


RESEARCH

Open Access



SRSF3 and SRSF7 modulate 3'UTR length through suppression or activation of proximal polyadenylation sites and regulation of CFIm levels

Oliver Daniel Schwich^{1,2†}, Nicole Blümel^{1†}, Mario Keller^{2,3†}, Marius Wegener^{1,2}, Samarth Thonta Setty², Melinda Elaine Brunstein⁴, Ina Poser⁵, Igor Ruiz De Los Mozos⁶, Beatrix Suess⁷, Christian Münch⁴, François McNicoll¹, Kathi Zarnack^{2,3*}  and Michaela Müller-McNicoll^{1*}

* Correspondence: kathi.zarnack@bmls.de; mueller-mcnicoll@bio.uni-frankfurt.de

[†]Oliver Daniel Schwich, Nicole Blümel and Mario Keller contributed equally to this work.

²Buchmann Institute for Molecular Life Sciences, Goethe University Frankfurt, Max-von-Laue-Str. 15, 60438 Frankfurt, Germany

¹Institute for Molecular Bio Science, Goethe University Frankfurt, Max-von-Laue-Str. 13, 60438 Frankfurt, Germany
Full list of author information is available at the end of the article

Abstract

Background: Alternative polyadenylation (APA) refers to the regulated selection of polyadenylation sites (PASs) in transcripts, which determines the length of their 3' untranslated regions (3'UTRs). We have recently shown that SRSF3 and SRSF7, two closely related SR proteins, connect APA with mRNA export. The mechanism underlying APA regulation by SRSF3 and SRSF7 remained unknown.

Results: Here we combine iCLIP and 3'-end sequencing and find that SRSF3 and SRSF7 bind upstream of proximal PASs (pPASs), but they exert opposite effects on 3' UTR length. SRSF7 enhances pPAS usage in a concentration-dependent but splicing-independent manner by recruiting the cleavage factor FIP1, generating short 3'UTRs. Protein domains unique to SRSF7, which are absent from SRSF3, contribute to FIP1 recruitment. In contrast, SRSF3 promotes distal PAS (dPAS) usage and hence long 3' UTRs directly by counteracting SRSF7, but also indirectly by maintaining high levels of cleavage factor Im (CFIm) via alternative splicing. Upon SRSF3 depletion, CFIm levels decrease and 3'UTRs are shortened. The indirect SRSF3 targets are particularly sensitive to low CFIm levels, because here CFIm serves a dual function; it enhances dPAS and inhibits pPAS usage by binding immediately downstream and assembling unproductive cleavage complexes, which together promotes long 3'UTRs.

Conclusions: We demonstrate that SRSF3 and SRSF7 are direct modulators of pPAS usage and show how small differences in the domain architecture of SR proteins can confer opposite effects on pPAS regulation.

Keywords: SRSF3, SRSF7, APA, FIP1, CFIm, pPAS, dPAS, iCLIP, MACE-seq, 3'UTR length



Background

Cleavage and polyadenylation (CPA) is an important step of gene expression that results in the endonucleolytic cleavage of pre-mRNAs and the subsequent addition of a non-templated poly(A) tail of ~150 nt by the poly(A) polymerase (PAP) [1]. CPA is mediated by the recruitment of *trans*-acting factors to *cis*-acting sequence elements within pre-mRNAs that are highly conserved and well defined [2]. At each polyadenylation site (PAS), the CPA machinery assembles from four subcomplexes—cleavage and polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CstF), cleavage factor Im (CFIm), and IIm (CFIIm) [3]. The CPSF complex is composed of six subunits (CPSF160 [CPSF1], CPSF100 [CPSF2], CPSF73 [CPSF3], CPSF30 [CPSF4], WDR33, and FIP1 [FIP1L1]) and CstF of three subunits (CstF50 [CSTF1], CstF64 [CSTF2], and CstF77 [CSTF3]). CPSF and CstF are sufficient to catalyze CPA *in vitro* [4, 5]. CPSF30 and WDR33 recognize the hexameric central sequence element (CSE), usually AAUAAA [6, 7], CstF64 binds to GU-rich downstream sequence elements (DSEs) [8], and CPSF73 acts as the endonuclease [9]. FIP1 binds to U-rich sequence elements (USEs) upstream of PASs and recruits PAP. FIP1 binding switches PAP from a slow into a fast, processive state [10]. CFIm is a tetramer consisting of two small 25 kDa subunits (CPSF5 [NUDT21]) and two large subunits of 68 kDa or 59 kDa (CPSF6 or CPSF7) that bind CPSF5 separately or together [11–13]. The large subunits play a structural role, bringing together the two CPSF5 subunits that each bind to a UGUA sequence upstream of most PASs, thereby providing sequence specificity to the complex [14, 15]. However, while CPSF6 enhances PAS usage through FIP1 recruitment, CPSF7 does not stimulate CPA [16, 17]. CFIIm contains two subunits, PCF11 and CLP1, and has recently been shown to play a role in transcription termination [18].

Around 70% of all mammalian genes contain more than one PAS and express transcript isoforms that differ in the PASs that are used through a process termed alternative polyadenylation (APA) [19]. APA affects either the coding potential of transcripts (CR-APA), when the PAS is located upstream of the stop codon, or the length of the 3' untranslated region (3'UTR-APA) and hence the potential of mRNAs to be regulated [20]. 3'UTR sequence elements regulate the stability, nuclear export, localization, and translation efficiency of transcripts as well as the localization of their protein products [21]. Dysfunctional APA has been implicated in various human diseases including cancer [22, 23].

3'UTR-APA is regulated in *cis* through the intrinsic strength of alternative PASs, which is determined by the sequence composition of their CSEs, DSEs, and USEs, the number of UGUA motifs and their distance to the cleavage site, as well as the presence of binding sites for additional APA regulators [24–26]. APA is also regulated in *trans* through the levels of CPA complex subunits. For example, FIP1 promotes the usage of proximal PASs (pPASs) and the generation of transcripts with short 3'UTRs, which is important during stem-cell self-renewal [27]. In contrast, CFIm binds preferentially to distal PASs (dPASs) and enhances their usage, promoting the expression of isoforms with long 3'UTRs [28]. Accordingly, depletion of the CFIm subunits CPSF5 and CPSF6 causes global 3'UTR shortening [29, 30].

Beside core CPA factors, other RNA-binding proteins (RBPs) have been identified as global 3'UTR-APA regulators. Interestingly, many of them are splicing factors, such as CELF2, TARDBP, FUS, HNRNPC, and NOVA2 [19, 24, 31]. We have previously

reported that two splicing factors, SRSF3 and SRSF7, connect APA to selective mRNA export via recruitment of the mRNA export factor NXF1 [32]. The APA function of SRSF3 was subsequently shown to regulate cellular senescence [33]. While SRSF3 is already known to regulate the splicing of terminal exons and thereby CR-APA [34], how SRSF3 and SRSF7 regulate 3'UTR-APA and whether this function depends on splicing remain to be determined.

SRSF3 and SRSF7 are members of the SR protein family, which comprises 12 canonical members [35]. Both proteins are very closely related and share a similar domain architecture with one RNA recognition motif (RRM, 80% identical), a short linker region for NXF1 interaction, and a region enriched in arginine-serine di-peptides called RS domain [36–38]. Unlike SRSF3, SRSF7 contains an additional CCHC-type zinc (Zn) knuckle domain that together with the RRM determines its distinct RNA-binding specificity [39–41]. In addition, the RS domain of SRSF7 is 66 amino acids longer and harbors a distinct hydrophobic stretch [41]. RS domains mediate protein-protein interactions and, through phosphorylation and dephosphorylation of their serine residues, the splicing activity, nuclear import, subnuclear localization, and mRNA export activity of SR proteins are regulated [42].

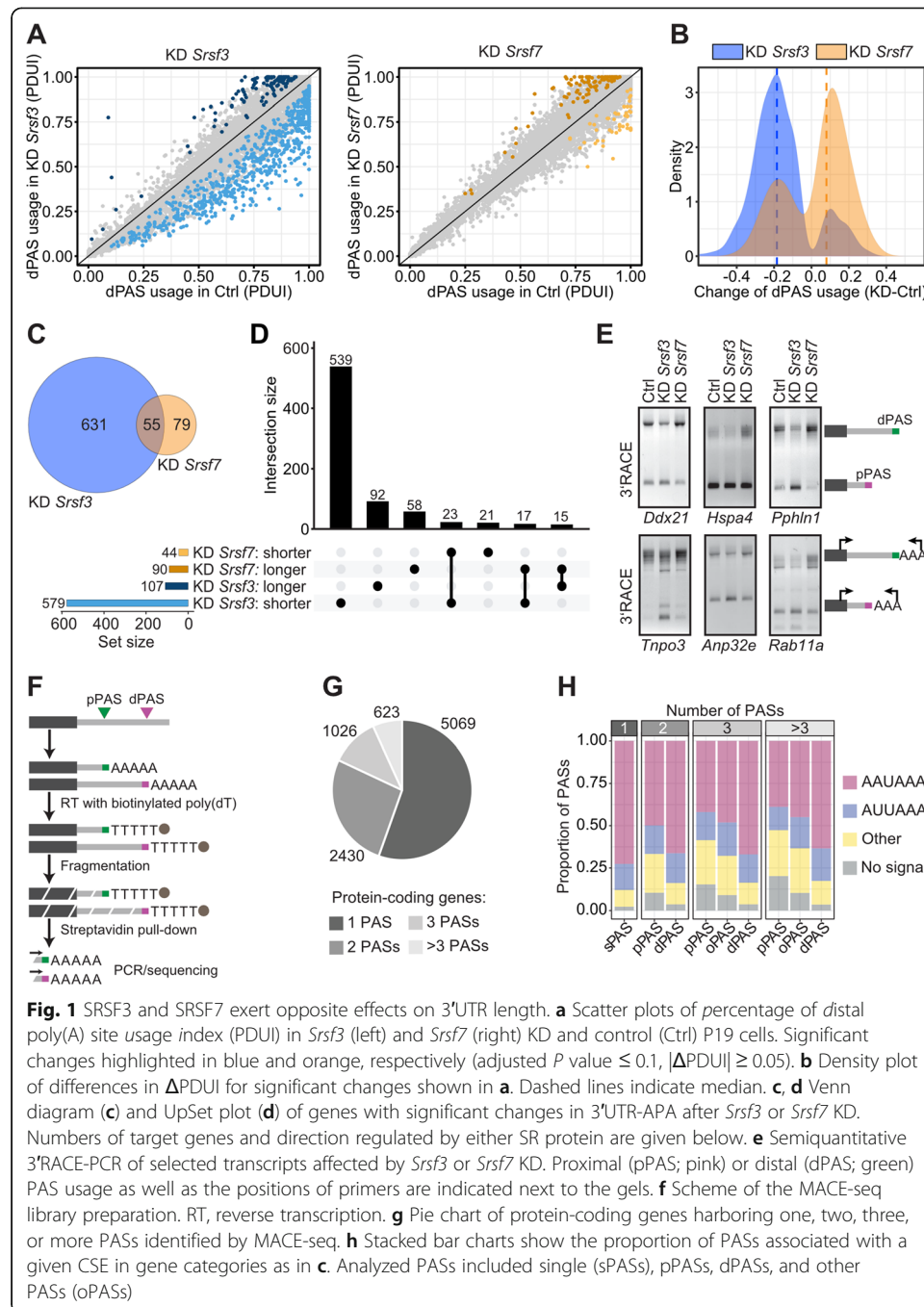
Here, we investigated the mechanisms underlying 3'UTR-APA regulation by SRSF3 and SRSF7. We found that both proteins bind upstream of pPASs but exert opposite effects on the length of 3'UTRs. SRSF7 activates pPAS usage directly in a splicing-independent manner via recruitment of FIPL1, thus promoting short 3'UTRs. In contrast, SRSF3 promotes long 3'UTRs by inhibiting pPAS usage either directly or indirectly by controlling the levels of active CFIm through alternative splicing. We found that SRSF3-regulated 3'UTRs are particularly sensitive to low CFIm levels, because CFIm inhibits pPAS usage and enhances dPAS usage in these transcripts. Given that SRSF3 and SRSF7 both recruit NXF1 for mRNA export [32], their binding and action at pPASs provide the means to sort mRNAs with different 3'UTR lengths for later steps in cytoplasmic gene expression, such as RNA localization and translation.

Results

SRSF3 and SRSF7 exert opposite effects on 3'UTR length

To study the mechanisms of 3'UTR-APA regulation by SRSF3 and SRSF7, we first acquired a comprehensive list of their APA targets in murine pluripotent P19 cells. We depleted each protein individually using esiRNAs (endoribonuclease-prepared siRNAs; Additional file 1: Fig. S1A) and sequenced poly(A) + RNA (50–60 million reads per sample). We used DaPars to identify pPASs and to quantify their usage based on changes in read coverage [43].

Knockdown (KD) of *Srsf3* significantly affected 3'UTR-APA of 686 genes (difference in percentage of distal poly(A) site usage index $|\Delta\text{PDUI}| \geq 0.05$, false discovery rate $[\text{FDR}] \leq 0.1$). Almost all targets showed an increased usage of the pPAS, resulting in 3'UTR shortening (579 genes, 84%; Fig. 1a, b, Additional file 2: Table S1). In contrast, *Srsf7* KD enhanced dPAS usage and 3'UTR lengthening (90 out of 134 genes, 67%). Despite the smaller number of genes affected by *Srsf7* KD, there was a considerable overlap (55 genes; Fig. 1c, P value = $2.5e-41$, Fisher's exact test), with 17 genes being regulated antagonistically by SRSF3 and SRSF7 (Fig. 1d). Analysis of the data with MISO TandemUTR annotations [44]



confirmed the trend of 3'UTR shortening and lengthening upon *Srsf3* and *Srsf7* KD, respectively (Additional file 1: Fig. S1B, Additional file 3: Table S2). SRSF3 target genes with shortened 3'UTRs were enriched for functions related to cell cycle progression and chromosome segregation (Additional file 1: Fig. S1C), indicating that SRSF3 regulates cell proliferation and growth via 3'UTR length control, as suggested in a previous study [33]. Changes in 3'UTR length were validated for six genes that were antagonistically regulated by SRSF3 and SRSF7 using semi-quantitative 3'RACE-PCR (rapid amplification of cDNA ends; Fig. 1e). As observed in the DaPars/MISO analyses, *Srsf3* KD promoted the use of their

pPASs and *Srsf7* KD the use of their dPASs. We concluded that SRSF3 and SRSF7 have opposite effects on 3'UTR-APA in hundreds of cellular transcripts and act antagonistically on a subset of targets.

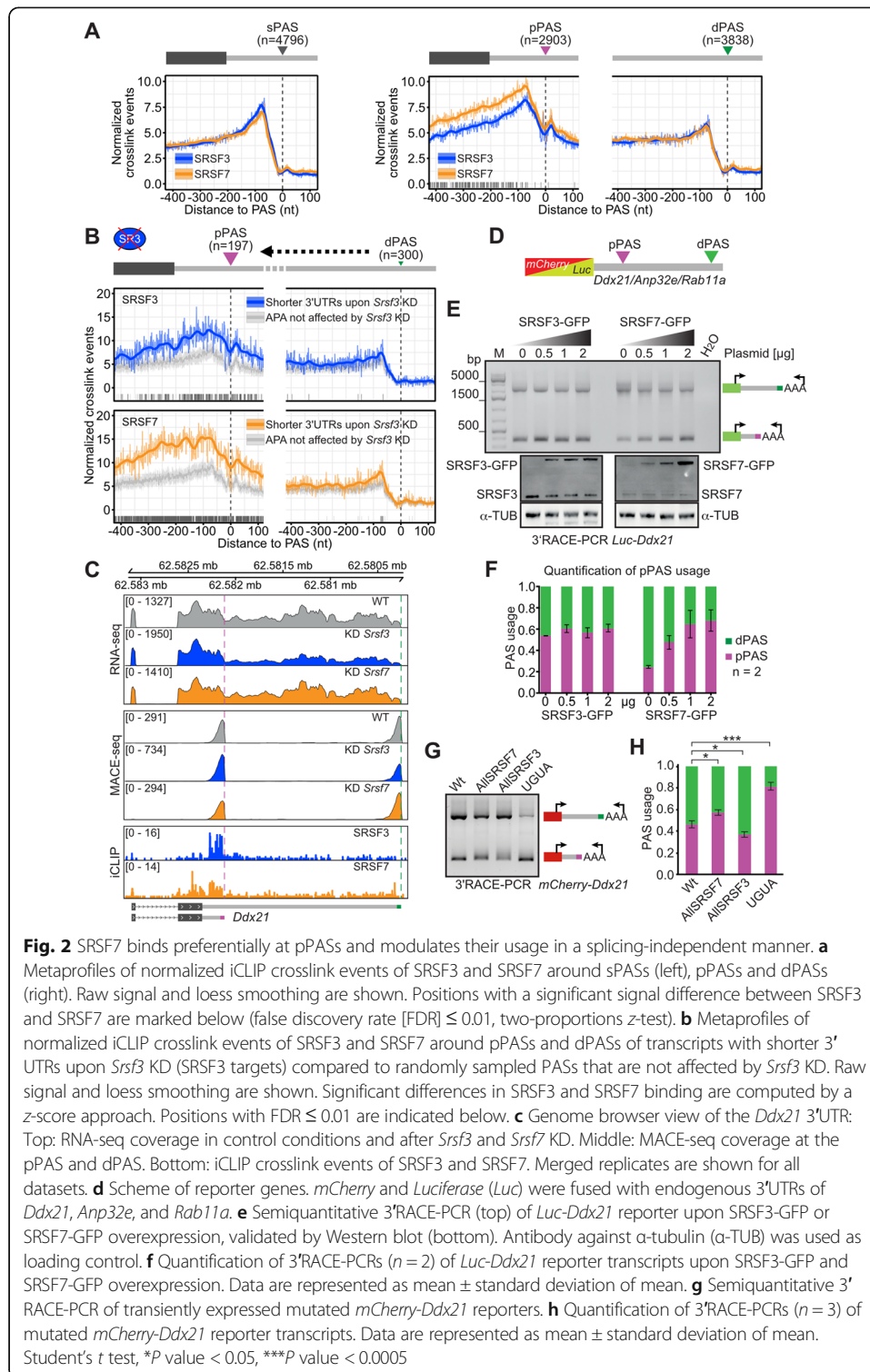
In order to precisely map the PASs used in murine pluripotent P19 cells, we performed MACE-seq (massive amplification of cDNA ends), a high-throughput 3' end sequencing method that enables the identification of PASs at single-nucleotide resolution [45, 46] (Fig. 1f). MACE-seq was performed with RNA from control, *Srsf3* and *Srsf7* KD cells to capture all PASs that are used in these conditions (Additional file 1: Fig. S1D). For PAS mapping, we pooled all MACE-seq samples and used a customized analysis pipeline to remove priming artifacts (Additional file 1: Fig. S1E, see the “Methods” section). We identified a total of 15,866 high-confidence PASs mapping to 9148 genes. Most PASs were found in protein-coding genes (15,805), within their 3' UTRs (13,706) (Additional file 1: Fig. S1F, S1G).

The majority of protein-coding genes (5069 genes, 55%) used only a single PAS (sPAS) in murine P19 cells. However, 4079 protein-coding genes (45%) showed clear evidence for APA, with 2430, 1026, and 623 genes harboring two, three, or more PASs, respectively (Fig. 1g). Depending on their relative position, we assigned the first, last, and intermediate PASs within each 3'UTR as proximal (pPAS), distal (dPAS), and other (oPAS), respectively. The PAS positions agreed well with GENCODE annotations (version M18), as most PASs (61%) mapped within 25 nt of an annotated transcript 3' end (Additional file 1: Fig. S1H, S1I). In addition, we detected 6207 PASs that were further downstream and likely belonged to non-annotated alternative isoforms (Additional file 1: Fig. S1H, S1I).

Screening for the 18 known CSE motifs [24] upstream of all PASs used in P19 cells, we found that 93% of all PASs contained a CSE variant (Additional file 1: Fig. S1J). The canonical AAUAAA hexamer was most common (59.5%), followed by the variants AUUAAA (19.9%), UAUAAA (7.4%), AAGAAA (6.3%), and AAUAUA (5.4%). In genes with multiple PASs, AAUAAA was predominantly found at dPASs, whereas alternative CSEs were predominant at pPASs and oPASs (Fig. 1h). This is in line with previous studies suggesting that pPASs are generally weaker and subject to regulation, while dPASs are stronger and used by default [47]. Similarly, almost 90% of sPASs harbored one of the two most frequent CSEs, indicating that sPASs are strong and well-defined in motif composition. Our data represent the first atlas of PAS positions and CSE composition in mouse pluripotent P19 cells and suggest that almost half of all expressed genes generate more than one 3'UTR isoform, indicating extensive gene expression regulation at the level of 3'UTR-APA.

SRSF7 binds preferentially at pPASs and modulates their usage in a splicing-independent manner

To investigate whether SRSF3 and SRSF7 directly affect 3'UTR-APA, we compared their binding at our mapped sPASs, pPASs, and dPASs using our previously published iCLIP (individual-nucleotide resolution UV crosslinking and immunoprecipitation) data sets obtained from P19 cells [32]. In line with a direct involvement in CPA, both proteins showed a pronounced binding peak ~ 75 nt upstream of sPASs, pPASs, and dPASs in a metaprofile analysis (Fig. 2a). An additional, minor binding peak was present ~ 20 nt downstream of all PASs and especially conspicuous at pPASs. In genes that undergo



3'UTR-APA, both proteins showed a strong preference for pPASs, where binding of SRSF7 exceeded that of SRSF3. In contrast, binding of both proteins was similar at sPASs and dPASs. To confirm that this is also true for genes whose APA is regulated by SRSF3 and SRSF7, we integrated the PAS coordinates from MACE-seq with 3'UTR

changes from the DaPars analysis and precisely mapped PASs for 361 out of 686 SRSF3 targets (Additional file 1: Fig. S2A, S2B). Indeed, binding of both SRSF3 and SRSF7 was much more frequent at SRSF3-regulated pPASs, compared to pPASs that are not affected by *Srsf3* KD, but there was no binding difference at dPASs (Fig. 2b). This suggests that the pPAS is the primary point of APA regulation by SRSF3 and SRSF7 and that both proteins might compete here for binding. To further strengthen this idea, we tested whether the binding motifs of SRSF3 (CNYC, C, cytosine; N, any nucleotide; Y, pyrimidine) and SRSF7 (GAY, G, guanine; A, adenine) are enriched around sPASs, pPASs, and dPASs [32, 48]. We generally observed an enrichment of both motifs upstream of sPASs and pPASs (Additional file 1: Fig. S2C, S2F), suggesting that both proteins can actively bind to pPASs rather than being recruited by other proteins. Of note, the SRSF7 binding motif GAY was particularly enriched at SRSF3-regulated pPASs compared to unregulated pPASs (Additional file 1: Fig. S2G), reflecting its enhanced binding at these sites in our iCLIP data (Fig. 2b). A similar pattern could not be observed for the SRSF3 binding motif CNYC motif (Additional file 1: Fig. S2D). In contrast, none of the motifs were enriched at dPASs (Additional file 1: Fig. S2E, S2H), underlining that pPASs are the regulatory hotspots of APA regulation by SRSF3 and SRSF7.

To test whether APA regulation by SRSF3 and SRSF7 is splicing-dependent, we selected three validated target genes—*Ddx21*, *Anp32e*, and *Rab11a* (Fig. 1e)—for a reporter gene study. All three genes showed increased pPAS usage upon *Srsf3* KD, increased dPAS usage upon *Srsf7* KD, and binding of both proteins upstream of the pPAS (Fig. 2c, Additional file 1: Fig. S2I, S2J). We fused the complete 3'UTRs (including 150 nt downstream of the dPASs) to two distinct reporter genes, encoding Firefly Luciferase (*Luc*) and mCherry (Fig. 2d). After transfection into P19 cells, pPAS usage was determined by 3'RACE-PCR. All reporter constructs produced alternative 3'UTR isoforms of similar length as the endogenous genes (Additional file 1: Fig. S3A, S3B). Importantly, *Srsf3* KD caused similar changes in pPAS usage in reporter transcripts and their endogenous counterparts (Additional file 1: Fig. S3C, S3D). The effect of the *Srsf7* KD could be recapitulated in some but not all cases. Since the reporter constructs did not contain any introns, the observed 3'UTR changes indicated that 3'UTR-APA regulation by SRSF3 and SRSF7 is independent of splicing of the respective transcripts.

To evaluate the dose-response relationship of this regulation, we co-transfected the *Luc-Ddx21* reporter with increasing amounts of SRSF3-GFP and SRSF7-GFP expression plasmids (Fig. 2e, f). Overexpression (OE) of SRSF7-GFP had the opposite effect of *Srsf7* KD and resulted in a concentration-dependent increase of the shorter *Luc-Ddx21* 3'UTR isoform (Fig. 2e, f). OE of SRSF3-GFP did not further increase the levels of the longer *Luc-Ddx21* isoform, but here the levels of SRSF3 might not have been sufficient.

We next tested whether the extent of binding upstream of pPASs by either SRSF3 or SRSF7 directly affects pPAS usage. To do this, we converted all SRSF3 binding motifs into SRSF7 binding motifs (allSRSF7) and vice versa (allSRSF3) within a region of 110 nt immediately upstream of the pPAS in the *mCherry*- and *LUC-Ddx21* reporter genes (Additional file 1: Fig. S3E). Of note, this region does not contain UGUA motifs. Consistent with the results presented above, shifting the binding potential towards SRSF7 (allSRSF7) increased pPAS usage and shifting it to SRSF3 (allSRSF3) decreased pPAS usage (Fig. 2g, h, Additional file 1: Fig. S3F). Importantly, single point mutations in the

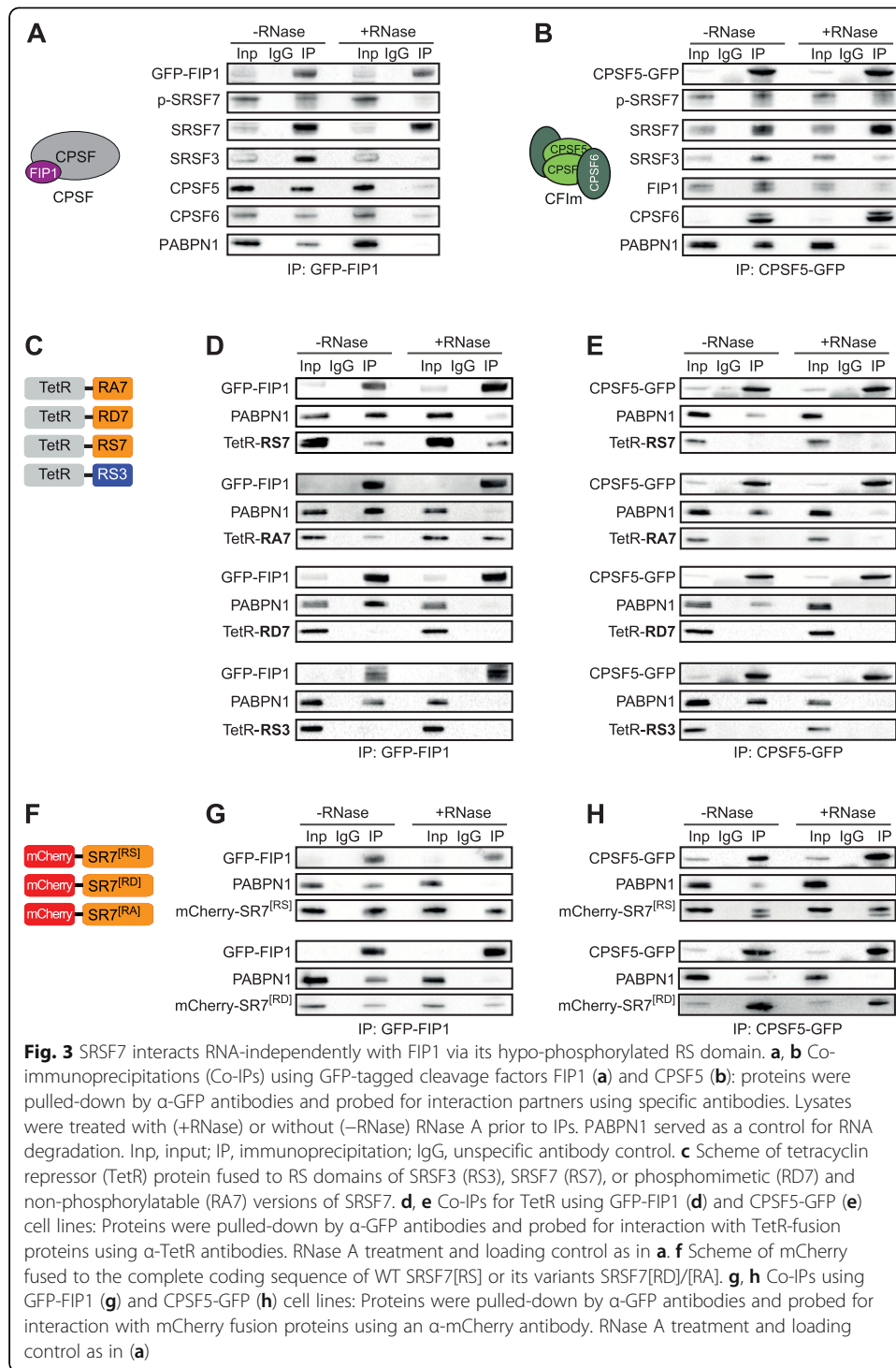
alternative CSE of the pPAS that strengthened it (AGUAAA to AAUAAA) completely abrogated dPAS usage, whereas its weakening (AGUAAA to AGUAAG) [49] abrogated pPAS usage (Fig. 2h, Additional file 1: Fig. S3F). Moreover, inserting one UGUA motif in the absence of any SRSF3 and SRSF7 binding sites also strongly increased pPAS usage (Fig. 2g, h). This suggests that an intermediate CSE and the absence of UGUA motifs at the pPAS allow SRSF7 and SRSF3 to modulate its usage in opposite directions. Whereas SRSF7 enhances pPAS usage, SRSF3 inhibits pPAS usage in a binding- and concentration-dependent manner.

SRSF7 interacts with FIP1 independently of RNA via its hypo-phosphorylated RS domain

The binding of SRSF3 and SRSF7 upstream of pPASs suggests that they might interact with the CPA machinery. In line with this possibility, quantitative mass spectrometry (MS) using TMT-labeling of immunopurified RNPs containing SRSF3-GFP identified several CPA factors, including components of the CPSF complex (FIP1, CPSF2, CPSF3, and WDR33) and CPSF5 from the CFIm complex (Additional file 1: Fig. S4A, Additional file 4: Table S3). To confirm these interactions, we performed semi-quantitative co-immunoprecipitations (Co-IPs) with and without RNase treatment. We focused hereby on CPSF5 and CPSF6 from the CFIm complex [28] and on FIP1 as representative of the CPSF complex, since CPSF6 and FIP1 both contain an RS-like domain (Additional file 1: Fig. S4B, S4C). We generated P19 cell lines expressing GFP-tagged CPSF5, CPSF6, and FIP1 from genomic loci (Additional file 1: Fig. S4D-G). Since CPSF5-GFP performed much better than CPSF6-GFP in initial pulldown tests (data not shown), we used CPSF5-GFP for the IP experiments.

CPSF5-GFP efficiently co-immunoprecipitated its CFIm complex partner CPSF6, while the association of both proteins with the CPSF complex protein FIP1 was RNase-sensitive (Fig. 3a, b). Both CPA factors co-immunoprecipitated SRSF3, but surprisingly, the signal disappeared after RNase treatment, suggesting an indirect association via co-bound (pre-)mRNAs. In contrast, SRSF7 was efficiently co-immunoprecipitated in the presence of RNase, supporting a direct interaction of SRSF7 with both CPSF5 and FIP1. The results were confirmed by reverse Co-IPs using GFP-tagged SRSF3 and SRSF7 [32] as baits (Additional file 1: Fig. S4H, S4I). These data suggest that binding of SRSF7 upstream of pPASs might enhance their usage through the recruitment of CPA factors, whereas competitive binding to the same sites by SRSF3 might impair their recruitment.

It was recently shown that the RS domain of CPSF6 is required for FIP1 recruitment and dPAS activation [17]. RS domains are also the main protein-protein interaction platforms of SR proteins [42], and their dimerization is regulated through phosphorylation/dephosphorylation of their serine residues [50]. To test whether SRSF7 recruits CFIm and FIP1 via its RS domain, we generated C-terminal fusions of the tetracycline repressor protein (TetR) with the RS domain of either SRSF3 (RS3) or SRSF7 (RS7). We included variants of SRSF7's RS domain in which all serine residues were exchanged for alanine (RA7) or aspartate (RD7) residues, thereby mimicking hypo- and hyper-phosphorylated RS domains, respectively (Fig. 3c). TetR constructs were transfected into GFP-FIP1 and CPSF5-GFP cell lines, and expression and phosphorylation of



all protein chimeras were verified by Western blot following phosphatase treatment (Additional file 1: Fig. S5A, S5B).

Co-IPs confirmed that the RS domain of SRSF7 (RS7) was sufficient to mediate interaction of TetR with GFP-FIP1 (Fig. 3d). Strikingly, this interaction was maintained with the RA7 variant but lost with the RD7 variant, suggesting that FIP1 interacts preferentially with the hypo-phosphorylated RS domain of SRSF7. In line

with this, SRSF7 that co-immunoprecipitated with GFP-FIP1 appeared hypo-phosphorylated as it did not react with mAb104, an antibody that recognizes exclusively hyper-phosphorylated domains of SR proteins (see p-SRSF7 in Fig. 3a) [51]. In contrast, the isolated RS domain of SRSF7 was not sufficient to promote interaction of TetR with CPSF5, regardless of its phosphorylation state (Fig. 3e). This suggested that interaction of SRSF7 with CPSF5 requires its RNA-binding domain. In line with the results described above, the isolated RS domain of SRSF3 (RS3) did not promote interaction of TetR with GFP-FIP1 or CPSF5-GFP in P19 cells (Fig. 3d, e).

To assess whether binding of SRSF7 to CPSF5 is also phosphorylation-dependent, we made full-length SRSF7 constructs (mCherry-SRSF7^[RS]) and replaced all serine residues within the RS domain with either alanine (mCherry-SRSF7^[RA]) or glutamate (mCherry-SRSF7^[RD]) residues (Fig. 3f). Expression of the fusion proteins was verified by confocal fluorescence microscopy and Western blot (Additional file 1: Fig. S5C, S5D). As expected, mCherry-SRSF7^[RS] and mCherry-SRSF7^[RD] localized to the nucleus, where they were enriched in nuclear speckles. However, mCherry-SRSF7^[RA] predominantly mislocalized to nucleoli likely due to the high positive charge of the RA domain and was therefore excluded from the Co-IP analyses (Additional file 1: Fig. S5D). Confirming that FIP1 preferentially associates with hypo-phosphorylated SRSF7, GFP-FIP1 co-immunoprecipitated with mCherry-SRSF7^[RD] less efficiently than mCherry-SRSF7^[RS] (Fig. 3g). Surprisingly, however, CPSF5-GFP co-immunoprecipitated more efficiently with mCherry-SRSF7^[RD] than mCherry-SRSF7^[RS], suggesting that CPSF5 rather associates with hyper-phosphorylated SRSF7 (Fig. 3h, see also p-SRSF7 in Fig. 3b).

The CFIm complex is a hetero-tetramer of two CPSF5 and two CPSF6/7 subunits [11]. We hypothesized that SRSF7 might replace CPSF6 in a subpopulation of CFIm complexes, because (i) CPSF6 and SRSF7 have similar domain structures, (ii) CPSF6 interaction with CPSF5 also requires its RRM domain, and (iii) CPSF6 also recruits FIP1 via its hypo-phosphorylated RS domain [16, 17]. To test this, we depleted CPSF6 from GFP-FIP1-expressing cell lines and performed Co-IPs. The interaction between FIP1 and SRSF7 was unchanged, while the interaction with CPSF5 was lost (Additional file 1: Fig. S5E-G). This confirmed that CPSF5 interacts with FIP1 via CPSF6 and suggested that SRSF7 and FIP1 interact directly and do not require CFIm.

Taken together, these results suggest that the RS domain of SRSF7 is sufficient for recruiting FIP1. The interaction does not require CFIm or bound RNA but dephosphorylation of SRSF7's RS domain. In contrast, interaction of SRSF7 with CPSF5 requires its RNA-binding domain and hyper-phosphorylation of its RS domain.

Two SRSF7-specific protein features promote its interaction with CPA factors

To understand why SRSF7 interacts directly with CPA factors while SRSF3 does not, we asked whether some features specific to SRSF7 might confer this recruitment. SRSF3 and SRSF7 are closely related and structurally similar, with nearly identical RRM domains [36]. However, SRSF7 contains an additional Zn knuckle for RNA binding [40] and has a longer RS domain, which is interrupted by a stretch of 27 amino acids (27aa) enriched in hydrophobic residues [52] (Fig. 4a). We have recently shown that

inclusion of this 27aa stretch is regulated by alternative splicing and promotes oligomerization of SRSF7, leading to the formation of nuclear bodies involved in auto-regulation of SRSF7 expression [41]. Moreover, inactivation of the Zn knuckle in SRSF7 changes its RNA-binding preference from GAY to CNYC, the binding motif of SRSF3 [39, 41]. To assess the impact of these SRSF7-specific protein features (27aa stretch, Zn knuckle) on the interaction with CPA factors, we used stable P19 cell lines expressing GFP-tagged SRSF7 variants either lacking the 27aa stretch (Δ 27aa) or containing an inactive Zn knuckle (mutZn) [41] (Fig. 4b). Expression, subnuclear localization and phosphorylation of the SRSF7 variants were verified by Western blot and confocal fluorescence microscopy (Additional file 1: Fig. S6A, B, D).

Strikingly, both mutants interacted much less with CPSF5 and FIP1 as determined by Co-IPs with RNase treatment (Fig. 4c), suggesting that the 27aa stretch and the Zn knuckle both contribute to SRSF7's interaction with the CPA factors. To test whether these features are sufficient to promote the interactions, we transferred the 27aa stretch and the Zn knuckle (separately and in combination) to the corresponding positions in SRSF3 in a domain-swap experiment (Fig. 4d). The chimeric constructs were transiently transfected into P19 cells, and their expression, subcellular localization, and phosphorylation were verified (Additional file 1: Fig. S6C, E, F). Chimeric SRSF3 proteins were often expressed at much lower levels than wild-type SRSF3-GFP in Co-IP transfection experiments (Fig. 4e), but taking this into account, the Co-IPs suggest that inserting the 27aa stretch alone into SRSF3 (SRSF3-27aa) increases its association with CPSF5, CPSF6, and FIP1 (Fig. 4e, Additional file 1: Fig. S7A). Insertion of the Zn knuckle alone (SRSF3-Zn) or in combination with the 27aa stretch (SRSF3-ZnF + 27aa) did not enhance CPA factor interactions visibly (Fig. 4e, Additional file 1: Fig. S7A). Despite this, transient OE of SRSF3-ZnF + 27aa chimeras increased pPAS usage in SRSF7 APA targets, similar to OE of SRSF7 (Figs. 2e and 4f, Additional file 1: Fig. S7C). When endogenous *Srsf7* was concomitantly depleted, this pPAS enhancement was abolished (Fig. 4f, Additional file 1: Fig. S7B, C). In contrast, OE of SRSF3 containing the 27aa hydrophobic stretch alone had no effect on pPAS usage, indicating that both sequence-specific binding via the Zn knuckle and enhanced FIP1 interaction via the 27aa stretch are required for pPAS activation. Based on these targeted deletions and domain-swap experiments, we conclude that the hydrophobic 27aa stretch and the Zn knuckle, which are absent in SRSF3, contribute to the recruitment of CPA factors and pPAS enhancement and therefore to the functional diversification of SRSF3 and SRSF7 in APA.

SRSF7 levels decrease and 3'UTRs are globally extended during neuronal differentiation

The CPA factor FIP1 was recently implicated in the maintenance of pluripotency and renewal of mouse embryonic stem cells. During differentiation, FIP1 is downregulated with a concomitant global 3'UTR lengthening [27]. To test whether SRSF3 and SRSF7 might also regulate pluripotency via APA, we differentiated pluripotent P19 cells into neuronal cells [53]. On day 8 of differentiation, the cells had adopted the characteristic neuronal morphology reflected by the presence of multiple neurites (Fig. 5a), expression of the neuronal marker Synapsin 1, and complete loss of the pluripotency transcription factor OCT4 (Fig. 5b). Importantly, the protein levels of FIP1 and SRSF7

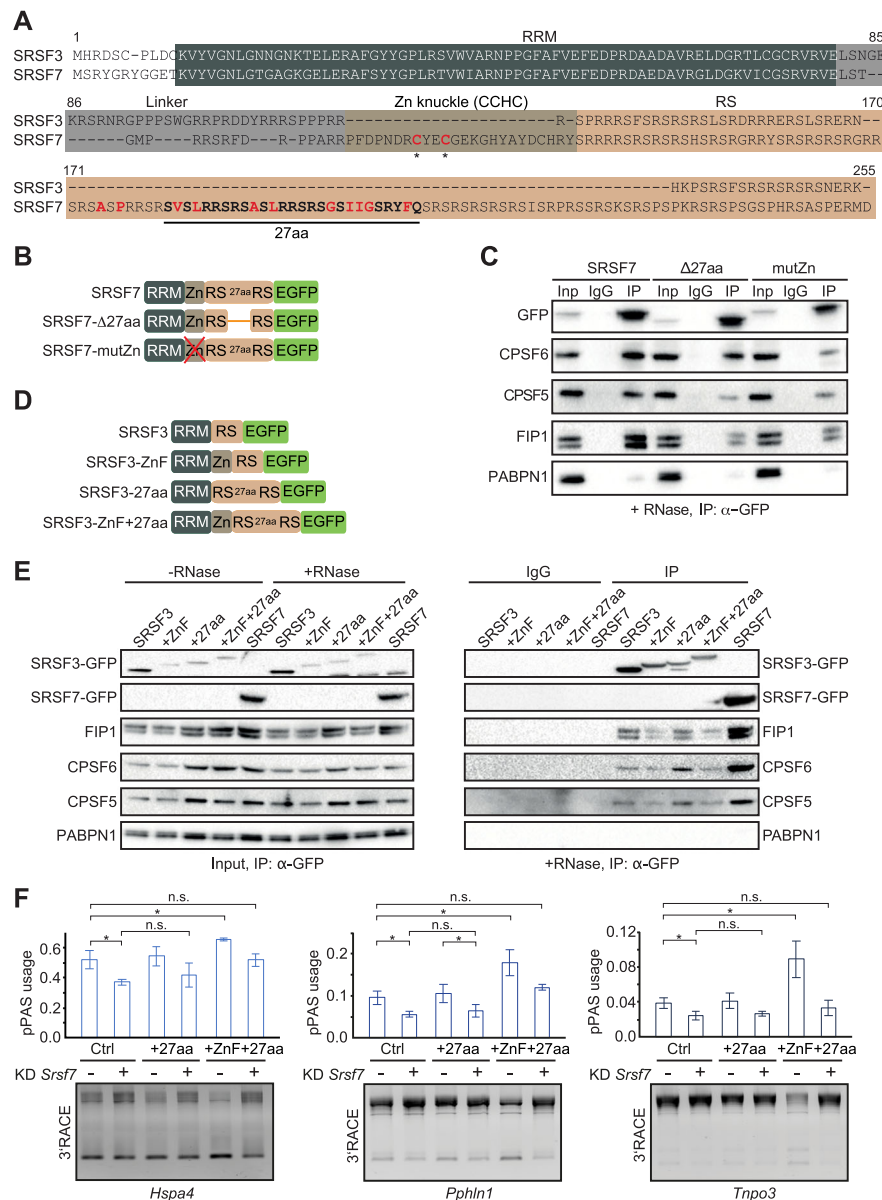


Fig. 4 Two SRSF7-specific protein features promote its interaction with CPA factors. **a** Alignment (Clustal Omega) of SRSF3 (amino acids [aa] 1–164) and SRSF7 proteins (aa 1–238). Protein domains are highlighted in dark gray (RRM), light gray (linker domain), olive (CCHC-type Zn knuckle), and light brown (RS domain). The 27 aa stretch is highlighted in bold with hydrophobic residues marked in red. Cysteine residues mutated to alanine in Zn knuckle in SRSF7-mutZn are marked with asterisks. **b** Scheme of GFP-tagged SRSF7 mutants stably expressed in P19 cells. **c** Co-IPs using GFP-tagged SRSF7/SRSF7-Δ27aa/SRSF7-mutZn cells with RNase A treatment: Proteins were pulled-down by α-GFP antibodies and probed for CPA factors using specific antibodies. PABPN1 served as a control for RNA degradation. IP, immunoprecipitation; IgG, unspecific antibody control. **d** Scheme of GFP-tagged SRSF3 chimeras harboring Zn knuckle, 27 aa stretch of SRSF7, or both. **e** Co-IPs using GFP-tagged SRSF3 chimeras: Proteins were pulled-down by α-GFP antibodies and probed for CPA factors using specific antibodies. Lysates were treated with (+RNase) or without (–RNase) RNase A prior to IPs. PABPN1 served as a control for RNA degradation. Inp, input. Samples without RNase A treatment are shown in Fig. S7A. **f** Quantification of 3'RACE-PCRs ($n = 3$) of *Hspa4*, *Pph1n1*, and *Tnpo3* transcripts in Ctrl and *Srsf7* knockdown P19 cells transiently overexpressing SRSF3-27aa and SRSF3-27aa + ZnF chimeric proteins. Data are represented as mean ± standard deviation of mean. Student's *t* test, **P* value < 0.05

decreased by ~3-fold, while CPSF6 and SRSF3 levels appeared unchanged (Fig. 5b). We performed RNA-seq from undifferentiated (Undiff) and differentiated (Diff) cells (3 replicates) and analyzed the data with DESeq2 [54] and DaPars [43]. *Srsf7* and *Fip1l1* (encoding the protein FIP1) mRNA levels were only slightly decreased in differentiated cells (1.65- and 1.33-fold; Fig. 5c, Additional file 5: Table S4), suggesting that their low protein levels are caused by reduced translation and/or protein stability. In agreement with previous studies [55, 56], we observed a global shift to dPAS usage in differentiated P19 cells (Fig. 5d, Additional file 6: Table S5). This suggested that in pluripotent P19 cells, pPASs are preferentially used. Consistent with SRSF7 being a pPAS-enhancing protein, the majority of transcripts with extended 3'UTRs in differentiated cells showed a similar trend after *Srsf7* KD, although most changes did not pass the significance thresholds of our DaPars analysis (Fig. 5e, f). Importantly, transcripts with extended 3'UTRs in differentiated cells were highly enriched for binding of SRSF7 at their pPASs in undifferentiated cells (Fig. 5g). In contrast, SRSF3 binding was not enriched at these pPASs and the targets seemed anti-correlated after *Srsf3* KD (Fig. 5g, Additional file 1: Fig. S8A).

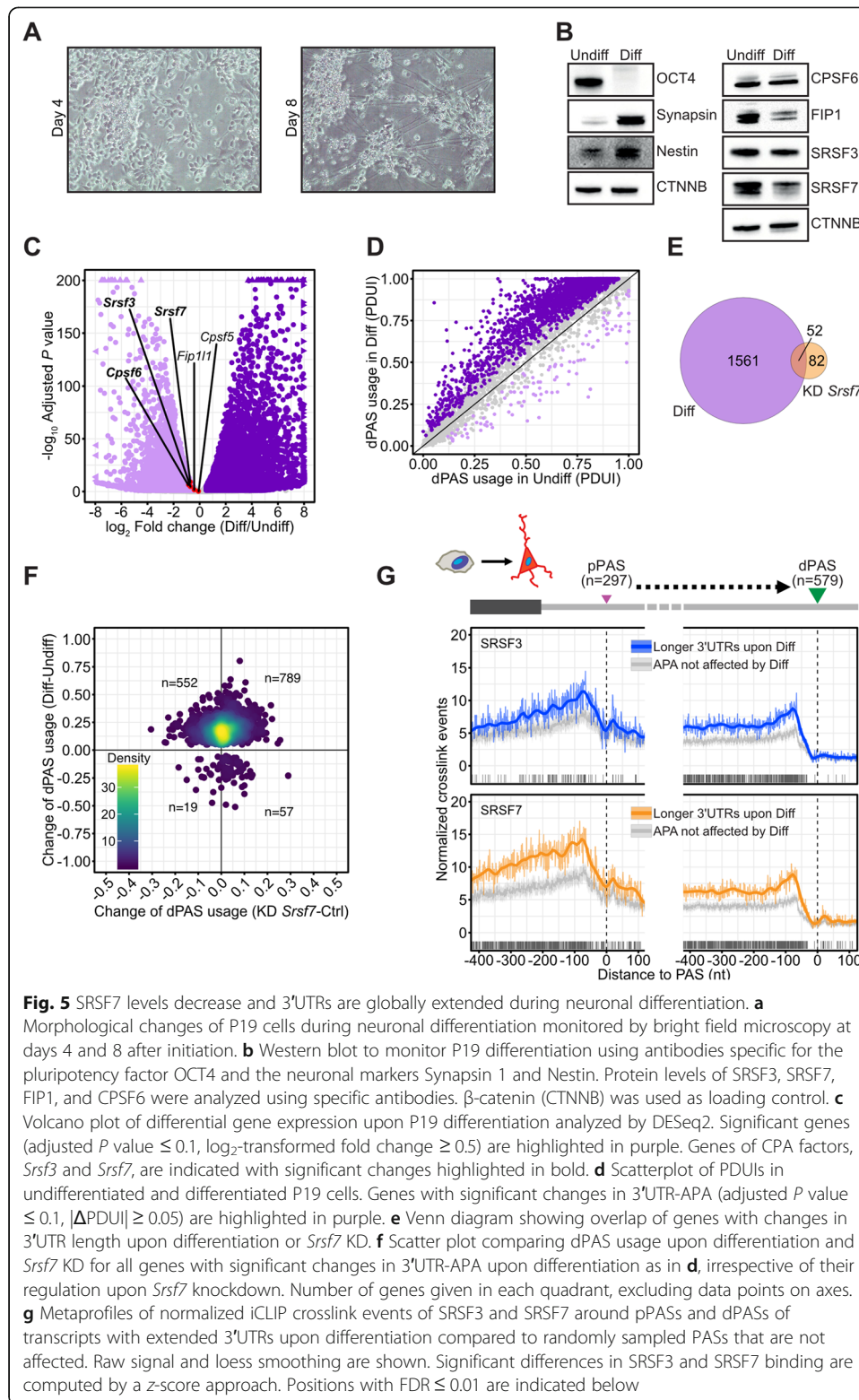
We conclude that high concentrations of factors that enhance pPAS usage, such as FIP1 and SRSF7, promote their usage under certain conditions, such as pluripotency. The downregulation of these factors in differentiated cells leads to the increased usage of the stronger dPASs and the expression of transcripts with longer 3'UTRs.

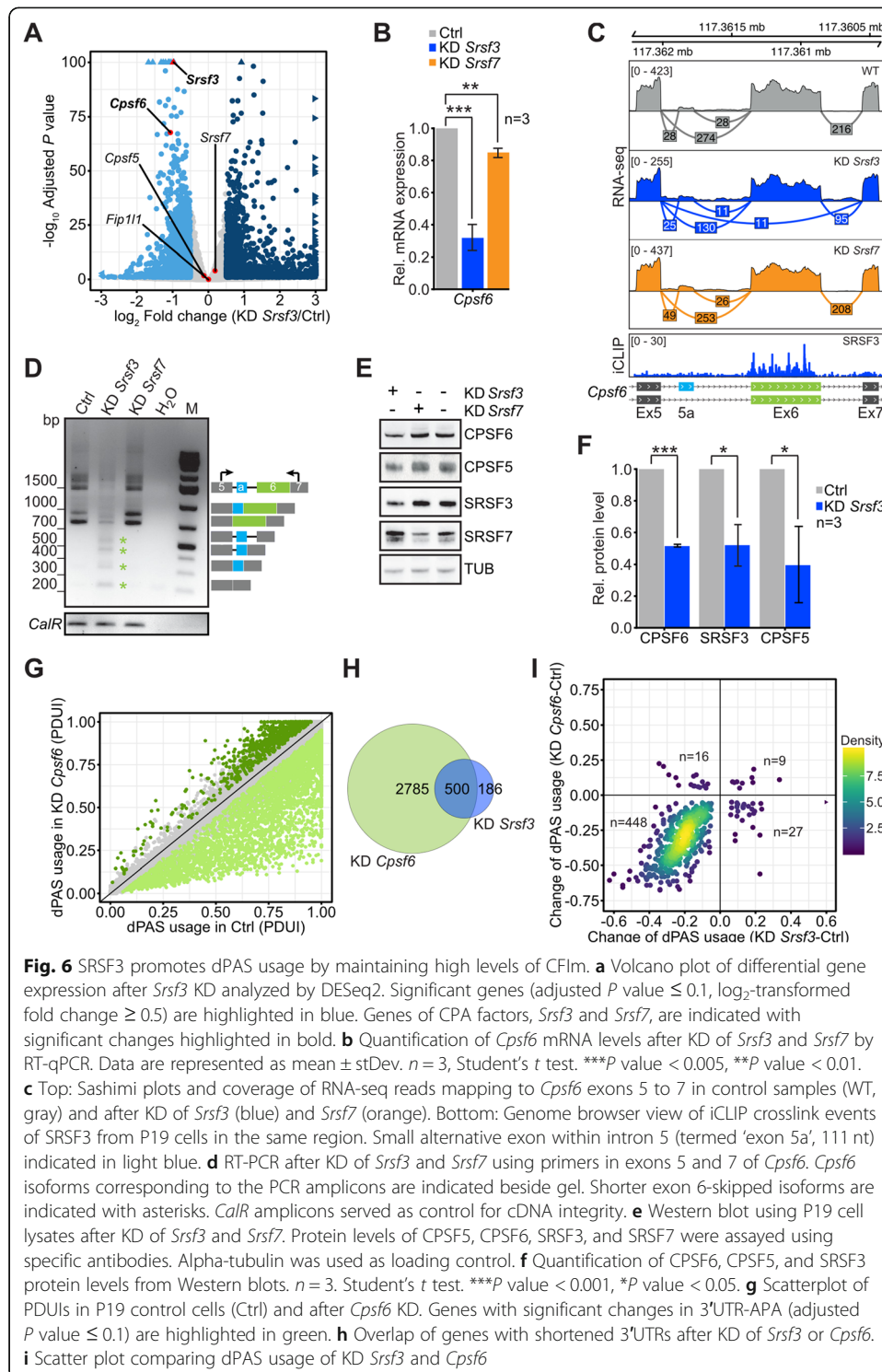
SRSF3 promotes dPAS usage by maintaining high levels of CFIm

Our data so far suggest that SRSF3 might prevent SRSF7 from recruiting CPA factors to some pPASs through competitive binding. But this is likely not the sole mechanism by which SRSF3 regulates 3'UTR-APA, since the number of genes with changes in 3' UTR length is 5-fold higher upon *Srsf3* KD than *Srsf7* KD (Fig. 1c). One possibility would be that SRSF3 regulates 3'UTR-APA also indirectly through modulating the levels of CPA factors.

To investigate this, we quantified changes in the expression of CPA factors in our RNA-seq data using DESeq2 (Additional file 1: Fig. S8B). Strikingly, *Srsf3* KD reduced the levels of *Cpsf6* transcripts by 2-fold, while no other CPA factor was affected (Fig. 6a). Decreased *Cpsf6* transcript levels were confirmed by RT-qPCR (~3-fold; Fig. 6b). *Srsf7* KD had only a minor effect on *Cpsf6* levels and did not affect the levels of any other CPA factor (Fig. 6b, Additional file 1: Fig. S8B).

To investigate how SRSF3 depletion decreases *Cpsf6* mRNA levels, we analyzed its splicing pattern using RNA-seq data (Fig. 6c). We observed multiple splicing alterations within the *Cpsf6* transcripts upon *Srsf3* KD, such as increased inclusion of a small alternative exon within intron 5 (termed 'exon 5a', 111 nt), the skipping of exon 6, and the retention of intron 6 (Fig. 6c). Transcripts including exon 5a likely encode an alternative CPSF6 protein isoform of 72 kDa (CFIm-72) whose functional difference from the main isoform of 68 kDa (CFIm-68) is not known. Skipping of exon 6 introduces a frameshift resulting in non-productive transcripts. iCLIP data showed that SRSF3 bound massively to exon 6, suggesting that it directly promotes its inclusion (Fig. 6c).





The appearance of various shorter *Cpsf6* transcript isoforms with skipped exon 6 after *Srsf3* KD was confirmed using semi-quantitative RT-PCR and sequencing (see asterisks in Fig. 6d). This was accompanied by a drastic reduction in the levels of both transcript isoforms encoding full-length CPSF6 proteins (68 and 72 kDa). Importantly, *Srsf3* KD—but not *Srsf7* KD—also resulted in ~ 2 -fold decreased CPSF6 protein levels

(Fig. 6e, f). CPSF5 was downregulated to a similar extent, in line with previous studies reporting that both proteins stabilize each other [11, 14].

In human cells, it was shown that CFIm binds preferentially upstream of dPASs and enhances their usage through a direct recruitment of FIP1 [17]. As a result, CPSF5 and CPSF6 depletion both lead to a global switch to pPAS usage [29, 30]. RNA-seq after KD of *Cpsf6* from mouse P19 cells also revealed a globally enhanced pPAS usage (Fig. 6g, Additional file 1: Fig. S8C, Additional file 8: Table S7) and clearly mimicked *Srsf3* KD by affecting the majority of SRSF3 APA targets in the same way (457 out of 686, 67%; Fig. 6h, i). Yet, a far greater number of transcripts got shorter 3'UTRs upon *Cpsf6* KD than *Srsf3* KD (2711 vs. 579; Additional file 1: Fig. S8D), presumably due to a more efficient depletion of CPSF6 upon *Cpsf6* KD compared to *Srsf3* KD (80% vs. 50% depletion). The similar regulation of 3'UTR-APA by SRSF3 and CPSF6 was validated for three target genes by 3'RACE-PCR (Additional file 1: Fig. S8E).

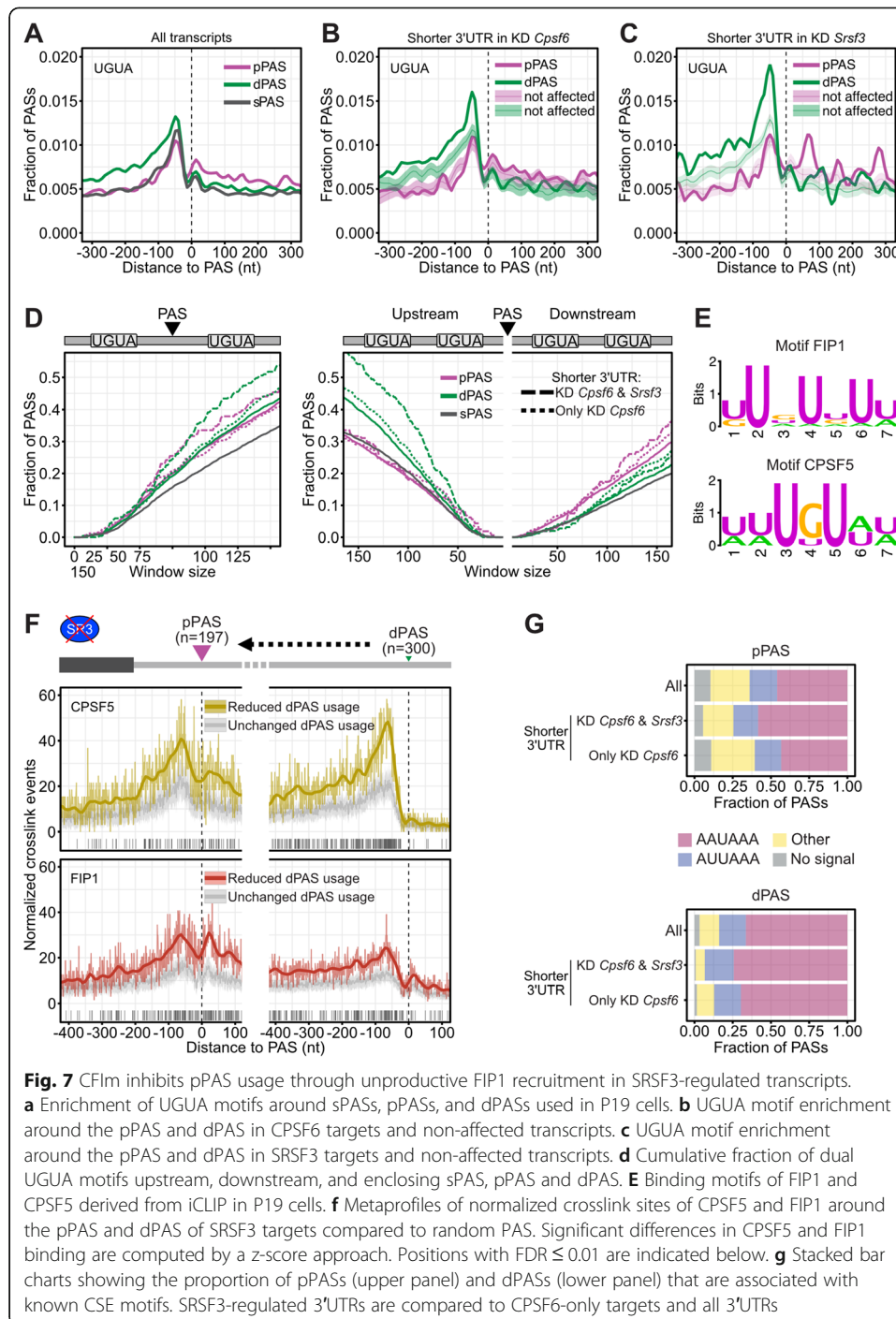
To confirm that SRSF3 affects 3'UTR length indirectly via CPSF6, we transiently overexpressed CPSF6-myc in SRSF3-depleted cells to rescue CPSF6 levels (Additional file 1: Fig. S8F). 3'RACE-PCR revealed that 3'UTR shortening of *Ddx21* and *Anp32e* could not be rescued by ectopic CPSF6 expression, suggesting that these transcripts are direct SRSF3 targets. *Rab11a*, *Phpln1*, and *Tnpo3* reacted to CPSF6 OE with increased dPAS usage, which partially counteracted the 3'UTR shortening triggered by the *Srsf3* KD (Additional file 1: Fig. S8G). However, the effects were small and the rescue was incomplete, likely because the CPSF5 levels remained low in the transient CPSF6 OE (Additional file 1: Fig. S8F).

These observations argue for two distinct modes of 3'UTR-APA regulation by SRSF3. On the one hand, SRSF3 inhibits pPAS usage directly through competitive binding with pPAS enhancer proteins. On the other hand, SRSF3 indirectly promotes dPAS usage by ensuring the productive splicing of mRNAs encoding CPSF6, a key component of the dPAS enhancer CFIm.

CFIm inhibits pPAS usage through unproductive FIP1 recruitment in SRSF3-regulated transcripts

SRSF3 3'UTR-APA targets appear especially sensitive to reduced levels of CPSF6. To better understand their regulation by CFIm, we analyzed some of their features. We first compared CFIm binding motifs at pPASs and dPASs, since CFIm binding to UGUA motifs upstream of dPASs (–80 nt to –40 nt) had been shown to enhance their usage [17]. In line with this, CPSF6 targets had a greater tendency than all transcripts to harbor at least one UGUA motif upstream of their dPAS (Fig. 7a, b, Additional file 1: Fig. S9A). But this tendency was even higher for SRSF3 targets, suggesting that they rely on CFIm to enhance dPAS usage (Fig. 7c).

Strikingly, in SRSF3 targets, UGUA motifs appeared equally distributed on either side of the pPASs, with a strong enrichment around 65 nt downstream. Given that each of the two CPSF5 subunits in a CFIm complex can bind one UGUA motif while the intervening RNA loops around the CFIm heterotetramer [15, 57], this raised the intriguing possibility that CFIm blocks these pPASs by excluding some of the sequences necessary for their activation such as USEs, CSEs, and DSEs. In line with this notion, we found



that SRSF3 targets indeed had a greater tendency to have UGUA pairs flanking their pPASs compared to CPSF6 targets and all transcripts (Fig. 7d, Additional file 1: Fig. S9B). SRSF3 targets also more frequently harbored UGUA pairs downstream of their pPASs, which could inhibit their usage, for instance by excluding DSEs [15, 57]. Finally, the SRSF3 targets also showed a much greater tendency to contain UGUA pairs upstream of their dPASs, which should enhance dPAS usage [17]. Altogether, the UGUA motif distribution suggests that SRSF3 targets are especially sensitive to reduced CFIm levels because

CFIm might play a dual role on these transcripts: it activates their dPAS and at the same time blocks their pPAS.

To assess whether the UGUA motif distribution is reflected by CFIm binding and FIP1 recruitment, we performed iCLIP of GFP-FIP1 and CPSF5-GFP from P19 cell lines (6–7 replicates; Additional file 1: Fig. S10A, S10B). Merging replicates, we obtained 1,851,266 unique crosslink events for CPSF5 and 3,759,237 for FIP1 (Additional file 1: Table S8). Pentamer enrichment analysis retrieved the expected binding motifs—UGUA for CPSF5, and UG-rich sequences for FIP1 (Fig. 7e). Indeed, cross-linking of both CPSF5 and FIP1 mirrored the UGUA motif distribution, with peaks directly upstream of pPASs and dPASs, in line with previous studies [14, 29]. But they also displayed an additional peak downstream of pPASs, which has not been reported previously (Additional file 1: Fig. S10C). Importantly, binding of both CPSF5 and FIP1 was strongly enriched on either side of pPASs and upstream of dPASs in SRSF3 targets, whereas CPSF6 targets showed only a slight enrichment for both proteins upstream of dPASs (Fig. 7f, Additional file 1: Fig. S10D). Given that many pPASs in SRSF3 targets are suppressed under normal conditions (where iCLIP was performed), recruitment of FIP1 to those pPASs might be unproductive, i.e., does not result in cleavage. To confirm this, we divided all transcripts with long 3'UTRs (unused pPAS) and short 3'UTRs (used pPAS) in control cells and compared their binding patterns of FIP1 and CPSF5 (Additional file 1: Fig. S10E). Interestingly, both CPSF5 and FIP1 bound downstream of suppressed pPASs, suggesting a mechanism of pPAS suppression through unproductive CFIm and FIP1 binding. Our data suggested that for many CPSF6 APA targets the promotion of dPAS usage is sufficient to generate transcripts with long 3'UTRs. However, in a subset of these (i.e., SRSF3 APA targets), additional mechanisms appear to be used to also suppress pPAS usage.

To test whether those suppressed pPASs are stronger than usual, we compared the frequency of CSE variants. Indeed, SRSF3 APA targets more often harbored the two main CSE variants (AAUAAA and AUUAAA) at their pPASs compared to CPSF6 APA targets (75% vs. 60%; Fig. 7g). Although the same was true for their dPASs (95%), pPASs are transcribed first; hence, they need to be suppressed to generate transcripts with long 3'UTRs.

Altogether, our data provide evidence for two modes by which CFIm can regulate 3' UTR-APA. CFIm binding can either inhibit or enhance PAS usage depending on the positioning of UGUA motifs and CFIm binding. CFIm complexes bound to suppressed pPASs still recruit FIP1 and likely other CPA factors, but cleavage does not occur, suggesting that the assembled cleavage complexes are inactive. Indirect SRSF3-regulated targets are more sensitive to low CFIm levels because they rely on pPAS inhibition and dPAS enhancement by CFIm to ensure the generation of long 3'UTRs. Additional pPAS inhibition might be necessary because SRSF3-sensitive pPASs are stronger compared to CPSF6-only targets, which do not show this dual CFIm dependency.

Discussion

The usage of alternative PASs can be modulated, either positively or negatively, by regulatory RBPs. RBPs either directly compete with core CPA factors for binding to regulatory elements or they bind to sequences outside of the PAS region (reviewed in [19, 22, 58]). We show here that SRSF3 and SRSF7 modulate PAS usage directly by

binding upstream of regulated pPASs. Despite being very closely related, SRSF3 and SRSF7 affect 3'UTR length in opposite directions. This is due to slight differences in the domain architecture of both SR proteins. We find that the direct modulation of pPAS usage by SRSF3 and SRSF7 requires CSEs of intermediate strength, is independent of splicing and occurs in a concentration- and binding-dependent manner. In addition, SRSF3 affects 3'UTR-APA also indirectly, by maintaining high levels of CFIm through alternative splicing of the *Cpsf6* mRNA. Based on the data presented here, we propose the following model for 3'UTR-APA modulation by SRSF3 and SRSF7.

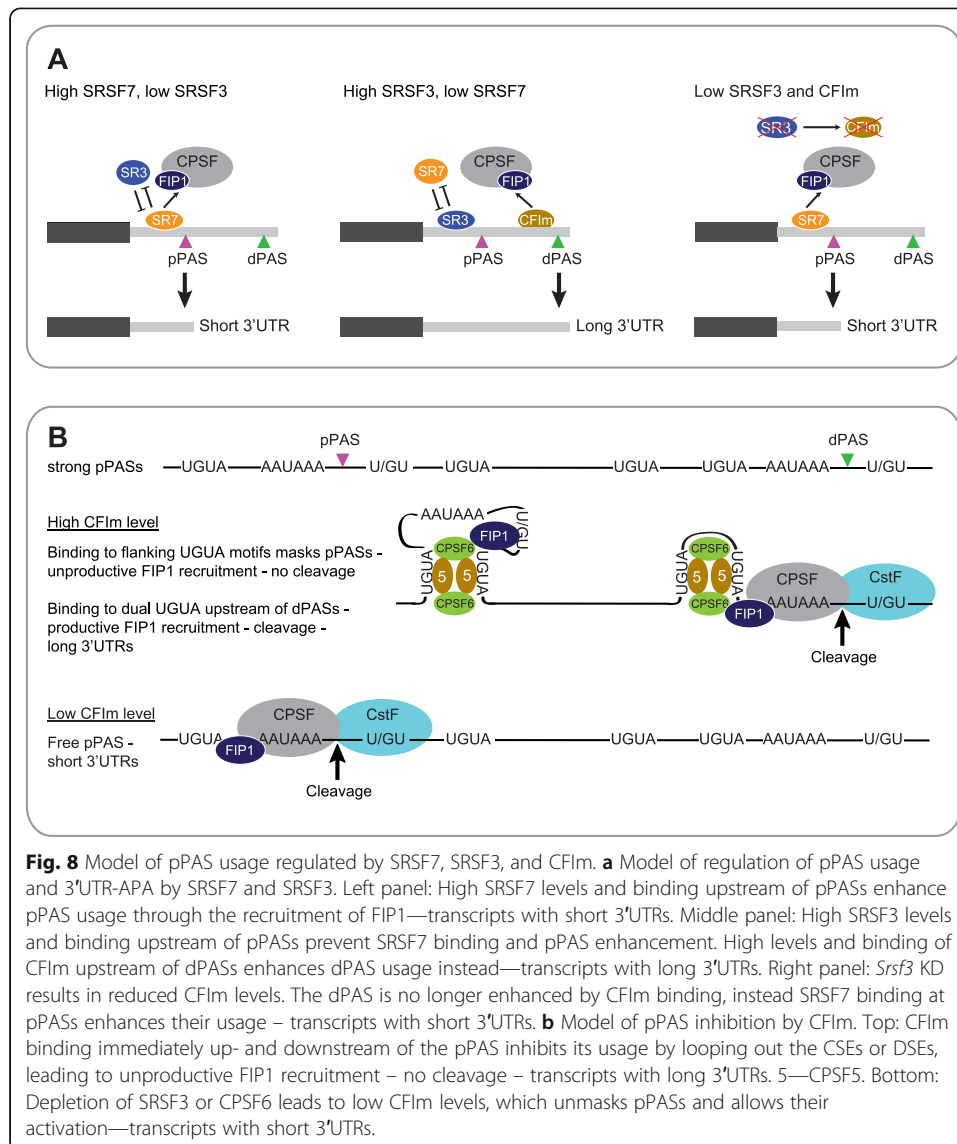
When SRSF7 is present at high levels, such as in pluripotent cells or certain cancer cells [59], it binds upstream of pPASs and recruits FIP1, thereby acting as a sequence-specific enhancer of pPAS usage and promoting transcripts with short 3'UTRs (Fig. 8a, left panel). SRSF3 also binds upstream of pPASs but, unlike SRSF7, it cannot recruit FIP1. Thus, SRSF3 binding prevents activation of these pPASs by impairing the association of SRSF7. When SRSF7 levels are low, e.g., in differentiated cells, increased SRSF3 binding to pPASs favors the usage of dPASs and the generation of long 3'UTRs (Fig. 8a, middle panel). When SRSF3 levels are low, as it occurs for example in progressive liver disease [60], SRSF7 can bind more to pPASs. In addition, the levels of CFIm decrease due to unproductive splicing of *Cpsf6*, which prevents the activation of dPASs. Together, this favors pPAS usage and the generation of transcripts with short 3'UTRs (Fig. 8a, right panel).

Apart from only preventing the association of SRSF7 with pPASs, SRSF3 binding might also actively inhibit their usage, similar to U1 snRNP [61, 62]. U1 snRNP inhibits polyadenylation through the inactivation of PAP by U1-70K and U1A, which both contain a PAP-inhibitory domain (PID) [63]. Interestingly, SRSF3 also contains a predicted PID within its RS domain and might inhibit PAP in these inactive CPA complexes [64], but further studies are required to test this hypothesis.

We show that FIP1 recruitment by SRSF7 occurs via direct protein-protein interaction. It is independent of CFIm, but dependent on the Zn knuckle and the hydrophobic 27aa stretch within the RS domain of SRSF7, which are absent in SRSF3. The interaction requires the RS domain of SRSF7 to be hypo-phosphorylated, but is independent of its RNA-binding domain, similar to what was shown for CPSF6 [17]. This means that SRSF7 can simultaneously bind to pPASs in pre-mRNAs and to FIP1 and thereby recruit FIP1 to bound pPASs, e.g., those without UGUA motifs like the *Ddx21* 3'UTR. Dephosphorylation of SRSF7 might occur during CPA, since our reporter experiments suggest that 3'UTR-APA regulation by SRSF7 is independent of splicing. In line with this, the SR protein phosphatase PP1 was also found in purified CPA complexes [65, 66].

We also find that indirectly SRSF3-regulated pPASs are more sensitive to reduced CFIm levels. This might be due to the fact that these pPASs are stronger than average pPASs and contain mostly the canonical CSE variants (Figs. 1h and 7g) [24, 47]. Since pPASs are transcribed first, strong pPASs should in principle be used by default when the core CPA machinery is present in sufficient levels. However, most of these SRSF3 targets do harbor long 3'UTRs in P19 cells under normal conditions. This suggests that in addition to the enhancement of their dPASs, their pPASs must also be inhibited.

Three models have been proposed for CFIm-mediated pPAS suppression. (i) CFIm binds to UGUA motifs that overlap with the cleavage site, thereby preventing PAS



usage [17, 67]. (ii) CFIm binds one UGUA motif upstream of a pPAS and a second UGUA upstream of a dPAS, thereby looping out a big part of the 3'UTR [15, 29]. (iii) CFIm binds to suboptimal target sites, e.g., non-UGUA motifs, that are more prevalent at pPASs, which would block productive CPSF recruitment [14]. Only the first model was experimentally verified [17, 67].

Our data suggest a fourth mode by which CFIm can inhibit pPAS usage when the pPAS is enclosed by or directly followed by a pair of UGUA motifs. This is in line with the second model implying that large regions of the 3'UTR including the pPAS loop around the CFIm tetramer [15, 29]. Our data suggest that CFIm binding at pPASs can also hide short 3'UTR regions that are important for pPAS activation, e.g., CSEs, DSEs, or the cleavage site (Fig. 8b). This is based on our findings that (i) SRSF3-sensitive pPASs are often flanked or followed by UGUA pairs, (ii) SRSF3-sensitive pPASs have enriched CPSF5 crosslinks flanking them, and (iii) these pPASs are used with low efficiency under the conditions where iCLIP was performed, but become activated when

SRSF3 and CFIm levels are reduced. This implies that for some transcripts CFIm plays a dual role in promoting the generation of long 3'UTRs; it binds to pairs of UGUA motifs upstream of dPASs and enhances their usage by recruiting FIP1 [17] (Fig. 8a, middle panel), but it also binds to UGUA pairs downstream of or enclosing strong pPASs to inhibit their cleavage (Fig. 8b). This differential sensitivity of pPASs to fluctuations in CFIm levels provides cells with the possibility to regulate 3'UTR-APA of specific subsets of transcripts, e.g., to regulate senescence [33].

Surprisingly, we found that FIP1 and likely other CPA factors are also recruited to unused pPASs, likely via CFIm, but cleavage does not occur at these sites. Thus, these pPAS-bound CPA complexes must be inactive. In favor of this model, it was recently shown that in HSV-1-infected cells, the immediate early protein ICP27 interacts with FIP1 and induces the assembly of a dead-end 3' processing complex on the mRNA [68]. Moreover, the U1 snRNP complex that assembles at intronic PASs to suppress premature polyadenylation also contains all CPA factors including FIP1 and CFIm, but cleavage also does not occur at these sites [69]. This suggests that the particular configuration of CFIm bound to suppressed pPASs causes the assembly of inactive CPA complexes that are unable to complete cleavage and polyadenylation, but they protect pPASs from recognition by pPAS enhancers (Fig. 8b). One possibility would be that suppressive pPAS CFIm complexes contain CPSF7, similar to suppressive U1-CPA complexes that are enriched in CPSF7 [69]. CPSF7 does not enhance CPA and its binding sites are much less predictive of PAS usage than CPSF6 binding [14, 29]. Moreover, CPSF7 has a slightly different mode of RNA binding than CPSF6 and does not interact with the pPAS enhancer SRSF7 [70]. Finally, depletion of CPSF7 does not cause a global switch to pPAS usage in HEK293 cells [13, 69]. Although this remains to be confirmed in other cellular systems, our data suggest that CPSF6 regulates fewer targets through pPAS inhibition than through dPAS enhancement.

Our study identifies SRSF3 as a critical regulator of CFIm levels. We show that SRSF3 binds massively to exon 6 in the *Cpsf6* pre-mRNA and enhances its inclusion. Upon *Srsf3* depletion, exon 6 is skipped and the resulting transcripts containing premature stop codons are unstable. Low levels of functional *Cpsf6* transcripts also reduce the levels of CPSF6 protein. Moreover, we find that CPSF5 is co-depleted, indicating that both proteins need to form a tetramer to stabilize each other [11, 14].

Both SRSF3 and SRSF7 were previously identified as interactors of CPSF6 using its RS-like domain in a yeast two-hybrid screen [16]. We confirm here the protein-protein interaction between CFIm and SRSF7 in P19 cells, but its interaction with SRSF3 requires simultaneous binding to the same (pre-)mRNAs. The interaction of CFIm with SRSF7 seems to be different from its interaction with FIP1. Here, binding requires the RNA-binding domain of SRSF7 and RS domain hyper-phosphorylation. This implies that SRSF7 cannot interact with CFIm when it is bound to RNA, and hence cannot recruit CFIm to pPASs. But given that SRSF7 and CPSF6 have a very similar domain structure, hyperphosphorylated SRSF7 might form heterotetramers with CPSF6 and CPSF5, whose function remains to be determined.

SRSF7 was previously implicated in the enhancement of polyadenylation of subgenomic transcripts of retroviruses, such as human immunodeficiency virus (HIV) and Rous sarcoma virus (RSV) [71, 72]. We show here that SRSF7 also enhances pPAS usage in cellular mRNAs in a sequence-specific and concentration-dependent manner

through the recruitment of FIP1. This is reminiscent of its functions in splicing, where sequence-specific binding of SRSF7 is required to recruit the spliceosome to splice sites to enhance their usage in a concentration-dependent manner [39]. The apparent discrepancy between the considerable pPAS and FIP1 binding of SRSF7 and relative few APA targets upon *Srsf7* KD is likely due to the partial redundancy of SR proteins that bind to purine-rich motifs, e.g., SRSF6 and SRSF1, which have also been suggested to enhance PAS usage [32, 71, 73].

Conclusion

Altogether, our data reveal novel mechanistic insights into the direct and indirect regulation of 3'UTR-APA by SRSF3 and SRSF7 in opposite directions, and into how CFIm regulates pPAS usage. Binding of SRSF3 at suppressed pPASs and binding of SRSF7 at activated pPASs followed by NXF1 recruitment might sort mRNAs with short and long 3'UTRs into distinct export-competent mRNPs. These mRNPs could follow distinct routes in the cytoplasm, for example being transported at different subcellular locations for their translation.

Methods

Generation and cultivation of stable BAC P19 cell lines

Murine P19 WT cells were purchased (Sigma-Aldrich) and grown under humidified conditions at 5% CO₂ and 37 °C in DMEM GlutaMAX (ThermoFisher Scientific), supplemented with 100 U/ml Penicillin-Streptomycin (ThermoFisher Scientific) and 10% (v/v) heat inactivated fetal bovine serum (ThermoFisher Scientific), on dishes coated with 0.1% gelatine-PBS (Sigma-Aldrich). Mouse BACs harboring GFP-tagged *Nudt21*, *Cpsf6* or *Fip1l1* genes were isolated from *Escherichia coli* DH10 cells using the Nucleo-Bond PC 20 kit (Macherey-Nagel). P19 cells were transfected with 1 µg purified BAC DNA using Effectene Transfection Reagent (Qiagen). Cells with stably integrated BACs were selected with 500 µg/ml Geneticin (G418, ThermoFisher Scientific) and regularly checked for mycoplasma contaminations.

P19 differentiation

P19 wild type cells were differentiated into neuronal cells using retinoic acid according to [53]. Briefly, 10 cm culture dishes were coated with 10 µg laminin diluted in 4 ml PBS overnight. Laminin solution was removed and the dishes were washed three times with 1x PBS (Sigma-Aldrich) before seeding P19 cells into 10 ml Gibco™ DMEM/F-12, GlutaMAX™ (ThermoFisher Scientific), supplemented with 1x Gibco™ N-2 Supplement (ThermoFisher Scientific) and 100 U/ml penicillin-streptomycin (ThermoFisher Scientific). To start the differentiation the medium was supplemented with 10 ng/ml FGF8 (Sigma-Aldrich), 10 µM DAPT (Sigma-Aldrich) and 500 nM retinoic acid (Sigma-Aldrich). Cells were grown under humidified conditions at 5% CO₂ at 37 °C. After 4 days, synaptogenesis was induced with 10 ml Gibco™ CTS™ Neurobasal® Medium (ThermoFisher Scientific) supplemented with 1x Gibco™ B-27™ Supplement (ThermoFisher Scientific). To remove all dividing cells, 8 µM Cytosine β-D-arabinofuranoside hydrochloride (Ara-C, Sigma-Aldrich) was added to the cultures. Cells were grown for

another 4 days and fully differentiated cells were harvested at day 8. Differentiation progress was monitored every second day by bright field microscopy.

Co-IPs, Western blot and antibodies

For Western blot experiments, protein concentrations were measured using Bradford 1x Dye Reagent (Bio-Rad) on a NanoDrop2000 (Thermo Scientific). Protein lysates were mixed with 5x Laemmli buffer, boiled at 95 °C for 5 min and 10–20 µg total protein per lane were separated either on homemade 10% SDS-PAGE (Bio-Rad) or on NuPAGE 4–12% Bis-Tris PAGE (ThermoFisher Scientific) gel electrophoresis. Proteins were transferred onto 0.1 µm nitrocellulose membrane (GE Healthcare). The transfer was evaluated by staining with Ponceau S (Amresco).

For Co-IPs, approximately 4×10^7 P19 cells were lysed using NET-2 buffer (150 mM NaCl, 0.05% NP-40, 50 mM Tris-HCl pH 7.5), supplemented with EDTA-free cComplete Protease Inhibitor Cocktail (Sigma-Aldrich) and 10 mM β-glycerophosphate (Fluka Bio-Chemica) and sonicated on ice (Branson). Lysates were cleared, split in two, and treated with or without 100 µg/ml RNase A for 20 min at 21 °C. 0.2% of total lysate served as input. 10 µg of goat IgG (Sigma-Aldrich) or goat α-GFP (provided by D. Drechsel, MPI-CBG, Dresden, Germany) were pre-incubated with Gammabind G Sepharose beads (GE Healthcare) for 2 h at 4 °C. Subsequently, they were mixed with RNase A-treated or untreated lysates and incubated for 1.5 h at 4 °C. The beads were washed and co-precipitated proteins were eluted with 1.32x NOVEX sample mix with 1/10 reducing agent (ThermoFisher Scientific). Proteins were separated on NuPAGE 4–12% Bis-Tris gels (ThermoFisher Scientific), blotted on nitrocellulose membranes (GE Healthcare), and probed with the following antibodies: rabbit α-CTNNB (Abcam ab2365), rabbit α-CPSF6 (Abcam ab99347), mouse α-FIP1L1 (Santa Cruz Biotechnology sc-398392), goat α-GFP (MPI-CBG), mouse α-CPSF5 (NUDT21; Santa Cruz Biotechnology sc-81109), rabbit α-PABPN1 (Abcam ab75855), mouse α-phosphoSR (mAb104, kindly provided by K. Neugebauer), mouse α-SRSF3 (m7B4, kindly provided by K. Neugebauer), rabbit α-SRSF7 (Assay Biotechnology C18943), rabbit α-mCherry (Invitrogen PA5-34974), and rabbit α-Tet-Repressor (MoBiTec TET01). Donkey α-mouse IgG-HRP (AP192P, Sigma-Aldrich), donkey α-rabbit IgG-HRP (AP182P, EMD Millipore), donkey α-goat IgG-HRP (AB324P, Sigma-Aldrich), and goat α-mouse IgM-HRP (A8786, Sigma-Aldrich) secondary antibodies were used respectively for immuno-detection in combination with ECL Prime Western Blotting Detection Reagent (GE Healthcare). Western blots were additionally probed with the following antibodies: rabbit α-tubulin (Abcam ab176560) and mouse α-GAPDH (Santa Cruz Biotechnology sc-32,233).

Shrimp alkaline phosphatase treatment

Total protein was extracted from approximately 1×10^7 P19 cells as described before using ice-cold NET2 buffer with 10 mM MgCl₂ supplemented with EDTA-free cComplete Protease Inhibitor Cocktail (Sigma-Aldrich) and 10 mM β-glycerophosphate (Fluka BioChemica). Protein concentrations were measured using Bradford 1x Dye Reagent (Bio-Rad) on a NanoDrop2000 (Thermo Scientific). Ten micrograms total protein was added to two 1.5 ml reaction tubes and mixed with either –SAP-Mix (1x CutSmart Buffer [New England Biolabs],

10 mM beta-phosphoglycerate) or + SAP-Mix (1x CutSmart Buffer [New England Biolabs], 5 U rSAP [New England Biolabs]). The volume was adjusted to 20 μ l using NET2/MgCl₂-buffer and the samples were incubated for 30 min at 37 °C with agitation (300 rpm). Subsequently, the samples were mixed with 5x Laemmli buffer, with and without glycerol, boiled at 90 °C for 5 min, and subjected to Western blot.

SRSF3 immunoprecipitation, sample preparation, and quantitative mass spectrometry (TMT)

P19 cells stably expressing SRSF3-GFP at physiological levels [32] or nuclear GFP (GFP-NLS) as control [48] were grown in 14-cm dishes and harvested after washing with ice-cold 1x PBS. Immunoprecipitations (IPs) were performed as described above. The IPs were eluted three times from magnetic beads with each 50 μ l 6 M Guanidinium-Hydrochloride (in Tris pH 8.0) at room temperature (RT). Eluted IP-Samples were reduced with DTT (final 5 mM) for 30 min at 55 °C and alkylated using chloroacetamide (final 15 mM) for 30 min in the dark at RT. The reaction was quenched by DTT (final 10 mM) for 15 min at RT. Samples were cleaned by methanol chloroform precipitation and dried protein pellets were taken up in 0.2 M EPPS pH 8.2 and 10% acetonitrile (ACN). For digestion, 0.4 μ g trypsin (Promega) was added and the samples were incubated over night at 37 °C. The amount of ACN was adjusted to 20% and peptides were incubated with 20 μ g of TMT-reagents (ThermoFisher Scientific) for 1 h at RT. TMT-labeling reaction was quenched by addition of hydroxylamine to a final concentration of 0.5% for 15 min at RT. Samples were pooled, acidified, and dried for further processing. Peptides were cleaned by stage tip using Empore C18 (Octadecyl) resin material (3 M Empore) and taken up in 0.1% formaldehyde and analyzed by LC-MS² using an Easy nLC 1200 (ThermoFisher Scientific) unit with a 22 cm long, 75 μ m ID fused-silica column, which has been packed in house with 1.9 μ m C18 particles (ReproSil-Pur, Dr. Maisch), and kept at 45 °C using an integrated column oven (Sonation). Peptides were eluted with a non-linear gradient from 5 to 38% ACN over 120 min and directly sprayed into a QExactive HF mass-spectrometer equipped with a nanoFlex ion source (ThermoFisher Scientific) at a spray voltage of 2.3 kV. Full scan MS spectra (350–1400 m/z) were acquired at a resolution of 120,000 at m/z 200, a maximum injection time of 100 ms, and an AGC target value of 3×10^6 . Up to 20 most intense peptides per full scan were isolated using a 1 Th window and fragmented using higher energy collisional dissociation (normalized collision energy of 35). MS/MS spectra were acquired with a resolution of 45,000 at m/z 200, a maximum injection time of 80 ms and an AGC target value of 1×10^5 . Ions with charge states of 1 and > 6 as well as ions with unassigned charge states were not considered for fragmentation. Dynamic exclusion was set to 20 s to minimize repeated sequencing of already acquired precursors.

Analysis of mass spectrometry (MS) data

Raw data were analyzed using Proteome Discoverer (PD) 2.4 software (ThermoFisher Scientific). Files were recalibrated using the *Mus musculus* SwissProt database (TaxID = 10090, v. 2017-07-05) with methionine oxidation (+ 15.995) as dynamic modification and carbamidomethyl (Cys,+ 57.021464), TMT6 (N-terminal, + 229.1629) and TMT6

(+ 229.1629) at lysines as fixed modifications. Spectra were selected using default settings and database searches were performed using SequestHT node in PD. Database searches were performed against trypsin digested *Mus musculus* SwissProt database and FASTA files of common contaminants (“contaminants.fasta” provided with MaxQuant) for quality control. Fixed modifications were set as TMT6 at lysine residues, TMT6 (N-terminal), and carbamidomethyl at cysteine residues. As dynamic modifications acetylation (N-terminal) and methionine oxidation were set. After search, posterior error probabilities were calculated and PSMs filtered using Percolator using default settings. The Consensus Workflow for reporter ion quantification was performed with default settings. Results were then exported to Excel and protein levels were normalized to GFP (UniProtKB - P42212).

Cloning and transfection

TetR cloning

Fusion constructs of the tetracycline repressor (TetR) protein and the RS domains of SRSF3 and SRSF7 as well as their phosphomimetics were generated via restriction and ligation cloning. For the phosphomimetics, each serine was replaced by either alanine or aspartic acid residues. The sequences were purchased as gBlocks DNA fragments (Integrated DNA Technologies) and amplified using specific primers (Additional file 1: Table S9) to add overhangs including *Xma*I restriction sites. The plasmid containing a single-chain TetR gene and the PCR amplicons were digested using *Xma*I (New England Biolabs) and ligated in a 1:6 ratio using T4 DNA Ligase (Promega).

Multiple sequence alignments to identify domain boundaries were performed using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

mCherry/eGFP cloning

Fusion constructs of the complete coding sequence of SRSF3 and SRSF7 with eGFP or mCherry, including phosphomimetics and chimeric constructs were generated by restriction and ligation cloning. All sequences were purchased as gBlocks DNA fragments (Integrated DNA Technologies) and amplified using specific primers (Additional file 1: Table S9) to add overhangs including *Nhe*I and *Kpn*I restriction sites. Plasmid containing either eGFP-N1 (Clontech) or mCherry-N1 (Clontech) and the respective PCR amplicons were digested using *Nhe*I-HF and *Kpn*I-HF (New England Biolabs) and ligated in a 1:6 ratio using T4 DNA Ligase (Promega).

Gibson Assembly cloning

Gibson assembly was used to generate mCherry and Luciferase reporter genes with different 3'UTR variants. Therefore, the backbones of the mCherry-N1 (Clontech) and Luciferase (Promega) plasmids were linearized and inserts were amplified by PCR using specific primers (Additional file 1: Table S9). Backbones and purified PCR inserts were mixed in a 1:3 ratio and ligated using a Gibson Assembly Mastermix (1.3x Isothermal Mastermix [100 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 200 nM dNTPs, 10 mM DTT, 1 mM NAD, 5% (w/v) PEG-8000], 0.1 U T5 exonuclease, 0.5 U Phusion DNA Polymerase, 0.1 U Taq DNA Ligase) at 50 °C for 15 min.

All ligation products were transformed into *E. coli* TOP10 cells and positive colonies were identified by Sanger sequencing (Eurofins). For the reporter gene assays 5 µg of each plasmid were transfected per 10 cm cell culture dish for 24 h using JetOPTIMUS Transfection Reagent (Polyplus) according to the manufacturer's instructions. For CPSF6-myc (SinoBiological) and chimera expression experiments 1.5 µg of each plasmid were transfected per 6 cm cell culture dish for 24 h using Lipofectamine 2000 (ThermoFisher Scientific). Cells were starved with Opti-MEM® (ThermoFisher Scientific) for 4 h prior to transfection.

Design, preparation, and transfection of esiRNAs

Suitable esiRNA target regions of approximately 400 nt were chosen using the DEQOR2 algorithm [74]. Template regions were amplified using primers that add a T7 promoter sequence (Additional file 1: Table S9) and in vitro transcribed using the HiScribe T7 High Yield RNA Synthesis Kit (NEB). Double-stranded RNAs were digested using RNase III (MPI-CBP, Dresden, Germany) at 37 °C for 2 h with agitation (1100 rpm). Digested esiRNAs were purified using Q Sepharose Fast Flow (Sigma-Aldrich), resuspended in TE buffer (pH 7.9), and stored at – 80 °C. Per 10 cm cell culture dish, 5 µg purified esiRNAs were transfected for 36 h (SRSF3&SRSF7) and 48 h (CPSF6) using JetPRIME Transfection Reagent (Polyplus) according to the manufacturer's instructions. EsiRNAs against GFP (SRSF3 and SRSF7) or Luciferase (CPSF6) were used as controls.

Reverse transcription (RT), rapid amplification of 3'cDNA ends (3'RACE), and qPCR

Total RNA was isolated using the TRIzol method (Invitrogen). Genomic DNA was removed by TURBO DNase (Invitrogen). RNA concentrations were measured using a NanoDrop2000 (ThermoFisher Scientific). For 3'RACE-PCR, 1 µg RNA was reverse transcribed into cDNA using an oligo (dT) primer or an anchored oligo (dT) primer including a platform sequence (Additional file 1: Table S9) and SuperScript III Reverse Transcriptase (Invitrogen). 3'RACE PCRs were done using gene-specific forward primers (Additional file 1: Table S9) and a reverse primer complementary to the platform sequence of the RT primer with 28 cycles. For qPCR, primers were chosen using Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). qPCRs were performed using ORA qPCR Green ROX H Mix, 2x (HighQu) on a PicoReal 96 (ThermoFisher Scientific).

RNA-seq and MACE-seq

For RNA-seq, 7 µg total RNA were subjected to poly(A) + selection and RNA-seq library preparation (Novogene). The libraries were sequenced on an Illumina HiSeq4000 instrument with either 75-bp single-end or 150-bp paired-end reads and 50–60 million reads per replicate.

MACE-seq libraries were prepared and sequenced at GenXPro GmbH (Frankfurt am Main, Germany) as described by [45, 46]. Total RNA was isolated as described before and 1 µg was submitted to GenXPro for downstream procedures. Briefly, first poly(A) + RNA was isolated from total RNA using the Invitrogen™ Dynabeads™ mRNA Purification Kit (ThermoFischer Scientific) followed by reverse transcription into cDNA using

the Invitrogen™ SuperScript™ Double-Stranded cDNA Synthesis Kit (ThermoFischer Scientific) with an anchored biotinylated poly (dT) primer. Next, the cDNA was randomly fragmented to an average size of 250 bp by sonication using a Bioruptor (Diagenode). The biotinylated cDNA ends were captured by Invitrogen™ Dynabeads™ M-270 Streptavidin Beads (ThermoFischer Scientific) and ligated with T4 DNA Ligase 1 (NEB) to modified TrueQuant adapters (GenXPro). Libraries were amplified using KAPA HiFi Hot-Start Polymerase (KAPA Biosystems), followed by purification using Agencourt AMPure XP beads (Beckman Coulter). The libraries were sequenced on a Illumina HiSeq2000 platform yielding 75-bp single-end reads.

RNA-seq data analysis

RNA-seq reads were obtained from Novogene and mapped against the mouse genome (version mm10) with GENCODE gene annotation (version M18) using STAR (v2.6.1d) [75] with the following parameters: `--outSAMattributes All --outSAMtype BAM SortedByCoordinate --runThreadN 2 --outFilterMismatchNmax 2 --readFilesCommand zcat --quantMode GeneCounts`. Mapped reads were counted with the `summarizeOverlaps` function of the `GenomicAlignments` (version v1.18.1) R/Bioconductor package using the “Union” mode and the exons of all genes in the GENCODE annotation (version M18) as features. The count tables were used for differential gene expression analysis with DESeq2 (version 1.22.2) [54]. Bam files were converted to bedGraph files using `bedtools2` (v2.26.0) [76] and changes in 3'UTR length were analyzed with `DaPars` (version 0.9.1) using the GENCODE annotation [43]. For each reported gene with a change in 3'UTR-APA, the location of the pPAS and dPAS was extracted as well as the adjusted *P* value (false discovery rate, FDR) and the change in dPAS usage (Δ PDUI). Changes with an $FDR \leq 0.1$ and a Δ PDUI ≥ 0.05 (longer 3'UTR) or Δ PDUI ≤ 0.05 (shorter 3'UTR) were considered significant.

MACE-seq data analysis

The analysis of MACE-seq data was performed as described in [45]. In brief, low-quality regions were trimmed from both ends (Phred score < 16). PCR duplicates were removed based on unique molecular identifiers introduced during MACE-seq library preparation. Trailing A's were then detected using a 5-nt sliding window, allowing one non-adenosine per window to account for sequencing errors. Reads with at least 10 trailing A's (A_{10}) were considered to arise from a poly(A) tail, which was subsequently trimmed off. Reads were then mapped against the mouse reference genome (mm10) using `Novoalign` (<http://novocraft.com>), keeping only uniquely mapped reads, controlled by “-r none” and without soft clipping (“-o FULLNW”).

To detect inadvertent internal priming events, homopolymeric A-stretches were determined by mapping A_{10} to the mouse reference genome (mm10) with `Bowtie` [77], allowing two errors (“-v 2, -r all”), transformed to BED format and merging overlapping intervals with `BEDtools` [76]. Putative poly(A) tail reads were then excluded if adjacent to a homopolymeric A-stretch. To identify clusters of reads for subsequent PAS definition, we used the combined 3' end coordinates of the poly(A) tail-trimmed reads from all three conditions (WT, *Srsf3* KD, *Srsf7* KD, two replicates each). These were assigned

to clusters by iteratively extending the cluster to the next downstream 3' end coordinate if ≤ 25 nt away from the median 3' end coordinate of the cluster.

Identification of PASs from MACE-seq data

In order to precisely assign PAS positions, we piled up all cleavage events (i.e. the positions 1 nt downstream of the 3' end coordinates of the poly(A) tail-trimmed reads) and resized the clusters to 15-nt windows centered on mode of cleavage events per nucleotide within the cluster. Cleavage events from each condition and replicate were then recounted into the 15-nt windows (CE_{window}), and replicates averaged for each condition ($CE_{average}$). Next, the percent usage (PU) was calculated for each 15-nt window by dividing CE_{window} by to all cleavage events in the complete gene * 100, and averaged between replicates for each condition ($PU_{average}$). Next, a transcripts per million (TPM)-type metric was calculated for each 15-nt window, by dividing CE_{window} by window length (in kilobases; i.e., 0.015) by 1 million reads, averaged between replicates for each condition ($TPM_{average}$), and then taken as maximum across all windows in a given gene ($TPM_{maximum}$). For each 15-nt window, the maximum CE_{window} , $PU_{average}$, and $TPM_{maximum}$ across the three conditions was used to apply the following filters: $CE_{window} > 4$ cleavage events AND $PU_{average} > 5\%$ AND $TPM_{maximum} > 0.25$. Finally, we annotated the positions of the detected PASs based on GENCODE gene annotation (version M18). In this context, 15-nt windows, whose center overlapped more than one gene or were more than 50 nt downstream of a gene were removed. This procedure yielded a total of 15,866 PASs mapping to 9148 genes, whereby the center of the 15-nt windows was considered as the actual cleavage site.

To assign the PASs to transcript regions, each PAS was checked for overlaps with 5' UTR, CDS, 3'UTR, or intronic regions retrieved from the GENCODE annotation. Except for PASs overlapping with 3'UTR and intronic regions, all PASs overlapping with more than one region (e.g., 5'UTR and CDS) were assigned to category "Other." In addition, PASs located 1–50 nt downstream of a gene were assigned to the 3'UTR of that gene.

The spatial allocation of PASs as proximal PAS (pPAS), distal PAS (dPAS) or intermediate/other PAS (oPAS) was achieved by defining the most upstream PAS as pPAS for a given gene, the most downstream PAS as dPAS, and the remaining PASs located in between as oPAS. In the case of only a single PAS for a given gene, the PAS was defined as sPAS. This procedure yielded 5095 pPASs, 2589 oPASs, 4091 dPASs, and 5095 sPASs, whereby 2903 of the pPASs, 2169 of the oPASs, 3838 of the dPASs, and 4796 of the sPASs were located in annotated 3'UTRs.

Motif analysis around PASs

To identify the associated CSEs (Additional file 1: Fig. S1J), the start positions of all CSE hexamers reported by [24] were determined in a window of 50 nt upstream of each PAS. A unique CSE was then assigned by applying the following hierarchy: AAUAAA > AUUAAA > other hexamers.

The integration of PAS coordinates from MACE-seq and APA changes (3'UTRs getting shorter or longer) reported by DaPars was done by matching PASs obtained by MACE-seq to PASs reported by DaPars. A DaPars PAS was considered as a match to a

MACE-seq PAS if it was located at maximum 250 nt upstream or 50 nt downstream of the MACE-seq PAS (Additional file 1: Fig. S2A). In the case of multiple DaPars PASs matching the same MACE-seq PAS, the closer one was considered. Further, we restricted that the matched PASs were of the same type (pPAS-pPAS or dPAS-dPAS).

CNYC, GAY, and UGUA motif analyses were performed on the 13,706 PASs located in 3'UTRs. For each of these PASs, motif starts in a ± 300 -nt or ± 500 -nt window were identified, which served as basis for the subsequent analyses. Metaprofiles of motif distributions were generated by calculating the fraction of PASs with a motif start in a specific distance to the PAS, followed by a loess smoothing. In addition, for the metaprofiles of transcripts with shorter 3'UTRs in *Srsf3* or *Cpsf6* KD, matched metaprofiles of non-affected transcripts were generated. For this purpose, we randomly selected 100 times similar set sizes of non-affected transcripts and determined for each nucleotide around the PASs the average fraction of PASs with a motif start. The shown metaprofiles are loess smoothed, whereby the shaded area reflects the confidence interval. Regarding the fraction of PASs enclosed by tandem UGUAs (UGUA-PAS-UGUA) and those preceded (UGUA-UGUA-PAS) or followed (PAS-UGUA-UGUA) by tandem UGUAs were determined in windows of incrementing size (Win_{size} ; 1-nt steps). For PASs enclosed by tandem UGUAs, two windows were simultaneously incremented on either side of the PAS (Win_{size} 1–75 nt), whereby the first UGUA had to start in a range of $[-Win_{size}; -4$ nt] and the second UGUA in a range of $[-3$ nt; $Win_{size}-3$]. Regarding PASs preceded or followed by tandem UGUAs, the Win_{size} was in a range of 1–150 nt, whereby both UGUAs had to start in a range of $[-Win_{size}; -4$ nt] (preceded) or $[1$ nt; $Win_{size}-3$] (followed). For each PAS, the minimum distance between two UGUAs is reported.

iCLIP library preparation

Approximately 4×10^7 P19 cells were irradiated once with 300 mJ/cm^2 UV light at 254 nm (CL-1000, UVP) on ice. iCLIP was performed as described in [78] with minor modifications. Briefly, Dynabeads Protein G (Invitrogen) were coupled with goat anti-GFP antibody (provided by D. Drechsel, MPI-CBG, Dresden, Germany) and used for immunoprecipitation. Crosslinked RNA from the immunoprecipitated RNPs was digested into smaller fragments using RNase I (Invitrogen) and purified. RNA fragments were ligated to pre-adenylated DNA 3' adapters (Integrated DNA Technologies) and reverse-transcribed using barcoded RT primers by Invitrogen Superscript IV (ThermoFisher Scientific). cDNA fragments were size-selected and circularized by CircLigase II (Epicentre/Lucigen) before re-linearization using *Bam*HI HF (New England Biolabs). The final libraries were amplified using AccuPrime SuperMix I (ThermoFischer Scientific) and subjected to Illumina sequencing on a HighSeq2000 instrument with 75-bp single-end reads.

Analysis of iCLIP data

Analysis of iCLIP sequencing data was done using the iCount package (<http://icount.biolab.si>). Briefly, adapters and barcodes were removed from all reads before mapping to the mouse mm9 genome assembly (Ensembl59 annotation) using the

Bowtie aligner (version 0.12.7). To determine protein-RNA contact sites, all uniquely mapping reads were used, PCR duplicates were removed and crosslink events (X-links) were extracted (1st nucleotide of the read). To determine statistically significant X-links, a false discovery rate (FDR < 0.05) was calculated using normalized numbers of input X-links and randomized within co-transcribed regions [79–81]. To obtain comparable numbers of significant binding sites, replicates that correlated well were pooled according to their overall number of crosslink events.

For motif searching, a *z*-score analysis for enriched *k*-mers was performed as described previously [80]. Sequences surrounding significant X-links were extended in both directions by 30 nucleotides (windows: – 30 nt to – 5 nt, and 5 nt to 30 nt). All occurring *k*-mers within the evaluated interval were counted and weighted. Then, a control dataset was generated by randomly shuffling 100 times significant X-links within the same genes, and a *z*-score was calculated relative to the randomized genomic positions (<https://github.com/tomazc/iCount>). The top 25 *k*-mers were aligned to determine the *in vivo* binding consensus motif. Sequence logos were produced using WebLogo (<http://weblogo.berkeley.edu/logo.cgi>).

For metaprofiles of crosslink sites around pPASs, dPASs and sPASs, the positions with crosslink events in a window of – 400 nt to 100 nt around the PAS were extracted. Afterwards, for each PAS type and protein (SRSF3, SRSF7, CPSF5, and FIP1), two normalization steps were conducted. In the first normalization step, the summed crosslink sites were normalized to the number of PASs in this category to make signals around pPASs, dPASs, and sPASs comparable. In the second step, we normalized for the total number positions with crosslink sites in the iCLIP library to make signals of different iCLIP libraries comparable. Binding signals were smoothed with the loess function. Signal differences between two iCLIP libraries (e.g., SRSF3 and SRSF7) were determined nucleotide-wise by two-proportions *z*-tests, followed by a correction for multiple hypothesis testing using the Benjamini-Hochberg procedure. Positions with FDR ≤ 0.01 were considered as significant. To account for different sequencing depths, the proportions were normalized via a scaling factor before the test. The scaling factor was calculated as number of positions with crosslink sites in the first iCLIP library divided by the number of positions with crosslink sites in the second iCLIP library.

Metaprofiles around PASs with reduced or increased usage in KDs or during differentiation were compared to metaprofiles of matched background sets of PASs with unchanged usage. This procedure is exemplified for the iCLIP signal of SRSF3 around 197 pPASs with increased usage upon *Srsf3* KD (Fig. 2b, upper panel). For the background set, we randomly selected 100 times 197 pPASs with unchanged usage upon *Srsf3* KD and determined for each of the 100 sets the iCLIP signal of SRSF3 as described above. Based on the resulting 100 profiles, we calculated for each position the mean and standard deviation and used the two measures to calculate a *z*-score for each position in the SRSF3 binding signal of the 197 pPAS with increased usage. *z*-scores were then transferred into *P* values and corrected for multiple hypothesis testing using the Benjamini-Hochberg procedure. *z*-scores with an FDR ≤ 0.01% are shown and reflect a significant difference in iCLIP signal between pPAS with increased and unchanged usage upon *Srsf3* KD.

Immunofluorescence microscopy and image acquisition

Adherent cells were fixed with 4% paraformaldehyde (PFA; Sigma-Aldrich) for 20 min at RT, washed and permeabilized in 5% BSA, 0.1% Triton in PBS for 30 min at RT. DNA was counterstained with Hoechst 34580 (Sigma) in TBST (20 mM Tris-HCl, 150 mM NaCl, 0.1% Tween 20 pH 7.5; 1:4000) at RT for 30 min. After washing, the cells were dried and mounted on slides using ProLong™ Diamond Antifade Mountant (Thermo Fisher Scientific) and stored at 4 °C until imaging. Images were acquired using a confocal laser-scanning microscope (LSM780; ZEISS) with a Plan-Apochromat × 63 1.4 numerical aperture oil differential interference contrast objective equipped with two photomultiplier tubes and a gallium arsenite phosphate (GaAsPPMT) detector system. Fluorescence signal was detected with an Argon laser (GFP, 488 nm). Fiji was used to process all acquired images [82]. Pictures were cropped with the Image crop function and scale bars were added.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-021-02298-y>.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Additional file 5.

Additional file 6.

Additional file 7.

Additional file 8.

Additional file 9.

Additional file 10.

Additional file 11.

Acknowledgements

We thank A. Dahl for advice and sequencing of the iCLIP libraries, GeneXPro for preparing the MACE-Seq libraries, sequencing, and initial bioinformatic analyses; J. Ule and T. Curk for advice on iCLIP data analysis and access to the iCOUNT server; D. Staiger, J. Wöhnert, and E. Schleiff for helpful discussions and mentoring; and the Müller-McNicoll and Zarnack labs for discussions and support.

Review history

The review history is available as Additional file 11.

Peer review information

Tim Sands was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions

ODS, NB, MK, CM, MMM, and KZ designed the experiments and analysis pipelines. ODS, NB, MW, MEB, and FM performed the experiments, ODS, NB, MK, STS, IRdIM, CM, KZ, and MMM performed the analyses. IP and BS contributed advice and reagents. The figures were prepared by ODS, NB, MK, KZ, and MMM. The manuscript was written by ODS, NB, MK, FM, KZ, and MMM. All authors read and approved the final manuscript.

Funding

We are grateful for funding from the Deutsche Forschungsgemeinschaft (CEF-MC and SFB902-B13 to MMM; Emmy Noether to CM; and SFB902-B13 to KZ). Open Access funding enabled and organized by Projekt DEAL.

Availability of data and materials

All iCLIP, RNA-seq, and MACE-seq data generated and/or analyzed during the current study have been submitted to Gene Expression Omnibus (GEO) under the SuperSeries accession GSE151724. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD018090. The computational code for the motif analyses and the RBP binding maps is available via the GitHub repository [83].

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute for Molecular Bio Science, Goethe University Frankfurt, Max-von-Laue-Str. 13, 60438 Frankfurt, Germany. ²Buchmann Institute for Molecular Life Sciences, Goethe University Frankfurt, Max-von-Laue-Str. 15, 60438 Frankfurt, Germany. ³Faculty of Biological Sciences, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany. ⁴Institute of Biochemistry II, Medical School, Goethe University Frankfurt, Sandhofstr. 2-4, 60528 Frankfurt am Main, Germany. ⁵Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden, Germany. ⁶The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. ⁷Department of Biology, Technical University Darmstadt, Schnittpahnstr. 10, 64287 Darmstadt, Germany.

Received: 16 June 2020 Accepted: 11 February 2021

Published online: 11 March 2021

References

- Chan S, Choi EA, Shi Y. Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip Rev RNA*. 2011;2:321–35.
- Tian B, Graber JH. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA*. 2012;3:385–96.
- Kumar A, Clerici M, Muckenfuss LM, Passmore LA, Jinek M. Mechanistic insights into mRNA 3'-end processing. *Curr Opin Struct Biol*. 2019;59:143–50.
- Takagaki Y, Ryner LC, Manley JL. Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev*. 1989;3:1711–24.
- Zhao J, Kessler M, Helmling S, O'Connor JP, Moore C. Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor. *Mol Cell Biol*. 1999;19:7733–40.
- Chan SL, Huppertz I, Yao C, Weng L, Moresco JJ, Yates JR 3rd, Ule J, Manley JL, Shi Y. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev*. 2014;28:2370–80.
- Schonemann L, Kuhn U, Martin G, Schafer P, Gruber AR, Keller W, Zavolan M, Wahle E. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev*. 2014;28:2381–93.
- Yao C, Biesinger J, Wan J, Weng L, Xing Y, Xie X, Shi Y. Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc Natl Acad Sci U S A*. 2012;109:18773–8.
- Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature*. 2006;444:953–6.
- Kaufmann I, Martin G, Friedlein A, Langen H, Keller W. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J*. 2004;23:616–26.
- Kim S, Yamamoto J, Chen Y, Aida M, Wada T, Handa H, Yamaguchi Y. Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. *Genes Cells*. 2010;15:1003–13.
- Rueggsegger U, Beyer K, Keller W. Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *J Biol Chem*. 1996;271:6107–13.
- Rueggsegger U, Blank D, Keller W. Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. *Mol Cell*. 1998;1:243–53.
- Gruber AR, Martin G, Keller W, Zavolan M. Cleavage factor Im is a key regulator of 3' UTR length. *RNA Biol*. 2012;9:1405–12.
- Yang Q, Gilmartin GM, Double S. The structure of human cleavage factor I(m) hints at functions beyond UGUA-specific RNA binding: a role in alternative polyadenylation and a potential link to 5' capping and splicing. *RNA Biol*. 2011;8:748–53.
- Dettwiler S, Aringhieri C, Cardinale S, Keller W, Barabino SM. Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. *J Biol Chem*. 2004;279:35788–97.
- Zhu Y, Wang X, Forouzmand E, Jeong J, Qiao F, Sowd GA, Engelman AN, Xie X, Hertel KJ, Shi Y. Molecular mechanisms for CFIm-mediated regulation of mRNA alternative polyadenylation. *Mol Cell*. 2018;69:62–74 e64.
- Kamieniarz-Gdula K, Gdula MR, Panser K, Nojima T, Monks J, Wisniewski JR, Riepsaame J, Brockdorff N, Pauli A, Proudfoot NJ. Selective roles of vertebrate PCF11 in premature and full-length transcript termination. *Mol Cell*. 2019;74:158–72 e159.
- Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol*. 2017;18:18–30.
- Neve J, Patel R, Wang Z, Louey A, Furger AM. Cleavage and polyadenylation: ending the message expands gene regulation. *RNA Biol*. 2017;14:865–90.
- Mayr C. Regulation by 3'-Untranslated regions. *Annu Rev Genet*. 2017;51:171–94.
- Gruber AJ, Zavolan M. Alternative cleavage and polyadenylation in health and disease. *Nat Rev Genet*. 2019;20:599–614.
- Turner RE, Pattison AD, Beilharz TH. Alternative polyadenylation in the regulation and dysregulation of gene expression. *Semin Cell Dev Biol*. 2018;75:61–9.
- Gruber AJ, Schmidt R, Gruber AR, Martin G, Ghosh S, Belmadani M, Keller W, Zavolan M. A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res*. 2016;26:1145–59.

25. Li W, You B, Hoque M, Zheng D, Luo W, Ji Z, Park JY, Gunderson SI, Kalsotra A, Manley JL, Tian B. Systematic profiling of poly(A)⁺ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet.* 2015;11:e1005166.
26. Weng L, Li Y, Xie X, Shi Y. Poly(A) code analyses reveal key determinants for tissue-specific mRNA alternative polyadenylation. *RNA.* 2016;22:813–21.
27. Lackford B, Yao C, Charles GM, Weng L, Zheng X, Choi EA, Xie X, Wan J, Xing Y, Freudenberg JM, et al. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.* 2014;33:878–89.
28. Hardy JG, Norbury CJ. Cleavage factor Im (CFIm) as a regulator of alternative polyadenylation. *Biochem Soc Trans.* 2016;44:1051–7.
29. Martin G, Gruber AR, Keller W, Zavolan M. Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell Rep.* 2012;1:753–63.
30. Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu AB, Li W, Wagner EJ. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature.* 2014;510:412–6.
31. Chatrikhi R, Mallory MJ, Gazzara MR, Agosto LM, Zhu WS, Litterman AJ, Ansel KM, Lynch KW. RNA binding protein CELF2 regulates signal-induced alternative polyadenylation by competing with enhancers of the polyadenylation machinery. *Cell Rep.* 2019;28:2795–806 e2793.
32. Müller-McNicoll M, Botti V, de Jesus Domingues AM, Brandl H, Schwich OD, Steiner MC, Curk T, Poser I, Zarnack K, Neugebauer KM. SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev.* 2016;30:553–66.
33. Shen T, Li H, Song Y, Li L, Lin J, Wei G, Ni T. Alternative polyadenylation dependent function of splicing factor SRSF3 contributes to cellular senescence. *Aging (Albany NY).* 2019;11:1356–88.
34. Lou H, Neugebauer KM, Gagel RF, Berget SM. Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Mol Cell Biol.* 1998;18:4977–85.
35. Manley JL, Krainer AR. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev.* 2010;24:1073–4.
36. Busch A, Hertel KJ. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip Rev RNA.* 2012;3:1–12.
37. Hargous Y, Hautbergue GM, Tintaru AM, Skrisovska L, Golovanov AP, Stevenin J, Lian LY, Wilson SA, Allain FH. Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J.* 2006;25:5126–37.
38. Wegener M, Müller-McNicoll M. View from an mRNP: the roles of SR proteins in assembly, maturation and turnover. *Adv Exp Med Biol.* 2019;1203:83–112.
39. Cavaloc Y, Bourgeois CF, Kister L, Stevenin J. The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA.* 1999;5:468–83.
40. Cavaloc Y, Popielarz M, Fuchs JP, Gattoni R, Stevenin J. Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa factor of the serine/arginine protein family. *EMBO J.* 1994;13:2639–49.
41. Königs V, de Oliveira Freitas Machado C, Arnold B, Blümel N, Solovyeva A, Löbber S, Schafrank M, Ruiz De Los Mozos I, Wittig I, McNicoll M, et al. SRSF7 maintains its homeostasis through the expression of Split-ORFs and nuclear body assembly. *Nat Struct Mol Biol.* 2020;27:260–73.
42. Xiao SH, Manley JL. Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev.* 1997;11:334–44.
43. Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ, Li W. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat Commun.* 2014;5:5274.
44. Katz Y, Wang ET, Airolidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods.* 2010;7:1009–15.
45. Müller S, Rycak L, Afonso-Grunz F, Winter P, Zawada AM, Damrath E, Scheider J, Schmah J, Koch I, Kahl G, Rötter B. APADB: a database for alternative polyadenylation and microRNA regulation events. *Database (Oxford).* 2014;2014.
46. Zawada AM, Rogacev KS, Müller S, Rotter B, Winter P, Fliser D, Heine GH. Massive analysis of cDNA ends (MACE) and miRNA expression profiling identifies proatherogenic pathways in chronic kidney disease. *Epigenetics.* 2014;9:161–72.
47. Beaudoin E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* 2000;10:1001–10.
48. Ánkö ML, Müller-McNicoll M, Brandl H, Curk T, Gorup C, Henry I, Ule J, Neugebauer KM. The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.* 2012;13:R17.
49. Clerici M, Faini M, Muckenfuss LM, Aebersold R, Jinek M. Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. *Nat Struct Mol Biol.* 2018;25:135–8.
50. Ghosh G, Adams JA. Phosphorylation mechanism and structure of serine-arginine protein kinases. *FEBS J.* 2011;278:587–97.
51. Zahler AM, Neugebauer KM, Stolk JA, Roth MB. Human SR proteins and isolation of a cDNA encoding SRp75. *Mol Cell Biol.* 1993;13:4023–8.
52. Popielarz M, Cavaloc Y, Mattei MG, Gattoni R, Stevenin J. The gene encoding human splicing factor 9G8. Structure, chromosomal localization, and expression of alternatively processed transcripts. *J Biol Chem.* 1995;270:17830–5.
53. Nakayama Y, Wada A, Inoue R, Terasawa K, Kimura I, Nakamura N, Kurosaka A. A rapid and efficient method for neuronal induction of the P19 embryonic carcinoma cell line. *J Neurosci Methods.* 2014;227:100–6.
54. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
55. Ji Z, Lee JY, 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 U S A.* 2009;106:7028–33.
56. Shepard PJ, Choi EA, Lu J, Flanagan LA, Hertel KJ, Shi Y. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA.* 2011;17:761–72.
57. Yang Q, Coseno M, Gilmartin GM, Doublet S. Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(a) site recognition and RNA looping. *Structure.* 2011;19:368–77.
58. Zheng D, Tian B. RNA-binding proteins in regulation of alternative cleavage and polyadenylation. *Adv Exp Med Biol.* 2014;825:97–127.

59. Saijo S, Kuwano Y, Masuda K, Nishikawa T, Rokutan K, Nishida K. Serine/arginine-rich splicing factor 7 regulates p21-dependent growth arrest in colon cancer cells. *J Med Invest*. 2016;63:219–26.
60. Kumar D, Das M, Saucedo C, Ellies LG, Kuo K, Parwal P, Kaur M, Jih L, Bandyopadhyay GK, Burton D, et al. Degradation of splicing factor SRSF3 contributes to progressive liver disease. *J Clin Invest*. 2019;130:4477–91.
61. Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, Zhang Z, Cho S, Sherrill-Mix S, Wan L, Dreyfuss G. U1 snRNP determines mRNA length and regulates isoform expression. *Cell*. 2012;150:53–64.
62. Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*. 2010;468:664–8.
63. Gunderson SI, Beyer K, Martin G, Keller W, Boelens WC, Mattaj LW. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell*. 1994;76:531–41.
64. Ko B, Gunderson SI. Identification of new poly(A) polymerase-inhibitory proteins capable of regulating pre-mRNA polyadenylation. *J Mol Biol*. 2002;318:1189–206.
65. Aubol BE, Hailey KL, Fattet L, Jennings PA, Adams JA. Redirecting SR protein nuclear trafficking through an allosteric platform. *J Mol Biol*. 2017;429:2178–91.
66. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR 3rd, Frank J, Manley JL. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell*. 2009;33:365–76.
67. Brown KM, Gilmartin GM. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol Cell*. 2003;12:1467–76.
68. Wang X, Hennig T, Whisnant AW, Erhard F, Prusty BK, Friedel CC, Forouzmand E, Hu W, Erber L, Chen Y, et al. Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27. *Nat Commun*. 2020;11:293.
69. So BR, Di C, Cai Z, Venters CC, Guo J, Oh JM, Arai C, Dreyfuss G. A complex of U1 snRNP with cleavage and polyadenylation factors controls telescripting, regulating mRNA transcription in human cells. *Mol Cell*. 2019;76:590–9 e594.
70. Millevoi S, Loulergue C, Dettwiler S, Karaa SZ, Keller W, Antoniou M, Vagner S. An interaction between U2AF 65 and CFIm links the splicing and 3' end processing machineries. *EMBO J*. 2006;25:4854–64.
71. Maciolek NL, McNally MT. Serine/arginine-rich proteins contribute to negative regulator of splicing element-stimulated polyadenylation in rous sarcoma virus. *J Virol*. 2007;81:11208–17.
72. Valente TW, Zogg JB, Christensen S, Richardson J, Kovacs A, Operskalski E. Using social networks to recruit an HIV vaccine preparedness cohort. *J Acquir Immune Defic Syndr*. 2009;52:514–23.
73. Hudson SW, McNally LM, McNally MT. Evidence that a threshold of serine/arginine-rich (SR) proteins recruits CFIm to promote rous sarcoma virus mRNA 3' end formation. *Virology*. 2016;498:181–91.
74. Surendranath V, Theis M, Habermann BH, Buchholz F. Designing efficient and specific endoribonuclease-prepared siRNAs. *Methods Mol Biol*. 2013;942:193–204.
75. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
76. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
77. Langmead B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics* 2010, Chapter 11:Unit 11.17.
78. Huppertz I, Attig J, D'Ambrogio A, Easton LE, Sibley CR, Sugimoto Y, Tajnik M, König J, Ule J. iCLIP: protein-RNA interactions at nucleotide resolution. *Methods*. 2014;65:274–87.
79. König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol*. 2010;17:909–15.
80. Wang Z, Kayikci M, Briese M, Zarnack K, Luscombe NM, Rot G, Zupan B, Curk T, Ule J. iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol*. 2010;8:e1000530.
81. Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol*. 2009;16:130–7.
82. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9:676–82.
83. Keller M, Zarnack K: Source code RNA maps. *github* 2021, <https://doi.org/10.5281/zenodo.4457218>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

