We also noted numerous mRNAs with longer 3′ UTR were also preferentially retained in NM, as exemplified with *FUBP1* and *hnRNPA2B1* (Fig. 1c,d, dotted box). We next investigated the distribution of mRNA isoforms ended at the proximal PAS (pPAS) versus distal PAS (dPAS) in different cellular compartments. By analyzing the ratios of pPAS and dPAS usage in roughly 2,100 APA genes, we identified 201 (9.6%) and 654 genes

(31.3%) that had their mRNA isoforms with longer 3′ UTR (using the dPAS) selectively enriched in NP compared to CY (Fig. 1e, red dots) and in NM compared to CY (Fig. 1f, red dots), respectively, and very few genes showed the opposite trend (Fig. 1e,f, blue dots).

Furthermore, 143 out of the 201 (about 70%) genes with relative higher dPAS usage in NP also have relative higher dPAS usage in NM (Fig. 1g), as individually highlighted (Extended Data Fig. 2a). Meanwhile, 511 genes with relative higher dPAS usage are uniquely identified in NM (Fig. 1g), as illustrated in specific examples (Fig. 1c,d and Extended Data Fig. 2b). This pattern was unlikely caused by faster degradation of longer 3′ UTR isoforms in CY because most of these genes exhibited a higher percentage of longer 3′ UTR isoforms only in NM, but not in NP (Fig. 1c,d,g and Extended Data Fig. 2b). Based on the percentage of distal usage index (PDUI) (Extended Data Fig. 2c)[9], it became further evident that NM was enriched with a much larger population of longer 3′ UTR isoforms (Extended Data Fig. 2d). These data indicate that certain *cis*-acting elements in those mRNAs with longer 3′ UTR may mediate their active nuclear retention.

Association of 3′-end formation factors with NM polyA⁺ RNA. To understand the mechanism for retaining APA isoforms with longer 3′ UTR in NM, we determined the interacting proteome on NM polyA⁺ RNA. For this purpose, HeLa S3 cells were cultured in media containing 4sU overnight followed by ultraviolet (UV) 365-nm irradiation. NM was prepared from such treated cells and oligo-dT magnetic beads were then used to capture polyA⁺ RNA-associated proteins[30]. A set of proteins were captured in an UV 365 irradiation-dependent manner (Fig. 2a). Several canonical RNA binding proteins (RBPs), such as hnRNP C, hnRNP U and nucleolin, were captured while LaminA/C, lacking RNA binding capacity, were undetectable, as examined by western blotting (Fig. 2b). These data indicate the selectivity of our pulldown conditions.

We subjected the captured sample to mass spectrometric analysis and identified 281 putative RBPs, including most SR and hnRNP proteins known to be involved in various aspects of RNA metabolism (Supplementary Table 1). In particular, consistent with a large amount of polyA⁺ RNA associated with nuclear speckle[29], which is well known as part of NM[28], we found that roughly one-third (92/281) of NM-associated RBPs were also found in the published nuclear speckle proteome[31]. And roughly 55% (160/281) were identified to associate with Pol II[32] (Fig. 2c). These results indicate that many RBPs may bind cognate *cis*-acting RNA elements during transcription and may continue to influence the fate of mRNA posttranscriptionally, as envisioned earlier[29].

As expected, many identified proteins in NM are components of the spliceosome[33], including U2AF65, U2AF35, PRPF8 and so on (Fig. 2c), consistent with co- and posttranscriptional splicing in NM. We noted the enrichment of some core components of the 3′ end processing complex involved in direct RNA binding (Fig. 2d), including CFIm68 (CPSF6), which together with CFIm25 functions to recognize the UGUA motif upstream of the PAS[15,34], FIP1L1, a component of CPSF subcomplex for binding to the U-rich element flanking the PAS[15,35], and CstF64 (CstF2) responsible for recognizing U/GU-rich element downstream of the PAS[15,36]. The presence of these factors is unexpected as our oligo-dT beads are supposed to capture polyA⁺ mRNAs that have already gone through the polyadenylation reaction. This inspired us to hypothesize that longer 3′ UTR isoforms retained in NM may serve as substrates for the 3′ end processing machinery to posttranscriptionally generate shorter 3′ UTR isoforms.

Distal PASs are stronger than proximal PASs. Next, we examined the RNA-seq libraries from NM polyA⁺ and polyA⁻ RNA. As expected, NM polyA⁻ RNA have much more intronic reads than polyA⁺ RNA (Extended Data Fig. 3a). As shown on the representative examples, the NM polyA⁻ signals extend far beyond the dPAS of *FUBP1* and *hnRNPA2B1* (Extended Data Fig. 3b,c). Global analysis further showed that the polyA⁻ RNA signals are similar across both pPAS and dPAS, and in comparison, the polyA⁺ RNA signals show a decrease across both the pPAS and dPAS (Extended Data Fig. 3d). Notably, the signal decrease is more dramatic at dPAS relative to pPAS, consistent with predominant polyadenylation at dPAS in the NM fraction.

The strength of individual PASs may dictate the polyadenylation frequency at specific sites. To test this hypothesis, we constructed a bicistronic reporter to determine the strength of several pairs of pPAS and dPAS from several APA genes. In this reporter, the Firefly luciferase (Fluc) and the Renilla luciferase (Rluc) are driven with the same promoter, and Rluc's translation relies on an internal ribosome entrance site (IRES) (Extended Data Fig. 4a). The processing of a PAS, inserted between the two luciferase genes, would trigger transcription termination to weaken the downstream Rluc. Therefore, the expression ratio of Fluc over Rluc can be used to quantify the relative PAS strength. Testing the PASs from *FUBP1*, *hnRNPA2B1*, *MECP2* and *BDNF*, we observed a larger Fluc/Rluc ratio for dPAS except for *BDNF* (Extended Data Fig. 4b), suggesting the dPAS is stronger than pPAS in three out of four these APA genes.

In this luciferase reporter assay, insertion of different PASs will generate transcripts with different 3′ UTRs, which may influence the translation efficiency, thereby changing the Fluc/Rluc ratio. We thus used a more direct strategy to determine the PAS strength by constructing a set of reporters in which the pPAS and dPAS were both placed downstream of a green fluorescent protein (GFP) expression unit in a tandem fashion, but with different orders (Extended Data Fig. 4c). We found that all four APA genes tested above showed the same trend: when the pPAS was placed upstream of dPAS, both pPAS and dPAS were used, but when the order was reversed, the dPAS was exclusively used (Extended Data Fig. 4c,d). These results strongly indicate that most dPASs are stronger than pPASs, thus enabling cotranscriptional processing of dPASs, which then serve as precursors for posttranscriptional processing at competent pPASs.

Evidence for progressive PAS recognition. According to the sequential PAS recognition model posed above, we would anticipate the predominant use of dPASs in nascent RNA compared to steady-state RNA. Thus, to obtain global evidence for the model, we took advantage of the strategy in thiol (SH)-linked alkylation for metabolic sequencing of RNA (SLAM-seq)[37] to perform PAS-sequencing (PAS-seq). We first cultured cells in 4sU-containing media for 30 min and then extracted total RNA for treatment with iodoacetamide to attach the carboxyamidomethyl group to the thiol group in 4sU, which would induce thymine-to-cytosine (T-C) conversion during reverse transcription. This allows us to distinguish between nascent and preexisting RNA reads to compute pPAS and dPAS usages in nascent versus total RNA. As expected, the T-C conversion events were the most significant mutations and Pol II inhibition by 5,6-dichloro-1-beta-D-ribobenzimidazole (DRB) treatment specifically reduced the T-C conversion over other mutations or sequencing errors (Extended Data Fig. 5a). Moreover, DRB treatment had limited effects on mitochondrial RNAs, which are transcribed by the mitochondrial RNA polymerase (Extended Data Fig. 5b). Given highly reproducible PAS-seq libraries (Extended Data Fig. 5c), we combined the two replicates for further analyses.

We first examined our PAS-seq data on *FUBP1* and *hnRNPA2B1*, which showed selective enrichment of their longer 3′ UTR isoforms in NM (Fig. 1c,d). The dPAS was used more frequently compared to the pPAS in nascent RNA in both cases (Fig. 3a,b), which are consistently detected in biological replicates. Three adjacent dPASs were equally used compared to the pPAS in *hnRNPA2B1* (Fig. 3b), indicating that those three consecutive dPASs were each partially recognized during transcription. In contrast, the pPAS became the

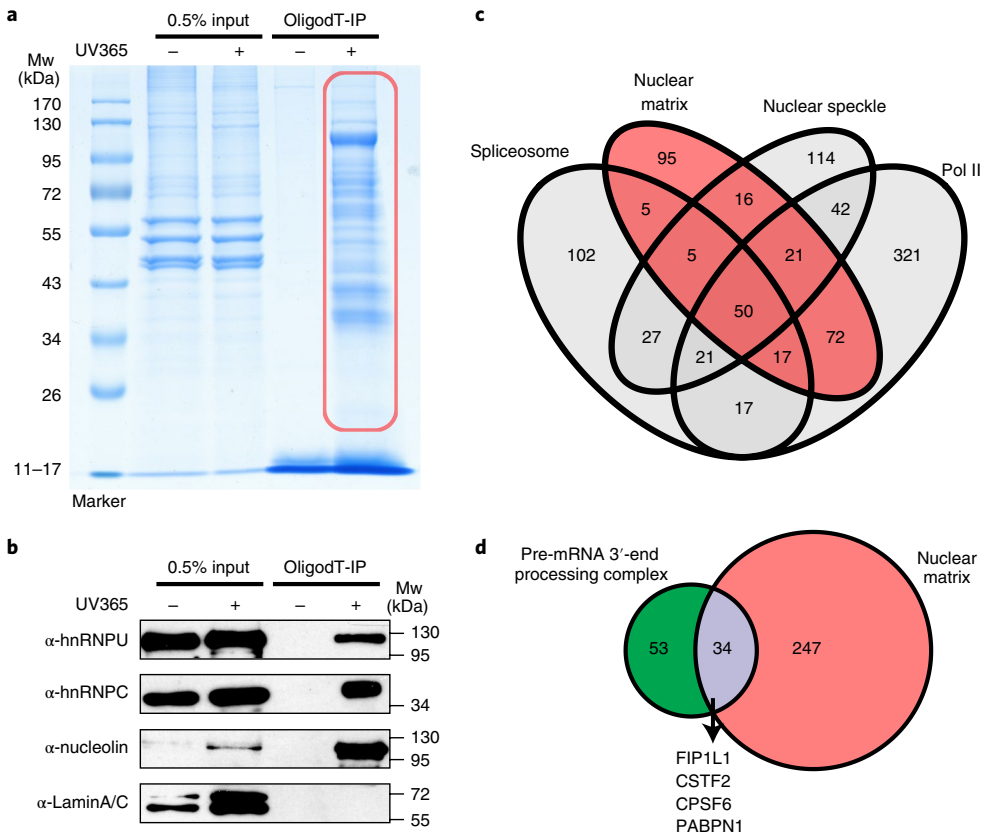

Fig. 2 | Characterization of the proteins associated with NM polyA⁺ RNA. a, Coomassie Brilliant Blue staining of captured proteins associated with polyA⁺ RNA in the NM. Data were from $n=1$ independent experiment. **b**, Western blotting of hnRNP U, hnRNP C, Nucleolin and Lamin A/C against captured proteins from the NM. Data were from $n=1$ independent experiment. **c**, A Venn diagram showing the intersection of proteins identified from four studies including our NM polyA⁺ RNA interactome, RNA Pol II interactome, nuclear speckle localized proteins and the spliceosome associated proteins. **d**, A Venn diagram showing the relationship between proteins in NM polyA⁺ RNA interactome and those in the 3′ end processing complex. The core 3′ end processing factors FIP1L1, CstF2, CPSF6 and PABPN1 are the common proteins directly binding to polyA⁺ RNA.

predominant one in total RNA in both cases. These specific events are representative of a global trend, as the dPAS/pPAS ratio (d/p ratio) of nascent RNA relative to steady-state RNA is the highest among genes with longer 3′ UTR enriched only in the NM fraction, as compared to other groups (genes with longer 3′ UTR enriched in both NM and NP or only in NP or genes without such enrichment) (Extended Data Fig. 5d). This does not seem to be caused by the modified Us, because cells cultured with 4sU did not distort the d/p ratios of PASs at the steady state compared with those in cells without 4sU treatment (Extended Data Fig. 5e). Together, these findings suggest a prevalent model where mRNAs using dPASs are selectively retained in the NM for further processing before nuclear export.

Limited stability difference for isoforms with varied 3′ UTRs. When the pPAS and dPAS pairs (APA pairs) were classified into four groups based on the d/p ratio at steady state and the difference in the d/p ratio between nascent RNA and total RNA (Fig. 3c), we found that the dPAS was more frequently used in nascent RNA compared to steady-state RNA ($\log_2$ fold-change of d/p ratio ≥ 1, nascent versus steady state) in 47.7% APA pairs (groups I and II, 1,022 out of 2,143 APA pairs). These 2,143 APA pairs are derived from 1,743 genes (Extended Data Fig. 5f), among which 850 (48.8%) genes have at least one APA pair belonging to group I or II (Extended Data Fig. 5g).

When transcription was inhibited with DRB, we found that the fold-change of the d/p ratio at the steady state for most genes was in a small range between −2 and 2 (median values of −0.45, 0.13, 0.29 and −0.35 for group I, II, III and IV genes categorized above, respectively, Extended Data Fig. 5h). These relatively small changes were unable to explain much larger d/p ratio differences between nascent and total RNA (cutoff as fold-change ≥ 2 or ≤ -2, Fig. 3c). Instead, we interpret these results to indicate the minimal impact of nuclear APA on mRNA stability, thus confirming and extending recent findings[12,13].

Strong PASs are processed before weak ones. An alternative explanation for the above observations (group I and II PAS pairs showing faster dPAS processing versus group III and IV PAS pairs exhibiting faster pPAS processing) is that the choice of nascent RNA PASs is just a reflection of kinetic difference in the 3′ end processing at different sites. As the kinetics of 3′ end formation is largely determined by many factors interacting with their cognate elements[15,16,38], especially the AAUAAA motif[39], we examined and found that this motif is more enriched upstream of dPASs than pPASs for PAS pairs in all groups, with some modest difference in group III (Fig. 3d). We hypothesized that stronger PASs would be processed before relatively weaker ones, and for PASs with similar strength, their processing would then follow the first come-first served rule.

Previous studies showed that the AAUAAA signal alone is insufficient to determine the 3′ end processing efficiency[15,16], as the downstream U/GU element[36] and the upstream UGUA motif[34] are also known to play critical roles in enhancing the 3′ end processing efficiency. Indeed, we observed an enrichment of U/GU-rich elements between 0 and 40 nucleotides (nt) downstream of the

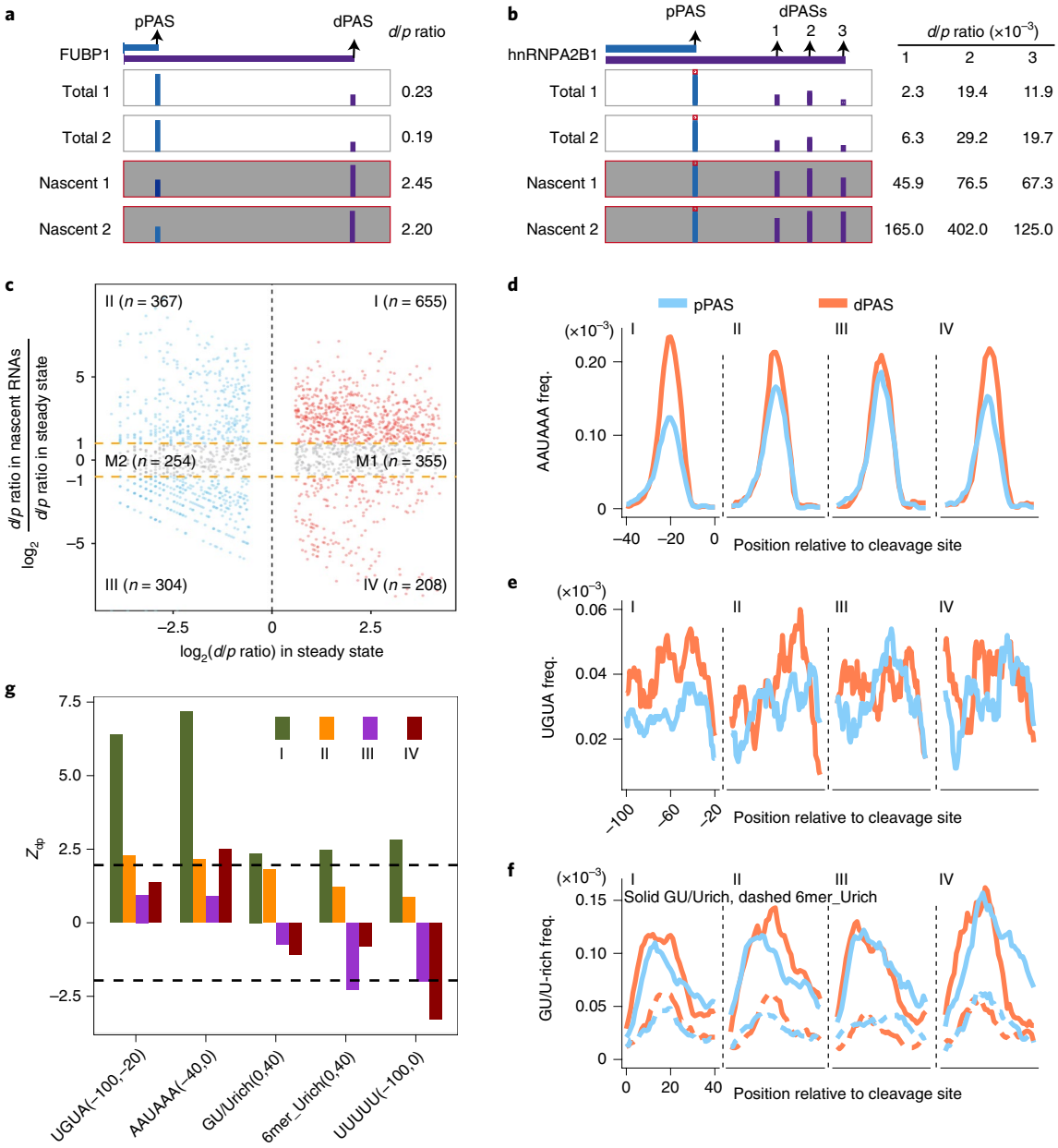

Fig. 3 | Distinct sequence features between pPAS and dPAS pairs. a,**b**, UCSC genome browser tracks showing the different *d*/*p* ratios from total reads and nascent reads from SLAM PAS-seq in duplicates of genes *FUBP1* (**a**) and *hnRNPA2B1* (**b**). The *d*/*p* ratio refers to the usage of dPAS divided by that of pPAS, and is shown on the right. **c**, Classification of the pairs of pPAS and dPAS into four groups (I–IV) by the *d*/*p* ratio from steady state and the relative value of *d*/*p* ratio from nascent RNAs over that of steady state. **d**–**f**, The distributions of different motifs at the flanking region of pPAS (blue) and dPAS (yellow) from different groups. **d**, AAUAAA motif in the region −40–0 nt upstream of the cleavage site. **e**, UGUA motif in the region −100 to −20 nt upstream of the cleavage site. **f**, GU/U-rich motif in the region 0–40 nt downstream the cleavage site, UGUG(C)U and UUUUU in solid line while other U-rich 6-mer motifs (described in the Methods) in dotted line. **g**, Z_{dp} values between dPAS and pPAS of different motifs in four groups of genes. The dotted line marks *P* value cutoff of 0.05.

cleavage site and UGUA motif between 20 and 100 nt upstream of all PASs (Extended Data Fig. 6a). Although the nucleotide composition around pPASs and dPASs is almost the same for all four groups (Extended Data Fig. 6b), the densities of those regulatory elements between pPASs and dPASs are different. Specifically, the upstream UGUA motif is more enriched in dPASs than pPASs in groups I and II, but not obvious in groups III and IV (Fig. 3e). The downstream GU/U-rich motif is more broadly enriched in dPASs than pPASs in group I and II, but reversed in groups III and IV (Fig. 3f). Given a previous report that GU-rich elements followed by a U-rich context can stabilize the interaction between CstF64 and RNA[36], we suspected CstF64 can bind to dPASs more efficiently relative to pPASs in group I and II, while the reverse is true in group III and IV.

We also found that the U-rich elements that flank the PAS, to which FIP1L1 binds and enhances the processing efficiency[35,38], are more closely associated with dPASs in group I and II, but with pPASs in group III and IV (Extended Data Fig. 6c). We further quantified the relative motif enrichment in dPASs versus pPASs using the Z_{dp} metric[38], and generally observed that polyadenylation enhancer elements are more strongly associated with dPASs in groups I and II compared with groups III and IV, or even negatively associated with dPASs in groups III and IV (Fig. 3g), suggesting that the strength of PAS is a

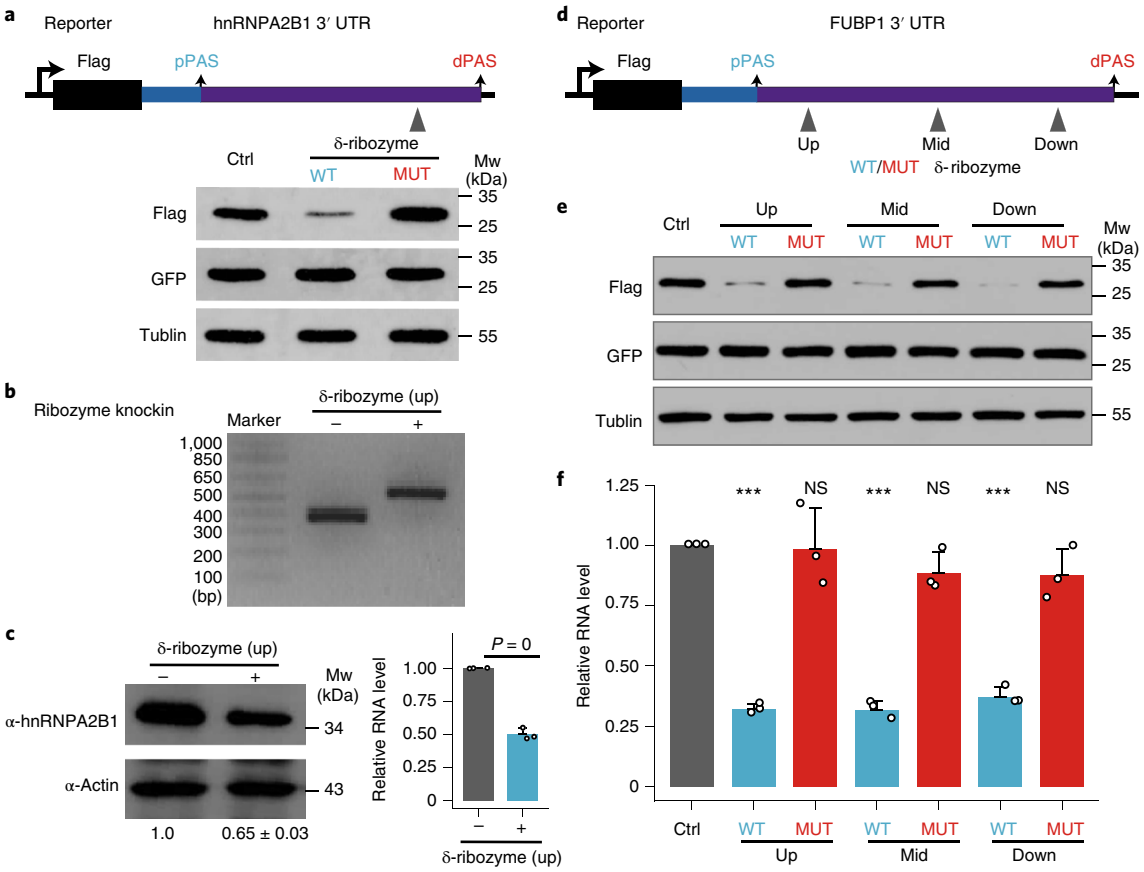

Fig. 4 | Isoforms with extended 3′ UTR are intermediates for shorter ones. a, Immunoblotting analysis of the cells transfected with different constructs with wild or mutation type ribozyme inserted in the 3′ UTR of *hnRNPA2B1*. Gray arrow is the position where wild or mutation type ribozyme is inserted. WT, wide type; MT, mutation type (C76U). GFP plasmid is used as the transfection control. Data were from $n=2$ independent experiments. **b**, Validation of the ribozyme inserted upstream of the dPAS of *hnRNPA2B1* by PCR. Data were from $n=1$ independent experiment. **c**, Effects of ribozyme insertion upstream of dPAS on endogenous *hnRNPA2B1* at the protein (left) and RNA (right) levels. The quantified values at the bottom represent mean ± s.d. ($n=2$). **d**, The diagram of the reporter that the ribozyme is inserted in 3′ UTR of *FUBP1* at different positions. **e**, Immunoblotting analysis of the cells transfected with different constructs, the same as **a**. Data were from $n=3$ independent experiments. **f**, Relative RNA level of different constructs. Data are presented as mean values ± s.d. ($n=3$). The *P* values are determined using the two-tailed unpaired *t*-test in comparing different ribozyme insertions to the control (Up WT $P=0$, MUT $P=0.8986$; Mid WT $P=0$, MUT $P=0.0861$; Down WT $P=0$, MUT $P=0.1233$; $^{***}P<0.001$; NS, not significant). Note that the reporter with inactive ribozyme seems to make more protein (Fig. 4a), which may result from experimental variation because of the lack of a similar change in the replicate of the same experiment (source data).

major determinant for the 3′ end processing kinetics, which may directly dictate the processing order for genes with multiple PASs.

Long 3′ UTR isoforms as intermediates for pPAS processing. For genes such as *FUBP1*, *hnRNPA2B1* and *CPSF6*, their longer 3′ UTR isoforms were almost all retained in the nucleus, and thus would not contribute to protein production. Indeed, when their constitutive UTR (cUTR) and alternative UTR (aUTR) were separately inserted into the 3′ UTR of a GFP reporter (Extended Data Fig. 7a), we only detected protein production with their cUTRs (Extended Data Fig. 7b). Taking advantage of this system, we directly addressed whether the longer 3′ UTR could act as intermediates for their shorter counterparts through progressive polyadenylation. As illustrated in Extended Data Fig. 7c, if the pPAS and dPAS were independently processed, we expect that the degradation of the longer 3′ UTR isoform would not affect the RNA level of the shorter 3′ UTR isoform, and thus, the protein production would not be altered. However, if the isoform with dPAS was processed and then served as a precursor for pPAS processing, the degradation of the longer 3′ UTR isoform would also diminish the generation of the shorter 3′ UTR isoform, thereby attenuating protein production.

We implemented this experimental scheme by inserting a self-cleavable ribozyme into the extended 3′ UTR of *hnRNPA2B1* to induce specific degradation of its long 3′ UTR isoform (Fig. 4a)[40]. We observed a substantial protein decrease from targeted transcript, while a mutated ribozyme, which lost its ribozyme activity, had no effect (Fig. 4a). We also inserted the ribozyme into the endogenous loci upstream of the dPAS of *hnRNPA2B1* by CRISPR (Fig. 4b and Extended Data Fig. 7d), and made the same observation at both the protein (Fig. 4c, left) and RNA (Fig. 4c, right) levels. Moreover, we detected similar reduction of both protein and RNA irrespective of the position (up, middle or down) for ribozyme insertion upstream of the dPAS in another gene (*FUBP*1)-based reporter (Fig. 4d–f). As critical controls for this set of experiments, we inserted the ribozyme downstream of the dPAS in both *FUBP1*-based and *hnRNPA2B1*-based reporters (Extended Data Fig. 8a). We detected slight decrease with *FUBP1* and *hnRNPA2B1* at the RNA level (likely due to a degree of ribozyme-induced nascent RNA degradation), but no difference at the protein level (Extended Data Fig. 8b,c). We also used the CRISPR to insert the ribozyme downstream of dPAS in the endogenous *hnRNPA2B1* gene and detected no difference at both the RNA and protein levels (Extended Data Fig. 8d–f). Together,

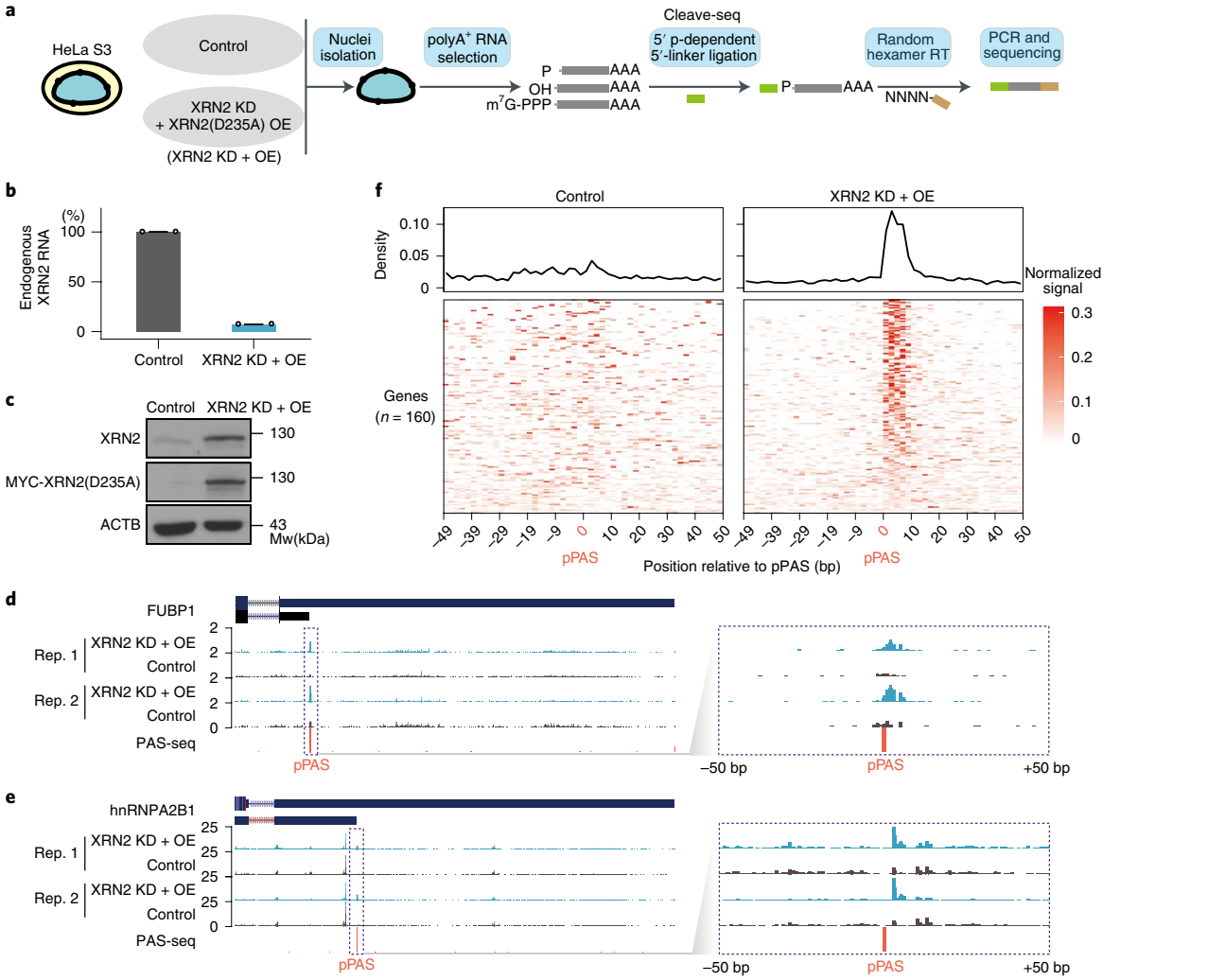

Fig. 5 | Identification of the 3′ cleavage fragments by Cleave-seq. a, Schematic illustration of Cleave-seq for detecting 3′ cleaved fragments of polyA+ RNA. **b**, Endogenous XRN2 RNA level in samples with XRN2 knockdown (KD) and myc-tagged XRN2(D235A) overexpression (OE) relative to that of controls, quantified by qPCR with reverse transcription (RT). Data are presented as mean values ± s.d. ($n = 2$). **c**, Immunoblotting showing the XRN2 protein level in control and treated samples with XRN2 knockdown and overexpression of myc-tagged XRN2(D235A). Data were from $n = 2$ independent experiments. **d**,**e**, UCSC genome browser view of the 5′ end signals of Cleave-seq libraries with replicates in the 3′ UTR region of *FUBP1* (**d**) and *hnRNPA2B1* (**e**). The zoomed-in views of the 100-nt region flanking the pPAS (left dotted) are shown in the dotted boxes (right). **f**, Meta profile (top) and heatmaps (bottom) of the 5′ end signals of Cleave-seq libraries in 160 genes with a minimum of ten reads in a 10-nt window downstream the proximal PAS. Genes are sorted by the degree of signal enrichment in the 10-nt window.

these results strongly indicate that longer 3′ UTR isoforms are intermediates for the shorter isoforms.

Detecting 3′ intermediates from sequential polyadenylation. According to our sequential polyadenylation model, cleavage at pPAS in longer 3′ UTR isoform would generate the shorter 3′ UTR isoform along with a 3′ polyadenylated fragment containing a 5′ mono-phosphorylate. Taking advantage of this property, we developed Cleave-seq by ligating nuclear polyA+ RNA to a linker with 3′-OH followed by randomly primed reverse transcription to generate complementary DNA libraries for deep sequencing, as schemed in Fig. 5a. To stabilize the intermediates, we inhibited the major nuclear 5′–3′ exonuclease *XRN2* by RNA interference (Fig. 5b). Simultaneously, we overexpressed a dominant negative, catalytic inactive XRN2(D235A) to further enhance the stabilization effect, which was validated in a previous study[41] (Fig. 5c). Under these conditions, we detected endonucleolytically cleaved products at or near the anticipated sites, as exemplified with a well-documented Drosha cleavage site in the *DGCR8* gene[42] (Extended Data Fig. 9a).

Globally, we identified 22,879 potential cleavage sites or peaks with significantly enriched signals in at least two out of four constructed libraries. We found 8,825 sites in all four samples and 3,289 sites only detected in the two samples with XRN2 depletion (Extended Data Fig. 9b). When focusing on the downstream 10-nt regions of pPAS and dPAS of 654 genes with longer 3′ UTR enriched in NM, we detected 26 cleavage sites, 20 out of which were located downstream of the pPAS only when XRN2 was depleted (Extended Data Fig. 9b, upper box), suggesting that these sites are more tightly associated with pPAS, which are sensitive to XRN2.

Notably, one of those 26 target sites is located downstream the pPAS of *FUBP1* (Fig. 5d). Although without sufficient statistical significance, we detected an increase in sequencing reads downstream of pPAS in *hnRNPA2*B1 when XRN2 was depleted (Fig. 5e). To assess the pervasiveness of this phenomenon, we identified 160 out of 654 genes with longer 3′ UTR enriched in NM, requiring a minimum

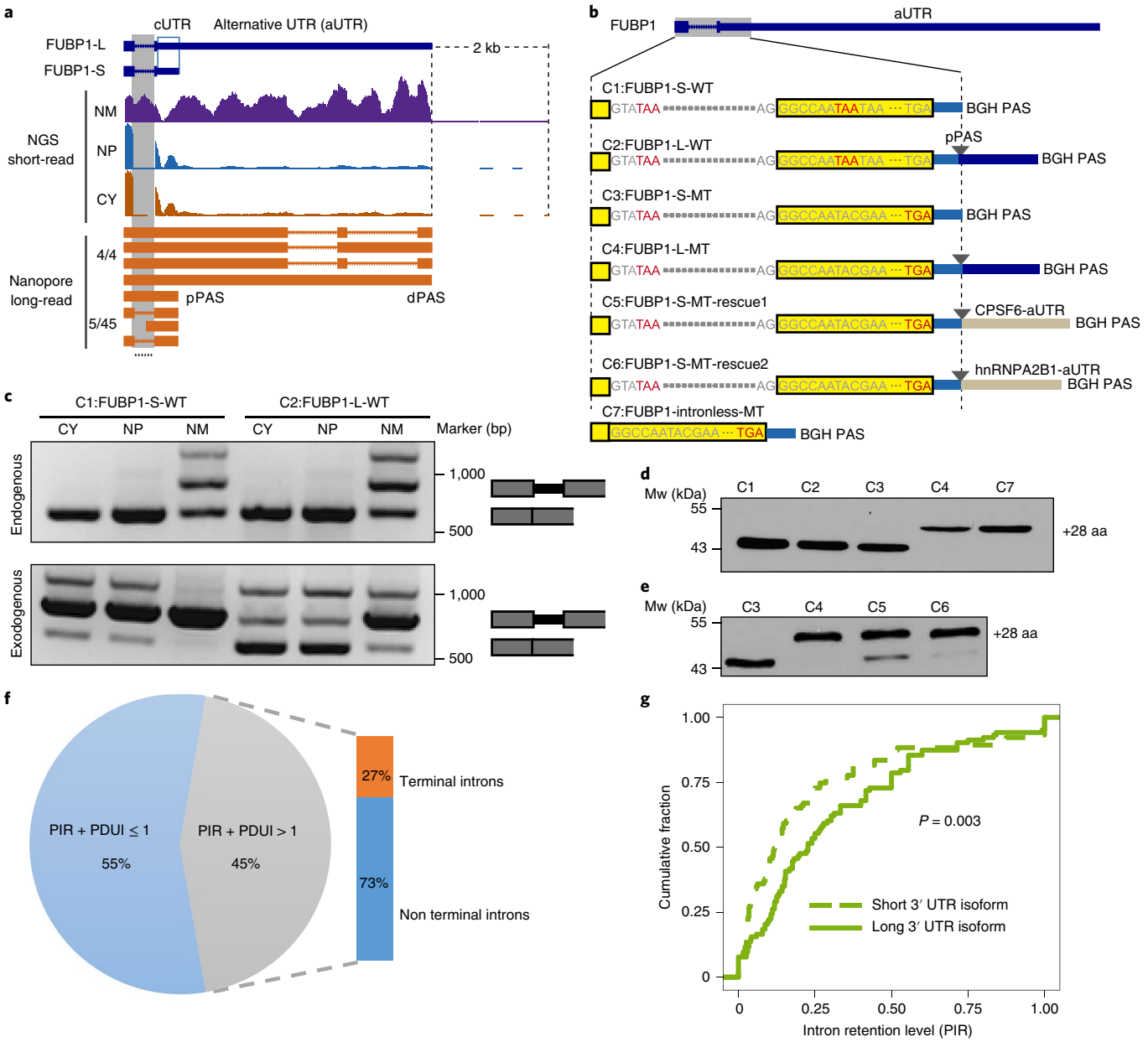

Fig. 6 | Progressive polyadenylation prevents the leakage of unspliced introns. a, UCSC genome browser tracks showing the even coverage across the terminal intron and extended 3′ UTR of FUBP1 in the NM from polyA+ RNA-seq. The short reads are sequenced on the Illumina platform for the NM, NP and CY RNAs, while the full-length long reads are sequenced on the Nanopore platform for the NM polyA+ RNA only. The long-read numbers (denominator, all isoforms that cover the last 3′ splicing site upstream of pPAS; numerator, the unspliced isoforms) are shown at the left of the tracks. **b**, Reporter constructs used in the downstream functional assays. The stop codons (TAA or TGA) are highlighted in red. **c**, Splicing analysis of the reporters construct C1 and C2 in **b** by PCR with reverse transcription. Data were from $n = 2$ independent experiments. **d**, Splicing analysis of the reporter constructs C1–4 in **b** by western blot. The intronless reporter C7 is a control. Data were from $n = 2$ independent experiments. **e**, Splicing analysis of the rescue reporter constructs C5–6 in **b** by western blot. Data were from $n = 3$ independent experiments. aa, amino acids. **f**, The statistics of the sum of PIR of the most unspliced intron and PDUI of the extended 3′ UTR. The data are from the short-read sequencing of the NM RNAs. **g**, Cumulative distribution of PIR of the introns from pPAS (short 3′ UTR isoform, dotted line) and dPAS (long 3′ UTR isoform, solid line) from long-read sequencing data. The significance for the shift is indicated by the *P* value determined using the two-sided Wilcoxon test.

of ten supporting reads in the downstream 10-nt region of pPAS on XRN2 depletion. We examined the reads distribution in the 100-nt region flanking the pPAS of these 160 genes and observed increased signals on XRN2 depletion with 121 genes that showed a more than twofold increase in the 10-nt region downstream of the pPAS (Fig. 5f). Moreover, this signal increase is specific for the pPAS, because at untreated state, the signals downstream of both pPAS and dPAS were identical but on XRN2 depletion, the signals were selectively increased in the pPAS downstream region compared with the dPAS downstream region (median ratio fold-change >10) (Extended Data Fig. 9c). Therefore, through the polyA selection, most cleaved but nonpolyadenylated products at dPASs were excluded, which enabled the detection of cleavage signals at pPASs in RNA already polyadenylated dPASs.

Together, these data provide strong evidence for the precursor–product relationship expected from the sequential polyadenylation model.

Sequential polyadenylation-mediated splicing regulation. Next, we wondered whether sequential polyadenylation could be involved in regulating gene expression. In fact, we observed that the terminal intron and the extended 3′ UTR (aUTR) of *FUBP1* had similar RNA-seq signals in NM, but the intron signals were efficiently removed in NP and CY (Fig. 6a). We further confirmed this phenomenon by using Nanopore long-read sequencing to directly count NM polyA$^+$ RNAs, showing that all transcripts with extended 3′ UTR (using dPAS) retained the upstream unspliced terminal intron (four out of four long reads up to 5 kb), while only five out of 54 transcripts with shorter 3′ UTR (pPAS usage from sequential polyadenylation) harbored unspliced intron (Fig. 6a). These results imply a coordination between splicing of this terminal intron and sequential polyadenylation in expressing *FUBP1*.

To substantiate this coupled processing, we detected splicing of the terminal intron in *FUBP1* on perturbing sequential polyadenylation by replacing the weak pPAS with the strong PAS from the *BGH* (Bovine growth hormone) gene (Fig. 6b, construct C1). We found that this caused the mRNA retaining this terminal intron to be efficiently exported to the CY (Fig. 6c, construct C1). In contrast, if the property of sequential polyadenylation was unaltered (Fig. 6b, construct C2), we observed the same level of intron retention as with the endogenous gene (Fig. 6c, construct C2). As the molecular mass of the protein product from the spliced isoform differed from the unspliced isoform only by one amino acid, we were unable to distinguish between their protein products based on the molecular mass (Fig. 6d, left two lanes C1 and C2). We therefore engineered another pair of reporters, in which we mutated the stop codon UAA on the exon so that the spliced isoform would use the downstream stop codon UGA to produce a larger protein with an additional 28 amino acids (Fig. 6b, construct C3 and C4).

As expected, we detected anticipated protein products from both spliced mRNA with construct C4 and unspliced mRNA with construct C3 (Fig. 6d, lane C3 and C4). These data further enforced a coupling between sequential polyadenylation and terminal intron removal. Moreover, we demonstrated that the splicing pattern could be rescued by using the extended 3′ UTR (aUTR) from another gene *CPSF6* or *hnRNPA2B1* (Fig. 6b, constructs C5 and C6; Fig. 6e, lanes C5 and C6 and Extended Data Fig. 10a,b), indicating that this effect is sequence independent. Based on these findings, we propose a model where nuclear retention conferred by some long 3′ UTRs potentiate posttranscriptional removal of certain slowly spliced introns (Extended Data Fig. 10d), representing a general strategy for coupling between separate posttranscriptional RNA processing events.

To obtain global evidence for this type of coupling, we quantified intron retention by percentage of intron retention (PIR) of the most unspliced introns and the dPAS usage by PDUI for genes with extended 3′ UTR retained in the NM fraction (Extended Data Fig. 10c). We found that unspliced intron and extended 3′ UTR coexisted in the same transcript for at least 45% of those genes, when requiring the sum of the PIR and PDUI values larger than 1 (Fig. 6f). The data from long-read sequencing also supported the idea that transcripts with longer 3′ UTRs are more likely to be associated with unspliced intron compared with those with shorter 3′ UTR (Fig. 6g). This is in line with the possibility that intron reads may arise from nascent RNA or retained intron with slow splicing kinetics, and dPASs are processed ahead of pPASs from nascent RNA. The coexistence of extended 3′ UTR and unspliced intron in the same transcripts may offer an opportunity for cross-regulation between polyadenylation and splicing.

Discussion

Our study uncovered an unexpected phenomenon that APA generates longer 3′ UTR isoforms specifically enriched in the NM. One mechanism is that potential nuclear retention elements within the long 3′ UTR isoform help to retain it in NM, which serve as the

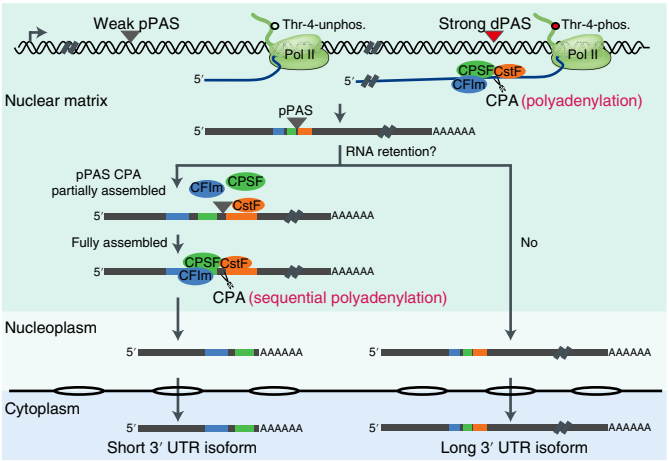

Fig. 7 | Progressive polyadenylation model for polyadenylation. The RNA polymerase II will travel cross the upstream weak PAS until it encounters the downstream strong PASs, which can be well recognized by the CPA complex and efficiently be processed. During this process, if the proximal PAS is not bound by any CPA component, the long 3′ UTR isoform will be released (right). On the other hand, bound by CPA component or other potential nuclear retention factors, the long 3′ UTR isoform will be restricted in the NM and undergo further processing to generate the short 3′ UTR isoform (left). See the text for details.

intermediates to generate shorter 3′ UTR isoform. Consistently, we found that the core 3′ end processing factors all bind to NM-associated polyA$^+$ RNA. We designed an autocleavable ribozyme strategy to directly demonstrate this sequential polyadenylation model. We provided transcriptome-wide evidence for the broad applicability of this regulatory model using SLAM PAS-seq and Cleave-seq. Together, these findings establish a general principle for regulated APA in a sequential manner with cotranscriptional selection of dPAS and posttranscriptional use of the pPAS.

Transcription termination is tightly associated with 3′ end formation, which ensures transcriptional termination near the end of genes, thereby preventing Pol II from running into downstream genes. Two models, the 'allosteric model'[43] and 'torpedo model'[44], have been proposed to explain how 3′ end processing triggers transcription termination. Both models assume the recognition of a functional PAS as a prerequisite for transcription termination. Conversely, a PAS may not be efficiently recognized when transcription termination is not elicited. This process appears to be marked by Pol II CTD phosphorylation at the Thr-4 position, which is tightly associated with transcription termination at dPASs[45,46]. These pieces of information are consistent with our findings that the dPASs are recognized first during transcription followed by the selection of the pPASs.

In our study, we reconfirmed that dPASs are usually stronger than the proximal ones[47] and the 3′ end processing kinetics are tightly associated with the PAS strength per se, which is reasonable given that pPASs are transcribed first. From the evolutionary perspective, only in this way, the dPASs can outcompete with their proximal counterparts. We thus conclude that, just like splicing, cotranscriptional processing is not an obligatory process for polyadenylation. Relatively weaker pPASs potentiate the cotranscriptional recognition and processing of strong dPASs.

Taking all into account, we propose a 'sequential polyadenylation' model for APA (Fig. 7). The Pol II will first travel across the weak pPAS until it encounters the strong dPAS, which is well recognized by the CPA (cleavage and polyadenylation) complex, in turn promotes polyadenylation and transcription termination in company with Thr4 of Pol II CTD phosphorylation[45,46]. During this process, if no CPA factor binding to the pPASs and no binding

of other nuclear retention factors to the RNA, the longer 3′ UTR isoform will be quickly released into the NP and then exported into the CY. On the other hand, if some CPA factors are assembled around the pPASs and/or the involvement of nuclear retention factors precedes the polyadenylation of the dPASs, it will restrict the longer 3′ UTR isoform in the nucleus, thus enabling the full assembly of CPA complex to process the pPAS to generate the shorter 3′ UTR isoform.

Our finding that polyadenylation can occur in a posttranscriptional manner offers new insights into the regulation of APA. It has been established that 3′ end processing is kinetically coupled with transcription[21,22]. Pol II pausing[48] or a slower polymerase[49] has been shown to promote the usage of pPASs, which is thought to enlengthen the window of opportunity for RNA processing. Our results indicate that any mode leading to extended nuclear exposure time to the 3′ end processing factors would confer APA regulation. This is in fact supported by a previous observation that increasing the distance between two identical PASs is able to enhance the use of the pPAS[50]. Our model can also fully explain CFIm25-mediated APA regulation: CFIm25 has been reported to bind its cognate UGUA motif, which is more frequently distributed upstream of the dPASs, thereby enhancing dPAS processing[34]. Thus, decrease in CFIm25 may cause less efficient processing of the dPASs, which may in turn lead to inefficient transcription termination. Our results now provide direct evidence supporting that tethering of RNA to elongating Pol II appears to provide a longer dwell time for the posttranscriptional processing of the pPASs[34].

About 70% of human genes can produce at least two different isoforms through APA[1], but the biological significance is still largely unknown[11]. Here, we report a new function for 3′ UTR in regulating intron removal through sequential polyadenylation. We show that the retained terminal intron of *FUBP1* gene will leak into the CY if its weak pPAS without the extended 3′ UTR is placed directly upstream of the strong BGH PAS, and this effect can be rescued with other extended 3′ UTRs. Early studies have documented the coordination of terminal intron excision with polyadenylations[51,52], but it has been thought that such coordination is mediated by protein–protein interactions between splicing and polyadenylation factors[22]. Our sequential polyadenylation model suggests a different mechanism, where the mRNA isoforms with longer 3′ UTRs are retained in the nucleus to enable additional posttranscriptional splicing and cleavage at a proximal PAS(s) before releasing the fully processed transcripts into the CY. We further speculate that such sequential polyadenylation may also be coupled with additional modes of posttranscriptional processing, such as RNA modification and RNA editing, through prolonged nuclear dwell time, such that transcripts with shortened 3′ UTR may have distinct fate and/or function compared to the precursor transcripts with longer 3′ UTRs.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-021-00709-z.

Received: 7 January 2021; Accepted: 29 November 2021;
Published online: 10 January 2022

References

1. Derti, A. et al. A quantitative atlas of polyadenylation in five mammals. *Genome Res* **22**, 1173–1183 (2012).
2. Brumbaugh, J. et al. Nudt21 controls cell fate by connecting alternative polyadenylation to chromatin signaling. *Cell* **172**, 106–120 (2018).
3. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science* **320**, 1643–1647 (2008).
4. Alt, F. W. et al. Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3′ ends. *Cell* **20**, 293–301 (1980).
5. Ji, Z., Lee, J. Y., Pan, Z., Jiang, B. & Tian, B. Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl Acad. Sci. USA* **106**, 7028–7033 (2009).
6. Singh, I. et al. Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nat. Commun.* **9**, 1716 (2018).
7. Lee, S. H. et al. Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature* **561**, 127–131 (2018).
8. Mayr, C. & Bartel, D. P. Widespread shortening of 3′ UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, 673–684 (2009).
9. Masamha, C. P. et al. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* **510**, 412–416 (2014).
10. Mueller, A. A., van Velthoven, C. T., Fukumoto, K. D., Cheung, T. H. & Rando, T. A. Intronic polyadenylation of PDGFRalpha in resident stem cells attenuates muscle fibrosis. *Nature* **540**, 276–279 (2016).
11. Mayr, C. What are 3′ UTRs doing? *Cold Spring Harb. Perspect. Biol.* **11**, a034728 (2019).
12. Spies, N., Burge, C. B. & Bartel, D. P. 3′ UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Res* **23**, 2078–2090 (2013).
13. Gruber, A. R. et al. Global 3′ UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nat. Commun.* **5**, 5465 (2014).
14. Shi, Y. et al. Molecular architecture of the human pre-mRNA 3′ processing complex. *Mol. Cell* **33**, 365–376 (2009).
15. Shi, Y. & Manley, J. L. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev.* **29**, 889–897 (2015).
16. Proudfoot, N. J. Ending the message: poly(A) signals then and now. *Genes Dev.* **25**, 1770–1782 (2011).
17. Nevins, J. R. & Darnell, J. E. Jr. Steps in the processing of Ad2 mRNA: poly(A)$^+$ nuclear sequences are conserved and poly(A) addition precedes splicing. *Cell* **15**, 1477–1493 (1978).
18. Salditt-Georgieff, M., Harpold, M., Sawicki, S., Nevins, J. & Darnell, J. E. Jr. Addition of poly(A) to nuclear RNA occurs soon after RNA synthesis. *J. Cell Biol.* **86**, 844–848 (1980).
19. McCracken, S. et al. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357–361 (1997).
20. Hirose, Y. & Manley, J. L. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395**, 93–96 (1998).
21. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15**, 163–175 (2014).
22. Tian, B. & Manley, J. L. Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* **18**, 18–30 (2017).
23. Wuarin, J. & Schibler, U. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell. Biol.* **14**, 7219–7225 (1994).
24. Tilgner, H. et al. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22**, 1616–1625 (2012).
25. Pandya-Jones, A. et al. Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA* **19**, 811–827 (2013).
26. Herman, R., Weymouth, L. & Penman, S. Heterogeneous nuclear RNA-protein fibers in chromatin-depleted nuclei. *J. Cell Biol.* **78**, 663–674 (1978).
27. Xing, Y. G. & Lawrence, J. B. Preservation of specific RNA distribution within the chromatin-depleted nuclear substructure demonstrated by in situ hybridization coupled with biochemical fractionation. *J. Cell Biol.* **112**, 1055–1063 (1991).
28. Mortillaro, M. J. et al. A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc. Natl Acad. Sci. USA* **93**, 8253–8257 (1996).
29. Han, J., Xiong, J., Wang, D. & Fu, X. D. Pre-mRNA splicing: where and when in the nucleus. *Trends Cell Biol.* **21**, 336–343 (2011).
30. Castello, A. et al. System-wide identification of RNA-binding proteins by interactome capture. *Nat. Protoc.* **8**, 491–500 (2013).
31. Saitoh, N. et al. Proteomic analysis of interchromatin granule clusters. *Mol. Biol. Cell* **15**, 3876–3890 (2004).
32. Melnik, S. et al. The proteomes of transcription factories containing RNA polymerases I, II or III. *Nat. Methods* **8**, 963–968 (2011).
33. Hegele, A. et al. Dynamic protein-protein interaction wiring of the human spliceosome. *Mol. Cell* **45**, 567–580 (2012).
34. Zhu, Y. et al. Molecular mechanisms for CFIm-mediated regulation of mRNA alternative polyadenylation. *Mol. Cell* **69**, 62–74 (2018).

35. Lackford, B. et al. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.* **33**, 878–889 (2014).
36. Yao, C. et al. Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc. Natl Acad. Sci. USA* **109**, 18773–18778 (2012).
37. Herzog, V. A. et al. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* **14**, 1198–1204 (2017).
38. Hu, J., Lutz, C. S., Wilusz, J. & Tian, B. Bioinformatic identification of candidate *cis*-regulatory elements involved in human mRNA polyadenylation. *RNA* **11**, 1485–1493 (2005).
39. Sheets, M. D., Ogg, S. C. & Wickens, M. P. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res.* **18**, 5799–5805 (1990).
40. Chen, L. et al. R-ChIP using inactive RNase H reveals dynamic coupling of r-loops with transcriptional pausing at gene promoters. *Mol. Cell* **68**, 745–757 (2017).
41. Fong, N. et al. Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA Polymerase II termination suggest widespread kinetic competition. *Mol. Cell* **60**, 256–267 (2015).
42. Karginov, F. V. et al. Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. *Mol. Cell* **38**, 781–788 (2010).
43. Logan, J., Falck-Pedersen, E., Darnell, J. E. Jr. & Shenk, T. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc. Natl Acad. Sci. USA* **84**, 8306–8310 (1987).
44. Connelly, S. & Manley, J. L. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.* **2**, 440–452 (1988).
45. Kamieniarz-Gdula, K. et al. Selective roles of vertebrate PCF11 in premature and full-length transcript termination. *Mol. Cell* **74**, 158–172 (2019).
46. Eaton, J. D., Francis, L., Davidson, L. & West, S. A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev.* **34**, 132–145 (2020).
47. Legendre, M. & Gautheret, D. Sequence determinants in human polyadenylation site selection. *BMC Genomics* **4**, 7 (2003).
48. Enriquezharris, P., Levitt, N., Briggs, D. & Proudfoot, N. J. A pause site for RNA polymerase-Ii is associated with termination of transcription. *EMBO J.* **10**, 1833–1842 (1991).
49. Pinto, P. A. et al. RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO J.* **30**, 2431–2444 (2011).
50. Denome, R. M. & Cole, C. N. Patterns of polyadenylation site selection in gene constructs containing multiple polyadenylation signals. *Mol. Cell. Biol.* **8**, 4829–4839 (1988).
51. Bauren, G., Belikov, S. & Wieslander, L. Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3′-end formation and excision of the 3′-terminal intron. *Genes Dev.* **12**, 2759–2769 (1998).
52. Niwa, M., Rose, S. D. & Berget, S. M. In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev.* **4**, 1552–1559 (1990).