

# The Cellular EJC Interactome Reveals Higher-Order mRNP Structure and an EJC-SR Protein Nexus

Guramrit Singh,<sup>1,2,4</sup> Alper Kucukural,<sup>1,2,4</sup> Can Cenik,<sup>1,2</sup> John D. Leszyk,<sup>2</sup> Scott A. Shaffer,<sup>2</sup> Zhiping Weng,<sup>2,3</sup> and Melissa J. Moore<sup>1,2,\*</sup>

<sup>1</sup>Howard Hughes Medical Institute

<sup>2</sup>Department of Biochemistry and Molecular Pharmacology

<sup>3</sup>Program in Bioinformatics and Integrative Biology  
University of Massachusetts Medical School, Worcester, MA 01605

<sup>4</sup>These authors contributed equally to this work

\*Correspondence: melissa.moore@umassmed.edu

<http://dx.doi.org/10.1016/j.cell.2012.10.007>

## SUMMARY

In addition to sculpting eukaryotic transcripts by removing introns, pre-mRNA splicing greatly impacts protein composition of the emerging mRNP. The exon junction complex (EJC), deposited upstream of exon-exon junctions after splicing, is a major constituent of spliced mRNPs. Here, we report comprehensive analysis of the endogenous human EJC protein and RNA interactomes. We confirm that the major “canonical” EJC occupancy site *in vivo* lies 24 nucleotides upstream of exon junctions and that the majority of exon junctions carry an EJC. Unexpectedly, we find that endogenous EJCs multimerize with one another and with numerous SR proteins to form megadalton sized complexes in which SR proteins are super-stoichiometric to EJC core factors. This tight physical association may explain known functional parallels between EJCs and SR proteins. Further, their protection of long mRNA stretches from nuclease digestion suggests that endogenous EJCs and SR proteins cooperate to promote mRNA packaging and compaction.

## INTRODUCTION

In cells, all RNAs exist as ribonucleoprotein particles (RNPs). The protein components of these particles both add functionality to the RNA and serve as key structural elements of the overall complex. The most compositionally diverse RNPs are those containing messenger RNAs (mRNPs), and several hundred different polypeptides have now been identified as mRNP components (Baltz et al., 2012; Castello et al., 2012; Dreyfuss et al., 2002). Among these are proteins recognizing specific structures (e.g., the cap and polyA-tail-binding proteins) and others associating along the RNA's length in a largely sequence independent manner (e.g., YB-1 protein) (Moore, 2005). Other abundant

mRNP components are heterogeneous nuclear RNP (hnRNP) and serine/arginine-rich (SR) proteins, which recognize short sequence motifs via sequence-specific RNA-binding domains (Dreyfuss et al., 2002; Singh and Valcárcel, 2005). SR proteins characteristically contain one to two N-terminal RNA recognition motifs (RRMs) and a C-terminal domain rich in RS dipeptides subject to dynamic phosphorylation (Long and Caceres, 2009). These proteins are often loaded cotranscriptionally and many accompany the fully processed mRNA to the cytoplasm. The three best-studied shuttling SR proteins are SRSF1 (ASF/SF2), SRSF3 (SRp20), and SRSF7 (9G8). All three become dephosphorylated during the course of pre-mRNA processing and, in the dephosphorylated state, promote mRNP transit through the nuclear pore by interacting with export factors (Huang and Steitz, 2005). Also facilitating mRNP export is the transcription/export (TREX) complex (UAP56, Aly/REF, and THO subcomplex) loaded near the 5' cap during transcription and pre-mRNA splicing (Cheng et al., 2006).

Another set of proteins stably associated with both nuclear and cytoplasmic mRNPs is the exon junction complex (EJC). EJCs are deposited upstream of exon-exon junctions as a consequence of pre-mRNA splicing (Le Hir et al., 2000a). Unlike hnRNP and SR proteins, the EJC has no apparent sequence dependence. Rather, it binds RNA primarily through eIF4AIII, a sequence-independent DEAD box protein (Shibuya et al., 2004) that grasps RNA stably only when associated with its partners, Y14 and Magoh. A fourth protein, MLN51, can further stabilize eIF4AIII on RNA (Ballut et al., 2005; Bono and Gehring, 2011). This core serves as a binding platform for additional “peripheral” EJC factors. These include Pinin, Upf3, SKAR, SRm160, the “ASAP” complex (RNPS1, Acinus, and SAP18), and the TREX proteins UAP56 and Aly/REF (Bono and Gehring, 2011). When assembled *in vitro* splicing reactions, the EJC has been estimated to be ~350 kDa and protects 8–10 nt from nuclease digestion (Le Hir et al., 2000a).

In mammalian cells, core EJC proteins and peripheral factors interface with numerous machineries controlling mRNA export, translation, and decay (Giorgi and Moore, 2007; Tange et al., 2004). The EJC is known to promote export of spliced mRNAs

in vertebrates, perhaps through its close association with the TREX complex (Cheng et al., 2006). One of the best-understood EJC functions is its role in discriminating between premature and normal translation termination events. When translation terminates on a mammalian mRNA upstream of at least one EJC, Upf3 and its cofactors Upf1 and Upf2 orchestrate a series of events that destabilize the message by a process known as nonsense-mediated mRNA decay (NMD) (Rebbapragada and Lykke-Andersen, 2009). Another documented role for the EJC in mammalian cells is enhanced translation of newly synthesized mRNAs through interactions between SKAR and activated S6-kinase (Gudikote et al., 2005; Ma et al., 2008; Nott et al., 2004; Wiegand et al., 2003).

To date, all studies exploring mammalian EJC deposition, structure, and composition have been carried out either in vitro or using exogenously expressed proteins and/or reporter RNAs in vivo (Bono and Gehring, 2011; Tange et al., 2004). In this study, we report comprehensive analysis of the endogenous EJC interactome, including the complete EJC proteome and all tightly associated RNA fragments strongly protected from nuclease digestion. Surprisingly, endogenous EJCs multimerize both with one another and with numerous SR proteins to form megadalton sized complexes in which SR proteins are super-stoichiometric to EJC core factors. This intimate physical association may explain known functional similarities between EJCs and SR proteins. Additionally, their protection of extended mRNA regions from nuclease digestion suggests that cooperation between endogenous EJCs and SR proteins leads to higher-order mRNP structures, which may facilitate overall mRNA packaging and compaction.

## RESULTS

### An Unexpectedly Long Footprint for Endogenous Human EJCs

EJC deposition in vitro is strictly splicing dependent, and once formed, the complex is remarkably stable (Le Hir et al., 2000a). Therefore, we reasoned that a native RNA:protein immunoprecipitation in tandem (RIPiT) approach could be used to purify endogenous EJCs and their RNA footprints (Figure 1A and Figure S1A available online). Validating this approach, Myc-Magoh only copurified with FLAG-eIF4AIII when both were coexpressed in human embryonic kidney (HEK293) cells, but not when the two proteins were expressed separately and mixed following cell lysis (Figure 1B). Thus, our strategy is specific for pre-existing EJCs because no new EJC-like complexes appear to form after cell lysis (Mili and Steitz, 2004).

Two different RIPiTs (FLAG-eIF4AIII:Y14 and FLAG-Magoh:eIF4AIII), but not two control RIPiTs (FLAG only:PHGDH [3-phosphoglycerate dehydrogenase] and FLAG-only:Y14) yielded efficient and specific enrichment of the EJC core components Y14, eIF4AIII, and Magoh (Figure 1C and data not shown). Surprisingly, even under the most stringent nuclease digestion conditions, we always observed copurification of the endogenous untagged protein with its FLAG-tagged counterpart (Figures 1C, lanes 5 and 6, and S1B). This copurification was eliminated under conditions disrupting EJC core integrity (e.g., > 450 mM NaCl or 0.5 M urea; Figures S1B and S1C)

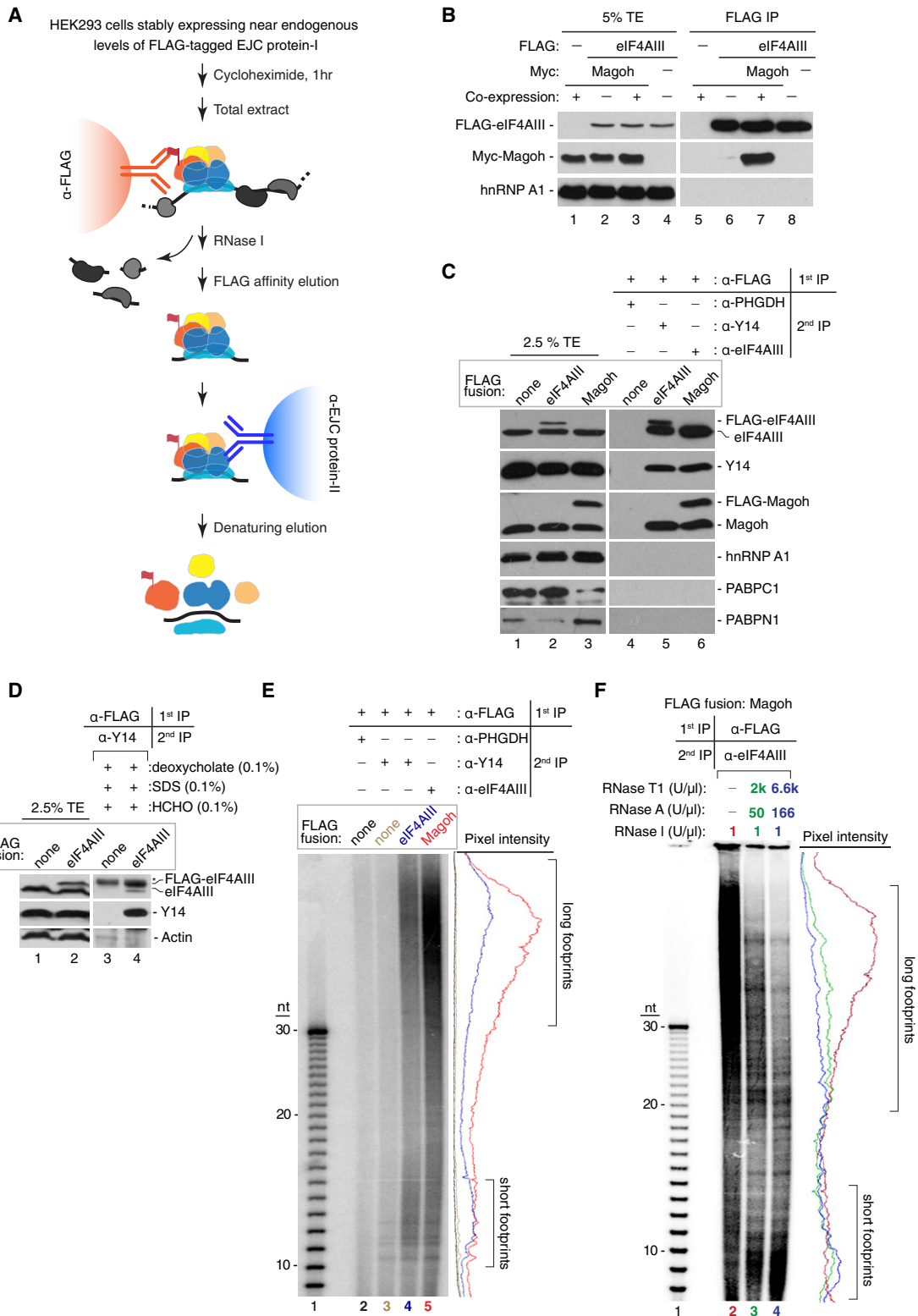
but persisted when cells were formaldehyde crosslinked prior to cell lysis and subjected to RIPiT under denaturing conditions (Figure 1D). This suggests that EJC monomers physically interact with one another in vivo (see below).

Another surprising finding was that, in addition to the small RNA footprints expected for monomeric EJCs (Ballut et al., 2005; Le Hir et al., 2000a; Stroupe et al., 2006), even more abundant and much longer ~30–150 nt RNase-resistant fragments were consistently observed in all of our EJC preps (Figure 1E). Both footprint sizes were observed in the absence or presence of cycloheximide with all nucleases tested but were lost if RNA-protein complexes were denatured with high salt or by incubation at 70°C prior to nuclease digestion (Figures S1D–S1F). The longer footprints persisted even after exposure to extreme nuclease conditions, although their size and intensity were reduced with a concomitant increase in shorter footprints (Figure 1F). The unexpectedly large RNA footprints, combined with the evidence above for EJC-EJC interactions, suggested greater complexity for endogenous EJCs than previously supposed.

### The EJC Proteome

To assess the full complement of proteins stably associated with EJC core factors in vivo, we performed comprehensive mass spectrometry analyses of samples subjected to extensive RNase I digestion and FLAG IP from cells expressing FLAG-eIF4AIII, FLAG-Magoh, or FLAG peptide only (to control for nonspecific interactions). Remarkably, ~70 different proteins were at least 10-fold enriched in both EJC core factor IPs compared to the control IP, and these enriched proteins could be organized into readily definable classes (Figures 2A and S2A and Table S1). Among the most abundant were the three EJC core factors: eIF4AIII, Y14, and Magoh. Also readily detectable were late-stage spliceosome components (i.e., U5 snRNP, NTC and NTC-related, and C complex proteins), consistent with reports that EJC assembly occurs concurrent with the second step of splicing and/or spliced exon release (Ideue et al., 2007; Reichert et al., 2002; Zhang and Krainer, 2007). The most highly enriched U5 snRNP component was the eIF4AIII recruitment factor Cwc22 (Barbosa et al., 2012; Steckelberg et al., 2012). The majority of known peripheral EJC proteins/protein complexes (such as MLN51, Pinin, Pym, the ASAP complex [Acinus, SAP18, and RNPS1], and TREX components [UAP56, UIF, URH49, and Aly/REF]) were also detected, but they were all sub-stoichiometric to the EJC core factors. Notably absent were Upf3 and SKAR, although they were marginally detectable by western blotting (Figures 2D and S2B).

Surprisingly, the entire complement of SR proteins was robustly enriched, with numerous members being either nearly stoichiometric (SRSF9, 10, 12) or even super-stoichiometric (SRSF1, 3, and 7) to the EJC core factors. Further, several other proteins containing SR domains were also enriched. These “SR-like” proteins (SRm160, SRm300, Trap150, Tra2 $\alpha$ , Tra2 $\beta$ , and SON) function as splicing coactivators and regulators in vivo (Ahn et al., 2011; Lee et al., 2010; Long and Caceres, 2009) and are also known mRNPs components (Baltz et al., 2012; Castello et al., 2012; Merz et al., 2007). Thus, the list of EJC-associated SR-like proteins extends significantly beyond the



**Figure 1. Purification of Endogenous EJCs and RNA Footprints**

(A) RIPIT schematic. EJC core factors, colored; non-EJC proteins, gray; RNA, black line.

(B) Copurification of EJC factors requires their coexpression. Western blot of total extract (TE, lanes 1–4) or FLAG IPs (lanes 5–8) from extracts in which FLAG-eIF4AIII and Myc-Magoh were either coexpressed (+) or expressed in separate cells (–) and mixed prior to IP.

previously known peripheral EJC components Pinin, Acinus, and RNPS1 (Le Hir et al., 2000a; Tange et al., 2005).

To assess the stability of SR protein and EJC core association, we performed western blots of RIPiT samples. Consistent with the EJC being a postsplicing complex, western blotting confirmed that only the hypophosphorylated forms of SRSF1 and SRSF3 copurified with the EJC (Figures 2B, 2C, S2C, and S2D). Further, even after extreme nuclease treatment, SRSF1 was readily detectable in native RIPiT and IP samples (Figures 2C and S2C). In contrast, SRSF1 was not detectable in samples that had been formaldehyde crosslinked prior to cell lysis, RNase digestion, and denaturing RIPiT (Figure 2C). This indicates that nuclease-resistant EJC and hypophosphorylated SR protein-containing complexes detected under native conditions are predominantly due to residence of EJC cores and SR proteins on the same highly protected fragments of RNA.

### Higher-Order EJC Interactions

To assess the size of endogenous EJCs, we performed gel filtration analysis of RNase-treated FLAG-eIF4AIII (Figure 2D) or FLAG-Magoh (data not shown) complexes subsequent to affinity elution with FLAG peptide (Figure 1A). Western blotting revealed two peaks containing the EJC core proteins eIF4AIII, Y14, and Magoh. These same two peaks were also observed upon gel filtration of RNase-A-treated cell lysates (data not shown), indicating that they were not formed as a consequence of the purification. The ~450 kDa peak contained only eIF4AIII, Y14, and Magoh and, in purified samples, only the FLAG-tagged protein used for affinity purification, but not its endogenous counterpart. This complex, which is similar in size to the in-vitro-assembled EJC (Le Hir et al., 2000a), thus likely represents EJC monomers, perhaps consisting of the core and other peripheral proteins undetected here.

The larger complex (high molecular weight [HMW] EJC) contained both the FLAG-tagged protein and its untagged endogenous counterpart. This complex also contained “peripheral” EJC proteins and other EJC-associated factors identified above, such as hypophosphorylated SRSF1. The larger complex eluted in the void volume of the gel filtration column, which had an exclusion limit of >2 MDa. Sucrose gradient fractionation revealed that the larger complex cosedimented with ribosomal subunits (Figure S2E), confirming their large size. In gel filtration experiments, the 30–150 nt RNA fragments almost exclusively fractionated with the HMW EJC, whereas the shorter fragments cofractionated with both the large and small complexes (Figure 2E). Consistent with the relative abundance of the long and

short RNA footprints in our RNase I digests (Figure 1E), the HMW EJC constituted ~70% of the total. Taken together, these results indicate that cells contain two EJC populations: monomeric EJCs with short footprints consisting primarily of the core factors and HMW EJCs protecting much longer stretches of RNA.

### Mapping of EJC- and Nuclear mRNP-Protected Sites

To investigate the nature of the short and long EJC-protected regions, we generated a series of libraries for deep sequencing on the SOLiD platform (Figure S3A). Short footprint EJC libraries were generated from 12–25 nt RNA fragments copurifying in FLAG-eIF4AIII:Y14 or FLAG-Magoh:eIF4AIII RIPiTs or in a FLAG-only:Y14 control RIPiT. Two libraries were generated from long EJC footprints: one from the 70–100 nt RNA fragments from the FLAG-Magoh:eIF4AIII RIPiT and another from 30–100 nt RNA fragments cofractionating with the HMW EJC (gel filtration fractions 2 and 3 in Figure 2D). To assess how the short and long EJC footprints compared to all protected regions on endogenous mRNAs, we also prepared a library of 24–36 nt fragments protected from RNase I digestion in native nuclear mRNPs. Finally, to quantify expression levels of endogenous mRNAs, we generated a polyA-selected RNA-seq library from one of our HEK293 cell lines. Raw reads were mapped to both the human genome and a reference exon junction data set (Figure S3B and Extended Experimental Procedures).

Among all libraries, reads mapping uniquely to the human genome and/or transcriptome were ~20- to ~37-fold enriched in exonic regions (Figure S3D). Exon mapping reads in our short footprint libraries proved highly reproducible, both between biological replicates and between different IP configurations (Figure S3E). As expected, these reads were significantly more abundant for mRNAs from intron-containing genes than for mRNAs from intronless genes (Figures S3F and S3G). In these libraries, ~40% of exonic reads mapped to the expected EJC deposition site (i.e., 15–31 nt upstream of exon-exon junctions). This read distribution led to readily identifiable peaks on individual exons (Figure 3A). For quantitative purposes, we developed an algorithm that defined “EJC peaks” as regions significantly enriched with short EJC footprint reads across multiple libraries (see Extended Experimental Procedures). Around half of such computationally identified EJC peaks mapped within exons (Figures S3D and S4A). In comparison, reads from the long EJC footprint and mRNP protection libraries were much more broadly distributed across exons (Figure 3A) and thus did not readily lend themselves to peak calling.

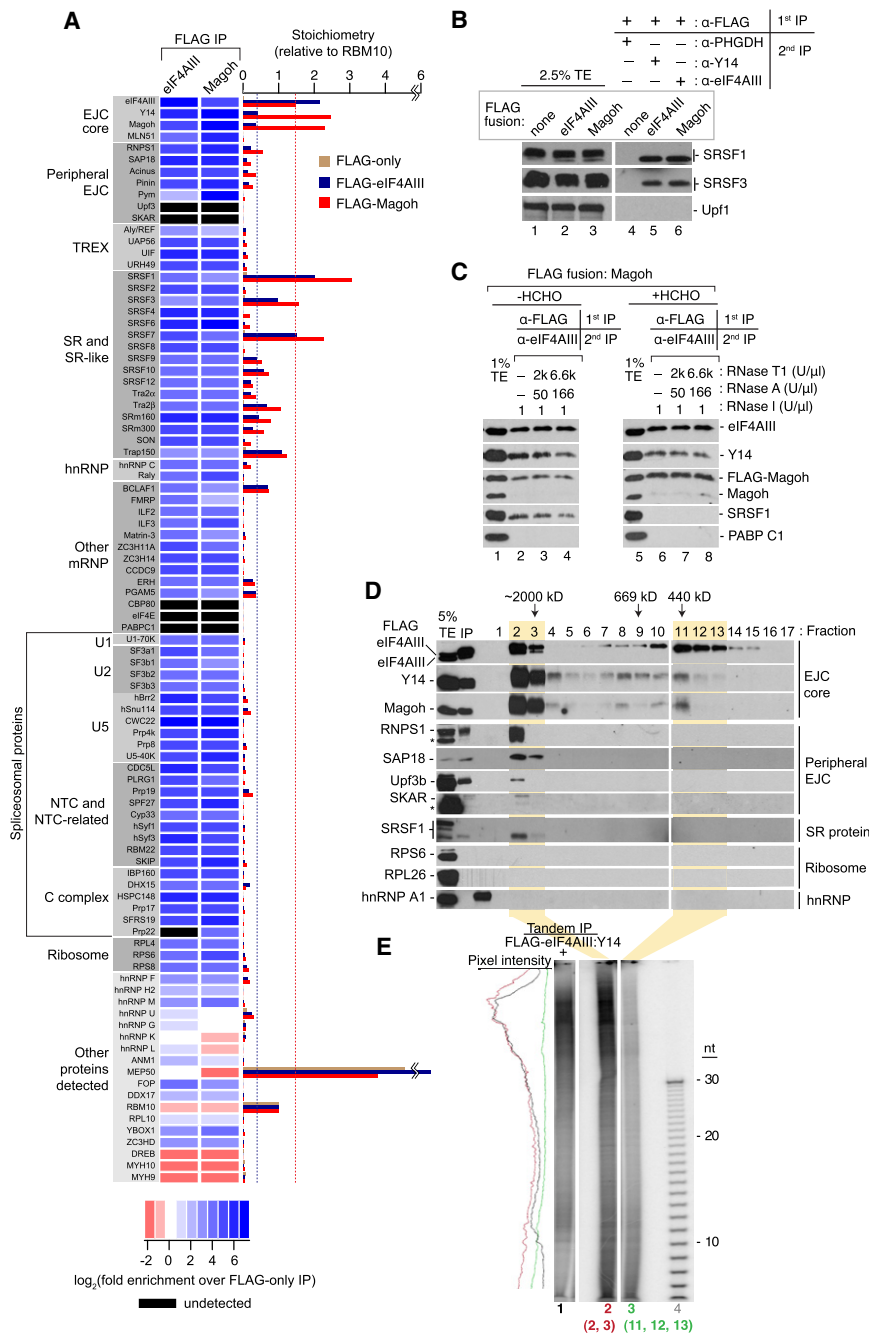
(C) Confirmation of EJC purification. Western blots of EJC or non-EJC RNA-binding proteins in total extracts (lanes 1–3), control RIPiT (lane 4), or EJC RIPiTs (lanes 5 and 6). For each lane, indicated are the FLAG-fusion protein (box above) and antibodies used for the first and second IPs (table above lanes 4–6).

(D) EJC purification under denaturing conditions after formaldehyde crosslinking. Western blots as in (C) showing protein levels in total extracts (lanes 1 and 2) or RIPiTs (lanes 3 and 4) from formaldehyde (HCHO)-crosslinked cells. SDS and deoxycholate were included during the  $\alpha$ -FLAG IP. Asterisk indicates IgG<sub>H</sub>.

(E) Length distribution of RNase I-resistant EJC footprints. Base-hydrolyzed synthetic polyU<sub>30</sub> RNA (lane 1) or immunoprecipitated RNA fragments from RIPiTs indicated at top (lanes 2–5) were 5' end <sup>32</sup>P labeled and separated by denaturing PAGE. Autoradiograph pixel intensity profiles of lanes 2–5 are at right.

(F) Same as (E) except that samples in lanes 3 and 4 were additionally incubated with RNases A+T1 after  $\alpha$ -eIF4AIII IP.

See also Figure S1.



**Figure 2. The EJC Interactome**

(A) The EJC proteome. (Left) Protein names organized into indicated classes. (Center) Heatmap showing fold enrichment of each protein in FLAG-eIF4AIII or FLAG-Magoh IPs over its level in a FLAG-only (control) IP (data not shown). (Black rectangles) Protein not detected. All proteins shown were at least 10-fold enriched in FLAG-EJC protein IPs except in the “Other proteins detected” class, which includes <10-fold enriched RNA-binding proteins, previously reported EJC-interacting proteins similarly enriched in our EJC and control IPs, and select proteins depleted in the EJC IPs compared to the control IP. (Bottom) Heatmap log<sub>2</sub> color scale. (Right) Bar graph showing protein stoichiometries relative to RBM10 in FLAG-eIF4AIII (blue), FLAG-Magoh (red), or FLAG-only (tan) IPs. (Dashed blue line) Level of copurifying Y14 and Magoh in FLAG-eIF4AIII IP. (Dashed red line) Level of copurifying eIF4AIII in FLAG-Magoh IP. These dashed lines indicate assembled EJC core stoichiometries in each IP.

(B) Confirmation of EJC proteomic analyses. Western blots of total extracts or RIPiTs as in Figure 1C showing levels of EJC-interacting proteins either detected or not in mass spec analysis.

(C) SRSF1 is tightly associated with but does not crosslink to EJC core factors. Western blots of indicated proteins in total extracts (lanes 1 and 5) or RIPiTs (lanes 2–4 and 6–8). Native and denaturing α-FLAG IPs were as in Figures 1C and 1D, respectively.

(D) Two distinct size complexes containing EJC core proteins. Western blots of indicated proteins (left) in total cell extract (TE), input FLAG-eIF4AIII IP sample (IP), and gel filtration fractions of RNase A-treated FLAG-eIF4AIII complexes (lanes 1–17). Peak elution fractions of known MW size markers are indicated at top. Asterisk indicates IgG<sub>H1</sub> (RNPS1 panel) or a cross-reacting unknown protein (SKAR panel).

(E) RNA footprints of EJC-core-containing complexes. Autoradiogram as in Figure 1D showing RNA profiles from FLAG-eIF4AIII:Y14 RIPiTs (lane 1) or from indicated gel filtration fractions in D (lane 2: fractions 2 and 3; lane 3: fractions 11, 12, and 13). Pixel intensity profile is on the left. See also Figure S2 and Table S1.

Nonetheless, the exonic read densities within transcripts were highly correlated between the long and short EJC footprint libraries (Figures S3H and S4B), indicating enrichment of similar RNA sequences in long and short footprints.

**The Canonical EJC Binding Site is 24 nt Upstream of Exon-Exon Junctions**

To map the predominant EJC deposition site in vivo, we calculated the distance from the center of each exon mapping read in all libraries to the 5’ or 3’ end of its parent exon. Whereas

no preferential read positioning was detected in the RNA-seq library, a composite plot revealed a striking accumulation of reads centered 24 nucleotides (nt) upstream of exon 3’ ends in all short EJC footprint libraries, with an apparent footprint width of ~14 nt (Figure 3B). No such accumulation was observed relative to any other mRNA feature such as exon 5’ ends, transcription start sites, or polyA sites (Figure 3B and data not shown). Reads in the long EJC footprint library were also preferentially skewed toward exon 3’ ends but were distributed more broadly than the short EJC footprints. Although protection at the –24 position (24 nt upstream of exon junctions) was evident on some individual exons in the nuclear mRNP

protection library (Figure 3A), this signal was only slightly enriched over other exonic sites in the composite plot (Figure 3B).

Preferential EJC association at the –24 position was readily apparent on individual spliced exons from both protein-coding and noncoding RNAs, as well as exons upstream of introns removed by both the “major” and “minor” (U2- and U12-dependent, respectively) spliceosomes (Figures 3C–3F). Further, the strict position of EJC deposition relative to exon 3' ends is maintained even on exons smaller than 24 nt. In such cases, the EJC deposition site lies within the preceding spliced exon (Figure 3G). Thus, the major, or “canonical,” EJC (cEJC) deposition site is centered 24 nt upstream of exon-exon junctions transcriptome-wide.

### EJC Occupancy Is Variable within mRNAs

We noticed that cEJC reads exhibited significant exon-to-exon variability within individual mRNAs (Figures 3A, S4B, and S4C). To test the validity of this observation, we calculated for each mRNA the coefficient of variation ( $C_v$ ) of the number of reads at cEJC sites. Consistent with profiles of individual mRNAs, the median  $C_v$  was much higher for the EJC libraries than for the RNA-seq library (Figure 4A). One possible explanation was differential EJC epitope accessibility at different exon junctions, leading to differential IP efficiencies. However, the nuclear mRNP protection library, which should be free of any IP bias, exhibited variability similar to the EJC libraries within the –15 to –31 window (Figure 4A). High variability at the cEJC site was also observed in a library generated from formaldehyde-crosslinked, FLAG-eIF4AIII cells lysed and FLAG-IPed under denaturing conditions (Figures 4A and S4C). Therefore, variability in cEJC reads is unlikely explained entirely by differences in epitope accessibility or EJC stability during purification.

Surprisingly, numerous exons completely lacked any EJC signal at the canonical site (e.g., *ENO1* exons 8 and 11, Figure 3A). We estimated the frequency of such occurrences in all mRNAs detectably expressed in our cell lines (reads-mapped per kilobase per million [RPKM] > 1 in our RNA-seq library). Within these, for each of the 4,156 mRNAs with 10 or more introns, we determined the fraction of cEJC sites with a detectable peak in our largest peak set (FLAG-Magoh:eIF4AIII). Among lower-expressed mRNAs, the fraction of exons with a cEJC peak was a direct function of transcript abundance (Figure 4B), suggesting that apparent EJC absence was mostly due to incomplete sampling. But at RPKM  $\geq 10$ , the fraction of cEJC-occupied sites plateaued at a median of  $\sim 80\%$ . Thus, even in the most highly expressed mRNAs, a fraction of cEJC sites appeared unoccupied.

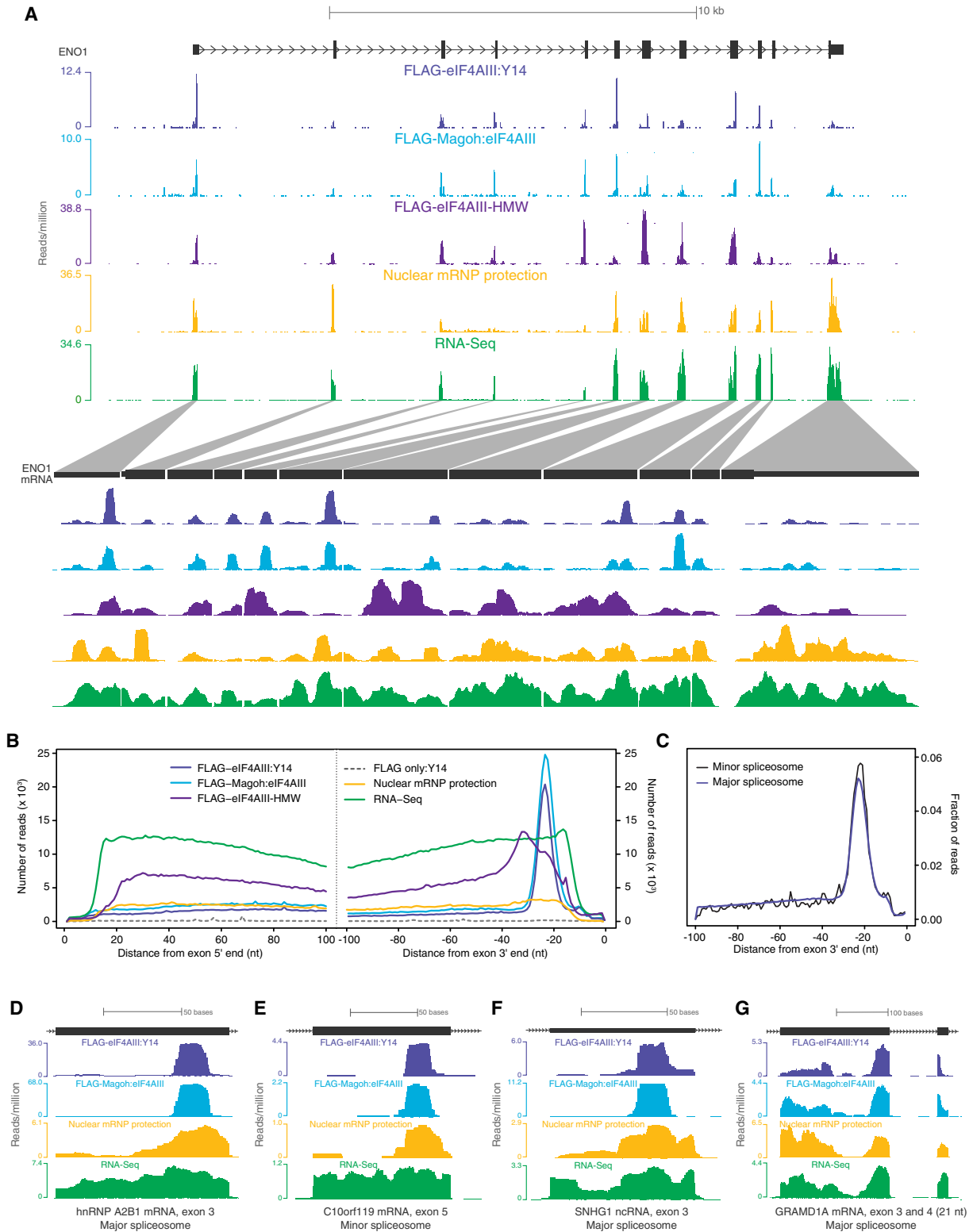
Some apparently cEJC-free sites in highly expressed mRNAs were undoubtedly due to poor mapping of the short EJC footprint reads. As compared to cEJC-occupied sites, cEJC-free sites exhibited a significantly lower ( $p < 1.35 \times 10^{-40}$ , Wilcoxon rank sum test) median mappability score (Figure 4C). Nonetheless, many canonical sites lacking a cEJC peak within highly expressed mRNAs (RPKM  $\geq 10$ ) also had high mappability scores ( $\geq 8$ ). When compared to occupied sites in the same mRNAs, these readily mappable but cEJC-free sites showed no significant difference with regard to either sequence compo-

sition (Figure 4D) or conservation (overall or wobble position only; data not shown) in the –15 to –31 window. We also detected no difference in predicted 5' or 3' splice site strengths between cEJC-free and occupied sites (Figures S4D and S4E). Nonetheless, there were small but significant differences in the exon and downstream intron length distributions, with occupied sites tending to occur on slightly shorter exons ( $\sim 10$  nt difference in median;  $p = 7.26 \times 10^{-5}$ ) or upstream of somewhat longer introns ( $\sim 150$  nt difference in median;  $p = 1.6 \times 10^{-3}$ ) than cEJC-free sites (Figures S4F and S4G). But the most significant difference between the two sets was the predicted propensity of nt –10 to –50 to form a secondary structure ( $p = 8.6 \times 10^{-7}$ ; Figure 4E). That no such difference was apparent for the adjacent window (–50 to –90 nt) strengthens the idea that EJC deposition is inhibited by local secondary structure (Mishler et al., 2008). In sum, our data indicate that cEJCs are deposited on the vast majority of exon-exon junctions in human mRNAs, but observed occupancy is influenced by exon and intron length and the availability of single-stranded RNA at the canonical deposition site.

### mRNAs Encoding RNA Processing Factors Display Higher Overall cEJC Occupancy

In addition to variable cEJC occupancy among exons within a single mRNA, we also wondered to what extent overall cEJC occupancy varied among mRNAs. We chose the 7,536 mRNAs from genes with one or more introns that had the most reproducible cEJC peaks, determined the average cEJC peak height for each mRNA, and then identified mRNAs with significantly lower or higher occupancy given their expression level (Figure 5A). Whereas no functionally related gene class (defined by Gene Ontology terms) was enriched in the low occupancy set, those pertaining to RNA metabolism were  $\sim 3$ -fold enriched in the higher occupancy set (Figure 5B). Of note, many mRNAs in this class regulate their own expression by alternative splicing linked to nonsense-mediated mRNA decay (AS-NMD), an EJC-dependent process (McGlinicy and Smith, 2008). Consistent with this, the most enriched category ( $\sim 8$ -fold) in the high cEJC occupancy class constituted known AS-NMD targets (Saltzman et al., 2008). This enrichment cannot be explained simply by a higher turnover rate for these mRNAs, as other classes of unstable mRNAs (e.g., transcription factors and cell-cycle proteins) were not similarly enriched (Figure S5A).

One possible explanation for the prevalence of AS-NMD targets in the high cEJC occupancy set could be enhanced cEJC occupancy downstream of the PTC-inducing alternative splicing event. To test this, we compared cEJC occupancies on exons flanking alternative splicing events that either did (AS-NMD mRNAs) or did not (AS-only mRNAs) introduce a PTC (Saltzman et al., 2008). For both data sets, there was no apparent preferential cEJC occupancy either 5' or 3' to the event (Figure 5C). What was readily apparent, however, was the higher overall cEJC occupancy on the AS-NMD mRNAs relative to the AS-only set. Thus, there may well be a propensity for increased cEJC occupancy on AS-NMD transcripts, but this occupancy is not differentially distributed relative to the PTC-inducing event.



**Figure 3. EJC Occupancy on Endogenous mRNAs**

(A) Distribution of short EJC footprints (FLAG-eIF4AIII:Y14 and FLAG-Magoh:eIF4AIII), HMW EJC footprints (FLAG-eIF4AIII-HMW), nuclear mRNP footprints, or RNA-seq reads on a representative gene (top half), *ENO1*, and its spliced mRNA (bottom half). Scale is indicated at top; reads per million at the highest position within each library window are indicated at left.

### Noncanonical EJC Peaks in Exons

In addition to the cEJC peaks centered at the  $-24$  position, we reproducibly observed peaks at other exonic locations (“noncanonical” or ncEJC peaks; Figures 3A and 6A). Such ncEJC peaks were also present in the library prepared from formaldehyde-crosslinked FLAG-eIF4AIII:Y14 lysates subjected to denaturing RIPiT, validating these as bona fide *in vivo* interactions (Figure S4C). Among our set of the most reproducible exonic peaks in the noncrosslinked sample libraries, these ncEJCs represented 40% of the total (Figure S6A). Unlike cEJCs, ncEJC peaks contained clearly identifiable sequence motifs (Figure 6). Highly enriched among first-exon ncEJC peaks was a CG-rich motif (Figures 6A, 6B, and S6C) almost identical to an mRNA-export-promoting sequence found in mRNAs lacking 5'UTR introns (Cenik et al., 2011). Two other motif classes were identifiable under ncEJC peaks in internal exons (Figures 6C and 6D). Several C-rich motifs (Figures 6D, top, S6B, and S6D) resemble known high-affinity sites for SRSF3, whereas an AG-rich motif (Figures 6D, bottom, and S6E) closely resembles the *in vivo* SRSF1-binding site determined by CLIP, as well as a reported binding site of the SR-like protein Tra2 $\beta$  (Long and Caceres, 2009 and references therein; Sanford et al., 2009). These internal exon motifs were also similarly enriched in the HMW EJC footprints (data not shown). Further, hexamers predicted to function as exonic splicing enhancers (ESEs) occurred more frequently as compared to exonic splicing silencers (ESSs) in the HMW EJC footprints (Figure S6F). ESEs are generally bound by SR proteins, whereas ESSs usually bind hnRNPs (Dreyfuss et al., 2002; Singh and Valcárcel, 2005). Thus, some ncEJC peaks likely represent short footprints from the SR proteins found to copurify with EJC core factors in our proteomics analysis (Figure 2).

Plotting ncEJC read densities versus transcript RPKM revealed that, like cEJCs, ncEJC signal is most enriched on RNA processing factor and AS-NMD mRNAs (Figure 6E). Further, there is a very high correlation ( $r = 0.81$ ) between average cEJC signal and ncEJC density across mRNAs (Figure S6G). Finally, as expected, these same gene classes are also enriched in the HMW EJC footprints (data not shown).

### cEJCs and ncEJCs Collaborate to Create Higher-Order mRNP Structure

Our biochemical analysis of cellular EJCs had revealed that EJC monomers physically interact *in vivo* (Figure 1). Such interactions most likely occur between neighboring EJCs. Supporting this, co-occupancy of adjacent cEJC sites was significantly higher than expected (Figure 7A). This effect is strongest between proximal sites and becomes progressively reduced at more distal sites (Figure S7A). A strong correlation was also observed for ncEJC peak co-occurrence on neighboring exons (Figure S7B).

Taken together, these data support the hypothesis of stable physical interactions between adjacent EJC-binding sites.

Also striking is the correlation between ncEJC and cEJC peaks. That is, ncEJC peaks are  $\sim 2$ - to  $\sim 4$ -fold more likely to occur adjacent to cEJC-occupied sites than unoccupied sites (Figure 7B). This is consistent with our biochemical data showing tight physical association between EJC cores and numerous SR proteins. To determine whether this physical interaction is of any functional consequence, we measured the crosslinking efficiencies of several mRNA binding proteins to polyA+ RNA upon knockdown of eIF4AIII (Figures 7C, 7D, and S7C). Whereas crosslinking of non-EJC-interacting proteins was unaffected by  $\sim 50\%$  eIF4AIII knockdown, crosslinking of SRSF1 and SRSF3 was reduced  $>2$ -fold. Thus, eIF4AIII abundance and likely EJC core deposition (Shibuya et al., 2004) functionally impact recruitment and/or stable association of SRSF1 and SRSF3 with spliced mRNA.

### DISCUSSION

A prevailing view based largely on *in vitro* studies is that EJCs are deposited upstream of every exon-exon junction on mammalian mRNAs, where they exist as singular units decorating the spliced mRNA much like beads on a string. Our new analysis of spliced mRNPs derived from cells, however, shows that, although EJCs are deposited on the majority of intron excision sites, this occupancy is not absolute. Further, our data indicate that EJCs exist in higher-order complexes containing multiple EJC cores in intimate contact with other mRNP components, especially SR proteins that are bound to their preferred target sites. These higher-order complexes protect long stretches of RNA from nuclease digestion. Our results thus provide a view of cellular mRNPs wherein the regions adjacent to sites of intron excision are protected and compacted by proteinaceous complexes consisting largely of EJCs and SR proteins (Figure 7E). The existence of these complexes explains numerous functional parallels between EJC and SR proteins.

### In Vivo EJC Occupancy

Our data indicate that EJCs are deposited  $\sim 24$  nt upstream of the vast majority of spliced exon junctions in human mRNAs (Figures 3 and 4). Nonetheless,  $\sim 20\%$  of cEJC sites appear unoccupied. Consistent with a previous *in vitro* study showing that local secondary structure can hamper EJC deposition (Mishler et al., 2008), the strongest contributor to EJC absence *in vivo* is the propensity of the  $-24$  region to engage in local secondary structure (Figure 4E). cEJC-occupied sites, on the other hand, are skewed toward shorter upstream exons and longer downstream introns, features that positively correlate with the presence of ESEs (Dewey et al., 2006) that generally

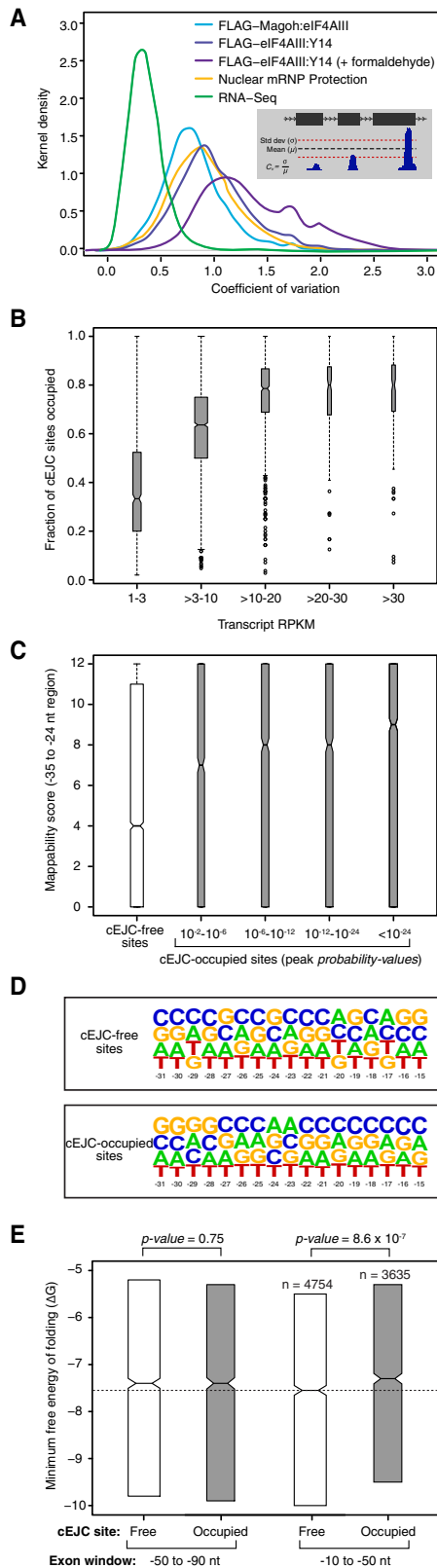
(B) Meta-exon analysis of read distribution. Composite plots of distances from the centers of all exon-mapping reads to the 5' (left) or 3' end (right) of each read's parent exon for the indicated libraries.

(C) Meta-analysis of exons upstream of major or minor spliceosome introns. Composite plots for FLAG-Magoh:eIF4AIII library reads mapping to exons upstream of introns excised by the “major” ( $\sim 1.7 \times 10^5$  introns) or “minor” (427 introns) spliceosome. In both plots, the fraction of reads at each nucleotide position out of all reads within 100 nt from exon 3' end was plotted.

(D–G) Profiles of short EJC footprints, nuclear mRNP footprints, and RNA-seq reads (as in A) on individual exons. Distinguishing features of each exon are indicated at bottom.

See also Figure S3.





**Figure 4. Variability in cEJC Occupancy and Contributing Factors**

(A) Signal variability among cEJC sites on individual mRNAs. Smoothed histograms of coefficient of variation ( $C_v$ ) in the number of reads at each cEJC site (–15 to –31 nt from exon junctions) within a single transcript in a set of 4,366 highly expressed transcripts. (Inset) Parameters used to calculate  $C_v$ .

(B) Fractions of occupied cEJC sites per transcript. Included were all mRNAs with >10 introns and representative transcript RPKM >1. All cEJC sites with a significant peak ( $p < 0.01$ ) overlapping the –24 position in the FLAG-Magoh:eIF4AIII set were considered occupied. The box plots show the interquartile range (IQR) of fraction-occupied cEJC sites in each RPKM bin. The whiskers are drawn at 1.5 times the IQR, and outliers are shown as open circles. Median (black horizontal line) and its confidence interval (notch) are also indicated within each box-plot.

(C) Range of mappability scores at cEJC-free (white) and cEJC-occupied (gray) sites from all spliced transcripts with RPKM >10. Occupied sites were binned by peak significance (probability values). Box plots are as in (B).

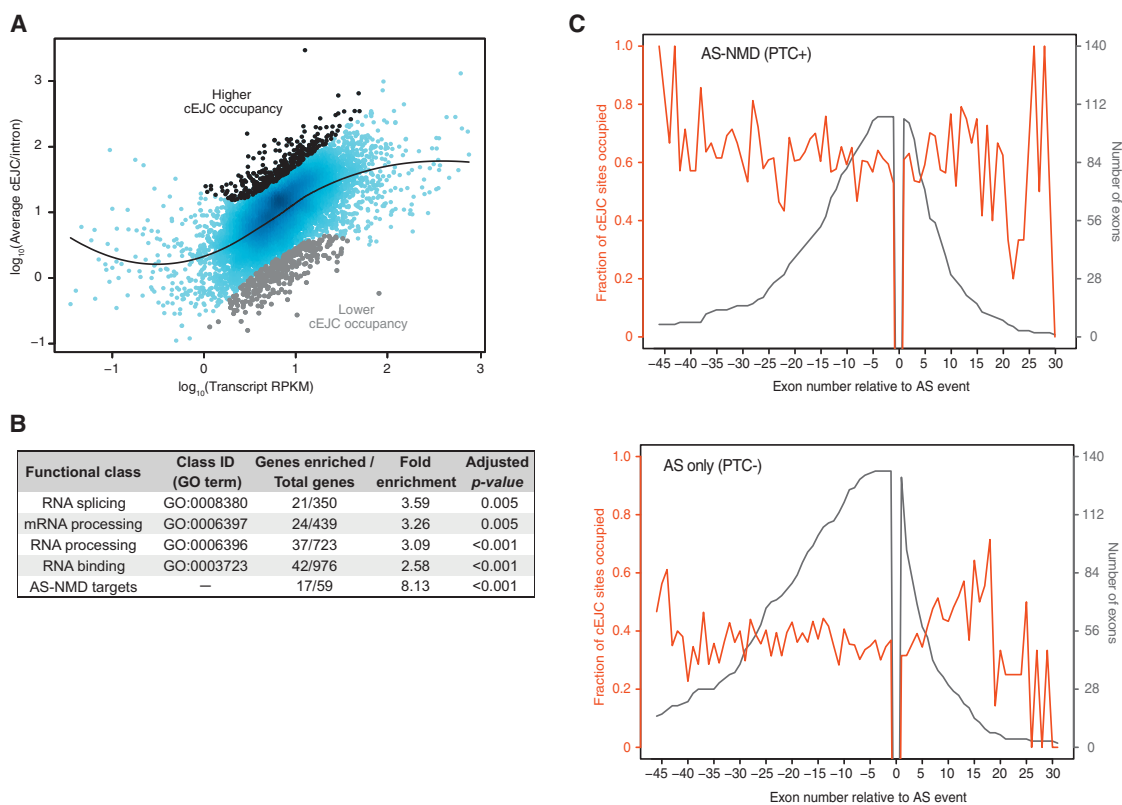
(D) Nucleotide frequency plot at cEJC-free and -occupied sites. All sites with mappability score  $\geq 8$  from all spliced transcripts with RPKM > 10 were used.

(E) Interquartile range of the minimal free energy of folding for two different 40 nt exonic windows (–50 to –90 nt from exon junctions, left; –10 to –50 nt from exon junctions, right) for cEJC-free (white) and cEJC-occupied (gray) sites from (D) except that included cEJC-occupied sites had a peak probability value  $< 10^{-12}$ . The total number of sites in each data set is indicated. Box plots are as in (B) except only the IQR is shown. The dashed line accentuates statistically significant (indicated by nonoverlapping notches) difference in median minimal free energy of cEJC-free and -occupied sites.

See also Figure S4.

bind SR proteins. As discussed below, this may reflect binding cooperativity between SR proteins and EJCs.

The high degree of cEJC occupancy in human cells contrasts with recent findings from *Drosophila melanogaster* in which only a few introns tested led to EJC deposition (Saulière et al., 2010). This apparent discrepancy could be due to either the different experimental approaches employed or real differences in EJC deposition and stability. That mammalian NMD is highly EJC dependent, whereas *Drosophila* NMD is largely EJC independent (Conti and Izaurralde, 2005), supports the latter view. In both organisms, numerous RNA-binding proteins regulate their own expression by AS-NMD. In *Drosophila*, it appears that AS-NMD is EJC independent and is accomplished primarily by modulating the distance between the stop codon and the polyA tail (i.e., 3'UTR length) (Hansen et al., 2009). In mammals, however, AS-NMD is most often accomplished by inserting a PTC upstream of at least one intron (i.e., EJC deposition site). Our data showing significant enrichment of both c- and ncEJCs on known AS-NMD targets (Figures 5 and 6) support this tight connection between EJC deposition and AS-NMD in mammalian cells. Although we observed no preferential cEJC enrichment downstream of the PTC-inducing AS event, higher overall EJC occupancy could enhance targeting of these transcripts to NMD by: (1) preferentially driving them into the translationally active pool (Gudikote et al., 2005; Nott et al., 2004; Wiegand et al., 2003) and then (2) enhancing PTC recognition due to higher cEJC levels downstream. Significant enrichment of ncEJC signal on AS-NMD targets is consistent with previous observations that SRSF1 overexpression stimulates NMD (Sato et al., 2008; Zhang and Krainer, 2004). One possible explanation is that increased SRSF1 occupancy on its preferred sequence motif



### Figure 5. Functional Classes Enriched in cEJCs

(A) Average cEJC peak height versus expression level (RPKM) for 7,536 mRNAs with highly reproducible cEJC peaks. Each dot represents an mRNA with two or more exons. (Black line) Locally weighted nonlinear fit (LOESS). The top and bottom 5% deviating furthest from the fit are indicated.

(B) Functional classes enriched in the higher cEJC occupancy set. The *p* values shown were adjusted for multiple hypotheses testing.

(C) Average cEJC occupancy up- and downstream of alternative spliced exons. (AS-NMD [PTC+]) Red line: fraction of cEJC occupied sites on successive exons flanking PTC-inducing alternative exon inclusion/exclusion events; black line: total number of exons at each position for AS-NMD transcript set (RPKM >1). (AS only [PTC-]): Same as above, except for a data set in which cassette exon inclusion/exclusion does not generate a PTC.

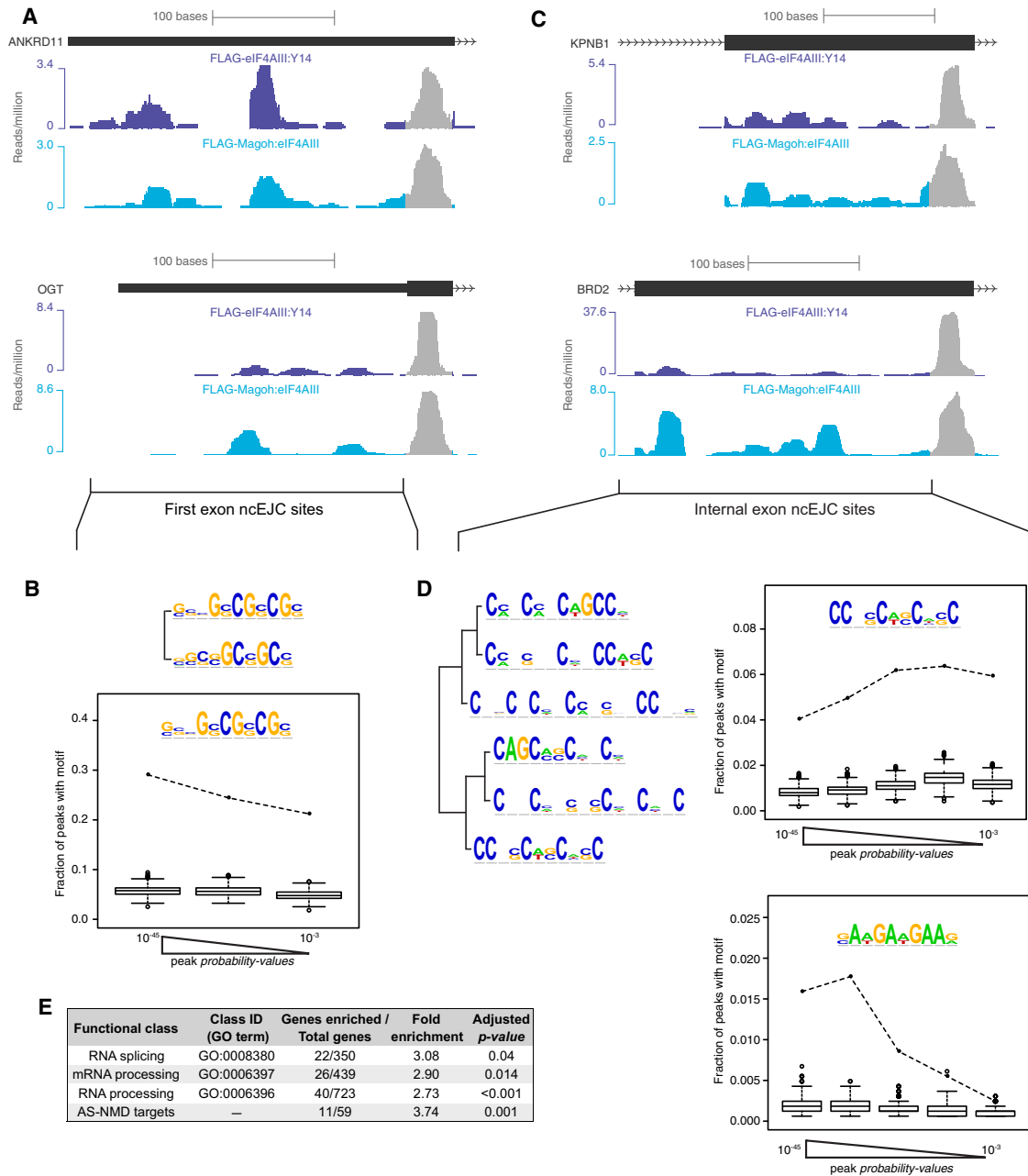
serves to promote or stabilize adjacent EJC deposition (see below).

### Extensive Interactions between EJCs and SR Proteins

A major surprise in this work is that the *in vivo* EJC proteome is much more extensive than previously imagined from *in vitro* studies (Figures 2 and S2). Particularly unexpected was our finding that, not only do numerous SR and SR-like proteins copurify with the EJC core, but several (SRSF1, 3, and 7) are superstoichiometric to the core. Consistent with this high abundance, preferred RNA binding sequences for SRSF1 and 3 were significantly enriched under ncEJC peaks in our deep sequencing libraries. Our stringent IP conditions (300 mM NaCl, 0.5% Empigen-BB) and extensive nuclease digestion conditions were specifically designed to eliminate all but the most tightly bound EJC-associated factors and RNA footprints. After nuclease digestion, subsequent purification steps usually required 4–5 additional hours. Therefore, any RNA sequence present in our short and long footprint libraries had to be very tightly associated with the EJC core. We conclude that many EJCs exist within larger complexes predominated by SR proteins bound to their specific recognition sequences.

Our finding of a tight physical association between EJCs and SR proteins *in vivo* is consistent with our previous results showing SR protein copurification with *in vitro* assembled EJCs (Le Hir et al., 2000b; Tange et al., 2005) and may help to explain the many known functional overlaps between these two protein sets. In addition to promoting NMD (Sato et al., 2008; Zhang and Krainer, 2004), both the EJC and SRSF1 can stimulate translation of bound mRNAs (Gudikote et al., 2005; Ma et al., 2008; Michlewski et al., 2008; Nott et al., 2004; Wiegand et al., 2003). Further, the EJC and multiple SR proteins all promote nucleocytoplasmic mRNA export (Huang and Steitz, 2005; Le Hir et al., 2001). In yet another functional parallel, knockdown of SRSF1 or any EJC core component leads to genomic instability (Li and Manley, 2005; Silver et al., 2010). Finally, both EJC core factors and the SR-like protein RNPS1 are required for splicing of long introns in *Drosophila* (Ashton-Beaucage et al., 2010; Roignant and Treisman, 2010), and the EJC, RNPS1, and other ASAP complex components are required for splicing of apoptosis gene introns in human cells (Michelle et al., 2012).

Our data strongly suggest that, rather than being attributable to functional redundancies, these functional parallels reflect tight



**Figure 6. Noncanonical EJC Sites**

(A and B) First exon noncanonical EJCs (ncEJCs).

(A) *ANKRD11* (top) and *OGT* exon 1 showing read distributions at canonical (gray) and noncanonical (colored) sites from indicated short footprint libraries.

(B) (Top) Two related sequence motifs enriched under first exon ncEJC peaks. (Bottom) Frequency of occurrence of inset motif among three equal size bins of first exon ncEJC peaks (black dots connected by dashed line) or in scrambled peak sequences. (Box plots: IQR and median [horizontal line within box plot] of the frequencies from 1,000 iterations; whiskers are at 1.5 times IQR, and outliers are as open circles.) Peaks were binned according to their probability values.

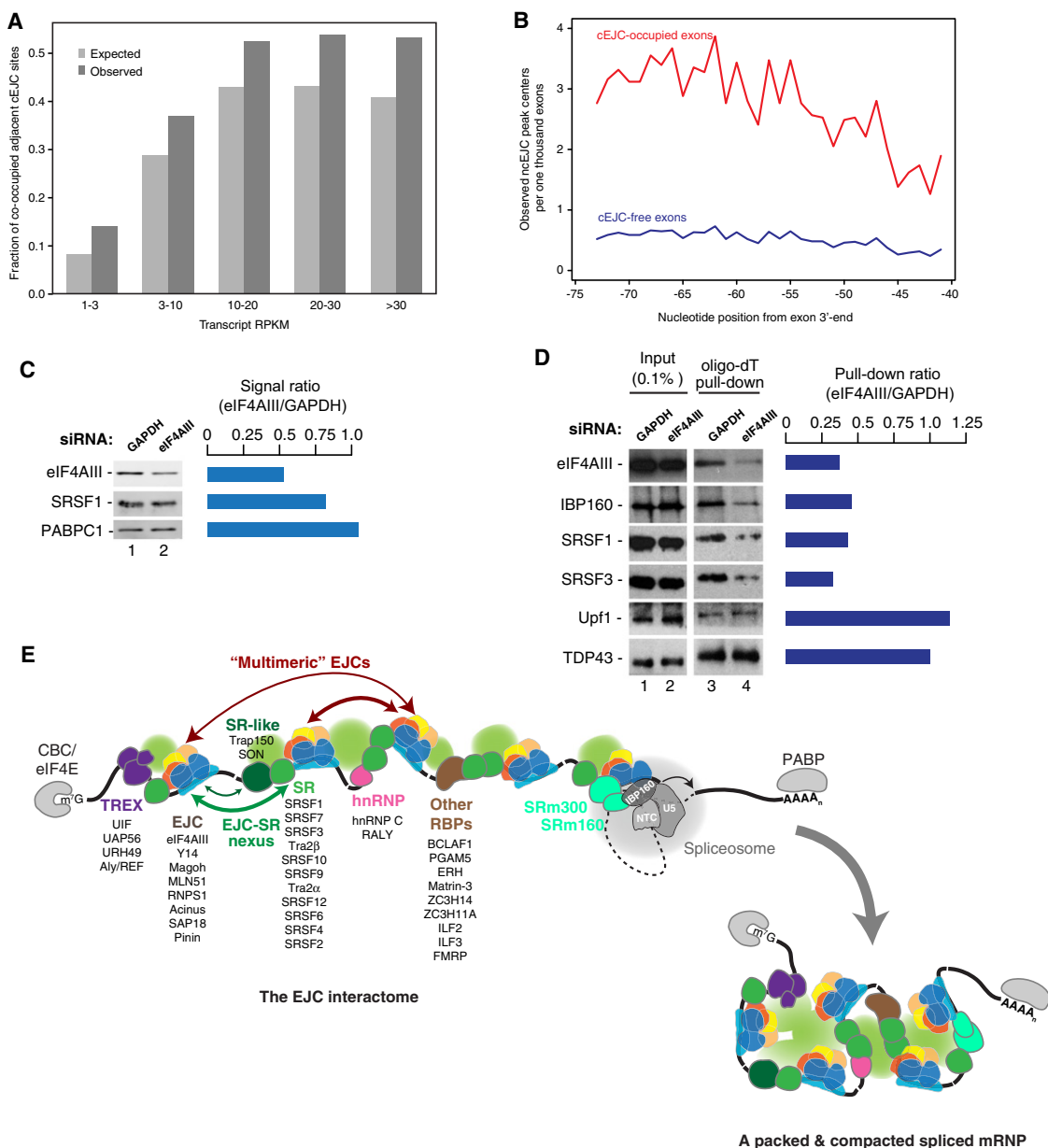
(C and D) Internal exon ncEJCs.

(C) cEJC and ncEJC sites in *KPNB1* exon 10 (top) and *BRD2* exon 11 (bottom, color scheme as in A).

(D) (Left) Dendrogram showing related motifs detected in internal exon ncEJC peaks. (Right) Same as in (B) except that internal ncEJC peaks were divided into five equal size bins.

(E) Functional classes enriched in ncEJC read densities.

See also Figure S6.



**Figure 7. A New View of mRNP Structure Incorporating EJC-EJC and EJC-SR Protein Collaboration**

(A) Co-occupancy of adjacent cEJC sites. Bar plots showing expected (light gray) and observed (dark gray) frequencies of co-occupied cEJC sites (mappability  $\geq 8$ ) at indicated transcript RPKMs. The two distributions are significantly different ( $p < 1 \times 10^{-15}$ , chi-square test).

(B) Occurrence of ncEJC peaks on cEJC-occupied (red line) and cEJC-free exons (blue line).

(C) Knockdown efficiency of eIF4AIII quantified by Odyssey imaging. (Bar plot) Lane 2:1 ratio for each protein.

(D) mRNA binding efficiency of proteins upon eIF4AIII knockdown. Western blots showing levels of indicated proteins in total extracts (lanes 1 and 2) and in oligo-dT-pull-downs (lanes 3 and 4) from UV-crosslinked HEK293 cells. (Bar plot) Lane 4:3 ratio for each protein.

(E) The EJC interactome and a new view of mRNP structure. (Solid black line) Exonic RNA; (dashed black line) a generic intron; (color ovals) proteins enriched  $>10$ -fold in the EJC proteome (Figure 2) listed in descending order of stoichiometry; (gray ovals) undetected proteins known to bind to mRNA ends; (green spheres) bridging protein-protein interactions.

See also Figure S7.

physical association between EJCs and SR proteins. Reduced SR protein crosslinking to polyA<sup>+</sup> RNA upon eIF4AIII knockdown suggests that EJC core deposition helps to stabilize SR protein association with spliced mRNA (Figure 7). Evidence that

sequences distal to the cEJC site can promote EJC deposition in *Drosophila* (Saulière et al., 2010) further suggests that this relationship is reciprocal (i.e., SR proteins may also stabilize EJC core association). Further, this model is consistent with the

cooperative mode of target recognition already proposed for the SR proteins themselves, which generally display weak binding with poor specificity in isolation (Singh and Valcárcel, 2005). The exact nature of the physical interactions between EJC core factors and SR proteins and the degree to which their many roles in modulating mRNA metabolism are functionally dependent on these physical interactions remain to be elucidated.

### Higher-Order Interactions within mRNPs

In comparison to other large RNPs such as the ribosome and spliceosome, rather little is known about overall mRNP architecture. Electron microscopy of purified *Saccharomyces cerevisiae* (*S. cerevisiae*) mRNPs has revealed them to be on average 10–15 times shorter than the calculated extended length of their resident mRNAs (Batisse et al., 2009). Even more compacted (~200-fold) are the gigantic Balbiani ring mRNPs assembled on long (~35–40 kB) *Chironomus tentans* salivary gland mRNAs (Skoglund et al., 1983). Extensive electron microscopy has shown these mRNPs to exist in the nucleus as 50 nm ring-like granules having a distinct coiled and folded structure containing several SR-like proteins (Björk et al., 2009).

In human mRNPs, EJC proteins are present on the vast majority of exons and form stable, high molecular weight multimers (Figures 1, 2, and 7). Further, we find a high degree of correlation between cEJC and ncEJC sites (Figure 7). Taken together, these results suggest that physical interactions between and among EJC cores and sequence-specific SR proteins could be major driving forces for mRNP compaction. Such compaction could be necessary for faithful and efficient execution of downstream events such as intranuclear mRNP mobility, mRNP export through the nuclear pore, and mRNP transport and translation in the cytoplasm. Of great interest for future studies will be the structural nature of the HMW EJC complexes and how spliced RNA sequences are packaged within them. Intriguingly, many spliceosomal and mRNP components, including EJC core factors and SR proteins, are enriched in inherently disordered regions (Castello et al., 2012; Korneta and Bujnicki, 2012), and the disordered region in another RNA-binding protein, TLS/FUS, was recently shown to mediate phase transition of RNPs into hydrogels (Han et al., 2012; Kato et al., 2012). Thus the presence of EJCs and SR proteins may give cellular mRNPs an intrinsic propensity to adopt hydrogel-like structures.

## EXPERIMENTAL PROCEDURES

Extended Experimental Procedures are provided in the Supplemental Information.

### RNA:Protein Immunoprecipitation in Tandem

Stable HEK293 cells expressing near endogenous levels of a FLAG-tagged EJC core protein were treated with cycloheximide (100  $\mu$ g/ml; 1 hr) and sonicated in a buffer containing 300 mM NaCl. RNA:protein complexes captured on FLAG-beads were washed and RNase I treated at 37°C for 10 min. Washed, FLAG-affinity eluted complexes were then subjected to a second IP with an antibody recognizing a different EJC core component.

### Mass Spectrometry

Stably expressed FLAG-eIF4AIII, FLAG-Magoh, or FLAG-peptide were IPed and RNase I digested as above in a lysis buffer also containing 0.5% Empigen

BB and 1  $\mu$ g/ml FLAG peptide. A short unresolved SDS-PAGE gel slice containing IPed, reduced, and alkylated proteins was subjected to in-gel trypsin digestion. Chromatographed peptides identified on a LTQ Orbitrap Velos mass spectrometer were searched against the human SwissProt index (09/21/11) using Mascot. Label-free quantitation via extracted ion chromatograms was performed using Mascot Distiller.

### Gel Filtration

FLAG-eIF4AIII-IPed complexes were fractionated on a 30 ml Sephacryl-S400 HR column. Proteins from 1 ml fractions were precipitated with trichloroacetic acid and western blotted. ~30–100 nt RNA fragments from HMW EJC fractions were extracted to make a deep sequencing library.

### Nuclear mRNP Footprints

HEK293 nuclei were lysed in a buffer containing 250 mM NaCl and 0.5% Triton X-100. PolyA-containing RNPs from cleared nuclear extract were captured on oligo-dT cellulose and released by a mild RNase I treatment in a low salt buffer. RNase-protected RNAs were extracted and size selected for deep sequencing.

### RNA-Seq and Generation of Deep Sequencing Libraries

4  $\mu$ g polyA+ RNA from FLAG-eIF4AIII-expressing cells was base-hydrolyzed at 90°C for 20 min. ~25–36 nt fragments were converted into a deep sequencing library. All short RNA fragment libraries were generated using the small RNA expression kit (Applied Biosystems) following manufacturer recommendations and were sequenced on the SOLiD platform.

### Mapping and Quantification of Deep Sequencing Data

SOLiD colorspace reads converted into sequence space using reference sequences (human genome and exon junction sequences) were aligned to reference sequences using Bowtie and were viewed on the UCSC genome browser. Uniquely mapping reads from biological replicates of short EJC footprint libraries were combined for peak calling (signal/background probability value < 10<sup>-2</sup>; peak height = number of reads at the most occupied position). The most reproducible peaks were those common between the FLAG-Magoh: eIF4AIII and FLAG-eIF4AIII:Y14 RIPIT sets. Data analysis was performed using the R statistical software package.

### siRNA Knockdown and Oligo-dT Pull-Down of polyA+ RNA-Crosslinked Proteins

HEK293 Flp-In cells transfected (Lipofectamine RNAiMAX) with 50 nM siRNA oligos were UV crosslinked at 800 mJ/cm<sup>2</sup> ~50 hr posttransfection and lysed in buffer containing 0.5% SDS and 0.5 M NaCl. PolyA+ RNA was captured on oligo-dT cellulose. RNA-crosslinked proteins were eluted by RNase A + T1 digestion, TCA-precipitated, and analyzed by western blots.

## ACCESSION NUMBERS

All sequencing data are available in NCBI's GEO database under accession number GSE41154.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2012.10.007>.

## ACKNOWLEDGMENTS

For valuable antibodies, we thank Elisa Izaurralde ( $\alpha$ -Magoh), Jens Lykke-Andersen ( $\alpha$ -Upf1, -Upf2, -Upf3, -Y14, and -eIF4AIII), Akila Mayeda ( $\alpha$ -RNPS1), and John Blenis ( $\alpha$ -SKAR). Frederick Roth and Hon Nian Chua generously provided a modified AlignACE algorithm. Ellie Kittler and the deep sequencing core are thanked for generating the deep sequencing data. We thank members of the Moore lab for discussions and critiques and Hervé Le Hir for communicating results prior to publication. This work was

supported by funding from HHMI and NIH RO1-GM53007 (to M.J.M.). G.S. was partially supported by the Charles A. King trust postdoctoral fellowship program. M.J.M. is an HHMI Investigator. G.S. and M.J.M. originally conceived the project, and G.S. executed all wet bench experiments. A.K. implemented all computational analyses with input from C.C. and Z.W. on computational method design and statistical analysis. J.D.L. and S.A.S. performed mass spec. All authors contributed to data analysis and interpretation. With input from all authors, G.S. and M.J.M. were primarily responsible for writing the paper.

Received: March 16, 2012

Revised: August 24, 2012

Accepted: October 3, 2012

Published online: October 18, 2012

## REFERENCES

- Ahn, E.Y., DeKaveler, R.C., Lo, M.C., Nguyen, T.A., Matsuura, S., Boyapati, A., Pandit, S., Fu, X.D., and Zhang, D.E. (2011). SON controls cell-cycle progression by coordinated regulation of RNA splicing. *Mol. Cell* **42**, 185–198.
- Ashton-Beaucage, D., Udell, C.M., Lavoie, H., Baril, C., Lefrançois, M., Chagnon, P., Gendron, P., Caron-Lizotte, O., Bonneil, E., Thibault, P., and Therrien, M. (2010). The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in *Drosophila*. *Cell* **143**, 251–262.
- 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.
- Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690.
- Barbosa, I., Haque, N., Fiorini, F., Barrandon, C., Tomasetto, C., Blanchette, M., and Le Hir, H. (2012). Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nat. Struct. Mol. Biol.* **19**, 983–990.
- Batisse, J., Batisse, C., Budd, A., Böttcher, B., and Hurt, E. (2009). Purification of nuclear poly(A)-binding protein Nab2 reveals association with the yeast transcriptome and a messenger ribonucleoprotein core structure. *J. Biol. Chem.* **284**, 34911–34917.
- Björk, P., Jin, S., Zhao, J., Singh, O.P., Persson, J.O., Hellman, U., and Wiestlander, L. (2009). Specific combinations of SR proteins associate with single pre-messenger RNAs in vivo and contribute different functions. *J. Cell Biol.* **184**, 555–568.
- Bono, F., and Gehring, N.H. (2011). Assembly, disassembly and recycling: the dynamics of exon junction complexes. *RNA Biol.* **8**, 24–29.
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406.
- Cenik, C., Chua, H.N., Zhang, H., Tarnawsky, S.P., Akef, A., Derti, A., Tasan, M., Moore, M.J., Palazzo, A.F., and Roth, F.P. (2011). Genome analysis reveals interplay between 5'UTR introns and nuclear mRNA export for secretory and mitochondrial genes. *PLoS Genet.* **7**, e1001366.
- Cheng, H., Dufu, K., Lee, C.S., Hsu, J.L., Dias, A., and Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* **127**, 1389–1400.
- Conti, E., and Izaurralde, E. (2005). Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol.* **17**, 316–325.
- Dewey, C.N., Rogozin, I.B., and Koonin, E.V. (2006). Compensatory relationship between splice sites and exonic splicing signals depending on the length of vertebrate introns. *BMC Genomics* **7**, 311.
- Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* **3**, 195–205.
- Giorgi, C., and Moore, M.J. (2007). The nuclear nurture and cytoplasmic nature of localized mRNPs. *Semin. Cell Dev. Biol.* **18**, 186–193.
- Gudikote, J.P., Imam, J.S., Garcia, R.F., and Wilkinson, M.F. (2005). RNA splicing promotes translation and RNA surveillance. *Nat. Struct. Mol. Biol.* **12**, 801–809.
- Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G., and McKnight, S.L. (2012). Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768–779.
- Hansen, K.D., Lareau, L.F., Blanchette, M., Green, R.E., Meng, Q., Rehwinkel, J., Gallusser, F.L., Izaurralde, E., Rio, D.C., Dudoit, S., and Brenner, S.E. (2009). Genome-wide identification of alternative splice forms down-regulated by nonsense-mediated mRNA decay in *Drosophila*. *PLoS Genet.* **5**, e1000525.
- Huang, Y., and Steitz, J.A. (2005). SRproteins along a messenger's journey. *Mol. Cell* **17**, 613–615.
- Ideue, T., Sasaki, Y.T., Hagiwara, M., and Hirose, T. (2007). Introns play an essential role in splicing-dependent formation of the exon junction complex. *Genes Dev.* **21**, 1993–1998.
- Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767.
- Korneta, I., and Bujnicki, J.M. (2012). Intrinsic disorder in the human spliceosomal proteome. *PLoS Comput. Biol.* **8**, e1002641.
- Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000a). The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.* **19**, 6860–6869.
- Le Hir, H., Moore, M.J., and Maquat, L.E. (2000b). Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev.* **14**, 1098–1108.
- Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20**, 4987–4997.
- Lee, K.M., Hsu, I.W., and Tarn, W.Y. (2010). TRAP150 activates pre-mRNA splicing and promotes nuclear mRNA degradation. *Nucleic Acids Res.* **38**, 3340–3350.
- Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365–378.
- Long, J.C., and Cáceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* **417**, 15–27.
- Ma, X.M., Yoon, S.O., Richardson, C.J., Jülich, K., and Blenis, J. (2008). SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. *Cell* **133**, 303–313.
- McGlincy, N.J., and Smith, C.W. (2008). Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem. Sci.* **33**, 385–393.
- Merz, C., Urlaub, H., Will, C.L., and Lührmann, R. (2007). Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment. *RNA* **13**, 116–128.
- Michelle, L., Cloutier, A., Toutant, J., Shkreta, L., Thibault, P., Durand, M., Garneau, D., Gendron, D., Lapointe, E., Couture, S., et al. (2012). Proteins associated with the exon junction complex also control the alternative splicing of apoptotic regulators. *Mol. Cell Biol.* **32**, 954–967.
- Michlewski, G., Sanford, J.R., and Cáceres, J.F. (2008). The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Mol. Cell* **30**, 179–189.
- Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694.
- Mishler, D.M., Christ, A.B., and Steitz, J.A. (2008). Flexibility in the site of exon junction complex deposition revealed by functional group and RNA secondary structure alterations in the splicing substrate. *RNA* **14**, 2657–2670.

- Moore, M.J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. *Science* *309*, 1514–1518.
- Nott, A., Le Hir, H., and Moore, M.J. (2004). Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes Dev.* *18*, 210–222.
- Rebbapragada, I., and Lykke-Andersen, J. (2009). Execution of nonsense-mediated mRNA decay: what defines a substrate? *Curr. Opin. Cell Biol.* *21*, 394–402.
- Reichert, V.L., Le Hir, H., Jurica, M.S., and Moore, M.J. (2002). 5' exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly. *Genes Dev.* *16*, 2778–2791.
- 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.
- Saltzman, A.L., Kim, Y.K., Pan, Q., Fagnani, M.M., Maquat, L.E., and Blencowe, B.J. (2008). Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. *Mol. Cell Biol.* *28*, 4320–4330.
- Sanford, J.R., Wang, X., Mort, M., Vanduyne, N., Cooper, D.N., Mooney, S.D., Edenberg, H.J., and Liu, Y. (2009). Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* *19*, 381–394.
- 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., Haque, N., Harms, S., Barbosa, I., Blanchette, M., and Le Hir, H. (2010). The exon junction complex differentially marks spliced junctions. *Nat. Struct. Mol. Biol.* *17*, 1269–1271.
- Shibuya, T., Tange, T.O., Sonenberg, N., and Moore, M.J. (2004). eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat. Struct. Mol. Biol.* *11*, 346–351.
- Silver, D.L., Watkins-Chow, D.E., Schreck, K.C., Pierfelice, T.J., Larson, D.M., Burnett, A.J., Liaw, H.J., Myung, K., Walsh, C.A., Gaiano, N., and Pavan, W.J. (2010). The exon junction complex component Magoh controls brain size by regulating neural stem cell division. *Nat. Neurosci.* *13*, 551–558.
- Singh, R., and Valcárcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. *Nat. Struct. Mol. Biol.* *12*, 645–653.
- Skoglund, U., Andersson, K., Björkroth, B., Lamb, M.M., and Daneholt, B. (1983). Visualization of the formation and transport of a specific hnRNP particle. *Cell* *34*, 847–855.
- Steckelberg, A.L., Boehm, V., Gromadzka, A.M., and Gehring, N.H. (2012). CWC22 connects pre-mRNA splicing and exon junction complex assembly. *Cell Rep.* *2*, 454–461.
- Stroupe, M.E., Tange, T.O., Thomas, D.R., Moore, M.J., and Grigorieff, N. (2006). The three-dimensional architecture of the EJC core. *J. Mol. Biol.* *360*, 743–749.
- Tange, T.O., Nott, A., and Moore, M.J. (2004). The ever-increasing complexities of the exon junction complex. *Curr. Opin. Cell Biol.* *16*, 279–284.
- Tange, T.O., 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.
- Wiegand, H.L., Lu, S., and Cullen, B.R. (2003). Exon junction complexes mediate the enhancing effect of splicing on mRNA expression. *Proc. Natl. Acad. Sci. USA* *100*, 11327–11332.
- Zhang, Z., and Krainer, A.R. (2004). Involvement of SR proteins in mRNA surveillance. *Mol. Cell* *16*, 597–607.
- Zhang, Z., and Krainer, A.R. (2007). Splicing remodels messenger ribonucleoprotein architecture via eIF4A3-dependent and -independent recruitment of exon junction complex components. *Proc. Natl. Acad. Sci. USA* *104*, 11574–11579.