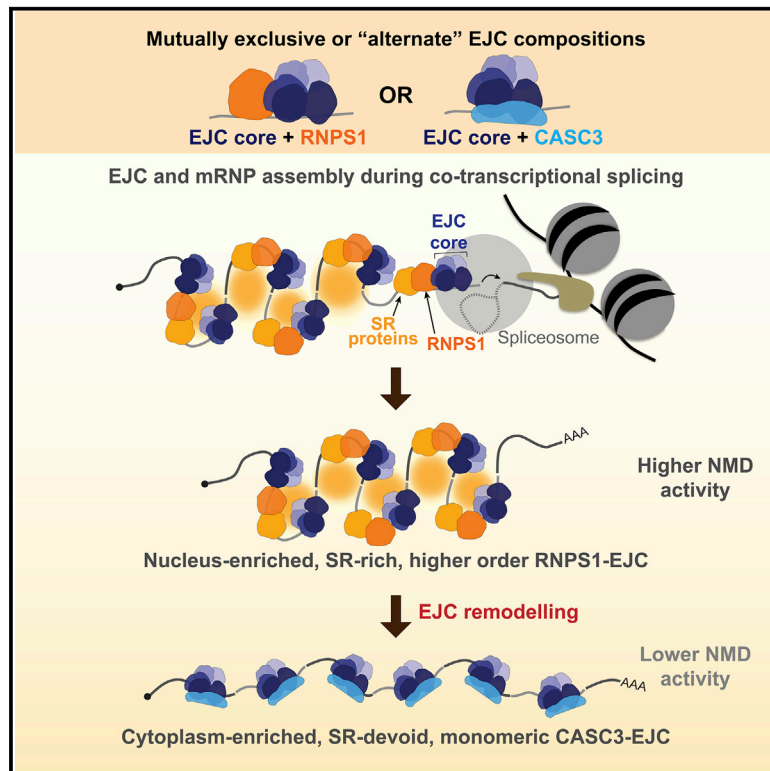


## The Exon Junction Complex Undergoes a Compositional Switch that Alters mRNP Structure and Nonsense-Mediated mRNA Decay Activity

### Graphical Abstract



### Authors

Justin W. Mabin, Lauren A. Woodward, Robert D. Patton, ..., Vicki H. Wysocki, Ralf Bundschuh, Guramrit Singh

### Correspondence

singh.734@osu.edu

### In Brief

The exon junction complex (EJC) is a mainstay of spliced mRNPs in mammalian cells. Mabin et al. discovered that the EJC undergoes a switch in composition from RNPS1-containing complex to CASC3-containing complex. This compositional switch changes higher-order EJC and mRNP structure and specifies two distinct phases of EJC-dependent NMD.

### Highlights

- Mammalian EJC core forms mutually exclusive complexes with RNPS1 and CASC3
- The early, higher-order RNPS1-EJC switches to the late, monomeric CASC3-EJC
- Nuclear RNA retention mechanisms impact the EJC compositional switch
- All NMD targets require RNPS1-EJC, while only select ones need CASC3-EJC



# The Exon Junction Complex Undergoes a Compositional Switch that Alters mRNP Structure and Nonsense-Mediated mRNA Decay Activity

Justin W. Mabin,<sup>1,5,6</sup> Lauren A. Woodward,<sup>1,5,6</sup> Robert D. Patton,<sup>3,5,6</sup> Zhongxia Yi,<sup>1,5</sup> Mengxuan Jia,<sup>2</sup> Vicki H. Wysocki,<sup>2,5</sup> Ralf Bundschuh,<sup>2,3,4,5</sup> and Guramrit Singh<sup>1,5,7,\*</sup>

<sup>1</sup>Department of Molecular Genetics, The Ohio State University, Columbus, OH 43210, USA

<sup>2</sup>Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210, USA

<sup>3</sup>Department of Physics, The Ohio State University, Columbus, OH 43210, USA

<sup>4</sup>Division of Hematology, Department of Internal Medicine, The Ohio State University, Columbus, OH 43210, USA

<sup>5</sup>Center for RNA Biology, The Ohio State University, Columbus, OH 43210, USA

<sup>6</sup>These authors contributed equally

<sup>7</sup>Lead Contact

\*Correspondence: [singh.734@osu.edu](mailto:singh.734@osu.edu)

<https://doi.org/10.1016/j.celrep.2018.11.046>

## SUMMARY

The exon junction complex (EJC) deposited upstream of mRNA exon junctions shapes structure, composition, and fate of spliced mRNA ribonucleoprotein particles (mRNPs). To achieve this, the EJC core nucleates assembly of a dynamic shell of peripheral proteins that function in diverse post-transcriptional processes. To illuminate consequences of EJC composition change, we purified EJCs from human cells via peripheral proteins RNPS1 and CASC3. We show that the EJC originates as an SR-rich mega-dalton-sized RNP that contains RNPS1 but lacks CASC3. Sometime before or during translation, the EJC undergoes compositional and structural remodeling into an SR-devoid monomeric complex that contains CASC3. Surprisingly, RNPS1 is important for nonsense-mediated mRNA decay (NMD) in general, whereas CASC3 is needed for NMD of only select mRNAs. The switch to CASC3-EJC slows down NMD. Overall, the EJC compositional switch dramatically alters mRNP structure and specifies two distinct phases of EJC-dependent NMD.

## INTRODUCTION

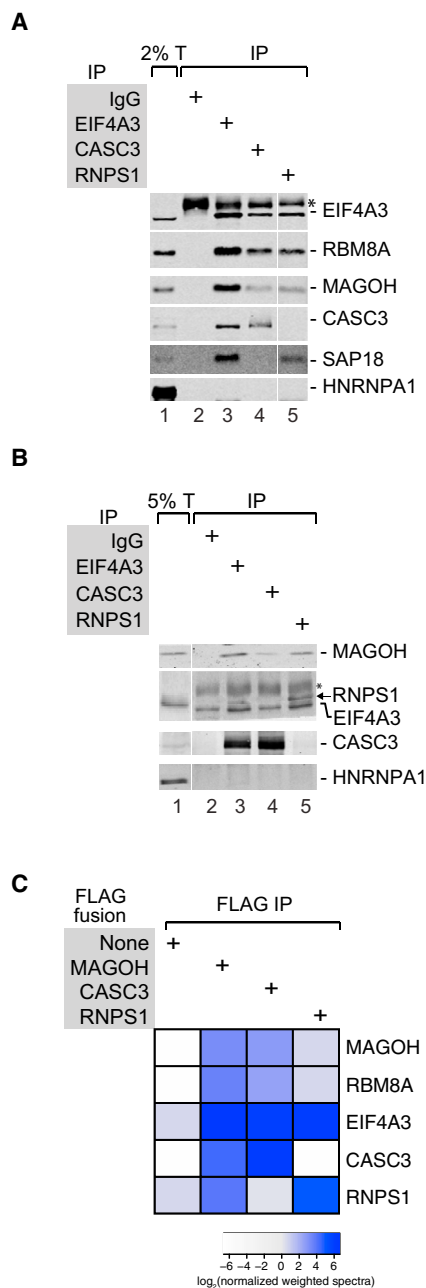
From the time of their birth until their eventual demise, messenger RNAs (mRNAs) exist decorated with proteins as mRNA-protein particles, or mRNPs (Singh et al., 2015). The vast protein complement of mRNPs has been illuminated (Henze et al., 2018) and is presumed to change as mRNPs progress through their life. However, the understanding of mechanisms and consequences of mRNP composition change remains confined to only a handful of its components. For example, mRNA export adapters are removed upon mRNP export to provide directionality to mRNP metabolic pathways, and the nuclear

cap and poly(A)-tail binding proteins are exchanged for their cytoplasmic counterparts after mRNP export to promote translation (Singh et al., 2015 and references therein). When, where, and how the multitude of mRNP components change during its lifetime and how such changes impact mRNP function remain largely unknown.

A key component of all spliced mRNPs is the exon junction complex (EJC), which assembles during pre-mRNA splicing 24 nucleotides (nt) upstream of exon-exon junctions (Boehm and Gehring, 2016; Le Hir et al., 2016; Woodward et al., 2017). Once deposited, the EJC enhances gene expression at several post-transcriptional steps, including pre-mRNA splicing, mRNA export, mRNA transport and localization, and translation. If an EJC remains bound to an mRNA downstream of a ribosome terminating translation, it stimulates nonsense-mediated mRNA decay (NMD). The stable trimeric EJC core forms when RNA-bound EIF4A3 is locked in place by RBM8A (also known as Y14) and MAGOH. This trimeric core is thought to be joined by a fourth protein CASC3 (also known as MLN51 or Barentsz) to form a stable tetrameric core (Boehm and Gehring, 2016; Hauer et al., 2016; Le Hir et al., 2016). However, more recent evidence suggests that CASC3 may not be present in all EJCs and may not be necessary for all EJC functions (Mao et al., 2017; Singh et al., 2012). Nonetheless, the stable EJC core interacts with a dynamic shell of peripheral EJC proteins such as pre-mRNA splicing factors (e.g., SRm160, RNPS1), mRNA export proteins (e.g., the TREX complex), translation factors (e.g., SKAR), and NMD factors (e.g., UPF3B) (Boehm and Gehring, 2016; Le Hir et al., 2016; Woodward et al., 2017). Some peripheral EJC proteins share similar functions and yet may act on different mRNAs; e.g., RNPS1 and CASC3 can both enhance NMD but may have distinct mRNA targets (Gehring et al., 2005). Thus, the peripheral EJC shell may vary between mRNPs leading to compositionally distinct mRNPs, an idea that has largely remained untested.

We previously showed that, within spliced mRNPs, EJCs interact with one another as well as with several SR and SR-like proteins to assemble into mega-dalton-sized RNPs (Singh et al., 2012). These stable mega-RNPs ensheath RNA well





**Figure 1. RNPS1 and CASC3 Exist in Mutually Exclusive EJCs in Mammalian Cells**

(A) Western blots showing proteins on the right in RNase-A-treated total HEK293 cell extract (T) (lane 1) or in the immunoprecipitates (IP) (lanes 2–5) of the antibodies listed on the top. Asterisk (\*) indicates IgG heavy chain.

(B) Western blots as in (A) from RNase-A-treated mouse cortical total extracts and IPs. Arrow: RNPS1 signal sandwiched between IgG heavy chain and EIF4A3.

(C) A heatmap of signal for normalized-weighted spectra observed for the proteins on the right in the bottom-up proteomics of the indicated FLAG IPs (top). Bottom: heatmap color scale.

See also Figure S1.

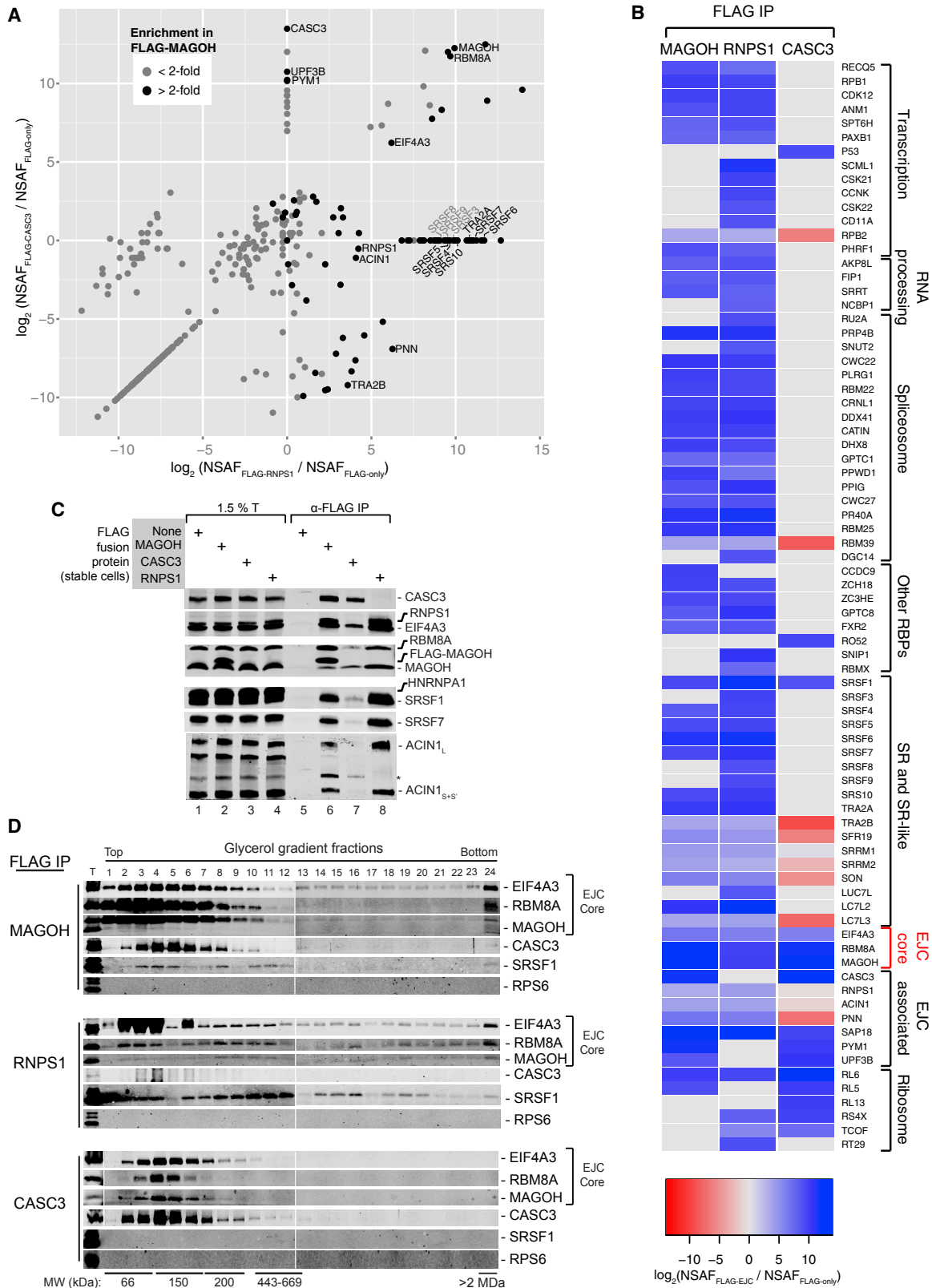
beyond the canonical EJC deposition site, leading to 150- to 200-nt-long RNA footprints, suggesting that the RNA polymer within these complexes is packaged into an overall compact mRNP structure. Such a compact structure may facilitate mRNP navigation of the intranuclear environment, export through the nuclear pore, and transport within the cytoplasm to arrive at its site of translation. Eventually, the mRNA within mRNPs must be unpacked to allow access to the translation machinery. How long mRNPs exist in their compact states and when, where, and how they are unfurled remains yet to be understood.

Our previous observation that, in human embryonic kidney (HEK293) cells, CASC3 and many peripheral EJC factors are substoichiometric to the EJC core (Singh et al., 2012) spurred us to investigate variability in EJC composition. Here, we use EJC purification via substoichiometric factors to reveal that EJCs first assemble into SR-rich mega-dalton-sized RNPs and then undergo a compositional switch into SR-devoid monomeric CASC3-containing EJCs. These EJC forms differ in their NMD activity; RNPS1, a component of the SR-rich EJCs, is crucial for all tested NMD substrates, whereas CASC3, a constituent of SR-devoid EJCs, is dispensable for NMD of some transcripts. Our findings reveal a step in the mRNP life cycle wherein EJCs, and by extension mRNPs, undergo a remarkable compositional switch that alters the mRNP structure and specifies two distinct phases of EJC-dependent NMD.

## RESULTS

### RNPS1 and CASC3 Associate with the EJC Core in a Mutually Exclusive Manner

We reasoned that substoichiometric EJC proteins may not interact with all EJC cores, and therefore, some of them may not interact with each other. To test this prediction, we performed immunoprecipitation (IP) for either endogenous core factor EIF4A3 or the substoichiometric EJC proteins RNPS1 and CASC3 from RNase-A-treated HEK293 total cell extracts. As expected, EIF4A3 IP enriches EJC core, as well as all peripheral proteins tested (Figure 1A, lane 3). In contrast, the IPs of substoichiometric factors enrich distinct sets of proteins. CASC3 immunoprecipitated EIF4A3, RBM8A, and MAGOH but not the peripheral proteins ACIN1 and SAP18 (Figure 1A, lane 4, and data not shown; RNPS1 could not be detected as it co-migrates with antibody heavy chain). Conversely, RNPS1 IP enriches the EJC core proteins and its binding partner SAP18 (Murachelli et al., 2012; Tange et al., 2005), but no CASC3 is detected (Figure 1A, lane 5). A similar lack of co-IP between RNPS1 and CASC3 even after formaldehyde crosslinking of cells prior to lysis (data not shown; see below) suggests that the lack of interaction is not due to their dissociation in extracts. In similar IPs from RNase-A-treated total extracts of mouse brain cortical slices, mouse embryonal carcinoma (P19) cells, and HeLa cells, RNPS1 and CASC3 efficiently co-IP with EIF4A3 but not with each other (Figures 1B, S1A, and S1B, compare lanes 4 and 5 with lane 3). Thus, in mammalian cells, the EJC core forms mutually exclusive complexes with RNPS1 and CASC3, which we refer to as alternate EJCs.



(legend on next page)

### Proteomic Analysis of Alternate EJCs

To gain further insights into alternate EJCs, we characterized RNPS1 and CASC3-containing complexes using bottom-up proteomics. We generated stable HEK293 cell lines to achieve tetracycline-inducible expression of FLAG-tagged RNPS1 and CASC3 at near-endogenous levels (Figure S1C). FLAG IPs from RNase-treated total extracts of these cells confirmed that the tagged proteins also exist in mutually exclusive EJCs (Figure S1D). Importantly, proteomic analysis of FLAG-affinity-purified alternate EJCs shows almost complete lack of spectra corresponding to RNPS1 and CASC3 in the IP of the other alternate EJC factor (Figures 1C and S1E). In comparison, a FLAG-MAGOH IP enriches both CASC3 and RNPS1 as expected (Figures 1C, S1D, and S1E).

We next analyzed the complement of proteins more than 2-fold enriched in FLAG-RNPS1, FLAG-CASC3, or FLAG-MAGOH IPs as compared to the FLAG-only control (Table S1). Of the 59 proteins enriched in the two FLAG-RNPS1 biological replicates (Figure S2A), 38 are common with the 45 proteins identified in the two FLAG-MAGOH replicates (Figures S2A and S2B), suggesting that RNPS1-containing complexes are compositionally similar to those purified via the EJC core. In comparison, among the FLAG-CASC3 replicates, EIF4A3 and MAGOH are the only common proteins that are also shared with FLAG-MAGOH (Figures S2A and S2B). Known EJC interactors such as UPF3B and PYM1 are identified in only one of the two FLAG-CASC3 replicates, indicating their weaker CASC3 association.

A direct comparison of FLAG-RNPS1 and FLAG-CASC3 proteomes reveals their stark differences beyond the EJC core (Figures 2A and 2B). Several SR and SR-like proteins are enriched in the RNPS1 and MAGOH proteomes but are absent from the CASC3 IPs. Among the SR protein family, SRSF1, 6, 7, and 10 are enriched in both RNPS1 and MAGOH samples, while all other canonical SR proteins, with the exception of non-shuttling SRSF2, are detected in at least one of the two replicates (Figures 2A–2C and Table S1). Several SR-like proteins such as ACIN1, PNN, SRRM1, and SRRM2 are also enriched in MAGOH and RNPS1 IPs but are absent in CASC3 IPs (Figure 2 and Table S1). A weak but erratic signal for SRSF1 and SAP18 is seen in FLAG-CASC3 IPs (Figure 2B). However, their enrichment with CASC3 is much weaker as compared to MAGOH or RNPS1. Thus, an interaction network exists between SR and SR-like proteins and the RNPS1-EJC but not the CASC3-EJC.

Other proteins specifically associated with RNPS1 (and MAGOH), but not with CASC3, include transcription machinery components (e.g., RPB1, RPB2), transcriptional regulators (e.g., CD11A, CDK12), and RNA processing factors (e.g.,

NCBP1, FIP1) (Figure 2B). Consistent with EJC core assembly during pre-mRNA splicing, MAGOH and RNPS1 interactors also include U2, U4, and U6 snRNP components and the nineteen complex subunits. None of the splicing components interact with CASC3. Considering that SR proteins also assemble onto nascent RNAs co-incident with transcription and splicing (Zhong et al., 2009), the EJC-RNPS1-SR interaction network likely originates during co-transcriptional mRNP biogenesis. In contrast, CASC3 most likely engages with the EJC only post-splicing, as previously suggested (Gehring et al., 2009a).

### Alternate EJCs Are Structurally Distinct

The enrichment of SR and SR-like proteins exclusively with RNPS1 suggests that RNPS1-EJCs are likely to resemble the previously described higher-order EJCs (Singh et al., 2012). Indeed, glycerol gradient fractionation of RNase-treated FLAG-RNPS1 complexes shows that, like FLAG-MAGOH EJCs, FLAG-RNPS1 EJCs contain both lower- and higher-molecular-weight complexes (Figure 2D). On the other hand, CASC3 is mainly detected in lower-molecular-weight complexes purified via FLAG-MAGOH (Figure 2D, compare CASC3 in fractions 2–10 and 22–24). Further, FLAG-CASC3 complexes are exclusively comprised of lower-molecular-weight complexes, likely to be EJC monomers (Figure 2D). Thus, compositional distinctions between the two alternate EJCs give rise to two structurally distinct complexes.

### RNPS1 and CASC3 Bind RNA via the EJC Core with Key Distinctions

We next identified the RNA binding sites for the two alternate EJC factors using RNA:protein immunoprecipitation in tandem (RIPiT) combined with high-throughput sequencing, or RIPiT-seq (Singh et al., 2012, 2014). RIPiT-seq entails tandem purification of two subunits of an RNP and is well suited to study EJC composition via sequential IP of its constant (e.g., EIF4A3, MAGOH) and variable (e.g., RNPS1, CASC3) components (Figure 3A). We carried out RIPiTs from HEK293 cells by either pulling first on FLAG-tagged alternate EJC factor followed by IP of an endogenous core factor or vice versa. EJC binding studies thus far have used translation elongation inhibitor cycloheximide (CHX) treatment to limit EJC disassembly by translating ribosomes (Hauer et al., 2016; Saulière et al., 2012; Singh et al., 2012). However, to capture unperturbed, steady-state populations of RNPS1- and CASC3-EJCs (as in Figures 1 and 2), we performed RIPiT-seq without translation inhibition.

As expected, RIPiTs for each of the alternate EJCs specifically purified the targeted complex along with the EJC core factors

### Figure 2. CASC3- and RNPS1-Containing Complexes Have Distinct Protein Composition and Hydrodynamic Size

(A) A scatterplot comparing fold-enrichment of proteins in FLAG-RNPS1 IP over FLAG-only control (x axis) to fold-enrichment of the same proteins in FLAG-CASC3 IP over FLAG-only control (y axis). Each dot represents a protein identified in FLAG-EJC or FLAG-only control samples by Scaffold. Black dots: proteins enriched > 2-fold over control in the FLAG-MAGOH IP.

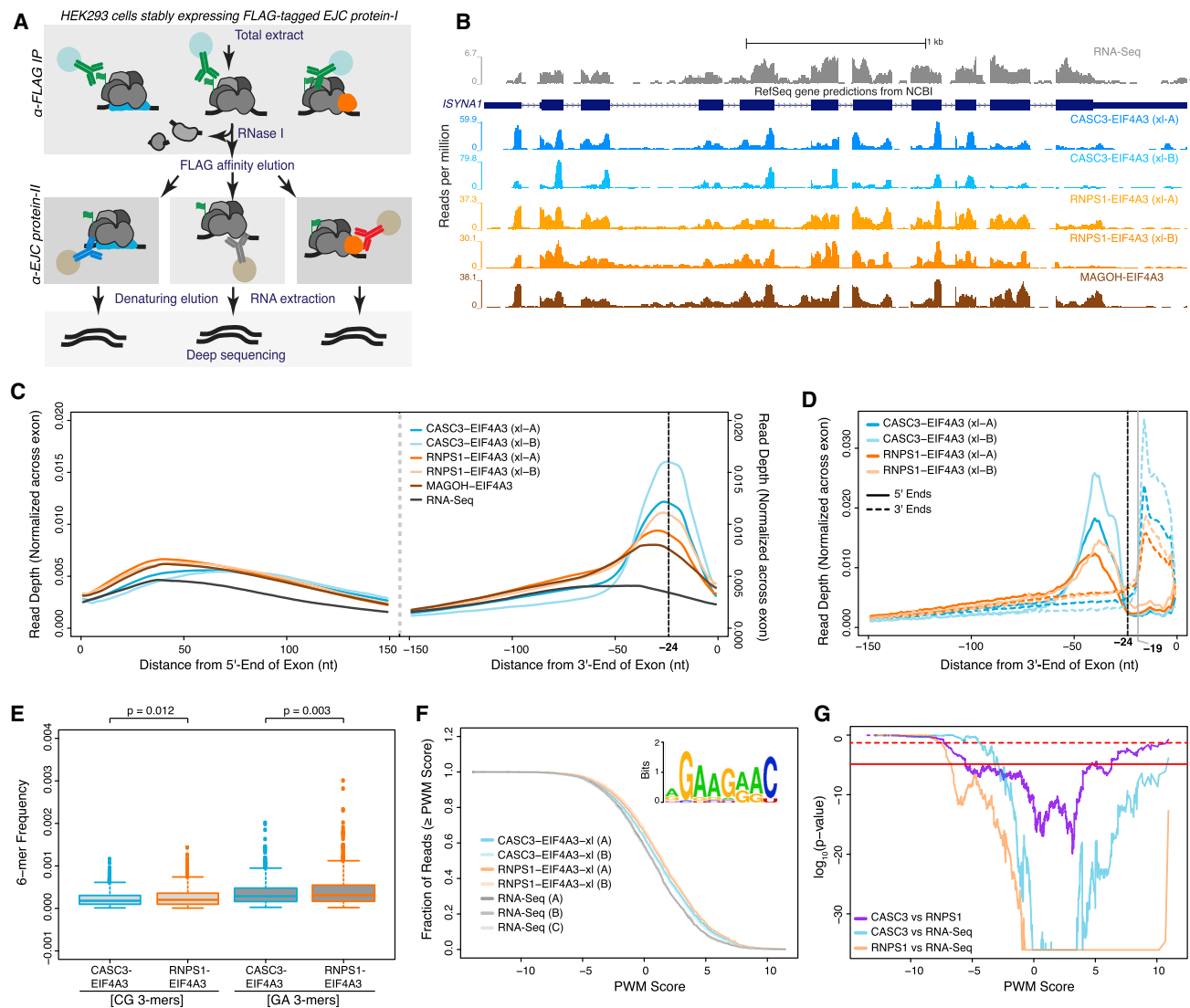
(B) A heatmap showing proteins >10-fold enriched in any of the FLAG-EJC IPs (indicated on the top) over FLAG-only control. Right: Protein groups by functions. Bottom: heatmap color scale.

(C) Western blots showing proteins (right) in total extract (T) or FLAG-IPs as on the top from HEK293 cells stably expressing FLAG-tagged proteins indicated on the top left. Note that RNPS1 migrates just above EIF4A3. Asterisk: CASC3 signal from an earlier probing.

(D) Western blots showing proteins on the right in glycerol gradient fractions of FLAG-IPs from HEK293 cells stably expressing FLAG-MAGOH, FLAG-RNPS1, or FLAG-CASC3 (far left). Top: gradient fraction. Bottom: molecular weight standards.

See also Figure S2 and Table S1.



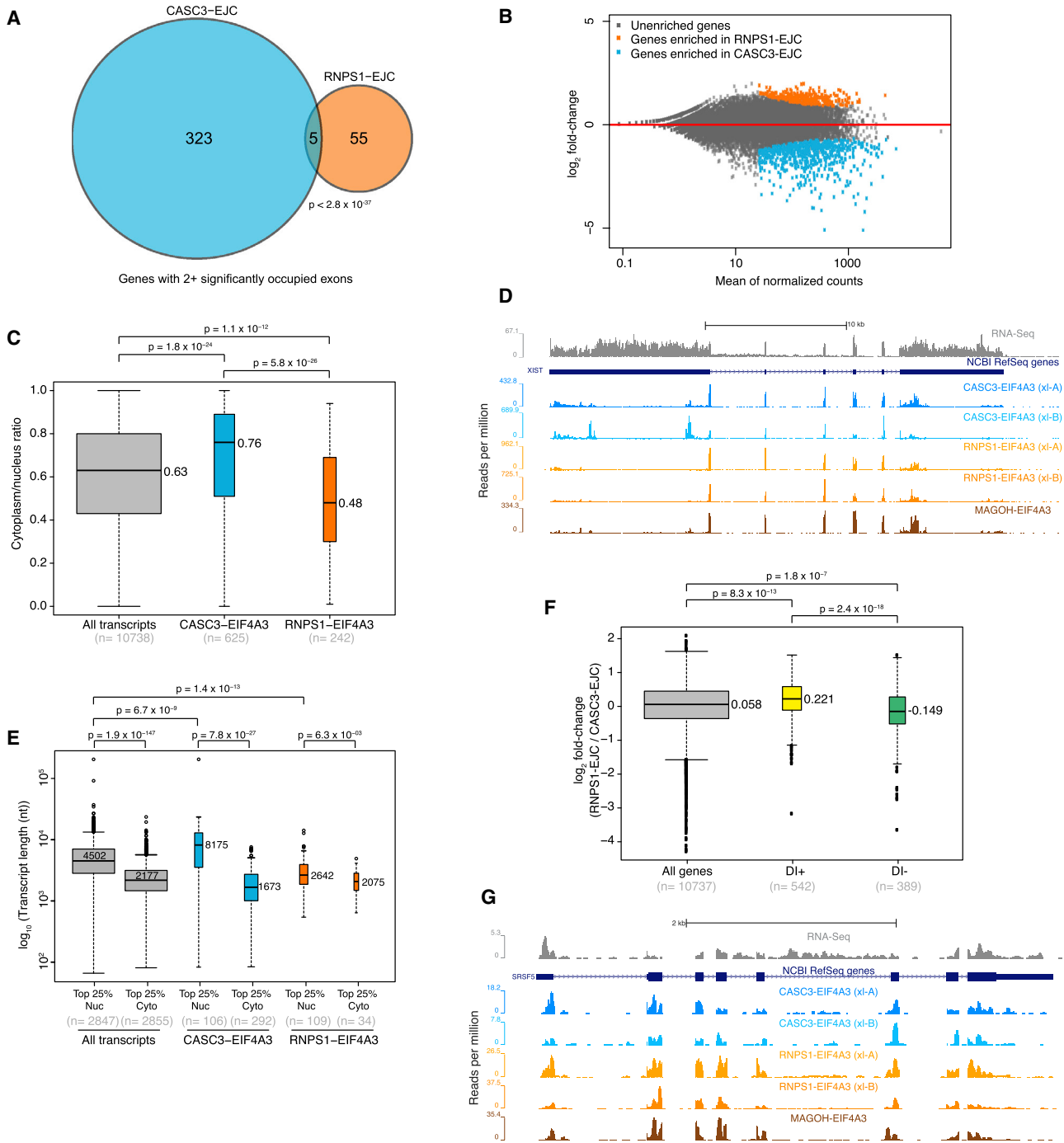


**Figure 3. RNPS1 Assembles with Canonical and Non-canonical EJCs, but CASC3 Is Mainly Part of Canonical EJCs**

(A) A schematic illustrating the main steps in RIPIT-seq.  
 (B) Genome browser screenshots comparing read coverage along the *ISYNA1* gene in RNA-seq or RIPIT-seq libraries (indicated on the right). Blue rectangles: exons; thinner rectangles: untranslated regions; lines with arrows: introns.  
 (C) Meta-exon plots showing read depth in different RIPIT-seq or RNA-seq libraries (indicated in the middle) in the 150 nt from the exon 5' (left) or 3' end (right). Vertical black dotted line: canonical EJC position at -24 nt.  
 (D) A composite plot of RNPS1 and CASC3 RIPIT-seq footprint read 5' (solid lines) and 3' (dotted lines) ends. Vertical dotted line: canonical EJC site (-24 nt). Vertical gray line: EJC 3' boundary (-19 nt).  
 (E) Boxplots showing frequencies of 6-mers that contained CG 3-mers (CGG, GCG, CCG, CGC) or GA 3-mers (GGA, GAA, AGG, GAG) in CASC3 or RNPS1 RIPIT-seq reads. Top: Wilcoxon rank-sum test p values.  
 (F) Cumulative distribution function plots showing frequency of reads in the indicated samples with the highest score for match to SRSF1 motif (inset) position weight matrix (PWM). Bottom left: sample identity legend. The SRSF1 motif in the inset is from Tacke and Manley (1995).  
 (G) A negative binomial based assessment of significance of differences in SRSF1 motif PWM scores in (F) between the two alternate EJC footprint reads or between alternate EJC and RNA-seq reads (legend on the bottom left). Horizontal dotted red line:  $p = 0.05$ . Horizontal solid red line: Bonferroni adjusted p value. See also Figure S3 and Table S2.

and yielded abundant RNA footprints (Figure S3A). Strand-specific RIPIT-seq libraries from ~35–60 nucleotide footprints yielded ~2.5–27 million reads, of which > 80% mapped uniquely to the human genome (Table S2). Genic read counts are highly correlated between RIPITs, where the order of IP of EJC core

and alternate factors was reversed (Figures S3B and S3C). Unlike CASC3-EJC, RNPS1-EJC interaction was susceptible to NaCl concentration > 250 mM (data not shown). Therefore, to preserve labile interactions, we performed alternate EJC RIPIT-seq from cells cross-linked with formaldehyde before



**Figure 4. RNPS1 and CASC3 Occupy the Same RNAs and Sites but within Different RNA Pools**

(A) A Venn diagram showing genes with at least two differentially enriched exons, for which all differentially enriched exons are preferentially bound to CASC3-EJC (blue region), RNPS1-EJC (orange region), or genes with at least one exon enriched in CASC3-EJC and one in RNPS1-EJC (overlap region). The probability of observing an overlap of 5 or fewer genes using a binomial distribution approximating each gene to contain two exons is shown.

(B) An MA-plot showing fold-change in RNPS1-EJC versus CASC3-EJC footprint reads (y axis) against expression levels (x axis). Each dot represents a canonical transcript for each known gene in GRCh38 from UCSC “knownCanonical” splice variant table. Transcripts differentially enriched ( $p$ -adjusted  $< 0.05$ ) in RNPS1-EJC (orange) and CASC3-EJC (blue) are indicated.

(C) Boxplots showing distribution of cytoplasmic/nuclear fraction (y axis) for all (gray), CASC3-EJC-enriched (blue), and RNPS1-EJC-enriched (orange) transcripts. The median values are to the right of each boxplot. Top:  $p$  values (Wilcoxon rank sum test). Bottom: number of transcripts in each group. Nuclear/cytoplasmic transcript level data are from [Neve et al. \(2016\)](#).

(D) Genome browser screenshots showing read coverage along the *XIST* gene in RNA-seq (top) or RIPit-seq libraries (labeled on right).

(legend continued on next page)

cell lysis (Figure S3D). A strong correlation is seen between crosslinked and uncrosslinked samples (Figures S3E and S3F). All analysis presented below is from two well-correlated biological replicates of formaldehyde crosslinked RIPit-seq datasets of RNPS1- and CASC3-EJC (Figures S3G and S3H).

Consistent with EJC deposition on spliced RNAs, alternate EJC footprints are enriched in exonic sequences from multi-exon genes (Figure S3I). Along exons, footprints of both alternate EJCs accumulate mainly at the canonical EJC binding site 24 nt upstream of exon junctions (Figures 3B and 3C) (Hauer et al., 2016; Saulière et al., 2012; Singh et al., 2012). Notably, the enrichment at the canonical EJC site is irrespective of whether the alternate EJC factors are IPed first or second during RIPit (Figure S3J). Thus, both RNPS1 and CASC3 bind to RNA via EJCs at the canonical site. The location of the 5'- and 3'-ends of the alternate EJC footprint reads shows that both alternate EJCs block positions –26 to –19 nt from RNase I cleavage (Figure 3D). Our finding that CASC3 mainly binds to the canonical EJC site is in agreement with the previous work (Hauer et al., 2016). However, in contrast to that study, RNPS1 binding to canonical EJC sites is readily apparent in our data (Figures 3B–3D), which could reflect differences in the UV-crosslinking-dependent CLIP-seq approach used by Hauer et al. (2016) and the photo-crosslinking independent RIPit-seq. Altogether, both RNPS1 and CASC3 exist in complex with EJC at its canonical site.

While read densities for both alternate EJC factors are highest at the canonical EJC site, 47%–62% of reads map outside of the canonical EJC site similar to previous estimates (Saulière et al., 2012; Singh et al., 2012). Non-canonical footprints are more prevalent in RNPS1-EJC as compared to CASC3-EJC (Figure S3K). As RNPS1-EJC is intimately associated with SR and SR-like proteins (Figures 2A and 2B), a k-mer enrichment analysis revealed a modest but significant enrichment of GA-rich 6-mers in RNPS1 over CASC3 footprints (Figure 3E). Such purine-rich sequences occur in binding sites of several SR proteins (SRSF1, SRSF4, Tra2a and b) (Änkö et al., 2012; Pandit et al., 2013; Take et al., 1998). A small but significant enrichment of SRSF1 and SRSF9 sequence motifs is seen in RNPS1 footprints as compared to CASC3 footprints or RNA-seq reads (Figures 3F, 3G, S3L, and S3M). Thus, within spliced RNPs, RNPS1-EJC is engaged with SR and SR-like proteins and other RNA binding proteins, which leads to co-enrichment of RNA binding sites of these proteins during RNPS1 RIPit.

### Subcellular mRNP Localization and Nuclear Retention Mechanisms Affect Relative RNPS1 and CASC3 Occupancy

Surprisingly, despite the mutually exclusive association of RNPS1 and CASC3 with the EJC core (Figures 1 and 2), the two proteins are often detected on the same sites on RNA, lead-

ing to their similar apparent occupancy on individual exons as well as entire transcripts (Figures 3B, S4A, and S4B). These results suggest that the two alternate EJC factors bind to two distinct pools of the same RNAs. To further reveal RNA binding patterns of the alternate EJC factors, we identified exons differentially enriched in one or the other factor. If two or more exons from the same gene are differentially enriched in RNPS1 or CASC3 footprints, these exons are almost always enriched in the same alternate EJC factor (Figure 4A). This tight linkage between EJC compositions of different exons of the same gene suggests that RNPS1 and CASC3 binding to the EJC is likely to be determined at the level of the entire transcript rather than at the level of the individual exon. Additionally, no RNA-dependent RNPS1 co-IP is observed with FLAG-CASC3, and only a weak RNA-dependent CASC3 co-IP is seen with FLAG-RNPS1 (Figures S4C and S4D). Thus, while some RNPs with a heterogeneous mix of alternate EJCs can be captured, at any given time, a transcript with multiple EJCs is likely to be more homogeneously associated with one or the other alternate factors.

At steady state, RNPS1 is mainly nuclear, whereas CASC3 is predominantly cytoplasmic, although both proteins shuttle between the two compartments (Daguenet et al., 2012; Degot et al., 2002; Lykke-Andersen et al., 2001). We reasoned that different concentrations of alternate EJC factors in the two compartments might mirror their EJC and RNA association. To test this possibility, we identified subsets of transcripts preferentially enriched in RNPS1- (242 transcripts) or CASC3-EJC (625 transcripts; Figure 4B) and compared their cytoplasm/nucleus ratios based on subcellular RNA distribution estimates in HEK293 cells (Neve et al., 2016). Indeed, the transcripts enriched in CASC3-EJC show higher cytoplasmic levels (median cytoplasm/nucleus ratio = 0.76), whereas those preferentially bound to RNPS1 show a higher nuclear localization (median cytoplasm/nucleus ratio = 0.48; Figures 4C and S4E). Thus, steady-state subcellular RNA localization is a key determinant of EJC composition.

Although CASC3-enriched RNAs show higher cytoplasmic levels as a group, a quarter of them are more nuclear (Figure 4C). Consistent with CASC3 shuttling into the nucleus (Daguenet et al., 2012; Degot et al., 2002), its footprints are abundantly detected on *XIST* RNA and several other spliced non-coding RNAs restricted to or enriched in the nucleus (Figures 4D and S4E). Interestingly, CASC3-EJC-enriched RNAs that fall within the top 25% nuclearly localized transcripts are characterized by significantly longer transcript, internal exon, and gene lengths (Figures 4E, S4F, and S4G). In contrast, neither the RNPS1-enriched transcripts nor the CASC3-enriched transcripts with higher cytoplasmic levels display these properties. We noted a direct correlation between transcript length and relative nuclear localization (Figure 4E and data not shown). Therefore, RNPs transcribed from long genes and containing long exons may

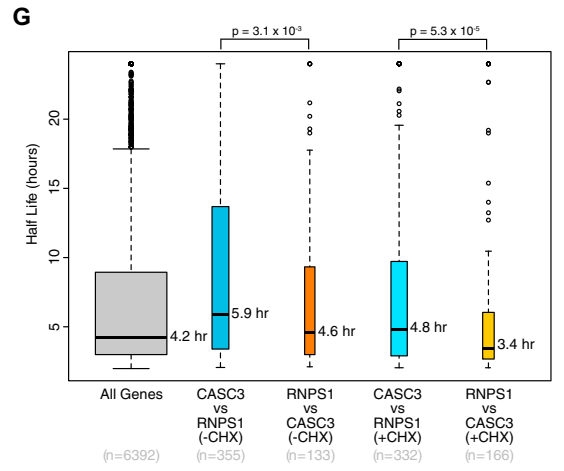
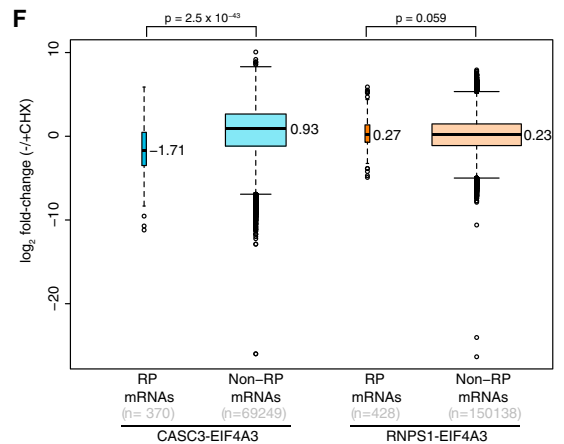
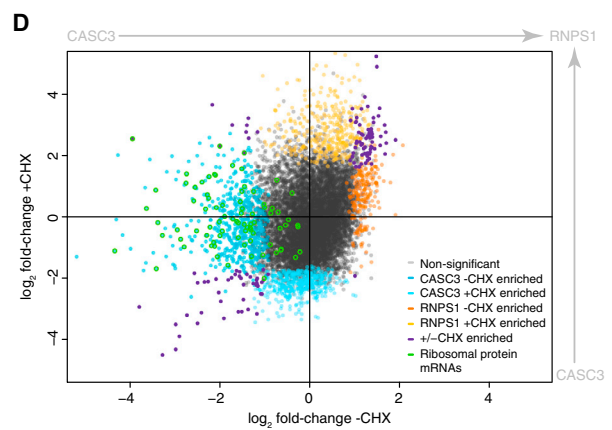
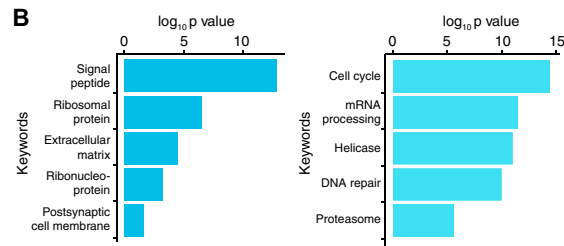
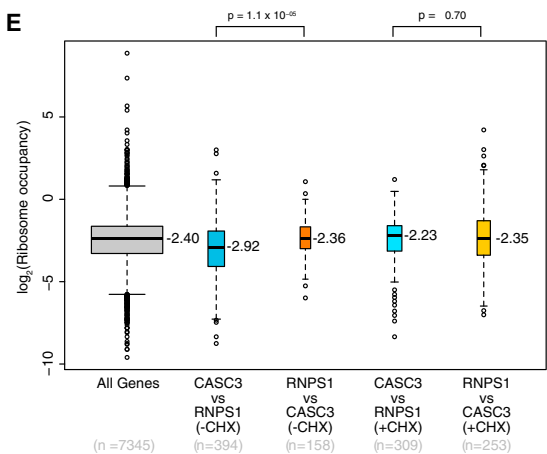
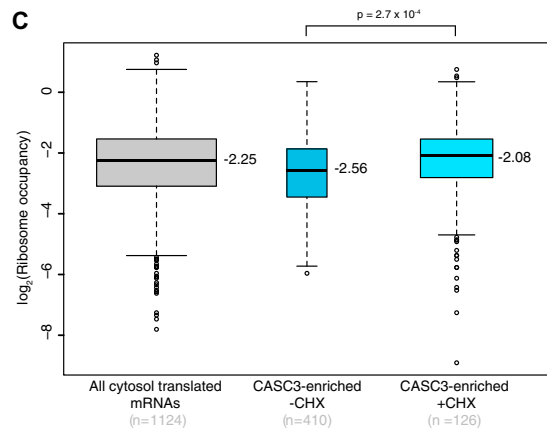
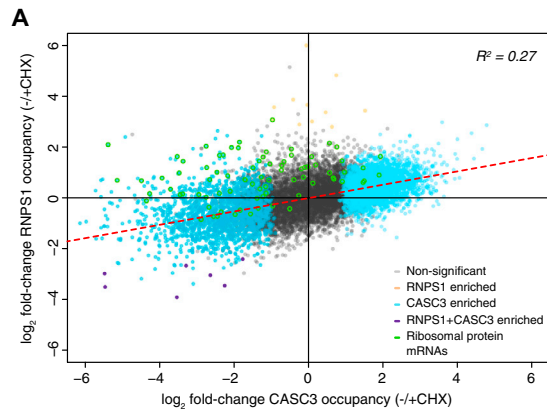
(E) Boxplots as in (C) showing transcript lengths (y axis) of the groups indicated on the x axis. Top: p values (Wilcoxon rank sum test). Bottom: number of transcripts in each group.

(F) Boxplots showing distribution of fold-change values (y axis) for RNPS1-EJC versus CASC3-EJC footprint read counts for all (gray), detained intron-containing (DI+; yellow) or detained intron-lacking (DI–; green) transcripts. (+ve values: RNPS1-EJC enriched; –ve values: CASC3-EJC enriched). Top: p values (Wilcoxon rank sum test). Bottom: number of transcripts in each group. DI+/DI– genes were defined based on Boutz et al. (2015).

(G) Genome browser screen-shot as in (E) for the *SRSF5* locus. Note increased RNA-seq reads in introns 4 and 5.

See also Figure S4.





(legend on next page)

persist for extended duration in the nucleus after their biogenesis and may eventually undergo the switch to CASC3-EJC within the nucleus itself.

We next compared alternate EJC occupancy of another class of transcripts retained in the nucleus due to presence of slow splicing introns (Boutz et al., 2015). Transcripts containing these detained introns (DIs) are restricted to the nucleus in a mostly pre-processed, poly(A)-tailed state until DIs are spliced. Of the genes that Boutz et al. (2015) found to contain DI in four human cell lines, 542 also contained DIs in HEK293 cells, whereas a set of 389 transcripts do not contain DI in any human cell lines, including HEK293 cells (data not shown). In contrast to the nuclear localized long transcripts, the DI-containing transcripts are significantly more enriched in RNPS1-EJC (Figures 4F and 4G), while DI-lacking transcripts show enrichment in CASC3-EJC. Therefore, transcripts that persist longer in the pre-processed state may remain preferentially associated with RNPS1-EJCs and switch to CASC3-EJC at a slower rate. Overall, distinct nuclear retention mechanisms can have opposing effects on EJC composition.

### Kinetics of Translation and mRNA Decay Impacts the Pool of Alternate EJC-Bound mRNAs

As EJCs are disassembled during translation (Dostie and Dreyfuss, 2002; Gehring et al., 2009b), abundant RNPS1 and CASC3 footprints at EJC deposition sites suggest that the bulk of mRNPs undergo the compositional switch before translation. Consistently, we detected RNA-dependent interactions between alternate EJC factors and nuclear cap binding protein CBP80 (Figure S4D). To test how translation impacts alternate EJC occupancy, and if this occupancy is influenced by the rate at which mRNAs enter the translation pool, we obtained RNPS1- and CASC3-EJC footprints from cells treated with CHX and compared them to alternate EJC footprints from untreated cells. When mRNAs bound to each alternate EJC are compared across the two conditions, CASC3-EJC occupancy shows a dramatic change (Figure 5A). The change in RNPS1-EJC occupancy trends in the same direction ( $R^2 = 0.27$ ) but is much more modest. Thus, the RNPS1-EJC likely precedes the CASC3-EJC. Also, either the RNPS1 to CASC3 switch is mildly affected by translation inhibition or RNPS1-EJCs are also subject to some translation-dependent disassembly. Nonetheless, translation inhibition leads to accumulation of mRNPs mainly with CASC3-EJC.

We predicted that poorly translated mRNAs will be enriched in CASC3-EJC under normal conditions, and more efficiently translated mRNAs will be differentially enriched upon translation inhibition. To test this idea, we inferred a measure of translation efficiency of human mRNAs based on their abundance normalized ribosome footprint counts in a human colorectal cancer cell line (Kiss et al., 2017). Presumably, this “ribosome occupancy” measure is comparable at least for ubiquitously expressed mRNAs across different human cell types and can be used as an indirect measure of HEK293 mRNA translation efficiencies. A comparison of the CASC3-EJC-enriched transcripts from untreated versus CHX-treated conditions, however, showed only a minor difference in their median ribosome occupancy (Figure S5A). A search for functionally related genes in the two sets revealed that each contains diverse groups (Figure 5B). Under normal conditions, the largest and most significant CASC3-EJC-enriched group encodes signal-peptide bearing secretory/membrane proteins (Figure 5B), which has significantly higher ribosome occupancy as compared to all transcripts (Figure S5B). We reason that, despite their higher ribosome occupancy, the “secretome” (as defined in Jan et al., 2014) transcripts may be enriched in CASC3-EJC because binding of the signal peptide to the signal recognition particle (SRP) halts translation until the ribosome engages with the endoplasmic reticulum (ER) (Walter and Blobel, 1981). Presumably, the time before translation resumption on the ER allows capture of mRNPs where EJC composition has switched but it has not yet been disassembled. Consistently, a weak RNA-dependent interaction is seen between CASC3 and SRP68, an SRP component (Figure S4D). When we considered only the cytosol-translated transcripts (as defined in Chen et al., 2011), a comparison of ribosome occupancy of CASC3-EJC-enriched transcripts from untreated versus CHX-treated conditions confirmed our initial hypothesis. The median ribosome occupancy of transcripts bound to CASC3-EJC in the absence of CHX is significantly lower ( $-2.56$ ) as compared to transcripts bound to CASC3-EJC in the presence of CHX ( $-2.08$ ,  $p = 2.7 \times 10^{-4}$ , Figure 5C).

Another functional group enriched in CASC3-EJC under normal conditions comprises the ribosomal protein (RP)-coding mRNAs (Figure 5B). Strikingly, transcripts encoding ~50% of all cytosolic ribosome proteins, as well as 13 mitochondrial ribosome subunits, are among this group (Figure 5A). Although RP mRNAs are among the most well translated in the cell, a sizeable fraction of RP mRNAs exist in a dormant untranslated state

### Figure 5. Kinetics of Translation and mRNA Decay Impacts Alternate EJC Occupancy

(A) A scatterplot of fold-change in CASC3-EJC occupancy with and without cycloheximide (CHX; x axis) and fold-change in RNPS1-EJC with and without CHX (y axis). Each dot represents a canonical transcript for each GRCh38 known gene and is colored as indicated in the legend (bottom right). Green outlined dots: ribosomal protein genes. Dotted red line: linear regression. Top left corner: coefficient of determination ( $R^2$ ).

(B) Top 5 GO term keywords and their enrichment p values in CASC3-EJC enriched transcripts (from A) in the absence (left) or presence of CHX (right).

(C) Boxplots showing distribution of ribosome occupancy (y axis) of transcript groups on y axis. Median values are shown to the right of each boxplot. Top: p values (Wilcoxon rank sum test). Bottom: number of transcripts in each group. Ribosome occupancy estimates were based on Kiss et al. (2017).

(D) Scatterplot as in (A) comparing fold-change in CASC3-EJC versus RNPS1-EJC footprint read counts in  $-CHX$  (x axis) and  $+CHX$  (y axis) conditions.

(E) Boxplots as in (C) showing distribution of ribosome occupancy estimates of transcript groups from (D) as indicated on the bottom.

(F) Comparison of fold-change in CASC3-EJC or RNPS1-EJC footprint reads at canonical EJC sites from ribosomal protein (RP)-coding mRNAs or non-ribosomal (non-RP)-protein coding mRNAs. Top: p values (Wilcoxon rank sum test).

(G) Boxplots as in (C) and (E) above comparing mRNA half-life of transcript groups from (D) as indicated on the bottom. mRNA half-life data are from Tani et al. (2012).

See also Figure S5.

(Geyer et al., 1982; Meyuhos and Kahan, 2015; Patursky-Polischuk et al., 2009). Consistently, RP mRNAs have significantly lower ribosome occupancy in human and mouse cells (Figures S5B and S5C; Ingolia et al., 2011). When transcripts differentially bound to RNPS1- versus CASC3-EJC are directly compared in normally translating cells, RP mRNAs are specifically enriched among CASC3-EJC-bound transcripts (Figure 5D). Therefore, RP mRNPs, and perhaps other translationally repressed mRNPs, switch to and persist in the CASC3-bound form of the EJC. Consistently, under normal conditions, CASC3-EJC-enriched RNAs have significantly lower ribosome occupancy as compared to RNPS1-EJC-bound transcripts (Figure 5E). When mRNPs are forced to persist in an untranslated state upon CHX treatment, cytosol-translated mRNAs show increased CASC3 occupancy, whereas their RNPS1 occupancy is not affected (Figure S5D).

As reported by Hauer et al. (2016), RP mRNAs are depleted of CASC3-EJC upon translation inhibition (Figures 5A and 5D). Upon CHX treatment, CASC3 occupancy is significantly reduced at canonical EJC sites of RP mRNAs as compared to non-RP mRNAs, which show an increase in CASC3 occupancy (Figure 5F). In comparison, RNPS1 occupancy on all transcripts modestly increases upon CHX treatment (Figure 5F). The decrease in CASC3-EJC occupancy on RP mRNAs upon CHX treatment suggests a paradoxical possibility that the untranslated reserves of RP mRNAs enter translation when the pool of free ribosomes is dramatically reduced upon CHX-mediated arrest of translating ribosomes. Intriguingly, a similar contradictory increase in ribosome footprint densities on RP mRNAs upon CHX treatment was recently reported in fission yeast (Duncan and Mata, 2017).

The alternate EJC occupancy landscape is also impacted by mRNA decay kinetics. CASC3-EJC-enriched RNAs have longer half-lives as compared to RNPS1-EJC-enriched transcripts under both translation conducive (median  $t_{1/2} = 5.9$  hr versus 4.6 hr,  $p = 3.1 \times 10^{-3}$ ) and inhibitory conditions (median  $t_{1/2} = 4.8$  hr versus 3.4 hr,  $p = 5.3 \times 10^{-5}$ , Figure 5G). Notably, RNAs enriched in both alternate EJCs upon CHX treatment have shorter half-lives as compared to the corresponding cohorts enriched from normal conditions. Thus, EJC detection is enhanced on transcripts that are stabilized after CHX treatment. Consistently, functionally related groups of genes encoding unstable transcripts (e.g., cell cycle, mRNA processing, and DNA damage; Schwahnhäuser et al., 2011) are enriched in CASC3-EJC upon translation inhibition (Figure 5B).

### **RNPS1 Is Required for Efficient NMD of All Transcripts, Whereas CASC3 Is Required Only for a Subset**

We next tested if alternate EJC factors were equally important for NMD and/or if they have distinct targets as previously suggested (Gehring et al., 2005). In HEK293 cells with ~80% of RNPS1 mRNA and proteins depleted, a majority of endogenous NMD targets tested are significantly upregulated (Figures 6A and S6A). Surprisingly, however, in cells with ~85% of CASC3 depleted, some NMD targets are largely unaffected, whereas others are only modestly upregulated. Notably, the RNAs that were upregulated upon CASC3 depletion are even more upregulated upon RNPS1 depletion. Most tested RNAs are significantly upregulated upon depletion of EIF4A3 and the central NMD factor UPF1 (Fig-

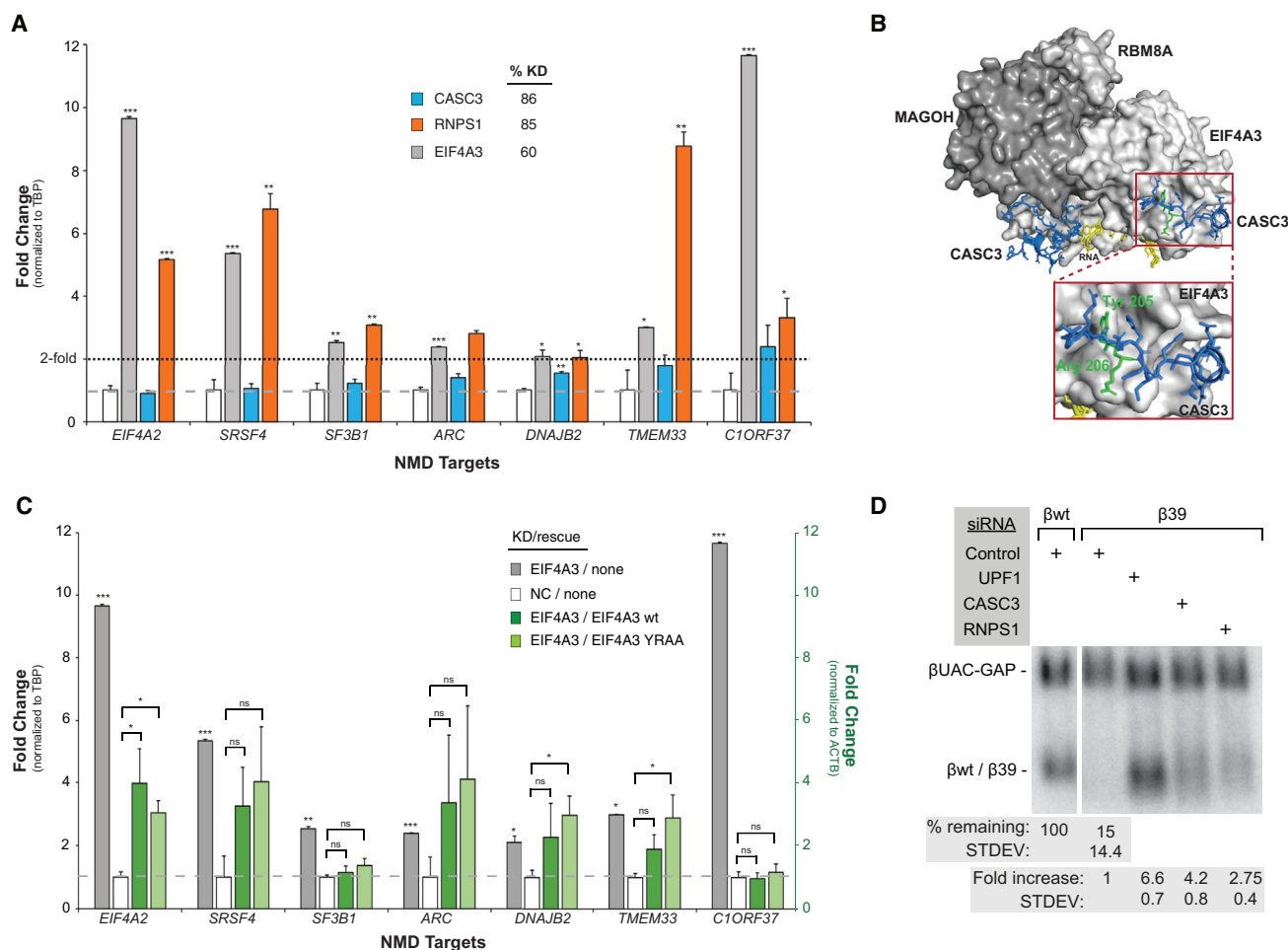
ures 6A and S6B). These data suggest that normal RNPS1 levels are required for efficient NMD of all endogenous NMD substrates tested, whereas even a large reduction in CASC3 levels leads to only small effects on NMD of this set of transcripts.

It is possible that either the residual amount of protein after CASC3 knockdown is sufficient to support NMD of the tested transcripts or CASC3 is dispensable for their NMD. To evaluate these possibilities further, we supplemented EIF4A3 knockdown cells with either a WT EIF4A3 (WT) or a mutant EIF4A3 (YRAA: Y205A, R206A) with much reduced CASC3 binding (Figures 6B and S6C–S6E) (Andersen et al., 2006; Ballut et al., 2005; Bono et al., 2006). Upon complementation of EIF4A3 knockdown cells with exogenous FLAG-EIF4A3 or FLAG-EIF4A3 YRAA proteins, the NMD targets tested can be divided into three groups based on the degree of rescue (Figure 6C). (i) *SF3B1* and *C1orf37* show almost complete rescue of their NMD upon expression of either wild-type or mutant EIF4A3. *ARC* and *SRSF4* also show a similar rescue as their transcript levels in the two rescue conditions are not significantly different than the control. Therefore, this group of transcripts can undergo NMD largely independently of CASC3. (ii) *EIF4A2* shows partial, but not complete, rescue under both conditions. (iii) *DNAJB2* and *TMEM33* levels were rescued by wild-type EIF4A3 but not by EIF4A3 YRAA. Therefore, at least these two transcripts depend on both RNPS1- and CASC3-EJC for their efficient NMD.

We next tested the effect of knockdown of RNPS1 and CASC3 on NMD of a well-known  $\beta$ -globin mRNA with a premature stop codon at codon 39 ( $\beta 39$ ). We observed that, like *DNAJB2* and *TMEM33*,  $\beta 39$  mRNA is upregulated upon knockdown of either alternate EJC factor in HeLa cells (Figures 6D and S6F). These results are also consistent with recent genome-wide screens that have identified CASC3 as an effector of NMD of exogenous reporters (Alexandrov et al., 2017; Baird et al., 2018). Overall, our results show that, in human cells, while all tested transcripts depend on RNPS1 for their efficient NMD, some NMD targets can undergo CASC3-independent NMD, and some endogenous transcripts and exogenously (over) expressed NMD targets may depend on both EJC compositions for their efficient NMD.

### **Increased CASC3 Levels Can Slow Down NMD**

We next tested if overexpression of RNPS1 or CASC3 can tilt EJC composition toward one of the two alternate EJCs, and if such a change can impact NMD. While CASC3 overexpression in HEK293 cells leads to a several-fold increase in CASC3 co-IP with EIF4A3 (Figures 7A and 7B) and RBM8A (Figures S7A and S7B), no concomitant decrease is seen in RNPS1 co-precipitation with EJC core proteins. Consistent with this, manifold overexpression of RNPS1 did not cause any detectable change in levels of the two alternate factors in the EJC core IPs (Figures 7A, 7B, S7A, and S7B). These results rule out a simple, direct competition between the two proteins for EJC core interaction. Surprisingly, upon CASC3 overexpression, all of the tested endogenous NMD targets show small but significant upregulation (with the exception of *EIF4A2* and *SF3B1*, Figure 7C). Similarly, the  $\beta 39$  mRNA exogenously expressed in HeLa Tet-off cells is also modestly stabilized upon CASC3 overexpression (Figures 7D and S7C). As previously reported (Viegas et al., 2007), RNPS1



**Figure 6. RNPS1 and CASC3 Differ in Their Requirement for Efficient NMD**

(A) Fold change in levels of endogenous NMD-targeted transcripts (bottom) in HEK293 Flp-In TRex cells depleted of core or alternate EJC factors. (EIF4A3 knockdown: 48 hr, alternate EJC factor knockdown: 96 hr). Shown are the average values normalized to TBP levels from three biological replicates  $\pm$  standard error of means (SEM). \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  (Student's *t* test). Percent knockdown of each protein in a representative experiment is in the legend. (B) An EJC core crystal structure (PDB: 2J0S) showing the EIF4A3 residues that interact with CASC3 (enlarged below).

(C) Real-time PCR analysis as in (A) from HEK293 cells depleted of EIF4A3. The EIF4A3 knockdown data are the same as in (A) and is plotted on the left y axis. Either wild-type EIF4A3 or a mutant with reduced CASC3 interaction were exogenously expressed in EIF4A3 knockdown cells as indicated in the legend on the top right, and these data are plotted on the right y axis.

(D) Northern blots showing levels of wild-type ( $\beta$ wt, lane 1) or PTC-containing (codon 39;  $\beta$ 39; lanes 2-5)  $\beta$ -globin mRNA and a longer internal control  $\beta$ -globin ( $\beta$ UAC-GAP) mRNA from HeLa Tet-off cells treated with siRNAs indicated on the top. Tables below indicate percentage of normalized  $\beta$ 39 mRNA as compared to normalized  $\beta$ wt mRNA (top) or fold-increase in  $\beta$ 39 mRNA upon knockdown as compared to control.

See also Figure S6.

overexpression further downregulates  $\beta$ 39 mRNA in HeLa cells (Figure S7C), indicating that NMD of this RNA occurs more efficiently when it is associated with early acting SR-rich EJC.

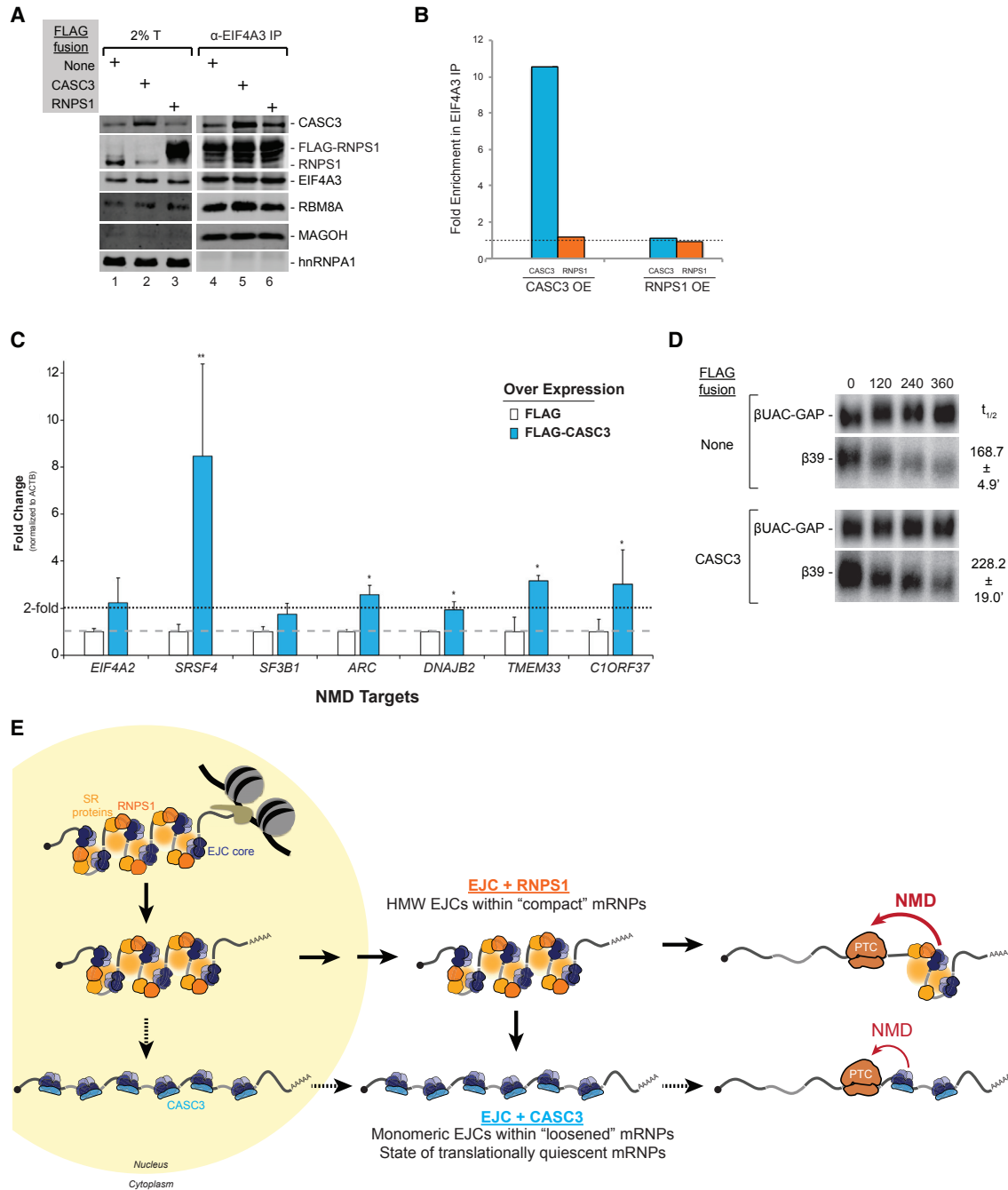
## DISCUSSION

The EJC is a cornerstone of all spliced mRNPs and interacts with >50 proteins to connect the bound RNA to a wide variety of post-transcriptional events. The EJC is thus widely presumed to be “dynamic.” By purifying EJC via key peripheral proteins, we demonstrate that a remarkable binary switch occurs in EJC's complement of bound proteins. Such an EJC composition

change has important implications for mRNP structure and function, including mRNA regulation via NMD (Figure 7E).

## EJC Composition and mRNP Structure

Our findings suggest that, when EJCs first assemble during co-transcriptional splicing, the core complex consisting of EIF4A3, RBM8A, and MAGOH engages with SR proteins and SR-like factors, including RNPS1 (Figure 2). Within these complexes, RNPS1 is likely bound to both the EJC core, as well as to the SR and SR-like proteins bound to their cognate binding sites on the RNA (Figure 3). This network of interactions bridges adjacent and distant stretches of mRNA, winding the mRNA up



**Figure 7. Promotion of Switch to Late-Acting CASC3-EJC Dampens NMD Activity**

(A) Western blots showing proteins on the right in total extract (T, lanes 1–3) or EIF4A3 IPs (lanes 4–6) from HEK293 cells overexpressing FLAG-fusion proteins at top left.

(B) A histogram showing the fold enrichment of alternate EJC factors in EIF4A3 IPs from (A). Overexpressed (OE) alternate EJC factors are at the bottom and CASC3 (blue) and RNPS1 (orange) levels in each OE sample (lane 5 or 6) were compared to the control IP in lane 4 (gray, set to 1).

(C) Fold change in levels of NMD-targeted endogenous transcripts (bottom) in HEK293 cells exogenously overexpressing FLAG or FLAG-CASC3. Average values from three biological replicates ± SEM are shown. \*p < 0.05 and \*\*p < 0.01 (Student's t test).

(D) Northern blots showing decay of Tetracycline-inducible β39 mRNA in HeLa Tet-off cells overexpressing FLAG-tagged proteins indicated on the left. Time after Tet-mediated transcriptional shut-off of β39 mRNA is indicated on the top. Right: β39 mRNA half-life (t<sub>1/2</sub>, average of three biological replicates ± standard deviation).

(E) A model depicting switch in EJC composition and its effect on mRNA structure and NMD activity. Key EJC proteins are indicated. Oval with radial orange fill in the high molecular weight (HMW) EJCs: unknown interactions that mediate EJC multimerization. Grey and black lines: exons.

See also Figure S7.



into a higher-order structure, which is characteristic of pre-translation RNPs purified from human cells via the EJC core or RNPS1 (Figure 2, (Singh et al., 2012; Metkar et al., 2018)). Such higher-order interactions are likely to be key in packaging spliced RNA into a compact RNP particle (Adivarahan et al., 2018). Presumably, the higher-order EJCs assemble via multiple weak interactions among low-complexity sequences (LCS) within EJC bound SR and SR-like proteins (Haynes and Iakoucheva, 2006; Kwon et al., 2014). RNPS1, which possesses an SR-rich LCS, could possibly act as a bridge between the EJC core and more distantly bound SR proteins. Our data also indicate that, at least on average-sized mRNPs, these SR-rich and RNPS1-containing higher-order EJCs persist through much of their nuclear lifetime (Figure 4). At some point before or during translation, the SR and SR-like proteins are evicted from all EJCs of an mRNP and the EJC is joined by CASC3 (Figures 4 and 5). It remains to be seen if CASC3's EJC incorporation drives EJC remodeling. Alternatively, active process(es) such as RNP modification via SR protein phosphorylation by cytoplasmic SR protein kinases (Zhou and Fu, 2013) or RNP remodeling by ATPases may precede CASC3 binding to EJC (Lee and Lykke-Andersen, 2013). What is clear is that CASC3-bound EJCs lose their higher-order structure and exist as monomeric complexes at the sites where EJC cores were co-transcriptionally deposited. Thus, the switch in EJC composition from RNPS1 and SR-rich complexes to CASC3-bound complexes changes the higher-order EJC, and possibly, mRNP structure (Figure 7E). The CASC3 bound form of the EJC is likely the main target of translation dependent disassembly, although RNPS1-EJC may also undergo similar disassembly (Figure 5).

### CASC3-EJC and Pre-translation mRNPs

Our findings support the emerging view that CASC3 is not an obligate component of all EJC cores. A population of assembled EJCs, especially those early in their lifetime, may completely lack CASC3 (Figures 1 and 2). Such a view of partial CASC3 dispensability for EJC structure and function is in agreement with findings from *Drosophila*, where the assembled trimeric EJC core as well as RNPS1 and its partner ACIN1 are required for splicing of long or sub-optimal introns, whereas CASC3 is not (Hayashi et al., 2014; Malone et al., 2014; Roignant and Treisman, 2010). Recent findings regarding EJC core protein functions during mouse embryonic brain development also support non-overlapping functions of CASC3 and the other core factors. While haploinsufficiency of EIF4A3, RBM8A, and MAGOH lead to premature neuronal differentiation and apoptosis in the mutant brains (Mao et al., 2016; McMahon et al., 2016), similar reduction in CASC3 levels (or even its near complete depletion) does not cause the same defects but leads to a more general developmental delay (Mao et al., 2017). We note that a view contrary to our findings is presented by the recently reported human spliceosome C\* structure, where CASC3 is seen bound to the trimeric EJC core (Zhang et al., 2017). As the spliceosomes described in these structural studies were assembled *in vitro* in nuclear extracts, it is possible that CASC3 present in extracts can enter pre-assembled spliceosomes and interact with EJC. Consistently, in the human spliceosome C\* structure, one of the two CASC3 binding

surfaces on EIF4A3 is exposed and available for CASC3 interaction. Still, it is possible that, at least on some RNAs or exon junctions, CASC3 assembly may occur soon after splicing within perispeckles (Daguenet et al., 2012).

What are the cellular functions of CASC3? Our data suggest that CASC3 is a prominent component of cytoplasmic EJCs within mRNPs that have not yet been translated or are undergoing their first round of translation (Figure 5). Previously described functions of CASC3 within translationally repressed neuronal transport granules (Macchi et al., 2003) and posterior-pole localized *oskar* mRNPs in *Drosophila* oocytes (van Eeden et al., 2001) further support CASC3 being a component of cytoplasmic pre-translation mRNPs. CASC3 plays an active role in *oskar* mRNA localization and translation repression (van Eeden et al., 2001), and a similar function is presumed in neuronal transport granules. These observations suggest that CASC3-EJCs may have a much more prominent role within longer-lived mRNPs that are transported to distant cytoplasmic locations. CASC3 is also known to activate translation of the bound RNA via EIF3 recruitment (Chazal et al., 2013). Although such a function appears to contradict with its association with translationally repressed mRNPs as discussed above, it is possible that, once localized mRNPs are relieved of repressive activity, EJC and CASC3 can promote translation activation.

### EJC Composition and NMD

Based on the binary EJC composition switch, EJC-dependent steps in the NMD pathway can be divided into at least two phases wherein the two alternate factors may perform distinct functions. The susceptibility of all tested NMD targets to RNPS1 levels suggests that this protein, and perhaps other components of SR-rich EJCs, serve a critical function in an early phase in the pathway. Such a function could be to recruit and/or activate other EJC/UPF factors either in the nucleus or even during premature translation termination as part of the downstream EJC. In a later phase of the EJC life cycle following the compositional switch, the EJC core maintains the ability to activate NMD as it can directly communicate with the NMD machinery via UPF3B (Buchwald et al., 2010). Still, following CASC3 incorporation into the EJC, its ability to stimulate NMD is reduced perhaps due to the loss of RNPS1 or SR proteins, which are known to enhance NMD (Figure S7) (Aznarez et al., 2018; Sato et al., 2008; Viegas et al., 2007; Zhang and Krainer, 2004). CASC3 overexpression can cause the compositional switch to occur at a faster rate or on a greater proportion of mRNAs, or both. Nevertheless, overexpressed NMD reporters (e.g.,  $\beta$ 39 mRNA) and some endogenous mRNAs depend on both early and late EJC compositions for their NMD. Notably, the  $\beta$ -globin NMD reporter was previously shown to undergo biphasic decay with faster turnover around the nuclear periphery and slower decay in more distant cytoplasmic regions (Trcek et al., 2013). More recently, single-molecule imaging of reporter RNPs showed that a fraction of their population diffuses for several minutes and micrometers away from the nucleus before undergoing the first round of translation (Halstead et al., 2015). It remains to be seen if mRNPs that are first translated in distant cytoplasmic locales, including those localized to specialized compartments such as neuronal dendrites and growth cones,

may rely more on a CASC3-dependent slower phase of NMD. The EJC compositional switch may also underlie the distinct NMD branches identified earlier via tethering of RNPS1 and CASC3 to reporter mRNAs (Gehring et al., 2005). Consistent with these observations, we find that RNPS1 co-purifies more strongly with UPF2, while CASC3 appears to interact more with UPF3B (Figure 2 and data not shown). The nature of the relationship between EJC composition and UPF2 and UPF3B-independent NMD branches is an important avenue for future work.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
  - Stable cell lines
  - Small-scale Immunoprecipitations
  - Mass spectrometry
  - Glycerol gradient fractionation
  - Manipulation of cellular protein levels for endogenous RNA analysis
  - Manipulation of cellular protein levels for  $\beta$ -globin reporter RNA analysis
  - Quantitative RT-PCR
  - RNA:protein immunoprecipitation in tandem (RIPit)
  - High-throughput sequencing library preparation
  - RNA-Seq data analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- **DATA AND SOFTWARE AVAILABILITY**

## SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and three tables and can be found with this article online at <https://doi.org/10.1016/j.celrep.2018.11.046>.

## ACKNOWLEDGMENTS

We thank Pearly Yan and the OSU Comprehensive Cancer Center genomics core for sequencing, Phillip Sharp and Paul Boutz for detained-intron data, Daniel Schoenberg for H1299 cell line Ribo-seq and RNA-seq data, Jens Lykke-Andersen for plasmids and antibodies, Akila Mayeda for anti-RNPS1 antibody, Can Cenik for advice, and Anita Hopper for critical comments. This work was supported in part by an allocation from the Ohio Supercomputer Center. Funding for this work was provided by the Ohio State University and National Institutes of Health (R01-GM120209) to G.S.

## AUTHOR CONTRIBUTIONS

Conceptualization, G.S. and R.B.; Investigation, J.W.M., L.A.W., R.D.P., Z.Y., M.J., and G.S.; Writing – Original Draft, G.S., J.W.M., and L.A.W.; Writing – Review & Editing, G.S., J.W.M., L.A.W., R.D.P., Z.Y., M.J., V.H.W., and R.B.; Funding Acquisition, G.S.; Resources, G.S., V.H.W., and R.B.; Supervision, G.S., V.H.W., and R.B.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 15, 2018

Revised: September 22, 2018

Accepted: November 12, 2018

Published: November 19, 2018

## REFERENCES

- Adivarahan, S., Livingston, N., Nicholson, B., Rahman, S., Wu, B., Rissland, O., and Zenklusen, D. (2018). Spatial organization of single mRNPs at different stages of the gene expression pathway. *Mol. Cell*. <https://doi.org/10.1101/237008>.
- Alexandrov, A., Shu, M.-D., and Steitz, J.A. (2017). Fluorescence Amplification Method for Forward Genetic Discovery of Factors in Human mRNA Degradation. *Mol. Cell* 65, 191–201.
- Andersen, C.B.F., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L.P., Pedersen, J.S., Séraphin, 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.
- Änkö, M.-L., Müller-McNicoll, M., Brandl, H., Curk, T., Gorup, C., Henry, I., Ule, J., and Neugebauer, K.M. (2012). The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.* 13, R17.
- Aznarez, I., Nomakuchi, T.T., Tetenbaum-Novatt, J., Rahman, M.A., Fregoso, O., Rees, H., and Krainer, A.R. (2018). Mechanism of Nonsense-Mediated mRNA Decay Stimulation by Splicing Factor SRSF1. *Cell Rep.* 23, 2186–2198.
- Baird, T.D., Cheng, K.C.-C., Chen, Y.-C., Buehler, E., Martin, S.E., Inglese, J., and Hogg, J.R. (2018). ICE1 promotes the link between splicing and nonsense-mediated mRNA decay. *eLife* 7, e33178.
- Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Séraphin, B., and Le Hir, H. (2005). The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat. Struct. Mol. Biol.* 12, 861–869.
- 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.
- Boutz, P.L., Bhutkar, A., and Sharp, P.A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.* 29, 63–80.
- Buchwald, G., Ebert, J., Basquin, C., Sauliere, J., Jayachandran, U., Bono, F., Le Hir, H., and Conti, E. (2010). Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. *Proc. Natl. Acad. Sci. USA* 107, 10050–10055.
- Chazal, P.-E., Dagueuet, E., Wendling, C., Ulyrick, N., Tomasetto, C., Sargueil, B., and Le Hir, H. (2013). EJC core component MLN51 interacts with eIF3 and activates translation. *Proc. Natl. Acad. Sci. USA* 110, 5903–5908.
- Chen, Q., Jagannathan, S., Reid, D.W., Zheng, T., and Nicchitta, C.V. (2011). Hierarchical regulation of mRNA partitioning between the cytoplasm and the endoplasmic reticulum of mammalian cells. *Mol. Biol. Cell* 22, 2646–2658.
- Dagueuet, E., Baguet, A., Degot, S., Schmidt, U., Alpy, F., Wendling, C., Spiegelhalter, C., Kessler, P., Rio, M.-C., Le Hir, H., et al. (2012). Perispeckles are major assembly sites for the exon junction core complex. *Mol. Biol. Cell* 23, 1765–1782.
- Degot, S., Régnier, C.H., Wendling, C., Chenard, M.-P., Rio, M.-C., and Tomasetto, C. (2002). Metastatic Lymph Node 51, a novel nucleo-cytoplasmic protein overexpressed in breast cancer. *Oncogene* 21, 4422–4434.
- Dostie, J., and Dreyfuss, G. (2002). Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr. Biol.* 12, 1060–1067.
- Duncan, C.D.S., and Mata, J. (2017). Effects of cycloheximide on the interpretation of ribosome profiling experiments in *Schizosaccharomyces pombe*. *Sci. Rep.* 7, 10331.
- Gangras, P., Dayeh, D.M., Mabin, J.W., Nakanishi, K., and Singh, G. (2018). Cloning and Identification of Recombinant Argonaute-Bound Small RNAs Using Next-Generation Sequencing. *Methods Mol. Biol.* 1680, 1–28.

- Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and Kulozik, A.E. (2005). Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol. Cell* 20, 65–75.
- Gehring, N.H., Lamprinak, S., Hentze, M.W., and Kulozik, A.E. (2009a). The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. *PLoS Biol.* 7, e1000120.
- Gehring, N.H., Lamprinak, S., Kulozik, A.E., and Hentze, M.W. (2009b). Disassembly of exon junction complexes by PYM. *Cell* 137, 536–548.
- Geyer, P.K., Meyuhas, O., Perry, R.P., and Johnson, L.F. (1982). Regulation of ribosomal protein mRNA content and translation in growth-stimulated mouse fibroblasts. *Mol. Cell. Biol.* 2, 685–693.
- Halstead, J.M., Lionnet, T., Wilbertz, J.H., Wippich, F., Ephrussi, A., Singer, R.H., and Chao, J.A. (2015). Translation. An RNA biosensor for imaging the first round of translation from single cells to living animals. *Science* 347, 1367–1671.
- Hauer, C., Sieber, J., Schwarzl, T., Hollerer, I., Curk, T., Alleaume, A.-M., Hentze, M.W., and Kulozik, A.E. (2016). Exon Junction Complexes Show a Distributional Bias toward Alternatively Spliced mRNAs and against mRNAs Coding for Ribosomal Proteins. *Cell Rep.* 16, 1588–1603.
- Hayashi, R., Handler, D., Ish-Horowitz, D., and Brennecke, J. (2014). The exon junction complex is required for definition and excision of neighboring introns in *Drosophila*. *Genes Dev.* 28, 1772–1785.
- Haynes, C., and Iakoucheva, L.M. (2006). Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res.* 34, 305–312.
- Hentze, M.W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341.
- Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.
- Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802.
- Jan, C.H., Williams, C.C., and Weissman, J.S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* 346, 1257521.
- Kiss, D.L., Baez, W., Huebner, K., Bundschuh, R., and Schoenberg, D.R. (2017). Impact of FHIT loss on the translation of cancer-associated mRNAs. *Mol. Cancer* 16, 179.
- Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S.L. (2014). Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* 345, 1139–1145.
- Le Hir, H., Saulière, J., and Wang, Z. (2016). The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.* 17, 41–54.
- Lee, S.R., and Lykke-Andersen, J. (2013). Emerging roles for ribonucleoprotein modification and remodeling in controlling RNA fate. *Trends Cell Biol.* 23, 504–510.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. (2001). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* 293, 1836–1839.
- Macchi, P., Kroening, S., Palacios, I.M., Baldassa, S., Grunewald, B., Ambrosino, C., Goetze, B., Lupas, A., St Johnston, D., and Kiebler, M. (2003). Barentsz, a new component of the Staufen-containing ribonucleoprotein particles in mammalian cells, interacts with Staufen in an RNA-dependent manner. *J. Neurosci.* 23, 5778–5788.
- Malone, C.D., Mestdagh, C., Akhtar, J., Kreim, N., Deinhard, P., Sachidanandan, R., Treisman, J., and Roignant, J.-Y. (2014). The exon junction complex controls transposable element activity by ensuring faithful splicing of the piwi transcript. *Genes Dev.* 28, 1786–1799.
- Mao, H., McMahon, J.J., Tsai, Y.-H., Wang, Z., and Silver, D.L. (2016). Haploinsufficiency for Core Exon Junction Complex Components Disrupts Embryonic Neurogenesis and Causes p53-Mediated Microcephaly. *PLoS Genet.* 12, e1006282.
- Mao, H., Brown, H.E., and Silver, D.L. (2017). Mouse models of Casc3 reveal developmental functions distinct from other components of the exon junction complex. *RNA* 23, 23–31.
- McMahon, J.J., Miller, E.E., and Silver, D.L. (2016). The exon junction complex in neural development and neurodevelopmental disease. *Int. J. Dev. Neurosci.* 55, 117–123.
- Metkar, M., Ozadam, H., Lajoie, B.R., Imakaev, M., Mirny, L.A., Dekker, J., and Moore, M.J. (2018). Higher-Order Organization Principles of Pre-translational mRNPs. *Mol. Cell.* <https://doi.org/10.1101/278747>.
- Meyuhas, O., and Kahan, T. (2015). The race to decipher the top secrets of TOP mRNAs. *Biochim. Biophys. Acta* 1849, 801–811.
- Murachelli, A.G., Ebert, J., Basquin, C., Le Hir, H., and Conti, E. (2012). The structure of the ASAP core complex reveals the existence of a Pinin-containing PSAP complex. *Nat. Struct. Mol. Biol.* 19, 378–386.
- Neve, J., Burger, K., Li, W., Hoque, M., Patel, R., Tian, B., Gullerova, M., and Furger, A. (2016). Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. *Genome Res.* 26, 24–35.
- Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G.W., Ares, M., Jr., and Fu, X.-D. (2013). Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol. Cell* 50, 223–235.
- Patursky-Polischuk, I., Stolovich-Rain, M., Hausner-Hanochi, M., Kasir, J., Cybulski, N., Avruch, J., Rüegg, M.A., Hall, M.N., and Meyuhas, O. (2009). The TSC-mTOR pathway mediates translational activation of TOP mRNAs by insulin largely in a raptor- or rictor-independent manner. *Mol. Cell. Biol.* 29, 640–649.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Roignant, J.-Y., and Treisman, J.E. (2010). Exon junction complex subunits are required to splice *Drosophila* MAP kinase, a large heterochromatic gene. *Cell* 143, 238–250.
- Sato, H., Hosoda, N., and Maquat, L.E. (2008). Efficiency of the pioneer round of translation affects the cellular site of nonsense-mediated mRNA decay. *Mol. Cell* 29, 255–262.
- Saulière, J., Murigneux, V., Wang, Z., Marquet, E., Barbosa, I., Le Tonquèze, O., Audic, Y., Paillard, L., Roest Crolius, H., and Le Hir, H. (2012). CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat. Struct. Mol. Biol.* 19, 1124–1131.
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342.
- Singh, G., Jakob, S., Kleedehn, M.G., and Lykke-Andersen, J. (2007). Communication with the exon-junction complex and activation of nonsense-mediated decay by human Upf proteins occur in the cytoplasm. *Mol. Cell* 27, 780–792.
- Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* 151, 750–764.
- Singh, G., Ricci, E.P., and Moore, M.J. (2014). RIPIT-Seq: a high-throughput approach for footprinting RNA:protein complexes. *Methods* 65, 320–332.
- Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu. Rev. Biochem.* 84, 325–354.
- Tacke, R., and Manley, J.L. (1995). The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* 14, 3540–3551.

- Tacke, R., Tohyama, M., Ogawa, S., and Manley, J.L. (1998). Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* 93, 139–148.
- Tange, T.Ø., Shibuya, T., Jurica, M.S., and Moore, M.J. (2005). Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* 11, 1869–1883.
- Tani, H., Mizutani, R., Salam, K.A., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., and Akimitsu, N. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* 22, 947–956.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
- Trcek, T., Sato, H., Singer, R.H., and Maquat, L.E. (2013). Temporal and spatial characterization of nonsense-mediated mRNA decay. *Genes Dev.* 27, 541–551.
- van Eeden, F.J., Palacios, I.M., Petronczki, M., Weston, M.J., and St Johnston, D. (2001). Barentsz is essential for the posterior localization of oskar mRNA and colocalizes with it to the posterior pole. *J. Cell Biol.* 154, 511–523.
- Viegas, M.H., Gehring, N.H., Breit, S., Hentze, M.W., and Kulozik, A.E. (2007). The abundance of RNPS1, a protein component of the exon junction complex, can determine the variability in efficiency of the Nonsense Mediated Decay pathway. *Nucleic Acids Res.* 35, 4542–4551.
- Walter, P., and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. *J. Cell Biol.* 91, 557–561.
- Woodward, L.A., Mabin, J.W., Gangras, P., and Singh, G. (2017). The exon junction complex: a lifelong guardian of mRNA fate. *Wiley Interdiscip. Rev. RNA* 8, e1411.
- Zhang, Z., and Krainer, A.R. (2004). Involvement of SR proteins in mRNA surveillance. *Mol. Cell* 16, 597–607.
- Zhang, Y., Wen, Z., Washburn, M.P., and Florens, L. (2010). Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal. Chem.* 82, 2272–2281.
- Zhang, X., Yan, C., Hang, J., Finci, L.I., Lei, J., and Shi, Y. (2017). An Atomic Structure of the Human Spliceosome. *Cell* 169, 918–929.e14.
- Zhong, X.-Y., Wang, P., Han, J., Rosenfeld, M.G., and Fu, X.-D. (2009). SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol. Cell* 35, 1–10.
- Zhou, Z., and Fu, X.-D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* 122, 191–207.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
CASC3	Bethyl Laboratories	A302-472; RRID: AB_1944210
RNPS1	Sigma Aldrich	HPA044014; RRID: AB_10966286
RNPS1	Akila Mayeda	NA
EIF4A3	Bethyl Laboratories	A302-980A; RRID: AB_10749033
HNRNPA1	Santa Cruz Biotechnology	SC-32301; RRID: AB_627729
ACIN1	Bethyl Laboratories	A300-999A; RRID: AB_2223628
SAP18	Santa Cruz Biotechnology	SC-25377; RRID: AB_2183526
RBM8A	Sigma Aldrich	Y1253 (4C4) ; RRID: AB_477643
MAGOH	Santa Cruz Biotechnology	SC-271365; RRID: AB_10611651
SRSF1	Santa Cruz Biotechnology	SC-33652; RRID: AB_628248
FLAG M2	Sigma Aldrich	F1804; RRID: AB_262044
EIF4E	Bethyl Laboratories	A301-154A; RRID: AB_2097708
CBP80	Bethyl Laboratories	A301-793A; RRID: AB_1211224
SRP68	Bethyl Laboratories	A303-955A; RRID: AB_2620304
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
Protein A Dynabeads	Life Technologies	10002D
Protein G Dynabeads	Life Technologies	10003D
Anti FLAG M2-Agarose Beads	Sigma Aldrich	A2220
FLAG peptide	Sigma Aldrich	F3290
MWCO 7,000 Da Dialysis Column	Pierce	88242
4X Lammelli SDS load buffer	Bio-Rad	161-0737
4%–15% Mini-PROTEAN TGX	Bio-Rad	4561086
Trypsin Gold	Promega	V5280
ProteaseMAX™ Surfactant	Promega	V2071
TransIT-X2 transfection reagent	Mirus	MIR 6003
RNAiMAX transfection reagent	Thermo Fisher	13778030
JetPrime transfection reagent	PolyPlus	114-01
oligo-(dT)15 Primer	Promega	C1101
Superscript III	Invitrogen	18080044
Ribonuclease H	Promega	M4281
2X SYBR Green Master Mix	Applied Biosystems	4309155
RNase I	Promega	M4261
Ribo-Zero rRNA Removal Kit (H/M/R)	Illumina	MRZH116
CircLigase	Lucigen	CL4115K
<b>Deposited Data</b>		
RNA-seq and RIPiT-seq data	This Study	GEO: GSE115977
<b>Experimental Models: Cell Lines</b>		
Human: HEK293 Flp-In TRex	Thermo Fisher Scientific	R78007; RRID:CVCL_U427
Human: HeLa CCL2	ATCC	CCL-2; RRID:CVCL_0030
Human: HeLa Tet-off	Takara	631156 RRID:CVCL_V352
Mouse: P19	ATCC	CRL-1825; RRID:CVCL_2153
<b>Oligonucleotides</b>		
siRNA: AllStars Negative Control	QIAGEN	SI03650318
siRNA: EIF4A3_187: CGA GCA AUC AAG CAG AUC AUU	This Study	N/A

(Continued on next page)



**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
siRNA: CASC3_534: CCA GCA UAC AUA CCU CGG AUU	This Study	N/A
siRNA: RNPS1_2: GCA UCC AGC CGC UCA GGA AUU	This Study	N/A
siRNA: UPF1_1879: AAG AUG CAG UUC CGC UCC AUU	This Study	N/A
Primers for RT-PCR, see <a href="#">Table S3</a>	This study	N/A
Primer: RT primer for HTS library prep (5 nt barcode is denoted by XXXXX): 5'-GGXXXXXXXXNNNAGATCGGAAGAGCGTCGTG TAGGGAAAGAGTGT-SPACER18-CTCGGCATTCTGCTGAA CCGCTCTTCGATCTCCTTGGCACCCGAGAATTCCA-3'	<a href="#">Gangras et al., 2018</a>	N/A
<b>Recombinant DNA</b>		
Plasmid: pc $\beta$ wt $\beta$	<a href="#">Lykke-Andersen et al., 2001</a>	N/A
Plasmid: pc $\beta$ 39 $\beta$	<a href="#">Lykke-Andersen et al., 2001</a>	N/A
Plasmid: pcTET2- $\beta$ wt $\beta$	<a href="#">Singh et al., 2007</a>	N/A
Plasmid: pcTET2- $\beta$ 39 $\beta$	<a href="#">Singh et al., 2007</a>	N/A
Plasmid: pc $\beta$ wtGAP3UAC	<a href="#">Lykke-Andersen et al., 2001</a>	N/A
Plasmid: pcDNA3	<a href="#">Singh et al., 2012</a>	N/A
Plasmid: pcDNA3 YFP	This study	N/A
Plasmid: pcDNA3 FLAG	<a href="#">Singh et al., 2012</a>	N/A
Plasmid: pcDNA3 FLAG-EIF4A3	<a href="#">Singh et al., 2012</a>	N/A
Plasmid: pcDNA3 FLAG-EIF4A3 YRAA	This study	N/A
Plasmid: pcDNA3 FLAG-MAGOH	<a href="#">Singh et al., 2012</a>	N/A
Plasmid: pcDNA3 FLAG-RNPS1	This study	N/A
Plasmid: pcDNA3 FLAG-CASC3	This study	N/A
Plasmid: pcDNA3 FLAG-HNRNPA1	This study	N/A
Plasmid: pcDNA5 FLAG-MAGOH	<a href="#">Singh et al., 2012</a>	N/A
Plasmid: pcDNA5 FLAG-RNPS1	This study	N/A
Plasmid: pcDNA5 FLAG-CASC3	This study	N/A
Plasmid: pcDNA5-FRT/TO	Thermo Fisher Scientific	V652020
<b>Software and Algorithms</b>		
Tophat v2.1.1	<a href="#">Trapnell et al., 2009</a>	<a href="https://tophat2-and-bowtie-compatibility.readthedocs.io/en/latest/">https://tophat2-and-bowtie-compatibility.readthedocs.io/en/latest/</a>
Bedtools	<a href="#">Quinlan and Hall, 2010</a>	<a href="https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html">https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html</a>
RNA Binding Protein Database		<a href="http://rbpdb.ccb.utoronto.ca/">http://rbpdb.ccb.utoronto.ca/</a>
DESeq2	<a href="#">Love et al., 2014</a>	<a href="http://bioconductor.org/packages/release/bioc/html/DESeq2.html">http://bioconductor.org/packages/release/bioc/html/DESeq2.html</a>
DAVID gene ontology tool	<a href="#">Huang et al., 2009</a>	<a href="https://david.ncicrf.gov">https://david.ncicrf.gov</a>
ImageQuant	GE Healthcare Life Sciences	<a href="https://www.gelifesciences.com/en/ch/shop/protein-analysis/molecular-imaging-for-proteins/imaging-software/imagequant-tl-8-1-p-00110">https://www.gelifesciences.com/en/ch/shop/protein-analysis/molecular-imaging-for-proteins/imaging-software/imagequant-tl-8-1-p-00110</a>
ImageStudio	LI-COR	<a href="https://www.licor.com/bio/products/software/image_studio/">https://www.licor.com/bio/products/software/image_studio/</a>

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Guramrit Singh ([singh.734@osu.edu](mailto:singh.734@osu.edu)).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

The following cell lines were used: HEK293 Flp-In TRex (human cells; sex: female; RRID:CVCL\_U427), HeLa CCL2 (human cells; sex: female; RRID:CVCL\_0030), and HeLa Tet-off (human cells; sex: female; RRID:CVCL\_V352), mouse P19 (mouse cells; sex: male; RRID:CVCL\_2153). The human cell lines were cultured at 37°C under 5% carbon dioxide in a humidified chamber in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. The mouse P19 cells were cultured at 37°C under 5% carbon dioxide in a humidified chamber in DMEM supplemented with 5% FBS and 1% penicillin-streptomycin.

## METHOD DETAILS

### Stable cell lines

Stable cell lines expressing tetracycline-inducible FLAG-tagged proteins were created using HEK293 Flp-In TRex cells as described previously (Singh et al., 2012). Briefly,  $\sim 1 \times 10^6$  HEK293 Flp-In TRex cells were seeded on 3.5-cm plates for 16 hr. A plasmid mix (0.2  $\mu$ g of a pcDNA5-FRT/TO-FLAG construct along with 1.8  $\mu$ g of pOG44) was transfected following the TransIT-X2 procedure (Mirus) following manufacturer's instructions. After 24hr, cells were split at 1:10 dilution into 10-cm dishes. After overnight incubation Blasticidin (15  $\mu$ g/ml) and Hygromycin (100  $\mu$ g/ml) containing media was added to cells to select for stably transfected cells. Once individual transfected cells had grown into colonies visible to the naked eye, the clonal pool of stably-transfected cells was harvested. A titration of tetracycline (Tet; 0-625 ng/ml) was used to determine a concentration where exogenous FLAG-tagged protein levels were comparable to its endogenous counterpart.

### Small-scale Immunoprecipitations

For immunoprecipitation of endogenous proteins from HEK293, HeLa and P19 cells, cultured cells were lysed and sonicated in hypotonic lysis buffer (HLB) [20 mM Tris-HCl pH 7.5, 15 mM NaCl, 10 mM EDTA, 0.5% NP-40, 0.1% Triton X-100, 1  $\times$  Sigma protease inhibitor cocktail, 1 mM PMSF]. Lysates were sonicated using a microtip for 6-18 s, [NaCl] was increased to 150 mM and RNase A was added to 125  $\mu$ g/ml. Following five minute incubation on ice, cell lysates were cleared at 15,000  $\times$  g. Complexes were then captured on Protein A/G Dynabeads (Life Technologies) conjugated to IgG,  $\alpha$ -EIF4A3,  $\alpha$ -CASC3 or  $\alpha$ -RNPS1 antibodies for 2 hr at 4°C. Complexes were washed in isotonic wash buffer (IsoWB) [20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% NP-40] and eluted in clear sample buffer [100 mM Tris-HCl pH 6.8, 4% SDS, 10 mM EDTA, 100 mM DTT]. The eluted proteins were separated via SDS-PAGE and analyzed by western blotting. Cortical neurons were isolated from FBV wild-type male mice in HBSS buffer. Once isolated, lysate preparation and IP was carried out as above.

For immunoprecipitation of FLAG-tagged proteins, stable HEK293 cells expressing FLAG-tagged EJC protein were lysed in HLB and prepared for IP in the same way as the endogenous protein IPs above. Lysates were then incubated with anti-FLAG M2-agarose beads (Sigma) for 1-2 hr at 4°C, washed 8  $\times$  with 1 mL IsoWB and eluted in IsoWB supplemented with 125  $\mu$ g/ml FLAG-peptide with gentle shaking for 2 hr at 4°C. Total cell extracts from mammalian cells prepared in an isotonic buffer supplemented with 125  $\mu$ g/ml RNase A were incubated at 4°C for 2 hr with anti-FLAG M2-agarose beads. Solid phase captured RNA:protein complexes were washed multiple times, and eluted by FLAG peptide affinity elution. Eluates were analyzed by SDS-PAGE followed by western blotting.

For IPs following alternate EJC protein overexpression, HEK293 cells stably expressing FLAG-tagged CASC3 or RNPS1, or FLAG-peptide as a control, were seeded in 10-cm plates and induced with 625 ng/ml of Doxycycline for 48 hr. Endogenous RBM8A or EIF4A3 IPs were carried out as described above.

### Mass spectrometry

#### FLAG Immunoprecipitation

HEK293 Flp-In TRex cells stably expressing FLAG-MAGOH, FLAG-CASC3, FLAG-RNPS1, or FLAG-peptide were lysed, RNase A digested and IPed at 4°C for 2 hr as above. The lysis buffer was supplemented with 1  $\mu$ g/ml FLAG peptide to improve IP specificity. Following FLAG-IP, the elution was completely dried by vacuum evaporation, re-suspended in 100  $\mu$ L of water and dialyzed (dialysis buffer: 10 mM Tris-HCl pH7.5, 75 mM NaCl, 0.01% Triton X-100) for  $\sim$ 6 hr at 4°C in a MWCO 7,000 Da mini dialysis column (Pierce). The dialyzed sample (90-100  $\mu$ L) was again completely dried by vacuum evaporation and re-suspended in 20  $\mu$ L of 0.1% SDS and 10 mM DTT (prepared from a fresh stock solution). The sample was heated at 95°C for 5 min and cooled to room temperature. The reduced thiol groups were alkylated by incubating with 0.8  $\mu$ L of freshly prepared 1M iodoacetamide at room temperature for 45 min in dark. The resulting samples were mixed with 5  $\mu$ L of 4  $\times$  Lammelli SDS load buffer (Bio-Rad) and loaded on 4%-15% Mini-PROTEAN TGX gel (Bio-Rad). The samples were migrated until the dye-front had run approximately 1 cm into the gel from the bottom of the well. The gel was washed three times with  $\sim$ 200 mL HPLC grade water for 5 min each. The gel piece containing protein was excised and processed for in gel digestion of proteins.

#### In-Gel Digestion

The gel slices were cut into 1  $\times$  5  $\times$  1mm<sup>3</sup> dimension and transferred into a 1.5 mL microcentrifuge tube. Samples were washed with water and dehydrated in 1:1 (v/v) mix of acetonitrile and 50 mM NH<sub>4</sub>HCO<sub>3</sub> for 5 minutes and 100% acetonitrile for 30 s. Following

drying with a vacuum concentrator (Savant), gel slices were rehydrated in 25 mM dithiothreitol in 50 mM  $\text{NH}_4\text{HCO}_3$  and incubated for 20 min at 56°C. With the removal of the supernatant, 55 mM iodoacetamide in 50 mM  $\text{NH}_4\text{HCO}_3$  was added and the samples were incubated in the dark for 20 min at room temperature. Gel slices were washed and dehydrated as described before. Following removal of the liquid by vacuum concentrator, the sample was rehydrated in 12 ng/ $\mu\text{L}$  Trypsin Gold (Promega) in 0.01% Protease-MAX<sup>TM</sup> Surfactant (Promega): 50 mM  $\text{NH}_4\text{HCO}_3$  and incubated at 50°C for one hr. Condensate was collected by centrifuging at 14,000  $\times$  g for 10 s and the solution was transferred to a new tube. Trifluoroacetic acid was added to a final concentration of 0.5% to inactivate trypsin and the solution was dried by vacuum concentrator.

### LC/MS/MS

The LC-MS/MS tryptic digested peptides were dissolved in 0.1% formic acid (v/v) and loaded at 15  $\mu\text{L}/\text{min}$  for 10 min using a nano-ACQUITY C18 Trap column (20  $\times$  0.18 mm i.d., 5  $\mu\text{m}$  100Å C18, Waters, 186007238). Peptides were separated using an EASY-Spray LC Column (150  $\times$  0.075 mm i.d., 3  $\mu\text{m}$  100Å C18, ThermoFisher, ES800) with a gradient using solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in acetonitrile) at a 0.5  $\mu\text{L}/\text{min}$  flow rate. The gradient started with 1% B for 5 min, followed by 85 min with a linear increase to 35% B. The gradient was increased to 85% B in 5 min followed by decreasing back to 1% B in 5 min as the final wash step. A nanoACQUITY UPLC (Waters) was coupled to a Velos Pro Dual-Pressure Linear Ion Trap mass spectrometer (Thermo Scientific) for data acquisition. A data-dependent acquisition routine was used for a mass spectrum from  $m/z$  300 to 2000 and followed by ten tandem mass spectrometry scans.

### Data Analysis

Raw data files were processed using Proteome Discoverer (ThermoFisher). The data were searched against the human Swiss-Prot index (09/01/15) using Sequest HT with precursor mass tolerances of 1.5 Da and fragment mass tolerances of 0.8 Da. Maximum missed cleavage sites of full tryptic digestion was two and dynamic modifications of acetylation (N terminus), carbamidomethylation (cysteine), propionamidation (cysteine) and oxidation(methionine) were considered. The processed mass spectrometry data was analyzed using the Scaffold software (4.4.5). Quantifications from replicate 1 are presented in main figures and from replicate 2 in supplementary figures. For the analysis presented in Figure 1 and S1, spectral quantification was done using the normalized weighted spectral counts with minimum peptide identification threshold at 95% and protein identification threshold of 95% with 1 minimum peptide. For analysis presented in Figure 2 and S2, spectral quantification was done using the Normalized Spectral Abundance Factor (NSAF) (Zhang et al., 2010). A total of 341 (replicate 1) or 259 (replicate 2) proteins were identified at 99% threshold each with a minimum of 2 peptides. When Scaffold grouped together similar proteins into clusters, only those proteins were retained from a cluster that had non-zero NSAF value in at least one of the four samples. There are several proteins that are detected only in one or the other alternate EJC, and thus had zero values in samples where they were undetected. Also, many EJC specific proteins had zero values in FLAG-only control IPs. A pseudocount of 0.000001 was added to all NSAF values to prevent loss of such proteins when dividing by zero. Fold-enrichment over FLAG-only control was calculated for each of the three EJC samples as follows:

$$\text{Fold enrichment} = \log_2[\text{NSAF}(\text{FLAG} - \text{EJC} + 0.000001)] / [\text{NSAF}(\text{FLAG} - \text{only} + 0.000001)]$$

The comparison of fold-enrichment values for all proteins in the EJC core or alternate EJCs was carried out in the R package ggplot2 (via scatterplots). Proteins that were > 10-fold enriched in one of the three EJC IPs were included in the heatmap (using R package gplots). In this analysis, proteins that most likely were contaminants (e.g., cytoskeletal proteins, histones, metabolic enzymes) were not included in the heatmap but are shown in Table S1.

### Glycerol gradient fractionation

Five 15-cm plates with ~90% confluent HEK293 Flp-In TRex cells expressing FLAG-tagged MAGOH, CASC3, RNPS1 proteins, or FLAG-peptide as a control, were cultured and induced as above. Cell lysis, FLAG-immunoprecipitation, RNase A digestion and FLAG-elution steps were also carried out as described above. ~500  $\mu\text{L}$  FLAG-IP elution was layered onto pre-cooled continuous 10%–30% glycerol gradients prepared in 11 mL Beckman centrifuge tubes. Gradients were run at 32,000 rpm for 16 hours at 4°C. Gradients were fractionated by-hand into 500  $\mu\text{L}$  fractions. Proteins were precipitated using TCA and resuspended in 15  $\mu\text{L}$  of 1  $\times$  SDS loading buffer for analysis on 12% SDS-PAGE followed by western blotting.

### Manipulation of cellular protein levels for endogenous RNA analysis

#### siRNA knockdowns

HEK293 Flp-In TRex cells were seeded into 24-well plates. Cells were transfected 4 hours later following the Mirus TransIT-X2 procedure with 55–60 pmols of control, EIF4A3, CASC3, RNPS1, or UPF1 siRNA. Cells were harvested 48 hr later. Knockdown was checked by western blot and subsequently purified RNA was used for qRT-PCR analysis. For 96 hr knockdowns, HEK293 cells were transfected as above and were incubated for 24 hr. Cells were transfected again at roughly 70%–80% confluence following the Mirus TransIT-X2 procedure using 50–60 pmol of control, EIF4A3, CASC3, RNPS1, or UPF1 siRNA. Cells were harvested ~48 hr (~96 total hours after first transfection) later. All knockdowns were confirmed by western blotting.

For EIF4A3- or EIF4A3 YRAA-mediated rescue experiments, HEK293 cells seeded into 24-well plates were transfected 4 hr later following the Mirus TransIT-X2 procedure with 60 pmol of control, or EIF4A3 siRNA, and 300 ng of pcDNA3 empty, FLAG-EIF4A3 wt or FLAG-EIF4A3 YRAA and 200 ng of pcDNA3 vector. Cells were harvested ~48 hr later. Knockdown and EIF4A3 expression was

checked by western blot and subsequently purified RNA was used for qRT-PCR analysis. For FLAG IPs, cells transfected as above were harvested in 1 × PBS 48 hr after transfection, pelleted at 1000 × g for 10 min and used for FLAG IP as described above.

#### **FLAG-EJC protein overexpression**

HEK293 cells expressing FLAG-tagged CASC3, RNPS1 proteins, or FLAG-peptide as a control, were seeded into 6-well plates. Cells were induced with 625 ng/ml of Doxycycline for 48 hr. Overexpression was checked by western blot and subsequently purified RNA was used for qRT-PCR analysis.

#### **Manipulation of cellular protein levels for $\beta$ -globin reporter RNA analysis**

##### **siRNA knockdowns**

HeLa Tet-off cells in 12-well plates ( $1.2 \times 10^5$  cells/well) were reverse transfected with 15 pmol of control, UPF1, CASC3 and RNPS1 siRNA using RNAiMAX following the manufacturer's protocol. After 24 hr, cells were co-transfected with plasmids (100 ng of pcTET2- $\beta$ wt $\beta$  or pcTET2- $\beta$ 39 $\beta$ , 30 ng of pc $\beta$ wtGAP3UAC and 70 ng of pezYFP or carrier DNA) and a second dose of 15 pmol of siRNA using JetPrime (PolyPlus). Tetracycline (50 ng/ml) was included at the time of transfection to repress reporter RNA expression, and was removed 20 hr later to induce reporter RNA expression for 6–8 hr. Cells were harvested in clear sample buffer for western blot analysis and TRIzol RNA extraction. RNA levels were analyzed by Northern Blotting using an *in vitro* transcribed body labeled  $\beta$ -globin anti-sense RNA probe.

##### **FLAG-EJC protein overexpression**

For pulse-chase experiments, HeLa Tet-off cells growing in 12-well plates were transfected with 10 ng of pc $\beta$ wtGAP3UAC, 200 ng of pcTET2- $\beta$ 39 $\beta$  and 250 ng of pezFLAG-CASC3 (or pezFLAG empty vector) in the presence of 50 ng/ml tetracycline. 36 hr later, expression of  $\beta$ 39 mRNA was induced for 6 hr by removing Tet, and then suppressed again via addition of 1  $\mu$ g/ml Tet. Time points were collected starting ~30 min after addition of Tet for 0 hr time point. Total RNA was extracted using TRIzol and one-half of the RNA sample was analyzed by Northern blots.

For steady-state assays, HeLa CCL2 cells growing in 12-well plates were transfected with 100 ng of pc $\beta$ wtGAP3UAC, 100 ng of either pc $\beta$ wt $\beta$  or pc $\beta$ 39 $\beta$ , and 150 ng of pezFLAG plasmids (expressing RNPS1 or CASC3, or empty vector as a control). Cells were harvested 48 hr post-transfections and total RNA extracted was analyzed by Northern blots.

#### **Quantitative RT-PCR**

RNA was isolated from cells using Trizol, DNase treated, purified with Phenol:Chloroform:Isoamyl alcohol (25:24:1, pH 4.5) and re-suspended in RNase-free water. RNA was reverse transcribed using oligo-dT (Promega) and Superscript III (Invitrogen). After reverse transcription of RNA the samples were treated with RNase H (Promega) for 30 min at 37°C. The samples were then diluted to 5 ng/ $\mu$ L before proceeding to qPCR setup. For each qPCR 25 ng of cDNA was mixed with 7.5  $\mu$ L of 2 × SYBR Green Master Mix (ABS), 0.6  $\mu$ L of a 10 mM forward and reverse primer mix (defrosted once) in a 15  $\mu$ L reaction. The qPCRs were performed in triplicates (technical) using primers described in Table S3. Parallel with NMD targets,  $\beta$ -actin was used as internal control except under following condition. We found that in RNPS1 knockdown conditions, there was ~2-fold upregulation of  $\beta$ -actin mRNA levels. Therefore, in experiments that included RNPS1 knockdown, we used TATA-binding protein (TBP) as control. Fold-change calculations were performed by the delta-delta Ct method. Fold-changes from at least three biological replicates were used to determine the standard error of means. The p values were calculated using Student's t test.

#### **RNA:protein immunoprecipitation in tandem (RIPiT)**

RIPiTs were carried out with and without formaldehyde crosslinking following a detailed protocol available in (Singh et al., 2014). Briefly, for native RIPiTs, total extracts from four 15-cm plates were prepared by disrupting cells via sonication using Branson Digital Sonifier-250 at 30% amplitude using a Microtip for a total of 16 s (in 2 s bursts with 10 s intervals) in hypotonic lysis buffer [HLB: 20 mM Tris-HCl pH 7.5, 15 mM NaCl, 10 mM EDTA, 0.5% NP-40, 0.1% Triton X-100, 10  $\mu$ g/ml Aprotinin (Sigma, A1153-10MG), 1  $\mu$ g/ml Leupeptin (Sigma, L9783-5MG), 1  $\mu$ M Pepstatin (Sigma, L4265-5MG), 1 mM PMSF (Sigma, P7626-5G)] supplemented with 150mM NaCl. Extracts were cleared by centrifugation at 15,000 × g for 10 minutes at 4°C, and used as input into FLAG-IP with 1 mL of FLAG-agarose resin. For formaldehyde crosslinked RIPiTs, total extracts were prepared from six 15-cm plates in denaturing lysis buffer [HLB supplemented with 0.1% SDS and 0.1% sodium deoxycholate] supplemented with 150 mM NaCl-containing for input into FLAG-IP. Following IP for 2 hr, beads were washed 4 times in 10 mL IsoWB [20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% NP-40]. In the case of formaldehyde crosslinked RIPiTs, the first two washes were performed with IsoWB supplemented with 0.1% SDS and 0.1% sodium deoxycholate. FLAG-agarose beads were then incubated with 0.5 mL RNase I dilution (0.006 U/ml in Isotonic wash buffer (IsoWB)) at 4°C for 10 min. Beads were washed again 4 times in 1 mL IsoWB. FLAG-epitope containing complexes were affinity eluted from the beads in one bed volume of IsoWB containing 250  $\mu$ g/ml FLAG peptide with gentle shaking at 4°C for 2 hr. The recovered elution volume was adjusted to 400  $\mu$ L and its composition to that of the lysis buffer above with NaCl at 150 mM, and used for input into the second IP. The second IP was performed for 2 hr at 4°C using the following antibodies conjugated to protein-A Dynabeads: anti-EIF4A3 (Bethyl A302-980A, 10  $\mu$ g/RIPiT), anti-CASC3 (Bethyl A302-472A, 2  $\mu$ g/RIPiT), anti-RNPS1 (HPA044014-100UL, 2  $\mu$ g/RIPiT). RIPiTs were eluted in clear sample buffer and divided into two parts for RNA and protein analysis.

RIPiTs to enrich EJC footprints upon cycloheximide (CHX) treatment were carried out as above except that cells were incubated with 100  $\mu$ g/ml CHX for 3 hr prior to harvesting. CHX was included at the same concentration in PBS (for washes before lysis) and cell lysis buffers.

### High-throughput sequencing library preparation

For RIPiT-seq, RNA extracted from  $\sim$ 80% of RIPiT elution was used to generate strand-specific libraries. For RNA-seq libraries, 5  $\mu$ g of total cellular RNA was depleted of ribosomal RNA (RiboZero kit, Illumina), and subjected to base hydrolysis. RNA fragments were then used to generate strand-specific libraries using a custom library preparation method (Gangras et al., 2018). Briefly, a pre-adenylated miR-Cat33 DNA adaptor was ligated to RNA 3' ends and used as a primer binding site for reverse-transcription (RT) using a special RT primer. This RT primer contains two sequences linked via a flexible PEG spacer. The DNA with free 3' end contains sequence complementary to the DNA adaptor as well as Illumina PE2.0 primer sequences. The DNA with free 5' end contains Illumina PE1.0 primer sequences followed by a random pentamer, a 5 nt barcode sequence, and ends in GG at the 5' end. Following RT, the extended RT primer is gel purified, circularized using CirLigase (Illumina), and used for PCR amplification using Illumina PE1.0 and PE2.0 primers. All DNA libraries were quantified using Bioanalyzer (DNA lengths) and Qubit (DNA amounts). Libraries were sequenced on Illumina HiSeq 2500 in single-end format (50 and 100 nt read lengths).

### RNA-Seq data analysis

#### Adaptor trimming and PCR duplicate removal

After demultiplexing, fastq files containing unmapped reads were first trimmed using Cutadapt. A 12 nt sequence on read 5' ends consisting of a 5 nt random sequence, 5 nt identifying barcode, and a CC was removed with the random sequence saved for each read for identifying PCR duplicates down the line. Next as much of the 3'-adaptor (miR-Cat22) sequence TGGGAATTC TCGGGTGCCAAGG was removed from the 3' end as possible. Any reads less than 20 nt in length after trimming were discarded.

#### Alignment and removal of multimapping reads

Following trimming, reads were aligned with tophat v2.1.1 (Trapnell et al., 2009) using 12 threads to NCBI GRCh38 with corresponding Bowtie2 index. After alignment, reads with a mapping score less than 50 (uniquely mapped) were removed, i.e., all multimapped reads were discarded.

#### Removal of stable RNA mapping reads

Next, reads which came from stable RNAs were counted and removed as follows. All reads were checked for overlap against hg38 annotations for miRNA, rRNA, tRNA, scaRNA, snoRNA, and snRNA using bedtools intersect (Quinlan and Hall, 2010), and any reads overlapping by more than 50% were removed. Reads aligned to chrM (mitochondrial) were also counted and removed.

#### Human reference transcriptome

The primary reference transcriptome used in all post-alignment analysis was obtained from the UCSC Table Browser. CDS, exon, and intron boundaries were obtained for canonical genes by selecting Track: Gencode v24, Table: knownGene, Filter: knownCanonical (describes the canonical splice variant of a gene).

#### Read distribution assignment

Fractions of reads corresponding to exonic, intronic, intergenic, and canonical EJC and non-canonical EJC regions were then computed. Exonic regions were defined by the canonical hg38 genes, with intronic regions defined as the regions between exons in said genes. Bedtools intersect was used to compare reads against these exon and intron annotations, and reads which overlapped the annotation by more than 50% were counted. Any reads which did not overlap either the exon or intron annotations sufficiently were counted as intergenic. For classification of reads as canonical versus non-canonical EJC footprints, the canonical region for each library was defined using the meta-exon distribution at exon 3' ends (Figures 3C and S3J). All reads with their 5' ends falling within the window starting at the  $-24$  position till the 25% max height on the 5' side of the canonical EJC peak were counted as canonical reads. Similarly any read whose 5' end was found anywhere between the start of the exon and 10 bp upstream of that 25% max point was considered non-canonical.

#### k-mer analysis

Lists of all 6-mers and 3-mers present in reads mapping to exonic regions (as described above) were produced for each RIPiT-seq sample. The ratio of total 3-mer frequency in RIPiT-seq samples to RNA-seq samples was then used to identify 3-mers enriched in alternate EJCs.

#### Motif enrichment analysis

Motif enrichment analysis was performed by first selecting RNA binding proteins of interest - namely SR proteins - from the position weight matrices (PWMs) available on <http://rbpdb.ccb.utoronto.ca/>. For all reads mapped to exonic regions, a score was then generated representing the highest possible binding probability for each protein on that read. For visualization the cumulative distribution of these score frequencies was plotted for both pull down and RNA-seq replicates, with a relatively higher score frequency at a positive score implying greater binding affinity. The p values between RIPiT-seq and RNA-seq replicates were also computed for every score using a negative binomial based model, with significant values primarily in positive score regions implying a binding preference for that protein.



### **Differential enrichment analysis**

Differential analysis of exons and transcripts between CASC3 and RNPS1 pull down was conducted with the DESeq2 (Love et al., 2014) package in R. Exons and transcripts with significant differential expression ( $p < 0.05$ ) were selected. All the following analysis was conducted using only the lists of significantly differentially expressed transcripts, unless otherwise noted.

### **Estimation of nuclear versus cytoplasmic levels**

Nuclear and cytoplasmic RNA levels were estimated by first obtaining nuclear and cytoplasmic reads from (Neve et al., 2016). Reads were aligned and mapped to our exonic annotation as described above, and a ratio of nuclear to cytoplasmic reads was then calculated for all transcripts.

### **Comparison to genes with detained introns**

A list of detained and non-detained introns was obtained from (Boutz et al., 2015). To identify detained intron containing and lacking genes in our RNA-seq data, we carried out analysis using a DESeq2-based pipeline as described by Boutz et al. Briefly, using two RNA-seq replicates we first created artificial datasets containing the same numbers of total reads per replicate, but with counts originating from introns in a given gene spread evenly among those introns in the artificial set. By comparing the intron distributions of these artificial replicates to the experimental replicates using DESeq we were able to produce lists of detained introns (introns with significantly ( $p < 0.05$ ) higher expression levels compared to the artificially spread data) and non-detained introns (introns with non-significant ( $p > 0.1$ ) expression levels compared to the artificially spread data). The transcripts containing introns found to be detained/non-detained in both our analysis and the analysis done by Boutz et al. make up the stringent lists, which were used for analysis in Figure 4F. Of the 693 canonical genes containing detained introns reported by Boutz et al. we found 555 in our own analysis (80%) and of the 5294 canonical genes lacking detained introns we found 812 (15%).

### **Ribosome occupancy and mRNA half-life estimates**

Ribosome occupancy data for known canonical transcripts was obtained from (Kiss et al., 2017) with no further processing. mRNA half life data was similarly obtained from (Tani et al., 2012).

### **Gene ontology analysis**

DAVID gene ontology tool (Huang et al., 2009) was used to compare the set of genes (canonical Ensembl transcript IDs) predicted by DESeq2 analysis to be significantly enriched in CASC3 or RNPS1 EJs against a background list containing only those human genes that were reliably detected by DESeq2 (all genes for which DESeq2 calculated adjusted  $p$  values). Only non-redundant categories with lowest  $p$  value (with Benjamini-Hochberg correction) are reported.

## **QUANTIFICATION AND STATISTICAL ANALYSIS**

All western blots were performed using infrared fluorophore conjugated secondary antibodies and were scanned on a LI-COR Odyssey CLx imager. Protein quantification was performed using Image Studio software. Northern blot autoradiograms were scanned using Fuji FLA imager, and quantified using ImageQuant software. Average and standard deviation in the observed signal were determined for data from at least three biological replicates. For qRT-PCR, fold-change calculated using the delta-delta Ct method from at least three biological replicates was used to determine standard error of means. The  $p$  values were calculated using Student's  $t$  test. For mass spectrometry analysis, the quantification and statistical approaches used are described with methods above. Quantification of and statistical approaches for high-throughput sequencing data analysis are also described in methods above. When non-parametric distributions (not normally distributed) of numerical values of a particular parameter in a population were compared (using boxplots), the Wilcoxon rank sum test (also known as Mann-Whitney test) was used to test if any two distributions being compared are significantly different.

## **DATA AND SOFTWARE AVAILABILITY**

All RNA-seq data are available in GEO (accession number GEO: GSE115977) and under Bioproject PRJNA471492.