Exon junction complexes suppress spurious splice sites to safeguard transcriptome integrity

Volker Boehm[1], Thiago Britto-Borges[2,3], Anna-Lena Steckelberg[1,4], Kusum K. Singh[1,5], Jennifer V. Gerbracht[1], Elif Gueney[1], Lorea Blazquez[6,7], Janine Altmüller[8,9,10], Christoph Dieterich[2,3], Niels H. Gehring[1,11]

[1] Institute for Genetics, University of Cologne, 50674 Cologne, Germany

[2] Section of Bioinformatics and Systems Cardiology, Department of Internal Medicine III and Klaus Tschira Institute for Integrative Computational Cardiology, University of Heidelberg, 69120 Heidelberg, Germany

[3] DZHK (German Centre for Cardiovascular Research), Partner site Heidelberg/Mannheim, 69120 Heidelberg, Germany

[4] present address: Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, CO 80045, USA

[5] present address: Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, 781039-Guwahati, Assam, INDIA

[6] Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London, WC1N 3BG, UK

[7] The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, UK

[8] Cologne Center for Genomics (CCG), University of Cologne, 50931 Cologne, Germany

[9] Institute of Human Genetics, University of Cologne, 50931 Cologne, Germany

[10] Center for Molecular Medicine Cologne, University of Cologne, 50937 Cologne, Germany

[11] Lead Contact

Correspondence

Christoph Dieterich: christoph.dieterich@uni-heidelberg.de

Niels H. Gehring: ngehring@uni-koeln.de

Summary

Productive splicing of human pre-mRNAs requires the correct selection of authentic splice sites (SS) from the large pool of potential SS. Although SS consensus sequence and splicing regulatory proteins are known to influence SS usage, the mechanisms ensuring the effective suppression of cryptic SS are insufficiently explored. Here, we find that many aberrant exonic SS are efficiently silenced by the exon junction complex (EJC), a multi-protein complex that is deposited on spliced mRNA near the exon-exon junction. Upon depletion of EJC proteins, cryptic SS are de-repressed, leading to the mis-splicing of a broad set of mRNAs. Mechanistically, the EJC-mediated recruitment of the splicing regulator RNPS1 inhibits cryptic 5′SS usage, while the deposition of the EJC core directly masks reconstituted 3′SS, thereby precluding transcript disintegration. Thus, the EJC protects the transcriptome of mammalian cells from inadvertent loss of exonic sequences and safeguards the expression of intact, full length mRNAs.

Keywords

Alternative splicing, mRNA processing, exon junction complex, gene expression, cryptic splice sites

Introduction

The majority of eukaryotic pre-mRNAs undergo alternative splicing and produce an assorted set of mRNAs (Lee and Rio, 2015). Splicing of pre-mRNAs not only increases the coding potential of the genome, but is also a key regulatory step in gene expression (Kornblihtt et al., 2013; Nilsen and Graveley, 2010). In order to execute the two consecutive steps of splicing, the splicing machinery has to correctly identify the splice sites (SS) at the 5′ and 3′ ends of the intron, which have defined consensus sequences (Papasaikas and Valcarcel, 2016). Due to the degenerate nature of the splice consensus sequences, the cell is faced with the challenging task to discriminate between authentic and so-called cryptic SS, which exhibit consensus motifs but are not intended to be used. Therefore, many different mRNA-binding proteins assist the spliceosome in the accurate detection of introns and SS. These splicing regulators commonly bind to specific sequence motifs on the transcript and act as enhancers (e.g. SR proteins) or silencers (e.g. hnRNP proteins) of splicing (Han et al., 2010; Long and Caceres, 2009).

In addition to the removal of intronic sequences, splicing also alters the protein composition of the messenger ribonucleoprotein (mRNP). This phenomenon is documented in particular for the exon junction complex (EJC), which is assembled and deposited onto mRNAs during splicing (Boehm and Gehring, 2016; Le Hir et al., 2016). Binding of the EJC to its canonical site 24 nucleotides (nt) upstream of exon-exon junctions does not require a specific RNA-sequence and involves the phosphate-backbone of the RNA (Andersen et al., 2006; Bono et al., 2006). The core of the EJC consists of four proteins (EIF4A3, MAGOH, RBM8A and CASC3) that can be used as an assembly platform for other proteins, the so-called peripheral EJC components (Singh et al., 2012). The deposition of the EJC is initiated by the recruitment of the core factor EIF4A3 to the activated spliceosome by the splicing factor CWC22 (Alexandrov et al., 2012; Barbosa et al., 2012; Steckelberg et al., 2012).

On the cellular level, EJCs represent central mRNP components with diverse functions. Specifically, the EJC serves as the molecular memory of the splicing process and passes on this information to later steps of gene expression (Le Hir et al., 2016; Woodward et al., 2017). In particular, the dynamically recruited peripheral EJC proteins expand the functional impact of the EJC on gene expression. EJC components have been shown to stimulate mRNA transport, increase translation efficiency and support mRNA surveillance by nonsense-mediated mRNA decay (NMD) (Boehm and Gehring, 2016; Le Hir et al., 2016; Woodward et al., 2017). With these functions, the EJC helps to ensure that correctly processed and error-free transcripts are preferentially expressed. In addition to these post-splicing processes, the EJC has also been shown to influence the splicing process of selected mRNAs. For instance, correct splicing of mapk pre-mRNA and other long intron-containing transcripts in *Drosophila* was found to require EJC components (Ashton-Beaucage et al., 2010; Roignant and Treisman, 2010). Furthermore, depletion of EJC core factors resulted in the retention of a suboptimal intron in the piwi transcript (Hayashi et al., 2014; Malone et al., 2014). It has been suggested that in this case splicing of the neighboring introns leads to the deposition of EJCs, which subsequently function as splicing enhancers for the weak intron. In human cells, depletion of EJC core components caused widespread changes in alternative pre-mRNA splicing (Wang et al., 2014). Different types of alternative splicing events were observed, of which cassette exons represented the majority.

In mammals, the known splicing regulators ACIN1, PNN, RNPS1 and SAP18 are peripheral EJC components which have been shown to co-purify with the EJC core (Singh et al., 2012). Interestingly, RNPS1 and SAP18 form two alternative complexes with either ACIN1 (also known as Acinus) or PNN (also known as Pinin), referred to as ASAP (apoptosis- and splicing-associated protein) or PSAP complex, respectively (Murachelli et al., 2012). Evidence from studies in *Drosophila* suggests that ACIN1 and RNPS1 aid in definition and splicing of

neighboring introns and are involved in the EJC-mediated splicing regulation (Hayashi et al., 2014; Malone et al., 2014). Furthermore, RNPS1 is also required for the correct splicing of mapk and AURKB pre-mRNA in *Drosophila* and human cells, respectively (Ashton-Beaucage et al., 2010; Fukumura et al., 2018; Roignant and Treisman, 2010). Finally, recent transcriptome-wide studies identified alternative splicing changes in ACIN1- and PNN-depleted cells, suggesting that ASAP and PSAP complexes can regulate certain EJC-dependent and -independent splice events (Rodor et al., 2016; Wang et al., 2018). Despite the identification of various splicing alterations upon EJC depletion in mammalian cells, the molecular mechanism underlying the EJC-mediated splicing regulation is mostly uncharacterized. Hence, it is of fundamental importance to dissect the mechanistic role and functional dependency on core and peripheral EJC components for splicing regulation.

Here, we investigate the mechanism of splicing modulation by the EJC. We show that depletion of the EJC-associated splice factor RNPS1 caused widespread changes in splicing and led to the usage of cryptic and reconstituted 5′SS, which are efficiently repressed in the presence of RNPS1. Moreover, we identified an EJC-dependent, but RNPS1-independent mechanism that prevents splicing involving the usage of cryptic and re-constituted 3′SS. Taken together, the EJC, in cooperation with RNPS1, prevents the recognition of irregular SS within many transcripts and thus the formation of incorrect mRNAs.

Results

Given the lack of insight into the mechanism of splicing regulation by the EJC, we set out to investigate the function of the EJC-associated splice factor RNPS1, a common component of ASAP and PSAP complexes (Mayeda et al., 1999; Murachelli et al., 2012; Sakashita et al., 2004) (Figure 1A and S1A). First, we tested if RNPS1 is required for the correct splicing of the MRPL3 transcript, which shows robust exon 4 skipping upon knockdown of the EJC core protein EIF4A3 (Wang et al., 2014). Interestingly, siRNA-mediated depletion of RNPS1 quantitatively recapitulated the 60-70-fold increase of EJC-dependent skipping of MRPL3 exon 4 (Figure 1B and 1C), suggesting a functional link between EJC- and RNPS1 dependent splicing regulation.

RNPS1 recruitment to spliced mRNPs via the RRM domain is required for splicing regulation

Recent *in vivo* crosslinking and immunoprecipitation (iCLIP) experiments showed that RNPS1 displays a similar RNA binding pattern as the EJC core protein EIF4A3 (Hauer et al., 2016). Hence, we assumed that RNPS1 is positioned on spliced mRNPs via an interaction with RNA-bound EJCs. To better understand the interaction between RNPS1 and the EJC, we studied ASAP/PSAP complex formation and EIF4A3-binding using co-immunoprecipitation experiments. FLAG-tagged RNPS1 co-precipitated the ASAP/PSAP complex component SAP18, as well as the core EJC protein EIF4A3 (Figure S1B). Mutating a surface exposed patch on RNPS1 (termed 176; Figure S1C) that prevents ASAP/PSAP complex formation and co-precipitation of SAP18, also abolished the interaction with EIF4A3 (Figure S1B), suggesting that RNPS1 interacts with the EJC as part of the fully assembled ASAP or PSAP complex. We further investigated the recruitment of RNPS1 to spliced transcripts using immunoprecipitation of *in vitro* spliced mRNPs (Steckelberg and Gehring, 2014). While the C-terminal RS/P domain of RNPS1 conferred unspecific binding to spliced and unspliced RNA (Figure S1D-F), the isolated RRM domain of RNPS1, which is sufficient to form a minimal ASAP/PSAP complex

(Murachelli et al., 2012), co-immunoprecipitated exclusively spliced transcripts. Interestingly, RNPS1 (176), which is deficient in ASAP/PSAP complex formation and EIF4A3 binding (Figure 1D) also failed to co-precipitate spliced mRNA (Figure 1E and S1G), indicating that ASAP/PSAP complex formation and splicing-dependent mRNP interaction are functionally linked. Of note, MRPL3 exon skipping observed upon knockdown of RNPS1 could be rescued by expression of wildtype RNPS1, but not RNPS1 (176) (Figure 1F). We conclude that RNPS1 requires the RRM-mediated formation of an ASAP/PSAP complex and the interaction with the EJC for its specific association with spliced mRNPs to modulate splicing.

RNPS1 depletion causes transcriptome-wide loss of exonic sequences

Having established a molecular link between RNPS1 and the EJC, we next examined the global role of RNPS1 in splicing regulation. To this end, we sequenced RNA from control- and RNPS1-depleted cells as well as RNPS1 knockdown cells that were complemented with RNPS1 wildtype or RNPS1 (176) (Figure 2A). Using the MAJIQ algorithm (Vaquero-Garcia et al., 2016) to identify local splicing variations (LSV), we found that RNPS1 depletion substantially altered the splicing of 318 LSV in 243 genes (Figure 2A, Table S1). The affected genes represented a diverse group, as no specific enrichment was detectable for the gene ontology (GO) terms *molecular function*, *biological process* or *cellular compartment*.

Classification of the splicing alterations revealed that the predominant group represented exon skipping events, followed by exon inclusion, alternative 5′ or 3′SS usage and intron retention (Figure 2B and Figure S2A). Remarkably, exon-exon junctions that were barely or never detected under control conditions (herein called “spurious” junctions; see STAR Methods for details) constituted a substantial proportion in all classes of splicing alterations except exon inclusion (Figure 2B and 2C). In more than 30% of the splicing events we observed that RNPS1 depletion leads to the activation of irregular SS, skipping of constitutive exons and the formation of unusual transcript variants. Furthermore, none of the RNPS1-dependent splicing

alterations were found in a recently published atlas of alternative splicing events in multiple human tissues, cell types, and developmental stages (VAST-DB) (Tapial et al., 2017). We verified the RNA-seq results for selected transcripts with RNPS1-dependent exon skipping, exon inclusion, alternative SS usage and intron retention by RT-PCR (Figure 2D-G; Figure S2B-I). In all cases, the splicing change caused by RNPS1 knockdown was completely rescued by wildtype RNPS1, but not the RNPS1 (176) mutant, underscoring that RNPS1-dependent splicing events require ASAP/PSAP complex formation and recruitment by the EJC.

RNPS1 suppresses cryptic 5′ splice sites

The usage of cryptic and irregular SS in RNPS1-depleted cells suggests that the recruitment of RNPS1 by the EJC plays a pivotal role in suppressing these SS under normal conditions. To identify the molecular mechanism of RNPS1-dependent SS suppression, we investigated transcripts with RNPS1-dependent alternative 5′SS, because this class contained a large proportion of spurious junctions with robust fold-changes (Figure 2B and 2C). The majority of spurious 5′SS that were upregulated in RNPS1-depleted cells exhibited a good splice consensus sequence (Desmet et al., 2009) and were located close to the 5′ end of an exon (Figure 3A). Therefore, these 5′SS are near canonical EJC-binding sites, on which EJCs can be deposited during splicing of the preceding intron. We hypothesized that RNPS1 bound to an upstream EJC suppresses the usage of nearby 5′SS. To test this model, we took advantage of the TUFM transcript, in which usage of a cryptic 5′SS is upregulated upon RNPS1 depletion (Figure 2E and 2G). We constructed a reporter plasmid of the TUFM gene, including the RNPS1-dependent 5′SS as well as upstream and downstream introns (Figure 3B). In line with our hypothesis, normal splicing of the reporter mRNA was observed when the intron upstream of the cryptic 5′SS was present (Figure 3C). In contrast, the irregular 5′SS was preferentially used when EJC deposition was prevented by deleting the upstream intron (Figure 3C). These results, together with our data on the position and strength of RNPS1-dependent SS (Figure 3A),

194 suggest that EJC-bound RNPS1 protects a certain region of the downstream exon from the use
195 of irregular 5′SS.

196 We recognized that our model of RNPS1-dependent 5′SS regulation could also explain exon
197 skipping events, when the skipped exon starts with a potential 5′SS (i.e. contains a GU or GC
198 dinucleotide at the 5′ end as part of a reconstituted SS). In these cases, the same mechanism that
199 regulates cryptic 5′SS could also affect reconstituted SS at the beginning of exons. This idea
200 was supported by the observation that after RNPS1-sensitive exon 4 skipping of the RER1
201 mRNA, a single guanosine nucleotide remained (Figure 3D-F) which could only result from
202 the usage of a reconstituted 5′SS. Indeed, the exon-exon boundary between exon 3 and 4
203 contains a 5′SS, which can be ligated to the splice acceptor of the downstream intron causing
204 exon 4 skipping (Figure 3G). This re-splicing of a reconstituted exonic 5′SS resembles the
205 process of sequential multi-step splicing of introns termed recursive splicing (Sibley et al.,
206 2016). To test if exon 4 skipping in the RER1 mRNA results from re-splicing, we took
207 advantage of a set of RER1 reporter plasmids (Figure 3H). When both introns were present in
208 the reporter (e3-e5), exon 4 was constitutively retained, mimicking the constitutive exon
209 inclusion observed *in vivo* (Figure 3H and 3I). A reporter construct lacking the upstream intron
210 3 (Δi3), which simulates splicing of this intron without EJC-deposition, produced exclusively
211 mRNAs lacking exon 4 (Figure 3H and 3I). Together, these data suggest that intron 3 splicing
212 and concomitant EJC deposition suppresses the use of an adjacent reconstituted SS. When we
213 removed both introns (Δi3+i4) or mutated the GU dinucleotide of the reconstituted 5′SS (Δi3
214 GU-Mut), exon 4 was retained (Figure 3H and 3I). These findings confirm that in the absence
215 of EJC deposition, the RER1 transcript undergoes re-splicing using the reconstituted 5′SS and
216 the splice acceptor of intron 4. Strikingly, exon skipping occurred despite the presence of the
217 genuine 5′SS of intron 4 that exhibits a stronger SS consensus score compared to the
218 reconstituted 5′SS (Figure S3A). This observation raises the interesting question of which other

219 determinants affect this SS selection. Apart from the SS consensus sequence, exonic splicing
220 enhancer (ESE) or silencer (ESS) motifs greatly influence the balance of SS usage (Caceres and
221 Hurst, 2013). Upon inspection of RER1 exon 4, we suspected that an unfavorable arrangement
222 of ESE and ESS leads to the selection of the reconstituted 5′SS. In support of this hypothesis,
223 removal of several ESS from the central region of exon 4 and insertion of ESE motifs at the
224 exon's 5′ end gradually restored the usage of the canonical intronic SS in the RER1 Δi3 reporter
225 (Figure S3A-C). These results indicate that RNPS1 and the EJC directly or indirectly counteract
226 exonic splicing silencer motifs, effectively leading to the definition of exons and exon inclusion.

227 In contrast to RER1, where a residual guanosine nucleotide served as a molecular mark of two
228 consecutive splicing events, most exon skipping events observed upon RNPS1-knockdown
229 displayed seamless skipping of one or more exons. We therefore investigated whether these
230 splicing events equally relied on re-splicing or resulted from a direct definition of exon-exon
231 boundaries across multiple introns. To this end, we generated reporter constructs of the
232 HSD17B10 transcript, for which a moderate increase in exon skipping was observed upon
233 depletion of RNPS1 (Figure 3J). Analogous to RER1 splicing, near-complete exon skipping
234 occurred in an HSD17B10 reporter transcript lacking the upstream intron, indicating that this
235 event is indeed a result of re-splicing (Figure 3K and 3L). This remarkable conversion of the
236 splice pattern is surprising, considering that this HSD17B10 splicing event was near the
237 detection limit of the MAJIQ alternative splicing analysis (fourth last record, Table S1).
238 Importantly, the HSD17B10 re-splicing event is indistinguishable from regular exon skipping
239 by standard computational analyses, because the first two nucleotides of the skipped exon
240 represent the GU dinucleotide of the 5′SS. This observation suggests that other exon skipping
241 events observed upon RNPS1 knockdown potentially utilize the same re-splicing mechanism.
242 To estimate the probability of seamless re-splicing, we identified the 5′ terminal dinucleotide
243 of the first skipped exon in spurious exon skipping events. More than 75% of the skipped exons

exhibited a suitable 5′SS on their 5′ end, suggesting that re-splicing is likely responsible for these exon skipping events (Figure S3D).

The EJC confers local protection from cryptic splicing

Based on our observation that compromised EJC deposition results in aberrant splicing, we tested whether site directed positioning of EJC factors on de-repressed mRNAs can restore normal splicing. To this end, we generated a reporter transcript with MS2 tethering sites to recruit effector proteins to different positions upstream of the reconstituted 5′SS of HSD17B10 Δi4, which constitutively undergoes re-splicing (Figure 4A and 4B). Tethering of EJC proteins to this mRNA mimics EJC assembly upstream of exon-exon junctions and thus uncouples EJC deposition from pre-mRNA splicing. Indeed, tethering of RNPS1 or the EJC core component RBM8A promoted exon inclusion over a short distance, but gradually lost its effect when the distance between the tethering site and the 5′SS was increased (Figure 4B). These results suggest that MS2-mediated positioning of RNPS1 or EJC components on the mRNA functionally recapitulates splicing-dependent EJC-deposition. Also, our data suggest that the effect of the EJC and RNPS1 is confined to a region surrounding the EJC binding site.

RNPS1 in the PSAP complex is the functional component for 5′ splice site suppression

To uncover the mechanistic details of cryptic SS suppression by RNPS1, we set out to define the minimal protein domain required for this process. Using rescue assays, we determined that the expression of the isolated RRM domain of RNPS1 considerably rescued RER1 splicing in RNPS1-depleted cells (Figure 4C). This indicates that the isolated RRM domain of RNPS1 retains some functional activity, potentially mediated through the recruitment of other protein factors, such as ASAP/PSAP complex components. To test this, we examined the effects of siRNA-mediated depletion of ASAP/PSAP components. Surprisingly, only the knockdown of SAP18 and PNN, but not ACIN1 resulted in RER1 exon skipping, (Figure 4D, E). Furthermore,

expression of full-length SAP18 or truncated PNN constructs, which support PSAP formation, restored normal splicing of the RER1 transcript in SAP18 or PNN-depleted cells, respectively (Figure 4F). Moreover, tethering of SAP18 or PNN to the HSD17B10 reporter stimulated exon inclusion (Figure S4A and S4B), demonstrating that the PSAP complex is the functional entity involved in EJC-mediated cryptic SS suppression.

To further dissect the hierarchy of EJC- and PSAP-dependent splicing regulation and to identify the active components of the PSAP complex, we performed epistasis experiments using tethering assays in knockdown cells. Tethering of RNPS1 resulted in exon inclusion of the HSD17B10 reporter mRNA even in EJC- or PSAP-complex depleted cells (Figure 4G). In contrast, tethered PNN and SAP18 were inactive in the absence of RNPS1 (Figure 4H) and tethering of PSAP-incompatible RNPS1 (176) led to similar exon inclusion as tethering of the RNPS1 wildtype protein (Figure 4I). This result supports the hypothesis that RNPS1 is a key effector of EJC- and PSAP-mediated splicing regulation, while EJC and other PSAP components are required primarily for the correct recruitment of RNPS1 to the mRNA.

Based on these findings we postulate that the PSAP complex, consisting of PNN, SAP18 and RNPS1, contributes to exon definition, a function that suppresses exon skipping and resembles the activity of classical SR proteins (Ibrahim et al., 2005). Supporting this exon definition hypothesis, tethering of SRSF2 or SRSF11, a reported interaction partner of RNPS1 (Sakashita et al., 2004), altered the HSD17B10 splicing similar to RNPS1 or RBM8A tethering (Figure S4C). However, EJC-dependent exon definition functions independently of these SR proteins, because depletion of SRSF1 and SRSF2 (single or combined) or SRSF11 did not activate RNPS1-dependent cryptic SS (Figure S4D-F). Finally, tethered nuclear beta-galactosidase (NLS-LacZ, a homotetramer of ~ 120 kDa proteins) had no effect on exon skipping, indicating that the presence of a large protein complex is not sufficient to prevent exon skipping (Figure

S4C). Taken together, these results imply that RNPS1 is an essential effector of SS suppression and is recruited to spliced mRNPs via the formation of PSAP rather than ASAP complexes.

EJCs prevent 3′ splice sites usage in an RNPS1-independent manner

Transcripts harboring RNPS1-dependent cryptic or reconstituted SS also showed splicing defects upon depletion of the EJC core components EIF4A3 and RBM8A (Figure 1B and Figure S5A-I). Moreover, expression of a dominant negative splicing factor that is unable to recruit EIF4A3 to the spliceosome (CWC22 (171)) and thus enables pre-mRNA splicing without EJC deposition (Steckelberg et al., 2012), also caused RER1 exon skipping (Figure S5J and S5K). Hence, a clear link between EJC deposition and splice regulation by RNPS1 exists. However, comparing the identified transcriptome-wide RNPS1-dependent splicing events with previously published RNA-Seq data from EIF4A3-depleted cells (Wang et al., 2014), we noticed that many splicing events induced by EJC knockdown were not found in RNPS1-depleted cells (Figure S5L-P and Table S2).

Although it is expected that the EJC knockdown strongly inhibits downstream processes like NMD, leading to the overstabilization of mis-spliced mRNAs, this observation suggested that additional EJC-dependent (but RNPS1-independent) mechanisms act on these mRNAs. The depletion of the EJC core component EIF4A3 caused mainly exon skipping, similar to RNPS1 knockdown (Figure 5A and S6A). However, the second most frequent splicing alteration was the use of alternative 3′SS, an event that was less often observed upon RNPS1 depletion. This observation indicated a mechanistic difference between EJC- and PSAP-dependent splicing regulation and prompted us to investigate this type of splicing dysregulation in more detail. In search of a possible mechanistic explanation for EJC-dependent splicing regulation, we analyzed the ACIN1 transcript, in which a cryptic 3′SS was used for splicing upon EIF4A3 or RBM8A, but not RNPS1 depletion (Figure 5B and 5C). We reasoned that a splicing order in which splicing of intron 14 precedes splicing of intron 13, could explain the observed splicing

pattern (Figure 5D). Splicing of intron 14 generates a composite exonic 3′SS spanning the junction between exon 14 and 15, which led to exon skipping if used together with the 5′SS of intron 13. Consistent with our theory, an ACIN1 reporter lacking intron 14 (Figure 5E) produced almost exclusively the skipped transcript, whereas a reporter construct with two introns was normally spliced (Figure 5F). The same order of splicing events was also observed for ATP5B and ATP5F1 transcripts, for which mis-splicing is enabled due to the activation of a cryptic 3′ splice acceptor site (Figure 5G-I). Similarly, splicing of the downstream intron precedes splicing of the upstream intron in the CIAO1 mRNA (Figure S6B-J). As a result, the EJC helps to suppress two cryptic 3′SS downstream of the exon-exon junction by RNPS1-dependent and -independent mechanisms. For all transcripts analyzed in detail, the predicted EJC binding sites coincide with putative branch points or polypyrimidine tracts of the EJC-dependent alternative 3′SS. Furthermore, the EJC-suppressed alternative 3′SS are mainly clustered at EJC binding sites or at the 5′ end of the following exon (Figure S5Q). This finding suggests that the RNPS1-independent EJC function is due to direct masking of important splice-regulatory sequences. This hypothesis was supported by the insertion of HA tag sequences in the ATP5F1 and ATP5B reporter, which moved the alternative 3′SS sufficiently upstream of EJC deposition sites, thereby de-repressing and constitutively activating these SS (Figure 5H and 5I). Taken together, in addition to the PSAP-mediated splicing regulation, the deposition of EJCs prevents the loss of exonic sequences via usage of reconstituted or cryptic 3′SS.

EJCs maintain transcriptome integrity and cellular survival

The importance of EJC and RNPS1-dependent splice site protection is underscored by the observation that RBM8A-depleted cells and to a lesser extend RNPS1-depleted cells showed reduced proliferation (Figure 6A), presumably caused by the activation of several signaling cascades related to stress and apoptosis (Figure 6B). As the EJC is involved in a multitude of gene expression processes, for example, mRNA quality control by NMD, the cellular stress

cannot be solely attributed to the uncontrolled splicing upon EJC knockdown. However, compared to the combined depletion of RNPS1 and SMG1, which is a central component of the NMD machinery, downregulation of RBM8A caused a more severe stress phenotype (Figure 6B). We suspected that normal EJC deposition and concomitant cryptic SS repression maintains proper transcriptome integrity and is therefore required for cellular fitness. Indeed, increased production of several shorter mRNA variants was observed in EJC knockdowns, when studying genes that harbor more than one EJC-dependent cryptic SS (Figure 6C-F). We reasoned that this shift from productive to non-productive splicing and therefore the loss of intact, full length transcripts might have a significant effect on cellular survival. In addition, we searched for mis-splicing of genes essential for cellular fitness by comparing the change of junction usage of EIF4A3-dependent spurious splice events with the fitness score of the respective gene, determined by a high-resolution CRISPR-screen (Figure 6G; (Hart et al., 2015)). In total, we found 184 mis-spliced genes that were classified as essential for survival and proliferation. Although many of the splice changes affect only one exon, we found many essential genes that produce substantial amounts of mis-spliced mRNAs, which are expected to encode for non-functional proteins. One example is the transcript of the splicing and DNA damage repair component PRPF19, for which exon 15 skipping upon EJC depletion leads to the deletion of two WD40 domains (Figure 6H), rendering the protein incapable of supporting proper DNA damage response (Marechal et al., 2014). Another essential target is uridine monophosphate synthetase (UMPS), which is required for the UMP biosynthesis pathway. UMPS transcripts lacking exon 2 due to exon skipping generate an early frame-shift and thus lead to the expression of a truncated protein isoform (Figure 6I). Therefore, the deposition of EJC on nascent mRNAs has an instant protective effect on gene expression, maintains the expression of many essential genes and presumably represents an essential function of the EJC.

Discussion

The human genome contains many cryptic SS that usually are used only at a very low frequency despite having similar sequences as canonical SS (Buratti et al., 2011). It is assumed that cryptic SS are suppressed by stronger SS in their vicinity and are only activated by mutations of nearby authentic SS (Roca et al., 2003). However, our data suggest that SS selection is not solely the result of SS competing for the splicing machinery. We find that many cryptic and reconstituted SS are efficiently silenced under normal conditions by the deposition of an EJC on adjacent exon-exon junctions. We uncover several different mechanisms by which EJCs globally inhibit the use of nearby irregular 5′ and 3′SS (Figure 7A). Since the EJC itself is deposited on the RNA during splicing, this mechanism functions as a positive feedback loop to reinforce authentic SS and establishes a hierarchy of preferential SS usage.

Mechanism and consequences of 5′SS suppression

The position of the EJC at the 3′ end of an exon is well suited to oversee the 5′ end of the next exon and to suppress exonic SS. Our work establishes that the EJC recruits the PSAP complex to exert an exon inclusion effect that decreases proportionally with the distance. Consequently, PSAP functions mainly in the vicinity of previously spliced introns, so that cryptic SS usage within already ligated exons is prevented. In contrast to other splice factors such as SR proteins, for which several binding sequences can be present in an exon (Long and Caceres, 2009), EJCs recruit PSAP complexes in a splicing-dependent, but sequence-independent manner to a single site at the 3′ end of an exon. Despite its limited binding potential/possibilities, the efficiency of PSAPs cryptic splice site suppression is remarkable. In many cases, EJC-regulated cryptic splices are barely used under normal conditions. While the knockdown of EJC components often causes only partial mis-splicing of endogenous transcripts (presumably due to incomplete knockdown of the proteins), we observe almost 100% splicing defects in our reporter constructs when EJC binding was prevented by intron deletions. Interestingly, there are also similarities

between PSAP complex and SR proteins. For example, we observe a distance-dependent exonization effect by the PSAP complex as has been described for SR proteins (Graveley et al., 1998). While several components of an EJC-dependent exon definition complex are now characterized, more work will be needed to determine the exact mechanism of 5′SS suppression. It will be of particular importance to uncover the process by which the PSAP complex guides U1 snRNP binding to 5′SS.

Mechanism and consequences of 3′SS protection

In addition to the PSAP-dependent 5′SS suppression, we observed a different, merely steric mechanism in the control of 3′SS. We conclude that the binding of the EJC core prevents the correct recognition of the 3′SS by the spliceosome, likely because factors such as U2AF1/2 are not able to bind to the RNA if an EJC occupies their genuine binding site. A similar interference with the assembly of the early spliceosome has been reported for a diverse group of repressors of 3′SS usage, including PTB or Sxl (Izquierdo et al., 2005; Valcarcel et al., 1993). Many of these factors compete with U2AF for binding to the pyrimidine tract, and therefore the inhibitory potential of the repressor depends on its binding strength to the 3′SS (Sohail and Xie, 2015). In contrast, the EJC has a clear competitive advantage over U2AF because its co-spliceosomal deposition ensures that cryptic or reconstituted SS are nearly immediately rendered inaccessible. Furthermore, the strong EJC binding to mRNA is usually only resolved by a translating ribosome in the cytoplasm (Gehring et al., 2009b). This advantage explains the surprisingly robust effect of the EJC on the suppression of 3′SS.

Significance of the EJC for gene expression

It is evident that the complex and heterogeneous architecture of the human transcriptome demands a mechanism that marks the position of already spliced introns and prevents re-splicing in their vicinity. The EJC meets this demand by masking and suppressing aberrant SS and thereby protecting the bound transcript from unintentional loss of exonic sequences (Figure

7B). Conceptually, this protective effect of the EJC is comparable to the telescripting mechanism in which the U1 snRNP prevents premature transcription termination by cleavage and polyadenylation of mRNAs in order to ensure transcriptome integrity (Berg et al., 2012). Similar to U1 depletion resulting in mRNA shortening, EJC depletion results in transcript disintegration by de-repression of single or multiple SS in a given transcript (Figure 7B). Mechanistically, the EJC-dependent re-splicing events share characteristics with reported multi-step splice processes such as recursive splicing or intra-splicing (Parra et al., 2008; Sibley et al., 2016). Functionally, however, the EJC-related re-splicing events reported here result primarily in destructive splice patterns, whereas canonical recursive splicing is defined as a productive mechanism resulting in the correct excision of introns.

We speculate that reduced expression of the EJC core components EIF4A3 and RBM8A can lead to significant changes in the transcriptome, explaining why the EJC is important for embryonic development (Mao et al., 2016) and how its misregulation can cause serious human diseases (Albers et al., 2012; Favaro et al., 2014). In sensitive cells or tissues, these changes in gene expression will affect cellular fitness and eventually have a negative impact on tissue differentiation and maintenance. Hence, our work uncovers EJCs as essential components of a splice-regulatory pathway, which safeguard transcriptome integrity and protect mRNAs against disruptive splicing events.

Acknowledgements

We thank members of the Gehring lab for discussions and reading of the manuscript. We also thank Marek Franitza and Christian Becker (Cologne Center for Genomics, CCG) for preparing the sequencing libraries and operating the sequencer. We are grateful to Thomas Wiehe and Peter Heger (Institute for Genetics, Cologne) for sharing computational infrastructure. We also thank Jernej Ule for sharing unpublished data and discussing the mechanism of recursive splicing. This work was supported by grants from the DST to K.K.S. (ECR/2015/000166) and the Deutsche Forschungsgemeinschaft to A.-L.S. (STE 2509/2-1), C.D. (DI 1501/8-1) and N.H.G (GE 2014/6-1). V.B. was funded under the Institutional Strategy of the University of Cologne within the German Excellence Initiative. N.H.G. acknowledges support by a Heisenberg fellowship (GE 2014/5-1 and GE 2014/7-1) from the Deutsche Forschungsgemeinschaft.

Author Contributions

Conceptualization, N.H.G., V.B., and L.B.; Methodology, N.H.G., V.B., A.-L.S., K.K.S and J.V.G.; Software, T.B.B., C.D. and V.B.; Investigation, V.B., A.-L.S., K.K.S., E.G., J.A.; Resources and Data Curation, T.B.B. and C.D.; Writing – Original Draft, Review & Editing, N.H.G., V.B., A.-L.S., T.B.B., C.D., J.V.G. and K.K.S.; Visualization, V.B.; Supervision and Funding Acquisition, N.H.G. and C.D.;

Declaration of Interests

The authors declare no competing interests.

References

Albers, C.A., Paul, D.S., Schulze, H., Freson, K., Stephens, J.C., Smethurst, P.A., Jolley, J.D., Cvejic, A., Kostadima, M., Bertone, P., *et al.* (2012). Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. Nat Genet *44*, 435-439, S431-432.

Alexandrov, A., Colognori, D., Shu, M.D., and Steitz, J.A. (2012). Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. Proc Natl Acad Sci U S A *109*, 21313-21318.

Andersen, C.B., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L., Pedersen, J.S., Seraphin, B., Le Hir, H., and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. Science *313*, 1968-1972.

Ashton-Beaucage, D., Udell, C.M., Lavoie, H., Baril, C., Lefrancois, M., Chagnon, P., Gendron, P., Caron-Lizotte, O., Bonneil, E., Thibault, P., *et al.* (2010). The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in Drosophila. Cell *143*, 251-262.

Barbosa, I., Haque, N., Fiorini, F., Barrandon, C., Tomasetto, C., Blanchette, M., and Le Hir, H. (2012). Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. Nat Struct Mol Biol *19*, 983-990.

Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L., *et al.* (2012). U1 snRNP determines mRNA length and regulates isoform expression. Cell *150*, 53-64.

Boehm, V., and Gehring, N.H. (2016). Exon Junction Complexes: Supervising the Gene Expression Assembly Line. Trends Genet *32*, 724-735.

Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. Cell *126*, 713-725.

Buratti, E., Chivers, M., Hwang, G., and Vorechovsky, I. (2011). DBASS3 and DBASS5: databases of aberrant 3'- and 5'-splice sites. Nucleic Acids Res *39*, D86-91.

Caceres, E.F., and Hurst, L.D. (2013). The evolution, impact and properties of exonic splice enhancers. Genome Biol *14*, R143.

Corvelo, A., Hallegger, M., Smith, C.W., and Eyras, E. (2010). Genome-wide association between branch point properties and alternative splicing. PLoS Comput Biol *6*, e1001016.

Dale, R.K., Pedersen, B.S., and Quinlan, A.R. (2011). Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. Bioinformatics *27*, 3423-3424.

Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., and Beroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res *37*, e67.

494 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson,
495 M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*,
496 15-21.

497 Dodt, M., Roehr, J.T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR-Flexible Barcode and
498 Adapter Processing for Next-Generation Sequencing Platforms. Biology (Basel) *1*, 895-905.

499 Favaro, F.P., Alvizi, L., Zechi-Ceide, R.M., Bertola, D., Felix, T.M., de Souza, J., Raskin, S.,
500 Twigg, S.R., Weiner, A.M., Armas, P., *et al.* (2014). A noncoding expansion in EIF4A3 causes
501 Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. Am J
502 Hum Genet *94*, 120-128.

503 Fukumura, K., Inoue, K., and Mayeda, A. (2018). Splicing activator RNPS1 suppresses errors
504 in pre-mRNA splicing: A key factor for mRNA quality control. Biochem Biophys Res Commun
505 *496*, 921-926.

506 Gehring, N.H., Lamprinaki, S., Hentze, M.W., and Kulozik, A.E. (2009a). The hierarchy of
507 exon-junction complex assembly by the spliceosome explains key features of mammalian
508 nonsense-mediated mRNA decay. PLoS Biol *7*, e1000120.

509 Gehring, N.H., Lamprinaki, S., Kulozik, A.E., and Hentze, M.W. (2009b). Disassembly of exon
510 junction complexes by PYM. Cell *137*, 536-548.

511 Graveley, B.R., Hertel, K.J., and Maniatis, T. (1998). A systematic analysis of the factors that
512 determine the strength of pre-mRNA splicing enhancers. EMBO J *17*, 6747-6756.

513 Han, S.P., Tang, Y.H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present
514 and perspectives. Biochem J *430*, 379-392.

515 Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M.,
516 Zimmermann, M., Fradet-Turcotte, A., Sun, S., *et al.* (2015). High-Resolution CRISPR Screens
517 Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell *163*, 1515-1526.

518 Hauer, C., Sieber, J., Schwarzl, T., Hollerer, I., Curk, T., Alleaume, A.M., Hentze, M.W., and
519 Kulozik, A.E. (2016). Exon Junction Complexes Show a Distributional Bias toward
520 Alternatively Spliced mRNAs and against mRNAs Coding for Ribosomal Proteins. Cell Rep
521 *16*, 1588-1603.

522 Hayashi, R., Handler, D., Ish-Horowicz, D., and Brennecke, J. (2014). The exon junction
523 complex is required for definition and excision of neighboring introns in Drosophila. Genes
524 Dev *28*, 1772-1785.

525 Ibrahim, E.C., Schaal, T.D., Hertel, K.J., Reed, R., and Maniatis, T. (2005). Serine/arginine-
526 rich protein-dependent suppression of exon skipping by exonic splicing enhancers. Proc Natl
527 Acad Sci U S A *102*, 5002-5007.

528 Izquierdo, J.M., Majos, N., Bonnal, S., Martinez, C., Castelo, R., Guigo, R., Bilbao, D., and
529 Valcarcel, J. (2005). Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and
530 PTB on exon definition. Mol Cell *19*, 475-484.

531 Katz, Y., Wang, E.T., Silterra, J., Schwartz, S., Wong, B., Thorvaldsdottir, H., Robinson, J.T.,
532 Mesirov, J.P., Airoldi, E.M., and Burge, C.B. (2015). Quantitative visualization of alternative
533 exon expression from RNA-seq data. Bioinformatics *31*, 2400-2402.

534 Kornblihtt, A.R., Schor, I.E., Allo, M., Dujardin, G., Petrillo, E., and Munoz, M.J. (2013).
535 Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev
536 Mol Cell Biol *14*, 153-165.

537 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat
538 Methods *9*, 357-359.

539 Le Hir, H., Sauliere, J., and Wang, Z. (2016). The exon junction complex as a node of post-
540 transcriptional networks. Nat Rev Mol Cell Biol *17*, 41-54.

541 Lee, Y., and Rio, D.C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing.
542 Annu Rev Biochem *84*, 291-323.

543 Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet:
544 Visualization of Intersecting Sets. IEEE Trans Vis Comput Graph *20*, 1983-1992.

545 Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master
546 regulators of gene expression. Biochem J *417*, 15-27.

547 Malone, C.D., Mestdagh, C., Akhtar, J., Kreim, N., Deinhard, P., Sachidanandam, R., Treisman,
548 J., and Roignant, J.Y. (2014). The exon junction complex controls transposable element activity
549 by ensuring faithful splicing of the piwi transcript. Genes Dev *28*, 1786-1799.

550 Mao, H., McMahon, J.J., Tsai, Y.H., Wang, Z., and Silver, D.L. (2016). Haploinsufficiency for
551 Core Exon Junction Complex Components Disrupts Embryonic Neurogenesis and Causes p53-
552 Mediated Microcephaly. PLoS Genet *12*, e1006282.

553 Marechal, A., Li, J.M., Ji, X.Y., Wu, C.S., Yazinski, S.A., Nguyen, H.D., Liu, S., Jimenez,
554 A.E., Jin, J., and Zou, L. (2014). PRP19 transforms into a sensor of RPA-ssDNA after DNA
555 damage and drives ATR activation via a ubiquitin-mediated circuitry. Mol Cell *53*, 235-246.

556 Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M.Q., Gardiner, E.M., and Krainer, A.R.
557 (1999). Purification and characterization of human RNPS1: a general activator of pre-mRNA
558 splicing. EMBO J *18*, 4560-4570.

559 Murachelli, A.G., Ebert, J., Basquin, C., Le Hir, H., and Conti, E. (2012). The structure of the
560 ASAP core complex reveals the existence of a Pinin-containing PSAP complex. Nat Struct Mol
561 Biol *19*, 378-386.

562 Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative
563 splicing. Nature *463*, 457-463.

564 Papasaikas, P., and Valcarcel, J. (2016). The Spliceosome: The Ultimate RNA Chaperone and
565 Sculptor. Trends Biochem Sci *41*, 33-45.

566 Parra, M.K., Tan, J.S., Mohandas, N., and Conboy, J.G. (2008). Intrasplicing coordinates
567 alternative first exons with alternative splicing in the protein 4.1R gene. EMBO J *27*, 122-131.

568 Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L.
569 (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat
570 Biotechnol *33*, 290-295.

571 Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing
572 genomic features. Bioinformatics *26*, 841-842.

573 Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016).
574 g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids
575 Res *44*, W83-89.

576 Roca, X., Sachidanandam, R., and Krainer, A.R. (2003). Intrinsic differences between authentic
577 and cryptic 5' splice sites. Nucleic Acids Res *31*, 6321-6333.

578 Rodgers, P., Stapleton, G., Flower, J., and Howse, J. (2014). Drawing area-proportional Euler
579 diagrams representing up to three sets. IEEE Trans Vis Comput Graph *20*, 56-69.

580 Rodor, J., Pan, Q., Blencowe, B.J., Eyras, E., and Caceres, J.F. (2016). The RNA-binding
581 profile of Acinus, a peripheral component of the exon junction complex, reveals its role in
582 splicing regulation. RNA *22*, 1411-1426.

583 Roignant, J.Y., and Treisman, J.E. (2010). Exon junction complex subunits are required to
584 splice Drosophila MAP kinase, a large heterochromatic gene. Cell *143*, 238-250.

585 Sakashita, E., Tatsumi, S., Werner, D., Endo, H., and Mayeda, A. (2004). Human RNPS1 and
586 its associated factors: a versatile alternative pre-mRNA splicing regulator in vivo. Mol Cell Biol
587 *24*, 1174-1187.

588 Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative
589 C(T) method. Nat Protoc *3*, 1101-1108.

590 Sibley, C.R., Blazquez, L., and Ule, J. (2016). Lessons from non-canonical splicing. Nat Rev
591 Genet *17*, 407-421.

592 Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J.
593 (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR
594 protein nexus. Cell *151*, 750-764.

595 Sohail, M., and Xie, J. (2015). Diverse regulation of 3' splice site usage. Cell Mol Life Sci *72*,
596 4771-4793.

597 Steckelberg, A.L., Boehm, V., Gromadzka, A.M., and Gehring, N.H. (2012). CWC22 connects
598 pre-mRNA splicing and exon junction complex assembly. Cell Rep *2*, 454-461.

599 Steckelberg, A.L., and Gehring, N.H. (2014). Studying the composition of mRNPs in vitro
600 using splicing-competent cell extracts. Methods *65*, 342-349.

601 Tapial, J., Ha, K.C.H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A.,
602 Quesnel-Vallieres, M., Permanyer, J., Sodaei, R., Marquez, Y., *et al.* (2017). An atlas of
603 alternative splicing profiles and functional associations reveals new regulatory programs and
604 genes that simultaneously express multiple major isoforms. Genome Res *27*, 1759-1768.

605 Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer
606 (IGV): high-performance genomics data visualization and exploration. Brief Bioinform *14*,
607 178-192.

608 Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013).
609 Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol
610 *31*, 46-53.

611 Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg,
612 S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of
613 RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.

614 Valcarcel, J., Singh, R., Zamore, P.D., and Green, M.R. (1993). The protein Sex-lethal
615 antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA.
616 Nature *362*, 171-175.

617 Vaquero-Garcia, J., Barrera, A., Gazzara, M.R., Gonzalez-Vallinas, J., Lahens, N.F.,
618 Hogenesch, J.B., Lynch, K.W., and Barash, Y. (2016). A new view of transcriptome complexity
619 and regulation through the lens of local splicing variations. Elife *5*, e11752.

620 Wang, Z., Ballut, L., Barbosa, I., and Le Hir, H. (2018). Exon Junction Complexes can have
621 distinct functional flavours to regulate specific splicing events. Sci Rep *8*, 9509.

622 Wang, Z., Murigneux, V., and Le Hir, H. (2014). Transcriptome-wide modulation of splicing
623 by the exon junction complex. Genome Biol *15*, 551.

624 Woodward, L.A., Mabin, J.W., Gangras, P., and Singh, G. (2017). The exon junction complex:
625 a lifelong guardian of mRNA fate. Wiley Interdiscip Rev RNA *8*.

626 Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with
627 applications to RNA splicing signals. J Comput Biol *11*, 377-394.

628 Young, L., Sung, J., Stacey, G., and Masters, J.R. (2010). Detection of Mycoplasma in cell
629 cultures. Nat Protoc *5*, 929-934.

630

Figure titles and legends

Figure 1. RNPS1 regulates splice site selection via an interaction with the exon junction complex

(A) Schematic overview of splicing modulation by exon junction complex (EJC) and ASAP/PSAP-complex formation.

(B) RT-PCR analysis of MRPL3 exon 4 skipping with RNA from HeLa cells transfected with the indicated siRNA. MRPL3 exon architecture is depicted schematically, alternatively spliced features are highlighted.

(C) Quantitative RT-PCR analysis of MRPL3 exon 4 skipping normalized to exon3/4 splicing and compared to the Luc control knockdown. n=3.

(D) Co-immunoprecipitation of EJC core component EIF4A3 and ASAP/PSAP component SAP18 from *in vitro* splicing experiments with the indicated FLAG-tagged RNPS1 variants. n=3

(E) *In vitro* splicing of ^{32}P body-labeled MINX mRNA in the presence of FLAG-RNPS1 variants. n=3

(F) RT-PCR analysis of MRPL3 exon 4 skipping with RNA from stable HeLa cells expressing the indicated rescue proteins, transfected with the indicated siRNA. Western blot analysis of protein expression is shown at the bottom.

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

See also Figure S1.

Figure 2. Transcriptome-wide alternative splicing events upon RNPS1 depletion

(A) Simplified overview of experimental RNA-Seq pipeline.

(B) Classification of selected alternatively spliced junctions upon RNPS1 knockdown as exon skipping (ES), exon inclusion (EI), intron-retention (IR), alternative 5′ or 3′ splice sites (A5SS, A3SS) or exitron (EX) events (see inset). The quantity of spurious junctions for each class is shown in yellow. The change in junction usage (delta percent spliced in; dPSI) and percent spliced in (PSI) fold change for each class are shown as boxplots on the bottom. The full classification plot is shown in Figure S2A.

(C) Heatmaps representing junction usage (percent spliced in; PSI) of alternative splicing events in control (Luc) and RNPS1 knockdown are depicted on the left. The fraction of spurious junctions and the PSI fold change in the same samples are shown on the right. Only junctions with a strong response in RNPS1-depleted cells with a dPSI of < -0.1 were selected.

(D), (E) Sashimi-plots of RNA-Seq data of genes with exon skipping (OCIAD1, (D)) and alternative 5′ splice site (SS) usage (TUFM, (E)). Only selected splice junctions are depicted. The thickness and color of the depicted junction represent the junction usage (PSI and dPSI). The counts of reads spanning the indicated junctions are shown.

(F), (G) RT- PCR analysis of the alternative splice events shown in (D) and (E). Total RNA was isolated from stable HeLa cells expressing the indicated rescue proteins and transfected with the indicated siRNA. Quantified results from the indicated biological replicates are shown as mean ± SD and compared to the Luc control knockdown.

See also Figure S2 and Table S1.

Figure 3. RNPS1-depletion leads to re-splicing of cryptic and reconstituted 5′ splice sites

(A) Depiction of alternative 5′SS position of spurious junctions relative to exon boundaries as density plot (bottom) and the corresponding 5′ MaxEnt scores (top).

(B), (C) TUFM minigene reporter constructs (B) were expressed in HeLa Tet-Off cells and analyzed via RT-PCR (C).

(D) Sashimi-plots of RER1 exon 4 skipping from RNA-Seq data.

(E) IGV snapshot of the guanosine insertion for RER1 reads.

(F) RER1 RT-PCR of RNA from HeLa cells transfected with the indicated siRNA. Sanger sequencing of the exon 4-skipped RER1 PCR product is shown with the guanosine retained from exon 4 highlighted.

(G) Scheme of RER1 exon 3-5 re-splicing in the presence (top) or absence (bottom) of RNPS1.

(H), (I) RER1 minigene reporter constructs (H) were expressed in HeLa Tet-Off cells and analyzed via RT-PCR (I).

(J) HSD17B10 RT-PCR of RNA from HeLa cells transfected with the indicated siRNA. The 5′ terminal GU dinucleotide at the exon 5 of HSD17B10 is indicated.

(K), (L) HSD17B10 minigene reporter constructs (K) were expressed in HeLa Tet-Off cells and analyzed via RT-PCR (L).

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

See also Figure S3.

Figure 4. Functional suppression of cryptic splice sites by the PSAP complex

(A) Scheme of HSD17B10 Δi4 tethering reporter. Two ESE-optimized MS2 stem-loops were inserted at varying distances upstream of the reconstituted HSD17B10 cryptic splice site, allowing the direct tethering of MS2V5-tagged proteins.

(B) The indicated MS2V5-tethering proteins and HSD17B10 Δi4 reporter with varying spacers were expressed transiently in HeLa Tet-Off cells and the splice patterns analyzed by RT-PCR.

(C), (D), (F) RT-PCR analysis of RER1 exon skipping with RNA from stable HeLa cells, expressing the indicated rescue proteins, transfected with the indicated siRNA. Western blot analysis of expressed FLAG- and FLAG-emGFP-tagged proteins is shown. Tubulin served as loading control.

(E) Scheme depicting the components of the ASAP and PSAP complexes.

(G), (H), (I) Dual inducible stable HeLa cell lines expressing both the HSD17B10 Δi4 e4-15 reporter and the indicated MS2V5-tagged tethering protein were transfected with the indicated siRNA and reporter splicing was detected via RT-PCR. Endogenous RER1 splicing RT-PCR analysis upon knockdown is shown for (G)-(H).

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

See also Figure S4.

Figure 5. Inability to deposit EJCs on mRNA result in the usage of aberrant 3′ splice sites

(A) Classification of selected alternatively spliced junctions upon EIF4A3 knockdown. The quantity of spurious junctions for each class is shown. The change in junction usage (delta percent spliced in; dPSI) and PSI fold change for each class are shown as boxplots on the bottom. The full classification plot is shown in Figure S6A.

(B) Sashimi-plots of multiple RNA-Seq data sets (Luc and RNPS1 vs. GFP, EIF4A3 and RBM8A) of ACIN1 exon 14 skipping. The predicted EJC binding site in relation to the cryptic splice site (SS) and putative branch points (BP) is indicated.

(C) RT-PCR analysis of ACIN1 exon 14 skipping with RNA from HeLa cells transfected with the indicated siRNA.

(D) Scheme indicating the ACIN1 re-splicing mechanism.

(E), (F) ACIN1 minigene reporter constructs (E) were expressed in HeLa Tet-Off cells and analyzed via RT-PCR (F).

(G) ATP5F1 and ATP5B RT-PCR of cDNA samples obtained from HeLa cells transfected with the indicated siRNA.

(H), (I) ATP5F1 and ATP5B minigene reporter constructs (H) were expressed in HeLa Tet-Off cells and analyzed via RT-PCR (I). Insertion of 27 nucleotide HA sequences in exon 4 (ATP5F1) or exon 6 (ATP5B) is schematically depicted.

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

See also Figure S5-S6 and Table S2.

Figure 6. RNPS1 and EJC depletion leads to transcript disintegration and cellular stress

(A) Proliferation of HeLa cells was measured after knockdown using the indicated siRNA. n=3

(B) Induction of stress pathways upon transfection of HeLa cells with the indicated siRNA. The indicated signaling molecules were detected via target-specific capture antibodies in a sandwich immunoassay. n=3

(C) - (F) Examples for accumulated mis-spliced transcripts upon EJC depletion, analyzed by RT-PCR.

(G) Plot of spurious junctions identified in EIF4A3 knockdown RNA-Seq data, comparing the log Bayes factor (BF) gene fitness score against the change in junction usage (dPSI). More positive scores increase the confidence in the essentiality of the gene. Individual targets are highlighted.

(H), (I) Scheme of highlighted targets in (G), depicting the transcript architecture, the strongest alternative splicing event, the expressed protein variants and the ratio of mis-splicing in control or EIF4A3 knockdown RNA-Seq data.

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

Figure 7. EJC and RNPS1 protect spliced transcripts from the usage of cryptic splice sites

(A) Mechanism of splice site suppression by the EJC and the RNPS1-containing PSAP complex.

(B) Deposition of EJCs prevents loss of exonic sequences by masking and suppressing cryptic splice sites in the vicinity, consequently enforcing correct splicing hierarchy. The inability to assemble EJCs on spliced transcripts results in the activation of cryptic SS, leading to mis-splicing events and loss of exonic sequences.

STAR METHODS

KEY RESOURCES TABLE

See separate Key Resources Table file.

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents should be directed to and will be fulfilled by the Lead Contact, Niels H. Gehring (ngehring@uni-koeln.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

HEK 293 Flp-In T-REx (human; sex: female; Thermo Fisher Scientific; RRID:CVCL_U427), HeLa Flp-In T-REx (human; sex: female; established by Elena Dobrikova and Matthias Gromeier, Duke University Medical Center) and HeLa Tet-Off (human; sex: female; Clontech; RRID:CVCL_V352) cells were cultured in DMEM (Gibco) supplemented with 9% fetal bovine serum (Gibco) and 1x Penicillin Streptomycin (Gibco). All cells were cultivated at 37°C and 5% CO2 in a humidified incubator.

METHOD DETAILS

Stable cell lines and plasmids

Mammalian expression constructs for *in vitro* splicing assays were inserted into the pCI-neo vector (Promega) with an N-terminal FLAG tag. MINX and MINX Δintron *in vitro* splice substrates were described previously (Gehring et al., 2009a; Gehring et al., 2009b). The point and deletion mutants of RNPS1 were PCR amplified and inserted into pCI-neo-FLAG. Accordingly, GST, CWC22 WT and CWC22 NK171DE mutant (described in (Steckelberg et al., 2012)) were cloned into pCI-neo-FLAG. For transient tethering assays, the constructs were subcloned into pCI-neo containing an N-terminal MS2V5 tag. For generating stable tethering cell lines, the constructs together with the MS2V5 tag were inserted in the cumate-inducible PB-CuO-MCS-IRES-GFP-EF1-CymR-Puro vector (System Biosciences). All reporter

constructs were PCR amplified from either HeLa cDNA or genomic DNA and, if applicable, mutagenized by PCR. The 2xMS2-ΔESE binding sites were optimized via PCR to remove potential ESE sequences and inserted with varying spacers into the truncated exon 4 of the HSD17B10 mini-gene. All minigene constructs were cloned in-frame with an N-terminal FLAG-tag into the pcDNA5/FRT/TO vector (Thermo Fisher Scientific). For generating stable RNPS1, SAP18 or PNN rescue cell lines, the expression constructs were cloned into the pcDNA5/FRT/TO vector containing an N-terminal FLAG-tag. To ensure robust expression in rescue assays, selected RNPS1 or PNN constructs were also cloned into the pcDNA5/FRT/TO vector containing an N-terminal FLAG-emGFP-tag. Standard protocols were used to generate stable rescue or reporter HeLa Flp-In T-REx cell lines and positive clones were selected with 100-150 µg ml^{-1} hygromycin B (InvivoGen). Expression of stable cell lines was induced for minimum 24 h with 1 µg ml^{-1} doxycycline. Dual-inducible tethering cell lines were generated by integrating a PB-CuO-MS2V5 construct in stable Flp-In T-REx reporter cell lines. 2 µg ml^{-1} puromycin was used for the selection of positive clones. Expression of the dual-inducible cell lines was first induced with 30 µg ml^{-1} cumate for 24h, followed by both 1 µg ml^{-1} doxycycline and 30 µg ml^{-1} cumate for another 24 h. Mycoplasma contamination was tested by PCR amplification of mycoplasma-specific genomic DNA (Young et al., 2010) or by using the Mycoplasmacheck service (Eurofins Genomics).

In vitro **transcription, *in vitro* splicing and RNP immunoprecipitation**

In vitro transcription and *in vitro* splicing experiments were performed as described previously (Gehring et al., 2009a). The capped MINX transcripts were generated by run-off transcription with SP6 polymerase (Promega) in the presence of Ribo m7G Cap Analog (Promega) and α-^{32}P-GTP (Hartmann Analytic). *In vitro* splicing reactions were carried out in HeLa nuclear extracts (CIL Biotech) supplemented with HEK 293 whole cell extracts expressing FLAG-tagged proteins. After splicing, immunoprecipitations were carried out with EZview Red ANTI-

FLAG M2 Affinity Gel (Sigma-Aldrich) in EJC buffer (20 mM HEPES-KOH pH 7.9, 200 mM NaCl, 2 mM $MgCl_2$, 0.2% Triton-X-100, 0.1% Nonidet-P40, 0.05% sodium deoxycholic acid). Subsequently, RNA was extracted from the bound proteins via peqGOLD TriFast (VWR) and resolved by denaturing PAGE. For detection of co-immunoprecipitated proteins, splicing reactions with 3´-O-Me-m7G(5')ppp(5')G RNA Cap Structure Analog (NEB)-capped but non-radioactively labeled MINX transcripts were performed. Immunoprecipitations were carried out with Anti-FLAG M2 Magnetic Beads (Sigma-Aldrich) in EJC buffer and co-immunoprecipitated proteins were eluted with SDS-sample buffer, separated by SDS-PAGE, and analyzed by immunoblotting.

Co-immunoprecipitation

FLAG-tagged proteins were expressed in stable HeLa Flp-In T-REx cells induced for 48 h and immunoprecipitated from 1 mg cell lysate (in 50 mM Tris [pH 7.2], 150 mM NaCl, 0.5% Triton X-100) for 2 h using Anti-FLAG M2 Magnetic Beads (Sigma-Aldrich) in the presence or absence of RNase A (50 µg ml^{-1}). Beads were washed four times with lysis buffer and co-immunoprecipitated proteins were eluted with SDS-sample buffer, separated by SDS-PAGE, and analyzed by immunoblotting. Effectiveness of RNAse A treatment was confirmed by ethidium bromide staining of total RNA isolated after immunoprecipitation.

siRNA transfections

Cell lines were reverse transfected with 60 pmol siRNA per 2 x 10^5 cells using 2.5 µl Lipofectamine RNAiMAX (Thermo Fisher Scientific). Fresh medium was supplied 24 h after siRNA transfection. The used siRNA target sequences are listed in the Key Resources Table.

Transient plasmid transfections

2.8 x 10^5 HeLa Tet-Off cells were seeded in 6-well plates 24 h before transfection by calcium phosphate precipitation. For reporter assays, 0.5 µg of a mVenus expression plasmid, 1 µg reporter plasmid and 2 µg fill plasmid encoding for β-globin were transfected. For tethering

experiments 0.5 µg of a mVenus expression plasmid, 0.5 µg reporter plasmid and 1 µg plasmid encoding for MS2V5-tagged proteins were transfected. Overexpression of proteins in HeLa Flp-In T-REx cells was performed using jetPRIME (Polyplus Transfection), co-transfecting 0.5 µg of a mVenus expression plasmid and 2 µg of plasmid encoding for FLAG-tagged proteins.

Immunoblot analysis and antibodies

SDS-polyacrylamide gel electrophoresis and immunoblot analysis were performed using protein samples derived from peqGOLD TriFast extractions, parallel transfection harvested with RIPA buffer or samples eluted from Anti-FLAG M2 magnetic beads. All antibodies (see Key Resources List) were used at 1:3000 dilutions in 50 mM Tris [pH 7.2], 150 mM NaCl with 0.2% Tween-20 and 5% skim milk powder. Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare) in combination with the myECL Imager (Thermo Fisher Scientific) was used for visualization.

Reverse transcription, end-point and quantitative RT-PCR

Cells were harvested with peqGOLD TriFast and RNA extracted according to the manufacturer's instructions. 1-4 µg of total RNA was reverse-transcribed in a 20 µl reaction volume with 10 µM VNN-$(dT)_{20}$ primer using the ProtoScript II Reverse Transcriptase (NEB), GoScript Reverse Transcriptase (Promega) or the GoScript Reverse Transcription Mix, Oligo(dT) (Promega). 0.5 % of purified cDNA was used as template in end-point PCRs using the GoTaq Green Master Mix (Promega) and 0.2 µM final concentration of sense and antisense primer (see Table S3 for sequences). After 30 PCR cycles, the samples were resolved by electrophoresis on ethidium bromide-stained, 1% agarose TBE gels and visualized by trans-UV illumination using the Gel Doc XR+ (Bio-Rad). Representative gel images from at least three independent experiments are shown. Sanger sequencing of individual bands was performed using the service of Eurofins Genomics.

Quantitative RT-PCR were performed with the GoTaq qPCR Master Mix (Promega) using 0.5 % of cDNA in 10 µl reactions, 0.2 µM final concentration of sense and antisense primer (see Table S3 for sequences), and the CFX96 Touch Real-Time PCR Detection System (Bio-Rad).

Pathscan and Cell Survival assays

The PathScan Stress and Apoptosis Signaling Antibody Array Kit (Cell Signaling Technology) was used to detect cellular stress responses upon depletion of RNPS1, SMG1 or RBM8A. In brief, siRNA-mediated knockdowns of HeLa Flp-In T-REx were performed in 6-well plates as described above. After 3 days, the cells were harvested in 1x Cell Lysis Buffer and protein concentration was measured using the Bradford Protein Assay (Bio-Rad). Lysates were diluted to 0.5 mg ml^{-1} protein concentration and the sandwich immunoassay was performed according to the manufacturer's instructions. The cell lysate was incubated with nitrocellulose-coated glass slides on which target-specific capture antibodies have been spotted in duplicate. Biotinylated detection antibody cocktail in combination with streptavidin-conjugated HRP and LumiGLO Reagent are then used to visualize the bound detection antibody by chemiluminescence.

To measure cell proliferation and survival upon EJC component depletion, the CellTiter-Glo Luminescent Cell Viability Assay (Promega) was used. HeLa Flp-In T-REx were reverse transfected with siRNA as described above and 2000 cells were seeded in triplicates in 96-well plates suitable for sensitive luminescence measurements. 2 hours after seeding, the first time point was measured according to the manufacturer's instructions and subsequent measurements were performed in 24-hour intervals.

RNA-Seq

RNA-Seq analysis was carried out on normal or stable RNPS1-expressing HeLa Flp-In T-REx cells transfected with siRNAs targeting RNPS1 or the negative control luciferase. Three biological replicates were analyzed for each sample. Total RNA was extracted with peqGOLD

TriFast as described above. Ribosomal depletion and strand specific library preparation was carried out with the TruSeq R Stranded Total RNA LT (with Ribo-ZeroTM Human/Mouse/Rat) according to the manufacturer's instructions. After validation (Agilent 2200 TapeStation) and quantification (Invitrogen Qubit System) all 12 transcriptome libraries were pooled. The pool was quantified using the Peqlab KAPA Library Quantification Kit and the Applied Biosystems 7900HT Sequence Detection System and loaded on one lane of Illuminas HiSeq4000 sequencer with a 2×75bp protocol. The analysis produced 5.3 Gb/sample (4.8-6.6 Gb), corresponding to 35 Mread-pairs/sample. Basic read quality check was carried out using FastQC showing >97.5% of Q30 bases (PF) and a mean quality score of 39,8.

Read processing and mapping

Adapter sequences and low quality 3′ bases were removed with Flexbar 3.0. (Dodt et al., 2012). Short reads from the rRNA locus were subtracted by mapping against the 45S precursor (Homo sapiens, NR_046235.1) using Bowtie2 (Langmead and Salzberg, 2012). The remaining reads were aligned against the human genome (version 38, EnsEMBL 90 transcript annotations) using the STAR read aligner (2.5.3a)(Dobin et al., 2013)

Gene expression analysis

We employed the Cuffquant and Cuffdiff software (release 2.2.1) (Trapnell et al., 2013) to estimate gene expression abundance and differential gene expression for the EnsEMBL reference annotation. The R package Cummerbund (Trapnell et al., 2012) was subsequently used to inspect and visualize the results.

Local splicing variants identification

We carried out local splicing variant detection using MAJIQ (1.0.6, without GC correction in the build step)(Vaquero-Garcia et al., 2016). To this end, we first produced a transcriptome annotation with StringTie (1.3.3b)(Pertea et al., 2015) for each RNA-Seq replicate on the control groups (Luc for RNPS1 vs. Luc, GFP for EIF4A3 vs. GFP) to compare with the

knockdown conditions. Next, we combined the annotations using the merge command from StringTie, with the minimum isoform fraction (-f parameter) set to 0.5 to eliminate lowly transcribed isoform. Finally, we applied gffcompare (0.10.1) against the human genome annotation with the parameters –R (precision correction), -Q (sensitivity correction) and –M (discard single-exons transfrags). The stringent transcriptome annotation enabled us to contrast between conditions as it highlights differential exon usage.

The Voila tabular output was processed and analyzed with Python programming language (3.6). The dataset, which initially contains one local splice variations (LSV) per row, was expanded to provide one exon-exon junctions per row, enabling us to filter the exon-exon junction given the delta percent spliced in (dPSI) (< -0.1) and their posterior probability P(dPSI) (> 0.90). These cutoffs were applied to ensure the sensitive detection of alternative splicing events (dPSI < -0.1), while keeping only probable splicing events (P(dPSI) > 0.90). The dPSI between two conditions is calculated by estimating a posterior distribution for the change in the respective junction's relative inclusion level (Vaquero-Garcia et al., 2016). For all dPSI calculations the PSI of the RNPS1 or EIF4A3 knockdown sample was substracted from the PSI of the control (Luc or GFP). Furthermore, we used the bedtools intersect command (pybedtools 0.7.10, bedtools 2.26.0)(Dale et al., 2011; Quinlan and Hall, 2010) to extract the exon coordinates from the StringTie annotation overlapping the respective junction. This allowed us to re-classify all junctions based on calculating distances of each exon to junction connection. Furthermore, junctions were classified as spurious if the PSI in control samples was low (< 0.05) and the junction was sufficiently upregulated (fold change PSI-KD/PSI-control >10).

Gene ontology analysis

Enrichment test for biological process, cellular component and molecular function terms was carried out as proposed by (Reimand et al., 2016), with G:Profiler and Enrichment Map Cytoscape plugin. We have compared genes identified by MAJIQ after filtering for probable and spurious events against a background of 14,594 multi-exon genes that were expressed (FPKM > 1) in any of the analysed sample.

Calculation of MaxEnt scores, ESE/ESS composition, putative branch points and additional analyses

We used the MaxEnt algorithm (Yeo and Burge, 2004) implemented in the Human Splicing Finder online tool (3.0) (Desmet et al., 2009) and maxentpy (https://github.com/kepbod/maxentpy) to calculate splice site strengths. Using the same tools, we analyzed the exonic splicing enhancer and silencer composition of the RER1 reporter. Putative mammalian U2 branch points were predicted using the SVM-BPfinder online tool (Corvelo et al., 2010). The ASAP structure (PDB ID: 4A8X)(Murachelli et al., 2012) was visualized using PyMOL (1.8). The UpSet R package (Lex et al., 2014) and the Euler3 Applet (Rodgers et al., 2014) was used to visually compare the alternative splice analysis results. Sashimi-plots (Katz et al., 2015) were generated from data generated in this study, as well as available RNA-Seq (GEO accession number GSE63091) (Wang et al., 2014) data using the integrated function of the IGV (2.3.98) (Thorvaldsdottir et al., 2013). The fitness scores were obtained from available data on HeLa cells generated by a high-resolution CRISPR-screen (Hart et al., 2015). The fitness score is a log Bayes factor (BF) for each gene, which was calculated with a Bayesian Analysis of Gene Essentiality algorithm (Hart et al., 2015). The confidence in the essentiality (higher impact on fitness) of the gene increases with more positive scores. The Bayes Factor cutoff at 5% false discovery rate for HeLa cells (15.47) was used to discriminate essential and non-essential genes.

Code availability

For availability of codes that were developed for this project, please contact the corresponding authors.

QUANTIFICATION AND STATISTICAL ANALYSIS

p values < 0.05 were considered significant. Significance in all figures is indicated as follows: ns = not significant, ∗p = 0.01 to 0.05, ∗∗p = 0.001 to 0.01, ∗∗∗p < 0.001.

RNP immunoprecipitation

Signals of ^{32}P-labeled RNAs were scanned using a Typhoon FLA 7000 (GE Healthcare) and quantified using the ImageQuant TL software (GE Healthcare). Results are shown as mean ± SD. GraphPad Prism 5 was used to determine the statistical significance by one-way ANOVA with Bonferroni post-test.

End-point RT-PCR

Bands detected in agarose gels from the indicated biological replicates of end-point PCRs were quantified using the Image Lab 6.0.1 software (Bio-Rad). Results of the indicated band % per lane are shown as mean ± SD. GraphPad Prism 5 was used to determine the statistical significance by one-way ANOVA with Bonferroni post-test. For experiments with only two samples, two-tailed unpaired Student's t test was performed with GraphPad Prism 5.

Quantitative RT-PCR

The reactions for each biological replicate were performed in duplicates and the average Ct (threshold cycle) value for retained or skipped MRPL3 exon 4 was measured. Values for skipped exon were subtracted from values for retained exon to calculate the ΔCt. The fold changes were calculated using the ΔΔCt method, using the Luc knockdown as normalization (Schmittgen and Livak, 2008). The mean fold changes were calculated from three biologically independent experiments. Results are shown as mean ± SD. GraphPad Prism 5 was used to determine the statistical significance by one-way ANOVA with Bonferroni post-test.

Pathscan assay

Chemiluminescent signals were measured with the ChemiDoc MP Imaging System (Bio-Rad) and quantified using the ImageQuant TL software (GE Healthcare). All signals were first normalized to the α-tubulin control signal. Then, mean values of three biological and two technical replicates, as well as differences between knockdown and control samples were calculated. Finally, the mean $\log_2$ of absolute relative signal intensities were plotted using the superheat R package (arXiv:1512.01524v2 [stat.AP]). Results are shown as mean ± SD. Propagation of error calculations were performed, and GraphPad Prism 5 was used to determine the statistical significance by one-way ANOVA with Bonferroni post-test.

Survival assay

Background corrected mean luminescence was calculated and plotted using GraphPad Prism 5. Results are shown as mean ± SD. GraphPad Prism 5 was used to determine the statistical significance by two-way ANOVA with Bonferroni post-test.

DATA AND SOFTWARE AVAILABILITY

Data Resources

The accession number for the raw RNA-sequencing data reported in this paper is ArrayExpress: E-MTAB-6564.

Data Availability

The authors declare that all the data supporting the findings of this study are available within the article and its Supplementary Information files and from the corresponding authors upon reasonable request. The raw imaging data can be accessed via Mendeley: http://dx.doi.org/10.17632/wt7ybwz82g.1

Supplemental Information

Table S1. RNPS1-dependent alternative splicing events, Related to Figure 2

Table S2. EIF4A3-dependent alternative splicing events, Related to Figure 5

Figure 1

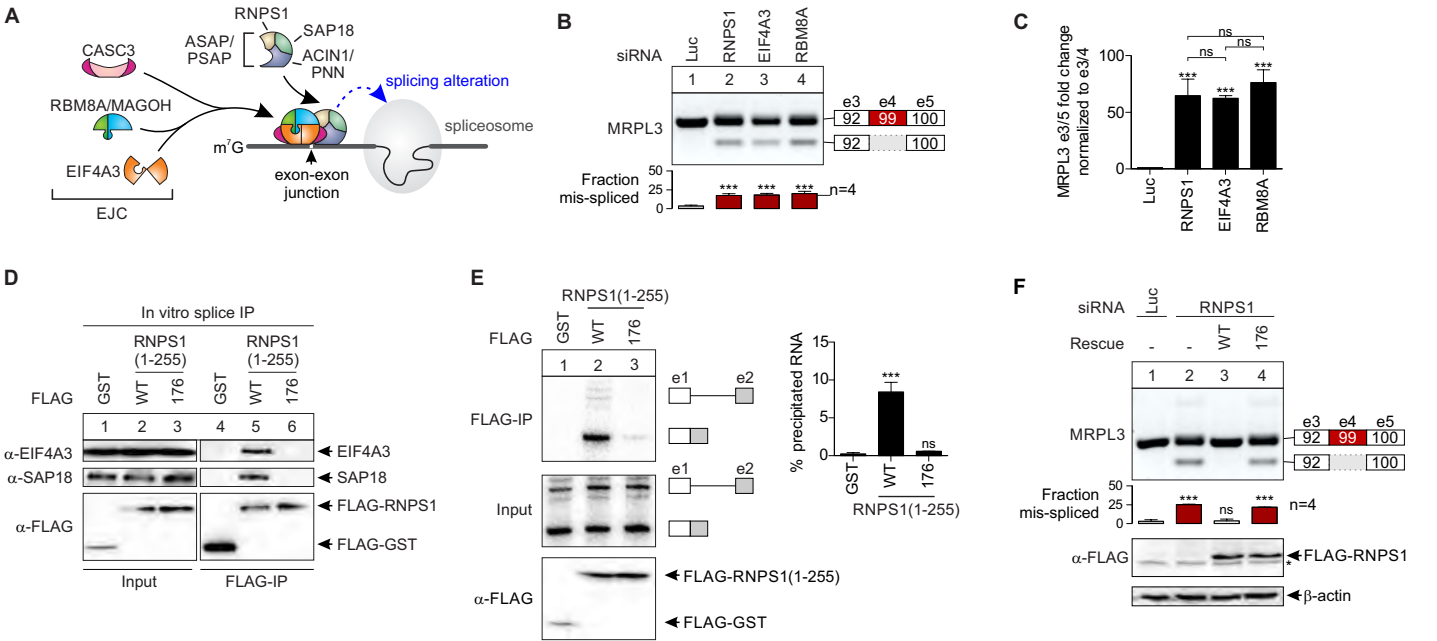

A
RNPS1
SAP18
ASAP/
PSAP
ACIN1/
PNN
CASC3
RBM8A/MAGOH
splicing alteration
spliceosome
m7G
EIF4A3
exon-exon
junction
EJC
B
siRNA
Luc
RNPS1
EIF4A3
RBM8A
1
2
3
4
MRPL3
e3
e4
e5
92
99
100
92
100
Fraction
mis-spliced
50
25
0
***
***
***
n=4
C
MRPL3 e3/5 fold change
normalized to e3/4
100
50
0
ns
ns
ns
***
***
***
Luc
RNPS1
EIF4A3
RBM8A
D
In vitro splice IP
RNPS1
(1-255)
RNPS1
(1-255)
FLAG
GST
WT
176
GST
WT
176
1
2
3
4
5
6
α-EIF4A3
α-SAP18
α-FLAG
EIF4A3
SAP18
FLAG-RNPS1
FLAG-GST
Input
FLAG-IP
E
RNPS1(1-255)
FLAG
GST
WT
176
1
2
3
FLAG-IP
Input
e1
e2
e1
e2
α-FLAG
FLAG-RNPS1(1-255)
FLAG-GST
% precipitated RNA
15
10
5
0
***
ns
GST
WT
176
RNPS1(1-255)
F
siRNA
Luc
RNPS1
Rescue
-
-
WT
176
1
2
3
4
MRPL3
e3
e4
e5
92
99
100
92
100
Fraction
mis-spliced
50
25
0
***
ns
***
n=4
α-FLAG
FLAG-RNPS1
*
β-actin

Figure 2

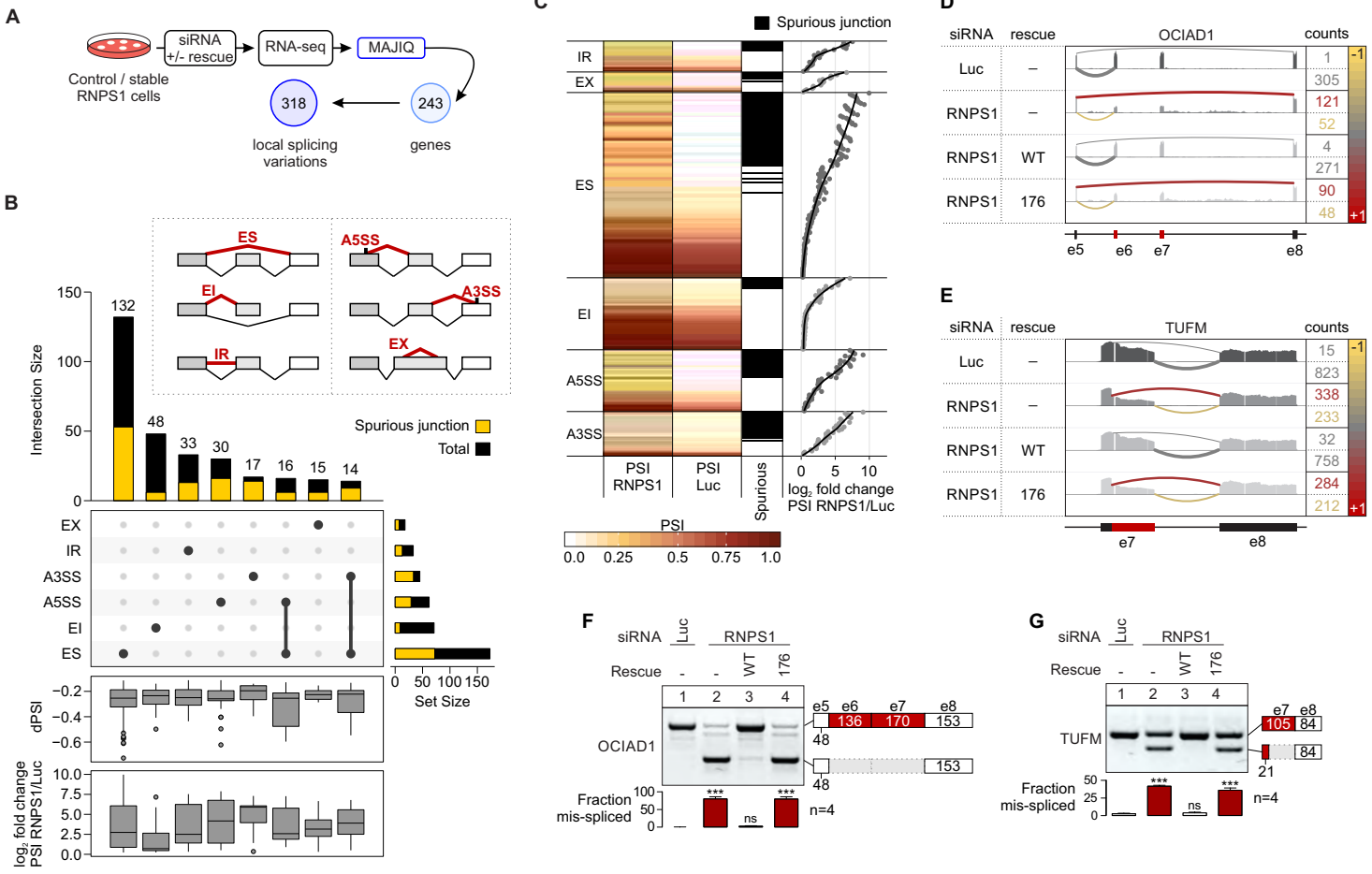

A
siRNA
+/- rescue
RNA-seq
MAJIQ
Control / stable
RNPS1 cells
318
local splicing
variations
243
genes
B
ES
A5SS
EI
A3SS
IR
EX
Intersection Size
150
100
50
0
132
48
33
30
17
16
15
14
Spurious junction
Total
EX
IR
A3SS
A5SS
EI
ES
0 50 100 150
Set Size
dPSI
-0.2
-0.4
-0.6
log2 fold change
PSI RNPS1/Luc
10.0
7.5
5.0
2.5
0.0
C
Spurious junction
IR
EX
ES
EI
A5SS
A3SS
PSI
RNPS1
PSI
Luc
Spurious
0 5 10
log2 fold change
PSI RNPS1/Luc
PSI
0.0 0.25 0.5 0.75 1.0
D
siRNA
rescue
OCIAD1
counts
Luc –
RNPS1 –
RNPS1 WT
RNPS1 176
1
305
121
52
4
271
90
48
-1
+1
dPSI
e5 e6 e7 e8
E
siRNA
rescue
TUFM
counts
Luc –
RNPS1 –
RNPS1 WT
RNPS1 176
15
823
338
233
32
758
284
212
-1
+1
dPSI
e7 e8
F
siRNA
Luc
RNPS1
Rescue
- - WT 176
1 2 3 4
OCIAD1
e5 e6 e7 e8
136 170 153
48
153
48
Fraction
mis-spliced
100
50
0
***
ns
***
n=4
G
siRNA
Luc
RNPS1
Rescue
- - WT 176
1 2 3 4
TUFM
e7 e8
105 84
84
21
Fraction
mis-spliced
50
25
0
***
ns
***
n=4

Figure 3

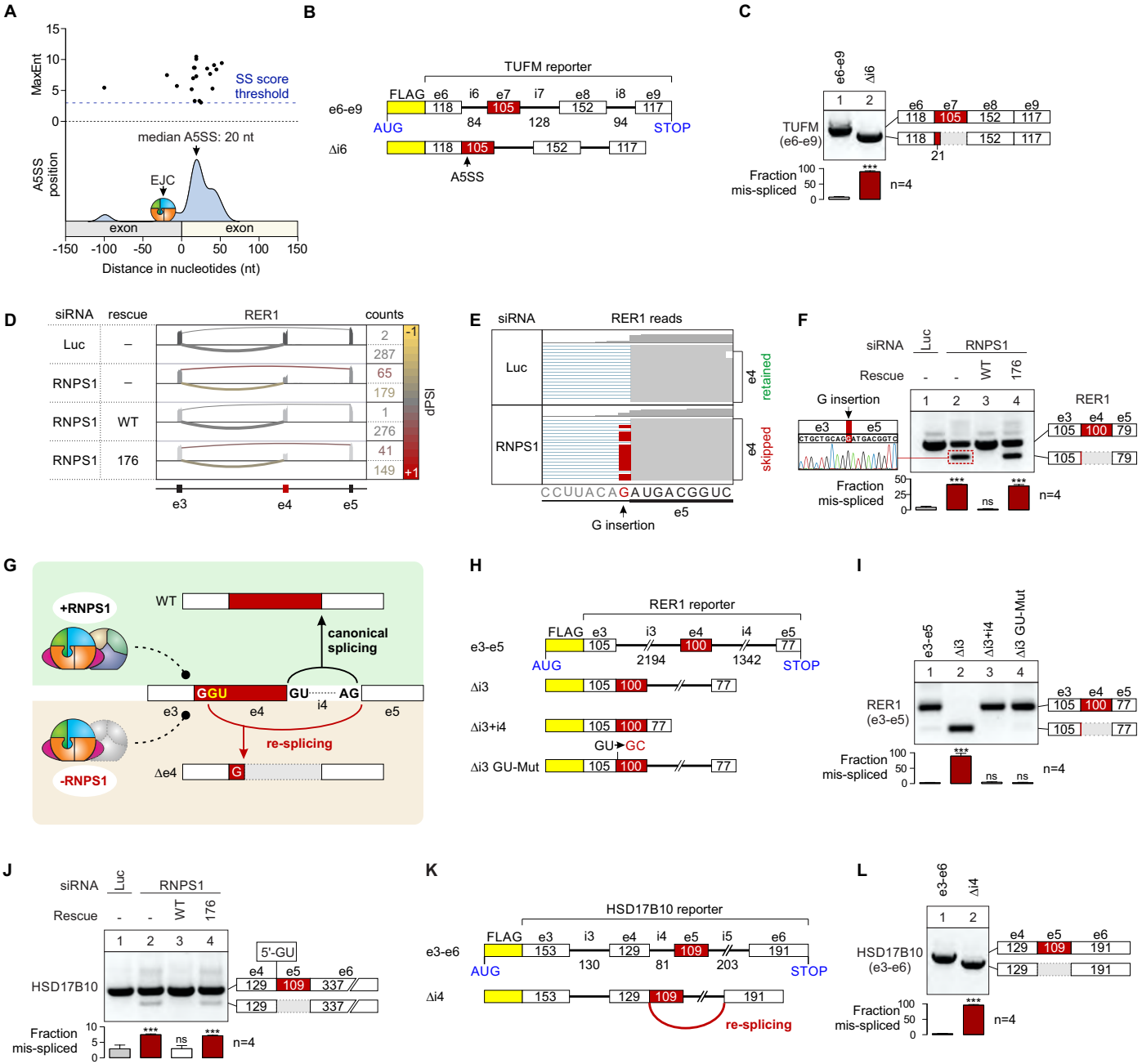

A
MaxEnt
SS score threshold
median A5SS: 20 nt
A5SS position
EJC
exon
exon
Distance in nucleotides (nt)
B
TUFM reporter
FLAG
e6-e9
Δi6
AUG
STOP
A5SS
C
e6-e9
Δi6
TUFM (e6-e9)
Fraction mis-spliced
n=4
D
siRNA
rescue
RER1
counts
Luc
RNPS1
WT
176
dPSI
e3
e4
e5
E
siRNA
RER1 reads
Luc
RNPS1
retained
skipped
G insertion
F
siRNA
RNPS1
Rescue
G insertion
RER1
Fraction mis-spliced
ns
n=4
G
+RNPS1
-RNPS1
WT
canonical splicing
GGU
GU
AG
re-splicing
Δe4
H
RER1 reporter
e3-e5
Δi3
Δi3+i4
Δi3 GU-Mut
GU→GC
I
e3-e5
Δi3
Δi3+i4
Δi3 GU-Mut
RER1 (e3-e5)
Fraction mis-spliced
ns
n=4
J
siRNA
RNPS1
Rescue
HSD17B10
5'-GU
Fraction mis-spliced
n=4
K
HSD17B10 reporter
e3-e6
Δi4
re-splicing
L
e3-e6
Δi4
HSD17B10 (e3-e6)
Fraction mis-spliced
n=4

Figure 4

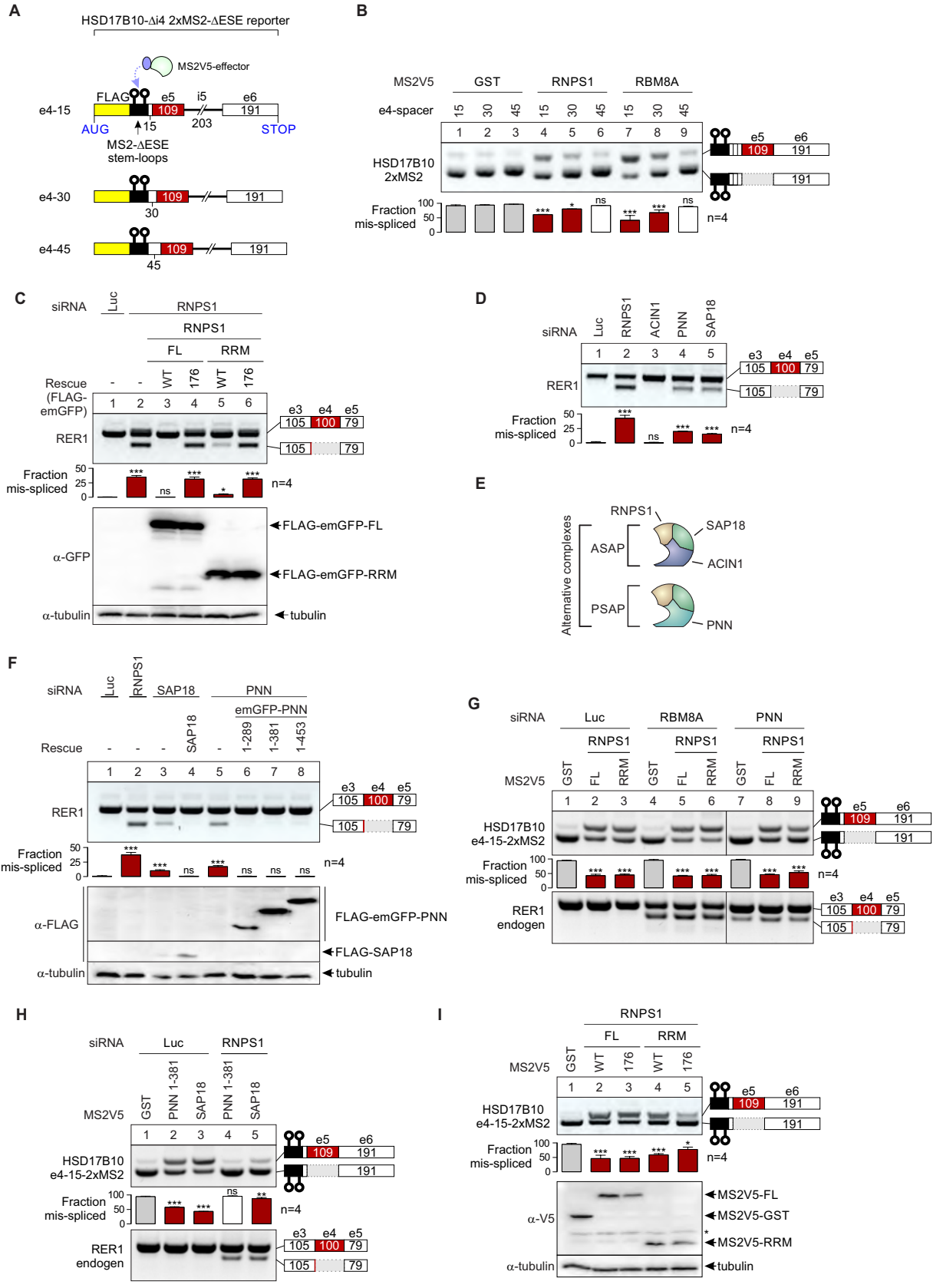

A
HSD17B10-Δi4 2xMS2-ΔESE reporter
MS2V5-effector
FLAG
e5
i5
e6
e4-15
109
191
AUG
15
203
STOP
MS2-ΔESE
stem-loops
e4-30
30
e4-45
45
B
MS2V5
GST
RNPS1
RBM8A
e4-spacer
15 30 45 15 30 45 15 30 45
1 2 3 4 5 6 7 8 9
HSD17B10
2xMS2
e5 e6
109 191
191
Fraction
mis-spliced
100
50
0
***
*
ns
***
***
ns
n=4
C
siRNA
Luc
RNPS1
RNPS1
FL
RRM
Rescue
(FLAG-
emGFP)
-
-
WT
176
WT
176
1 2 3 4 5 6
RER1
e3 e4 e5
105 100 79
105 79
Fraction
mis-spliced
50
25
0
***
ns
***
*
***
n=4
α-GFP
FLAG-emGFP-FL
FLAG-emGFP-RRM
α-tubulin
tubulin
D
siRNA
Luc
RNPS1
ACIN1
PNN
SAP18
1 2 3 4 5
RER1
e3 e4 e5
105 100 79
105 79
Fraction
mis-spliced
50
25
0
***
ns
***
***
n=4
E
RNPS1
SAP18
ASAP
ACIN1
PSAP
PNN
Alternative complexes
F
siRNA
Luc
RNPS1
SAP18
PNN
emGFP-PNN
Rescue
-
-
-
SAP18
-
1-289
1-381
1-453
1 2 3 4 5 6 7 8
RER1
e3 e4 e5
105 100 79
105 79
Fraction
mis-spliced
50
25
0
***
***
ns
***
ns
ns
ns
n=4
α-FLAG
FLAG-emGFP-PNN
FLAG-SAP18
α-tubulin
tubulin
G
siRNA
Luc
RBM8A
PNN
RNPS1
RNPS1
RNPS1
MS2V5
GST
FL
RRM
GST
FL
RRM
GST
FL
RRM
1 2 3 4 5 6 7 8 9
HSD17B10
e4-15-2xMS2
e5 e6
109 191
191
Fraction
mis-spliced
100
50
0
***
***
***
***
***
***
n=4
RER1
endogen
e3 e4 e5
105 100 79
105 79
H
siRNA
Luc
RNPS1
MS2V5
GST
PNN 1-381
SAP18
PNN 1-381
SAP18
1 2 3 4 5
HSD17B10
e4-15-2xMS2
e5 e6
109 191
191
Fraction
mis-spliced
100
50
0
***
***
ns
**
n=4
RER1
endogen
e3 e4 e5
105 100 79
105 79
I
RNPS1
FL
RRM
MS2V5
GST
WT
176
WT
176
1 2 3 4 5
HSD17B10
e4-15-2xMS2
e5 e6
109 191
191
Fraction
mis-spliced
100
50
0
***
***
***
*
n=4
α-V5
MS2V5-FL
MS2V5-GST
*
MS2V5-RRM
α-tubulin
tubulin

Figure 5

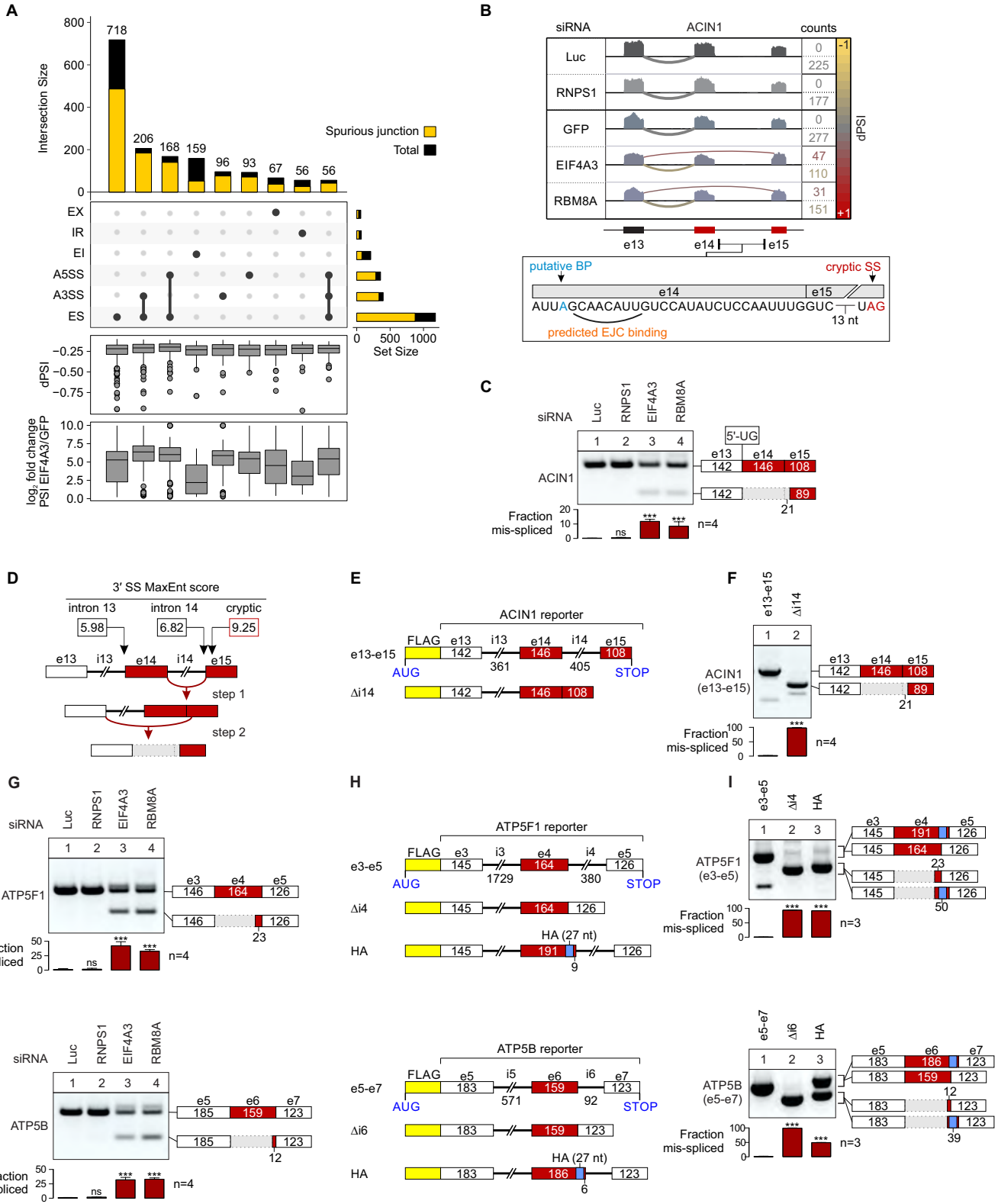

A
Intersection Size
718
206
168
159
96
93
67
56
56
Spurious junction
Total
EX
IR
EI
A5SS
A3SS
ES
Set Size
dPSI
log2 fold change PSI EIF4A3/GFP
B
siRNA
ACIN1
counts
Luc
RNPS1
GFP
EIF4A3
RBM8A
dPSI
e13
e14
e15
putative BP
cryptic SS
AUUAGCAACAUUGUCCAUAUCUCCAAUUUGGUC
UAG
13 nt
predicted EJC binding
C
siRNA
Luc
RNPS1
EIF4A3
RBM8A
ACIN1
5'-UG
Fraction mis-spliced
n=4
D
3′ SS MaxEnt score
intron 13
intron 14
cryptic
5.98
6.82
9.25
step 1
step 2
E
ACIN1 reporter
e13-e15
Δi14
FLAG
AUG
STOP
F
e13-e15
Δi14
ACIN1 (e13-e15)
Fraction mis-spliced
n=4
G
siRNA
ATP5F1
Fraction mis-spliced
n=4
H
ATP5F1 reporter
e3-e5
Δi4
HA
HA (27 nt)
I
ATP5F1 (e3-e5)
Fraction mis-spliced
n=3
siRNA
ATP5B
Fraction mis-spliced
n=4
ATP5B reporter
e5-e7
Δi6
HA
HA (27 nt)
ATP5B (e5-e7)
Fraction mis-spliced
n=3

Figure 6

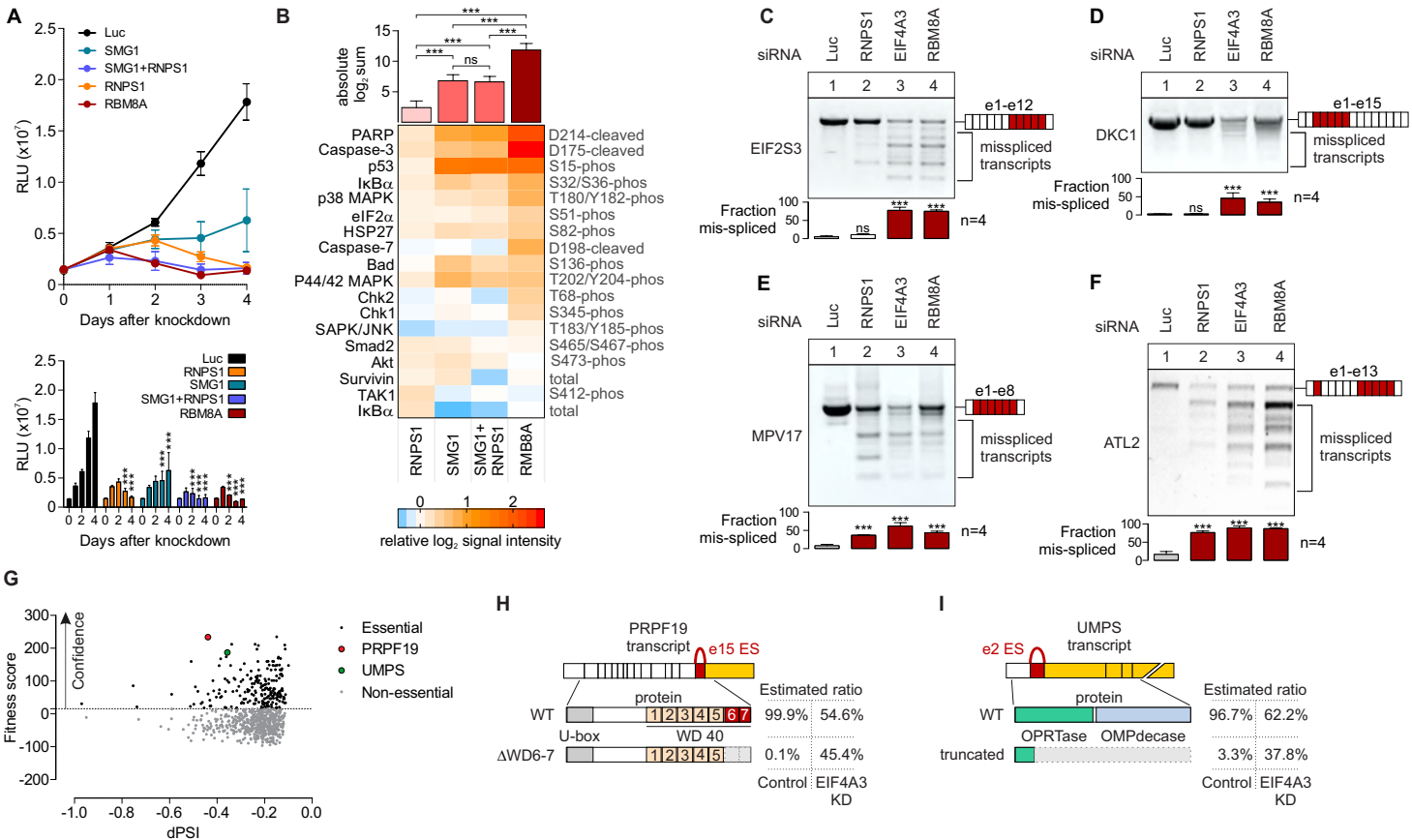
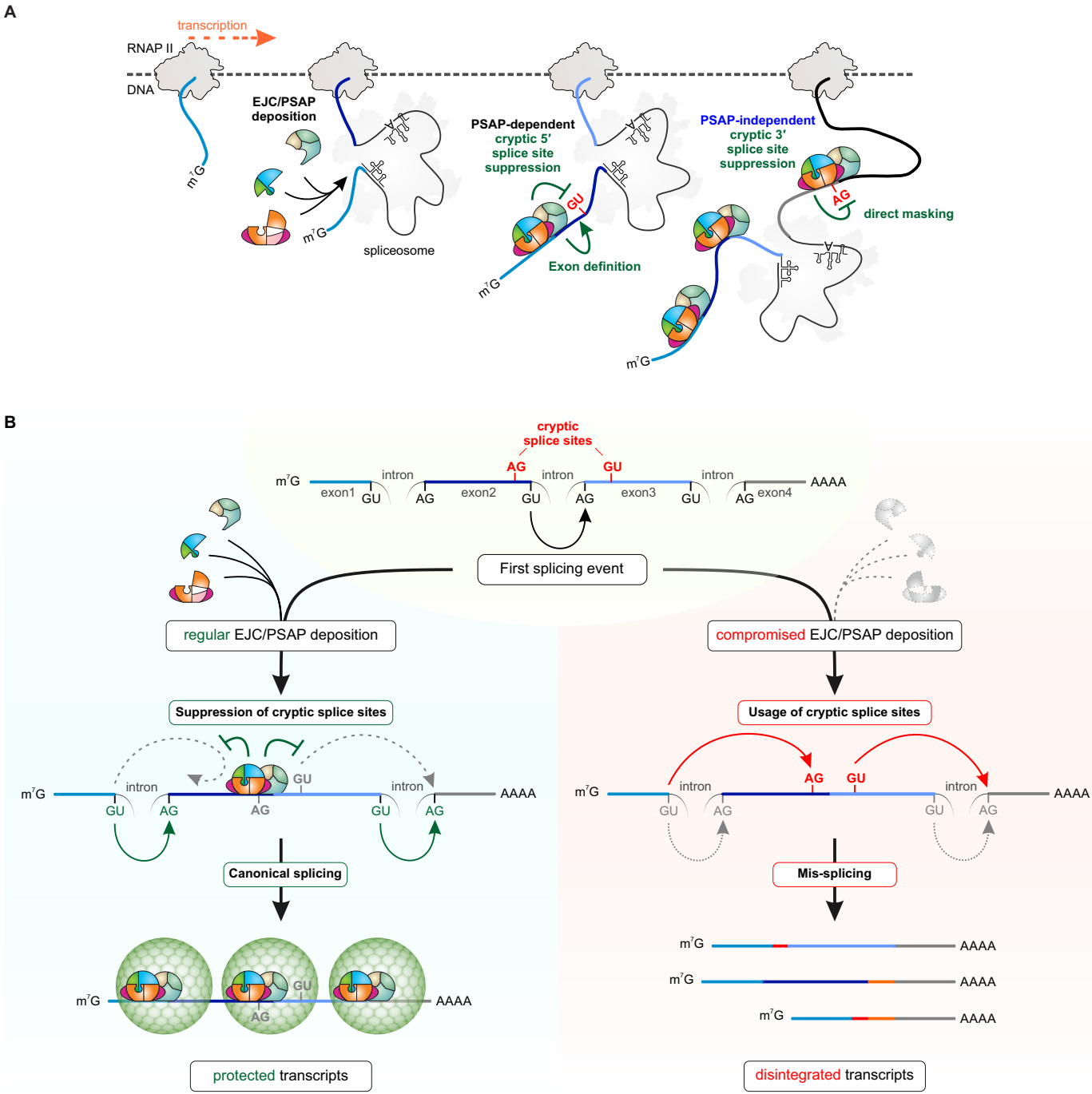

A
Luc
SMG1
SMG1+RNPS1
RNPS1
RBM8A
RLU (x10^7)
Days after knockdown
RNPS1
SMG1
SMG1+RNPS1
RBM8A
B
absolute log2 sum
ns
PARP
Caspase-3
p53
IκBα
p38 MAPK
eIF2α
HSP27
Caspase-7
Bad
P44/42 MAPK
Chk2
Chk1
SAPK/JNK
Smad2
Akt
Survivin
TAK1
IκBα
D214-cleaved
D175-cleaved
S15-phos
S32/S36-phos
T180/Y182-phos
S51-phos
S82-phos
D198-cleaved
S136-phos
T202/Y204-phos
T68-phos
S345-phos
T183/Y185-phos
S465/S467-phos
S473-phos
total
S412-phos
total
relative log2 signal intensity
C
siRNA
Luc
RNPS1
EIF4A3
RBM8A
EIF2S3
e1-e12
misspliced transcripts
Fraction mis-spliced
n=4
D
DKC1
e1-e15
E
MPV17
e1-e8
F
ATL2
e1-e13
G
Fitness score
Confidence
dPSI
Essential
PRPF19
UMPS
Non-essential
H
PRPF19 transcript
e15 ES
protein
WT
U-box
WD 40
ΔWD6-7
Estimated ratio
99.9%
54.6%
0.1%
45.4%
Control
EIF4A3 KD
I
UMPS transcript
e2 ES
OPRTase
OMPdecase
truncated
96.7%
62.2%
3.3%
37.8%

Figure 7


A
transcription
RNAP II
DNA
EJC/PSAP deposition
m7G
spliceosome
PSAP-dependent cryptic 5′ splice site suppression
GU
Exon definition
PSAP-independent cryptic 3′ splice site suppression
AG
direct masking
B
cryptic splice sites
AG
GU
m7G
exon1
intron
exon2
exon3
exon4
AAAA
First splicing event
regular EJC/PSAP deposition
compromised EJC/PSAP deposition
Suppression of cryptic splice sites
Usage of cryptic splice sites
Canonical splicing
Mis-splicing
protected transcripts
disintegrated transcripts

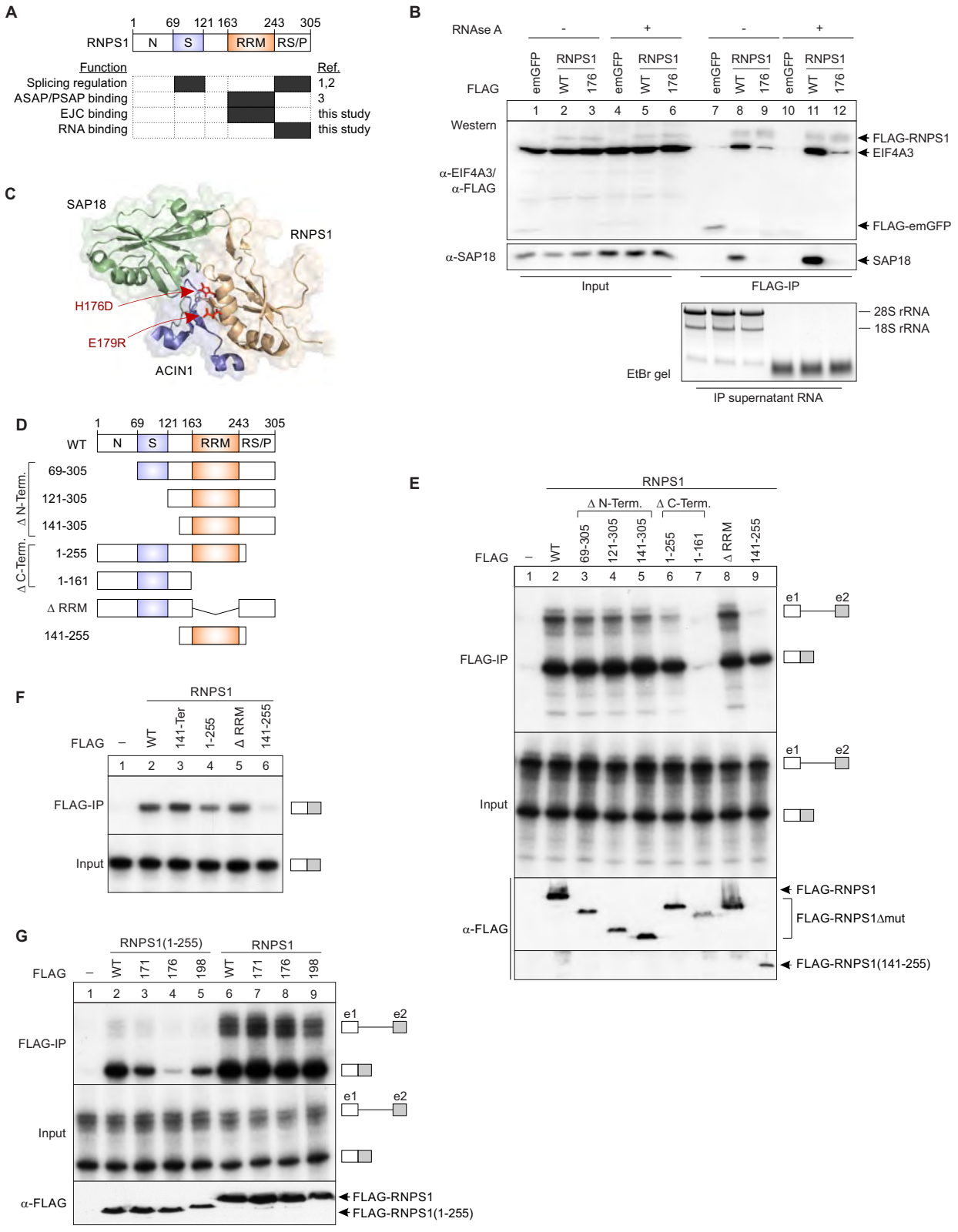


Figure S1. Characterization of RNPS1 binding to spliced mRNA, Related to Figure 1

(A) Schematic representation of RNPS1 domains and their functions. The indicated references are: 1 (Mayeda et al., 1999), 2 (Sakashita et al., 2004) and 3 (Murachelli et al., 2012).

(B) Co-immunoprecipitation of EJC core component EIF4A3 and ASAP/PSAP component SAP18 with the indicated FLAG-tagged RNPS1 variants in the presence or absence of RNAse A. n=1

(C) Location of the H176D-E179R mutation (red) on the surface of RNPS1 in the context of the ASAP complex components SAP18 and ACIN1.

(D), **(E)** In vitro splicing of 32P body-labeled MINX mRNA in the presence of FLAG-RNPS1 deletion variants **(D)**. After FLAG-immunoprecipitation of mRNPs, the RNA was extracted, resolved on a denaturing urea-gel and visualized by phosphorimaging **(E)**. Expression of the FLAG constructs was visualized by immunoblotting (bottom). n=1

(F) In vitro splicing of intron-less 32P body-labeled MINX mRNA in the presence of a subset of FLAG-RNPS1 deletion variants. n=1

(G) In vitro splicing of 32P body-labeled MINX mRNA in the presence of FLAG-RNPS1 WT or point mutants. n=1

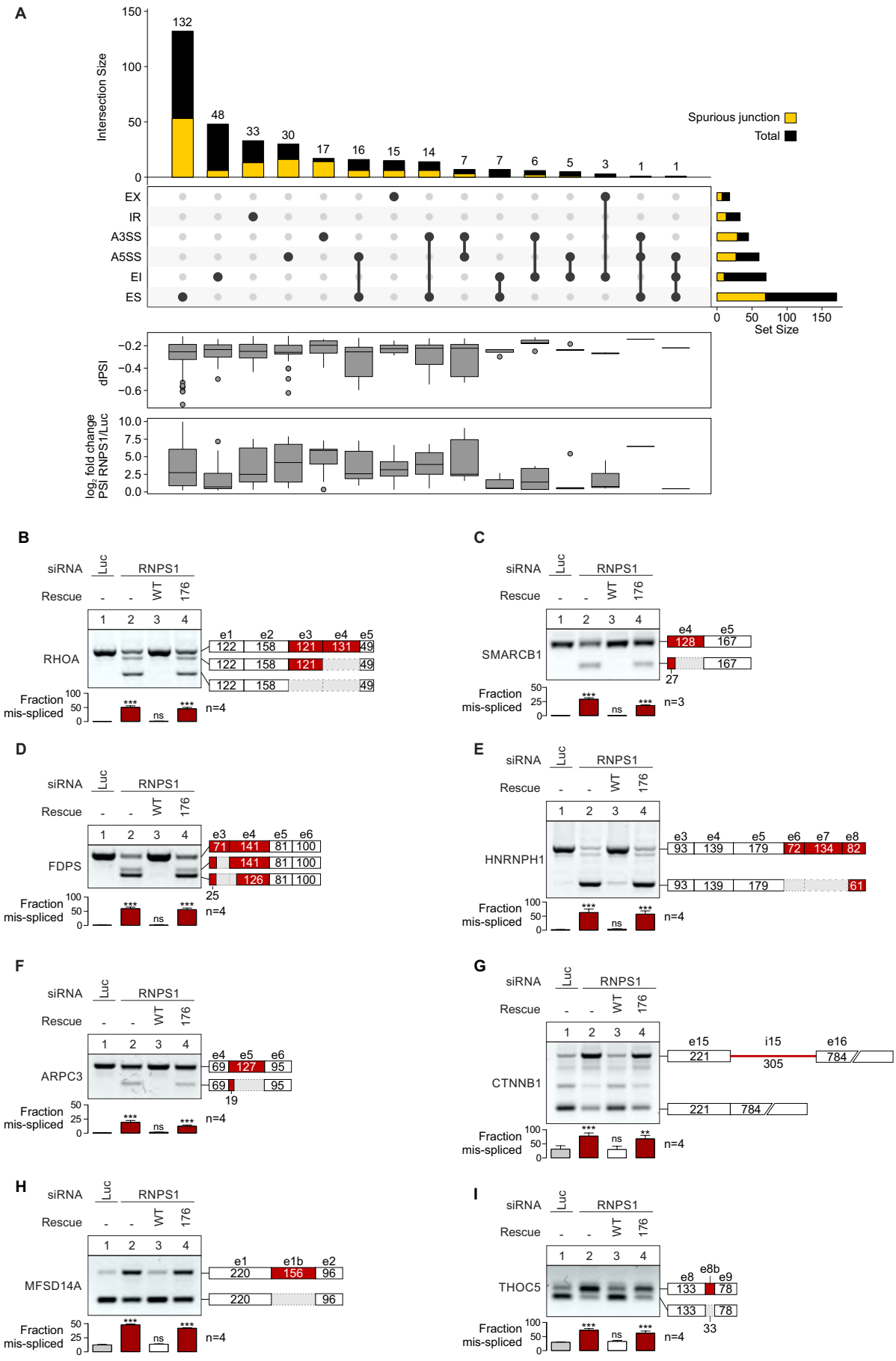


Figure S2. Validation of RNPS1-dependent alternative splice events, Related to Figure 2

(A) Complete intersection and classification of alternatively spliced junctions upon RNPS1 knockdown. The quantity of spurious junctions for each class is shown. The change in junction usage (delta percent spliced in; dPSI) and PSI fold change for each class is shown as boxplots on the bottom.

(B) - **(I)** RT-PCR analysis of the indicated splice events with RNA from HeLa cells expressing the indicated rescue proteins, transfected with the indicated siRNA. Exon-intron architecture for each target is depicted schematically, alternative sequences are highlighted. Quantified results from the indicated biological replicates are shown as mean ± SD and compared to the Luc knockdown control.

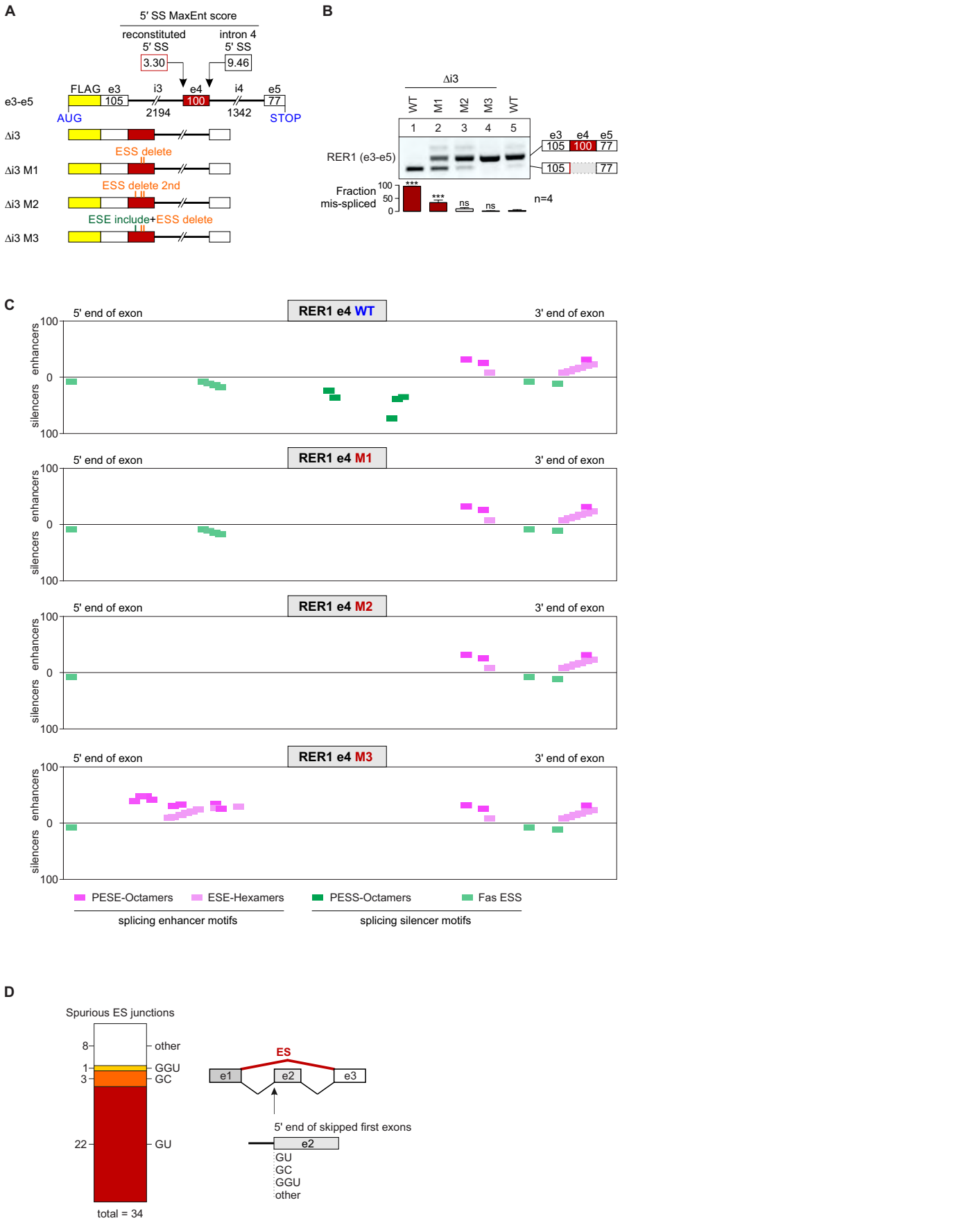


Figure S3. Characterization of the RER1 splice reporter, Related to Figure 3

(A) Overview of the RER1 Δi3 reporter with exonic splice silencer (ESS) and exonic splice enhancer (ESE) mutations in e4.

(B) RT-PCR analysis of the RER1 e4-mutated reporter constructs with RNA from HeLa cells. Quantified results from the indicated biological replicates are shown as mean ± SD and compared to the wildtype control.

(C) HSF analysis of ESS/ESE composition of the mutants depicted in **(A)**.

(D) Identification of 5′ terminal dinucleotides of first skipped exons in exclusive exon skipping events of spurious junctions.

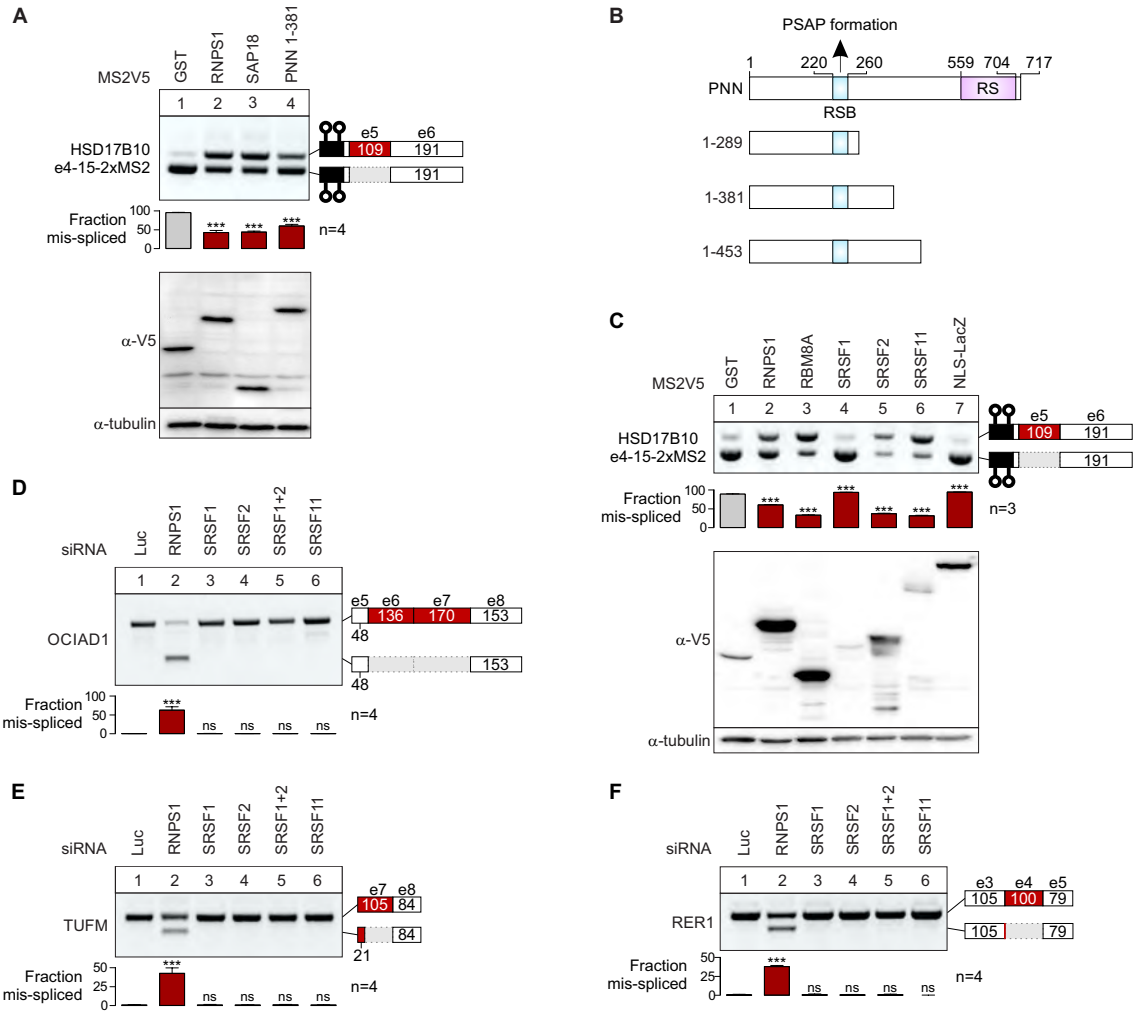


Figure S4. Insights into the mechanism of RNPS1- and PSAP-mediated splice site suppression, Related to Figure 4

(A) Reporter splicing from dual inducible stable HeLa cell lines expressing both the reporter (HSD17B10 e4-15) and the indicated tethering protein was detected via RT-PCR.

(B) Scheme of PNN domain architecture and constructs used for tethering and rescue assays.

(C) The indicated MS2V5-tethering proteins and HSD17B10 reporter were expressed transiently in HeLa Tet-Off cells and the splice patterns analyzed by RT-PCR.

(D) - **(F)** RT-PCR analysis of the indicated splice event with RNA from HeLa cells transfected with siRNA targeting various SR proteins.

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

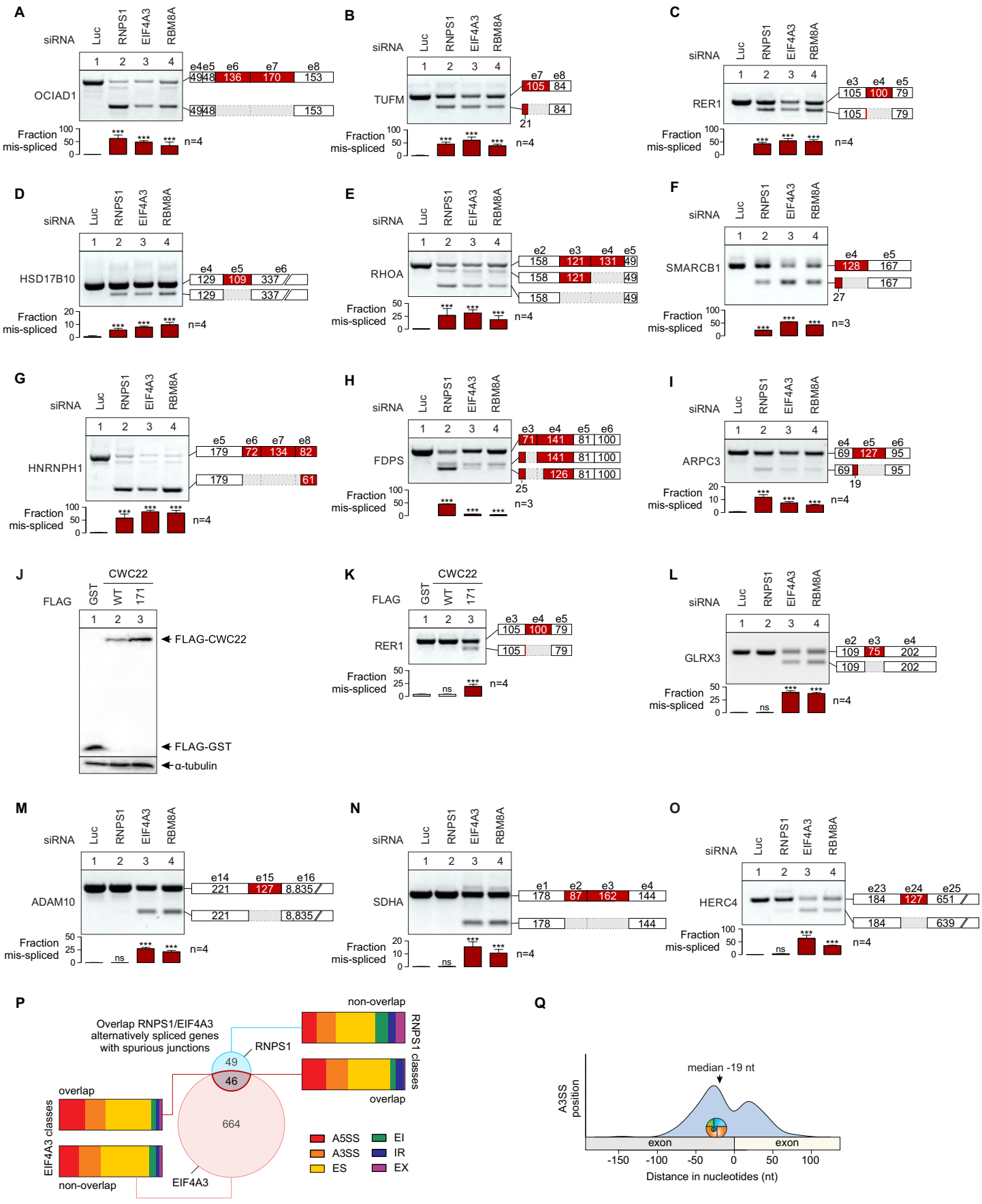


Figure S5. Validation of EJC-dependency of RNPS1-regulated splice events, related to Figure 5

(A) - **(I)** RT-PCR analysis of the indicated splice events with RNA from HeLa cells transfected with the indicated siRNA.

(J), **(K)** The effect of overexpression of CWC22 WT or NK171DE mutant in HeLa cells **(J)** on RER1 alternative splicing was analyzed by RT-PCR **(K)**.

(L) - **(O)** RT-PCR analysis of the indicated splice events with RNA from HeLa cells transfected with the indicated siRNA.

(P) Overlap of genes with at least one spurious junctions identified in RNPS1 and EIF4A3 knockdown RNA-Seq data. The proportion of identified classes for the alternatively spliced genes are shown.

(Q) Depiction of EJC-dependent alternative 3′SS position of spurious junctions relative to exon boundaries as density plot. An outlier at -645 nt distance is not shown in this plot.

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

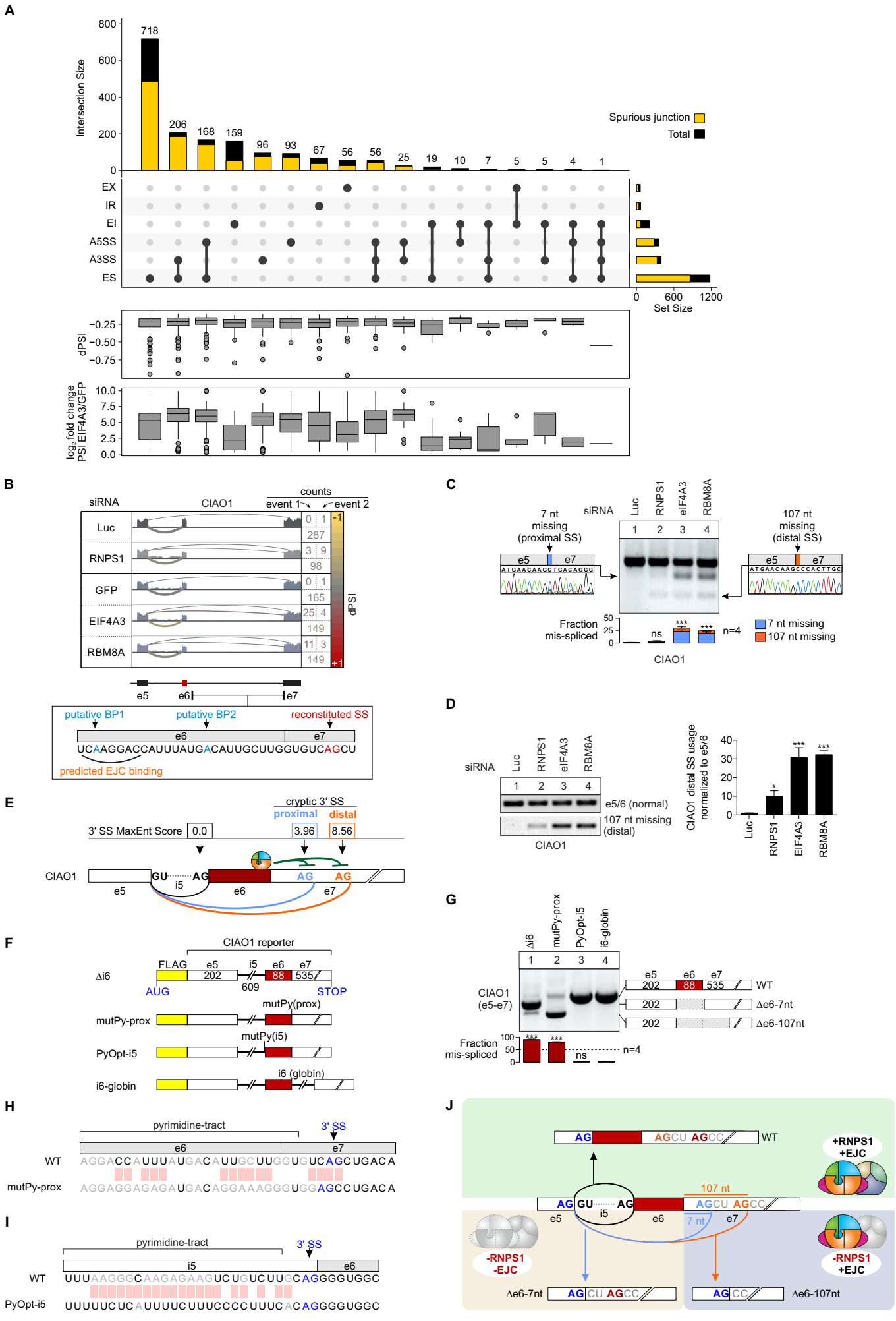

A
Intersection Size
718
206
168
159
96
93
67
56
56
25
19
10
7
5
5
4
1
Spurious junction
Total
EX
IR
EI
A5SS
A3SS
ES
Set Size
dPSI
log2fold change PSI EIF4A3/GFP
B
siRNA
CIAO1
counts
event 1
event 2
Luc
RNPS1
GFP
EIF4A3
RBM8A
dPSI
e5
e6
e7
putative BP1
putative BP2
reconstituted SS
UCAAGGACCAUUUAUGACAUUGCUUGGUGUCAGCU
predicted EJC binding
C
siRNA
Luc
RNPS1
eIF4A3
RBM8A
7 nt missing (proximal SS)
107 nt missing (distal SS)
Fraction mis-spliced
ns
n=4
7 nt missing
107 nt missing
CIAO1
D
siRNA
e5/6 (normal)
107 nt missing (distal)
CIAO1
CIAO1 distal SS usage normalized to e5/6
E
cryptic 3′ SS
proximal
distal
3′ SS MaxEnt Score
0.0
3.96
8.56
CIAO1
GU
AG
i5
e5
e6
e7
F
CIAO1 reporter
FLAG
Δi6
AUG
STOP
mutPy-prox
mutPy(prox)
PyOpt-i5
mutPy(i5)
i6-globin
G
Δi6
mutPy-prox
PyOpt-i5
i6-globin
CIAO1 (e5-e7)
WT
Δe6-7nt
Δe6-107nt
Fraction mis-spliced
n=4
H
pyrimidine-tract
3′ SS
WT
mutPy-prox
I
pyrimidine-tract
3′ SS
WT
PyOpt-i5
J
WT
+RNPS1 +EJC
107 nt
7 nt
-RNPS1 -EJC
Δe6-7nt
Δe6-107nt

Figure S6. Characterization of splice defects upon EJC-depletion, Related to Figure 5

(A) Complete intersection and classification of alternatively spliced junctions upon EIF4A3 knockdown. The quantity of spurious junctions for each class is shown. The change in junction usage (delta percent spliced in; dPSI) and PSI fold change for each class are shown as boxplots on the bottom.

(B) Sashimi-plots depicting two alternative splicing events for CIAO1. The predicted EJC binding site in relation to the cryptic splice site (SS) and putative branch points (BP) is indicated. Two different cryptic 3′SS in CIAO1 exon 7 can be activated, resulting in splicing of exon 6 together with additional 107 (distal event) or 7 (proximal event) nucleotides from exon 7.

(C) CIAO1 RT-PCR of cDNA samples obtained from HeLa cells transfected with the indicated siRNA. Sanger sequencing results of the individual PCR products are shown.

(D) RT-PCR of samples from **(C)** with CIAO1 specific exon 5/6 or distal cryptic splice site primers. Distal cryptic splice site usage was normalized to exon 5/6 splicing in three biological replicates.

(E) Scheme of CIAO1 canonical and cryptic splice sites highlighting the 3′ splice site scores. According to the hypothesis, the EJC deposited by intron 6 splicing suppresses the strong cryptic SS in exon 7 of CIAO1 and thereby promotes the regular splicing of the suboptimal intron 5.

(F), **(G)** CIAO1 minigene reporter constructs **(F) w**ere expressed in HeLa Tet-Off cells and analyzed via RT-PCR **(G)**. The CIAO1 reporter lacking intron 6 (Δi6) was primarily spliced at the proximal cryptic SS. Mutating the polypyrimidine tract of the proximal cryptic SS (mutPy-prox; **(H)**) activated the distal cryptic 3′SS. Normal splicing was restored by enhancing the polypyrimidine tract of the genuine 3′SS of intron 5 (PyOpt-i5; **(I)**), or by inserting an intron downstream of exon 6 (i6-globin).

(H), **(I)** Mutated splice site consensus for the proximal cryptic splice site **(H)** and intron 5 canonical splice site **(I)**

(J) Scheme of CIAO1 exon 5-7 intra-splicing in the presence of EJC without RNPS1 (bottom right) or absence of both RNPS1 and EJC (bottom left).

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

Table S3. List of oligonucleotides used in this study, Related to STAR Methods

Oligo	Sequence	Purpose
MRPL3_e3_se	CCTTGAAGCTGGGCATGATG	RT-PCR
MRPL3_e6_as	GCAGCATAAAGAGGAGTGCC	
OCIAD1_e4_se	CCTTTGGCTGCAACAAGTATG	
OCIAD1_e9_as	TGCAGGTAATCATGACCACCT	
TUFM_e6_se	TTAGGCCTGAAGTCTGTGCA	
TUFM_e8_as	GCTTGATGGAACCTGGCTTG	
RER1_e2_se	GGGGACAGTGTGGGAGAATC	
RER1_e6_as	ATCACCAGAATCGGCCAGAA	
HSD17B_e4_se	TGGGCCAGAATGAACCAGA	
HSD17B_e6_as	CAGGGAAAGGAAGGGCAGA	
ACIN1_e13_se	CCAGGTGTCAGTAGAGGTGG	
ACIN1_e15_as	CCACCAAGGTTCCTGTGCG	
ATP5F1_e3_se	CCCTGTACCACCTCTTCCTG	
ATP5F1_e5_as	CCAGTGCCTGTTGTGACTTC	
ATP5B_e5_se	TGTTTGCTGGTGTTGGTGAG	
ATP5B_e7_as	GGATTCGGCCCAATAATGCA	
EIF2S3_e1_se	CGGGGTGATTTCCTTCCTCT	
EIF2S3_e12_as	TCCAACTTCCAAATCCATCCG	
DKC1_e1_se	GTTCCCTCGGCTGTGGAC	
DKC1_e15_as	CCAGCTTCAAGTGGCCTTC	
MPV17_e2_se	ACCCGTGGAAAGTACAGGTC	
MPV17_e8_as	ATGGAGTGAGGCAGGCTTAG	
ATL2_e1_se	GACGGACCAGCGACCCAA	
ATL2_e13_as	CTGTCGCTGTGCTGATGAAA	
RHOA_e1_se	TCCGTCGGTTCTCTCGTTAG	
RHOA_e5_as	CGCCAATCCTGTTTGCCATA	
SMARCB1_e3_se	ACCCTGTTAAAAGCCTCGGA	
SMARCB1_e5_as	CCCATCGATCTCCATGTCCA	
FDPS_e3_se	TCGTTAGGGTGCTGACTGAG	
FDPS_e6_as	TCAGGTAATAGGGCTGCTCC	
HNRNPH1_e3_se	CAAATAGTCCTGACACGGCC	
HNRNPH1_e8_as	ACTCGACATCTGCTTCACCA	
ARPC3_e3_se	GCCATCTATTACTTCAAGGCCA	
ARPC3_e6_as	TCTCAGTCCAGTCTCTTGCC	
CTNNB1_e15_se	AGGATGCCTTGGGTATGGAC	
CTNNB1_e16_as	TCTTGTGATCCATTCTTGTGCA	
MFSD14A_e1_se	GGAAGAAGAAGAAACGGGCC	
MFSD14A_e2_as	AGGTGGGTGCTGTCAATAGT	
THOC5_e8_se	CATTCGACCAGGCTCACAAG	
THOC5_e9_as	CTTCATCCACACTGCCTTCG	
GLRX3_e2_se	CGAAGTTATGGCAGAGTTAGCT	
GLRX3_e4_as	GCGTGGTTCTTGAGGAGTTC	
ADAM10_e14_se	GACCCATCAACTTGTGCCAG	
ADAM10_e16_as	TTGATAACTCTCTCGGGGCC	
SDHA_e1_se	CGGCAACAGCAGACATGTC	
SDHA_e4_as	TTGTCCTCCTCCATGTTCCC	
HERC4_e23_se	GCTTTTCATGCGGGCTTTCA	
HERC4_e25_as	TACTCCTCACCACCTCCTGT	
CIAO1_e5_se	TTGAGGGCCATGAATCCACT	
CIAO1_e7_as	CAGGCCACACAGTTGACATC	
CIAO1_e4-5_se	CACCCAAGTCAGGAGCTCTT	
CIAO1_e5-7_dist_as	GATGCAAGTGGGCTTGTTCATTG	
FLAG_se	ACAAGGACGACGATGACA	
BGH_as	TAGAAGGCACAGTCGAGG	

MRPL3_e3-e4-splice_se	TGGTCACATTACTTCAGGTACAA	qPCR
MRPL3_e3-e5-splice_se	CACATTACTTCAGAAAGCTACATCC	
MRPL3_e6_as	GCAGCATAAAGAGGAGTGCC	
VNN-(dT)20	TTTTTTTTTTTTTTTTTTTVNN	cDNA synthesis
TUFM_e6_Xho_se	TTTTCTCGAGTTAGGCCTGAAGTCTGTG	Reporter construction
TUFM_e9_Not_as	TTTTTTTGCGGCCGCTTACTCTGGGGGCAGGATAAT	
RER1_e3_rep_Xho_frame_se	TTTTTCTCGAGATTTATCAGTCCTGGCTAGACAAGTCC	
RER1_e5_rep_Sal_Not_stop_as	TTTTTTGCGGCCGCCCGGGTCGACTTAAAATTTAAACTCTGGGAGCCTTCGAATGA	
RER1_del_intron3_as	CATAGGTCACAATGTACCAACCCTGCAGCAGGTAAACTCGAA	
RER1_new_5SS_as	CAGTGCACAGCCGCTAACTTACCTGAGTCTTCCATTAAGGAAGG	
RER1_new_GT_as	CCCCAAGGCATAGGTCACTACCAACCAAGCCTGCAGCAG	
RER1_mutESS_as	GGATCCACTTTGGGAGAAAGGTAAGCTATGACGAGATTTAGATGGTAGATCCCCAG	
RER1_e4_M2-as	TATGACGAGATTTAGATGGTAGATCGCGAAGGCATAGGTCACAATGTACC	
RER1_e4_M3-ESE-as	AGCTATGACGAGATTTAGATGGTTGATCGTCTGCTTTCTGGTCACAATGTACCAACCCT	
HSD17B10_e2_rep_Xho_frame_se	TTTTTCTCGAGGGCCTGGTGGCGGTAATA	
HSD17B10_e6_rep_Sal_Not_as	TTTTTTGCGGCCGCCCGGGTCGACTATTAGGCACAGAGGGCGAC	
HSD17B10_del_intron4_se	GCTGCCTTCGAGGGTCAGGTTGGACAAGCTGCATACTC	
HSD17B10_e4-15_Xho_se	TTTTTCTCGAGGCCTTCGAGGGTCAGGTT	
2MS2_inert_HSD17B10_e4-30_as	GGCAGCCACACTGGCAGTCTCGACCGACGGCTGAT	
2MS2_inert_HSD17B10_e4-45_as	CAGTGTTGATGATGACCCCCTCGACCGACGGCTGAT	
ACIN1_e13_Xho_se	TTTTTCTCGAGGTAGTACCTGCAGAGGGCC	
ACIN1_e15_Sal_Not_as	TTTTTTGCGGCCGCCCGGGTCGACTTACGTTACAAAGCAATGAGATTTG	
ATP5F1_e3_Xho_se	TTTTTCTCGAGGTATTGCAGGCAACAAGGAC	
ATP5F1_del-i4_se	GACTTTGCTGATAAACTCAATGAGCAAAAACTTGCCCAACTAGAA	
ATP5F1_e5_Sal_Not_as	TTTTTTGCGGCCGCCCGGGTCGACTTACCTTTGCACATCAAAAAGGTA	
ATP5F1_e4_HA_se	TTTGTTGCAGACTTTGCTGATTACCCATACGATGTGCCCGATTACGCTAAACTCAATGAGGTAAGAACCATAA	
ATP5B_e5_Xho_se	TTTTTCTCGAGCTTTTTGGTGGTGCTGGA	
ATP5B_del-i6_se	ACCCAGGCTGGTTCAGAGGTGTCTGCATTATTGGGC	
ATP5B_e7_Sal_Not_as	TTTTTTGCGGCCGCCCGGGTCGACTTACTGTACAGAGGTGATAGATCCCTT	
ATP5B_e6_HA_se	CGCTTCACCCAGGCTGGTTACCCATACGATGTGCCCGATTACGCTTCAGAGGTAAGAGGGAAGGC	
CIAO1_e5_Xho_se	TTTTCTCGAGCTCTTAGCTTCTGCCAGCTATG	
CIAO1_i5e6_as	CCGCTGCATGCCACCCCTGCAAGAC	
CIAO1_i5_se	TTGAGGTGCCCAGGACATAG	
CIAO1_rep_Not_Sal_as	TTTTTTGCGGCCGCCCGGGTCGACTC	
CIAO1_new_Py-tract_se	CCCAGGACATAGGAACGTTTTTCTCATTTTCTTTCCCCTTTCACAGGGGTGGCATGCAGC	