

Examining the Effects of Translation on the Exon Junction Complex

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy
in the Graduate School of The Ohio State University

By

Lauren Allison Woodward, B.S.

Graduate Program in Molecular Genetics

The Ohio State University

2019

Guramrit Singh, Advisor

Sharon Amacher

Paul Herman

Stephen Kolb

Copyrighted by
Lauren Allison Woodward
2019

Abstract

The Exon Junction complex, EJC, is an integral component of spliced mRNPs. Work in past decades has established that it is deposited 24 nt upstream of exon-exon junctions during the splicing reaction and consists of a trimeric core (made up of eIF4AIII, Y14, and Magoh) that serves as a binding platform for peripherally interacting proteins. The EJC influences steps in post-transcriptional regulation, such as splicing, export, translation and nonsense-mediated mRNA-decay (NMD). The EJC's role in regulating mRNA fate ends when EJCs are removed during translation.

Peripheral proteins that interact with the EJC core are involved in diverse processes, including: splicing, nuclear export, translation enhancement, and NMD. Most peripheral proteins—like the EJC—shuttle between the nucleus and cytoplasm, but have predominant steady-state localizations to specific subcellular compartments. Interestingly, some proteins with markedly different steady-state subcellular localization participate in overlapping EJC-related functions. Namely, RNPS1 and CASC3 are both involved in translation enhancement and NMD.

In Chapter 2, purification of EJC containing CASC3 or RNPS1 revealed EJCs with distinct complements of peripherally interacting proteins. The EJC trimeric core binds RNPS1 and CASC3 in a mutually exclusive manner. RNPS1-EJCs and CASC3-EJCs differ in their overall protein composition, subcellular localization, and mRNP structure. RNPS1-EJCs represent a more nuclear variation of the EJC and bind to SR- and SR-like proteins that assemble within high-molecular weight, compact mRNPs. CASC3 EJCs, however, are primarily nuclear, devoid of SR-proteins, and exist in more

monomeric structures within the mRNP, revealing that EJCs undergo a dramatic compositional change, likely shortly after their export into the cytoplasm.

Interestingly, CASC3 and RNPS1 also differ in their ability to trigger NMD of transcripts containing a premature termination codon. Human cells depleted of RNPS1 showed robust increase of NMD substrates, while the effect of CASC3 depletion was relatively modest. The differences in inhibition of NMD upon depletion of CASC3 and RNPS1 suggest the possibility of distinct branches of NMD.

RIPiT-seq (RNA Immunoprecipitation in Tandem) of CASC3-EJCs and RNPS1-EJCs revealed that these distinct complexes largely bind to the same transcripts, with some notable exceptions. Namely, CASC3-EJCs are significantly enriched on mRNA's encoding ribosomal proteins. Ribosomal protein mRNAs, containing a 5' terminal oligo pyrimidine motif (5' TOP), are under mTOR-dependent translational regulation via binding of LARP1. CASC3-EJC interaction with LARP1 may target untranslated 5' TOP mRNPs for translational repression in response to poor cellular conditions. Consistently, CASC3-EJC occupancy on mRNA is much more sensitive to active translation, suggesting CASC3-EJCs are directly removed by the translating ribosome.

EJCs must be removed from coding stretches of mRNA to permit ribosomal decoding of the transcript. EJC removal from mRNA has long been attributed to the action of the so-called EJC disassembly factor, PYM. However, the relative contribution of PYM interaction-mediated and ribosome-mediated EJC disassembly has not been discussed in literature. Examples of loss of PYM activity do not result in the translational defects one would expect of a necessary mechanism of EJC disassembly. PYM null

Drosophila melanogaster mutants and PYM knockout human cell lines have no obvious phenotype or observed translational defects. Thus, investigation in Chapter 3 compares the role of translation and PYM in EJC disassembly. Translation was found to be the major determinant of EJC occupancy, with PYM interaction being largely dispensable in terms of EJC occupancy on mRNA. Surprisingly, loss of PYM interaction resulted in more EJC binding to non-canonical stretches of mRNA (i.e. regions distant from the canonical -24 nt site of EJC deposition), including single exon transcripts.

If PYM function is unnecessary for EJC removal from mRNA, then what is PYM's primary function? Others have observed that PYM interaction with the EJC enhances translation of luciferase reporters, and Kaposi Sarcoma-associated herpes virus targets PYM to recruit translational machinery to unspliced viral transcripts. Perhaps PYM similarly functions to enhance translation of endogenous single exon mRNA. In support of this, RIPiTs of PYM-EJC interaction are enriched on single exon mRNAs. Consistently, bulk translation of endogenous mRNA is reduced upon PYM depletion.

Alternatively, PYM may function to prevent extra-splicing assembly of EJCs in the cytoplasm, resulting in the increased occupancy of EJCs at non-canonical regions and on unspliced mRNAs. Such spontaneous cytoplasmic EJC assembly would presumably result in misregulation of many transcripts. In agreement, knockdown of PYM resulted in significant up and down regulation of many transcripts. While PYM interaction with the EJC is unnecessary for EJC disassembly, PYM may have an as yet unappreciated role in translation enhancement and proper EJC placement on RNA.

Footprinting of the EJC with various approaches, including RIPiT- and CLIP-seq have revealed the presence of EJCs non-canonical positions, including in the last exon and 3'UTR of nearly all transcripts. These stretches of mRNA are ordinarily devoid of a downstream exon-exon junction, so it is not clear how the EJC comes to be bound to these regions. Chapter 4 investigates the origin of non-canonical EJCs in the 3' UTR. EJC 3' UTR occupancy is increased with active translation and in transcripts with a greater exon count, leading to a model where EJC are displaced during translation to the 3' UTR. EJC occupancy in the 3' UTR is also increased in mRNPs enriched in the cytoplasm, consistent with the role of translation in 3'UTR EJC occupancy. Surprisingly, these non-canonical 3'UTR EJCs do not result in EJC-mediated NMD.

In sum, this work extends our understanding of the EJC's interaction with diverse peripherally interacting proteins and the complexity of EJC-mediated post-transcriptional regulation, and lends new insight to current understanding of the role of translation on EJC occupancy on mRNA.

Dedication

This work is dedicated foremost to my daughter, Norah, who is my greatest inspiration, and my two guys: Adam and Jack. I also dedicate this work to my friends, family, and church family—it is sometimes difficult to distinguish between the categories. Without their continued help and prayer, none of this would have been possible.

Acknowledgments

I must begin by acknowledging my advisor, Dr. Guramrit Singh. Completion of this work has taken hard work, patience, perseverance, and creativity. For all of these traits and more, I could not have a better mentor from which to model. I count myself very fortunate to have an advisor who is so dedicated to training independent scientists yet is always available to chat and lend advice. In addition to being a good scientist, Amrit has also created a lab environment that is truly enjoyable to work in. The Singh lab is home to some amazing minds and very good friends; we have celebrated each other's successes and are supportive in the effort to make one other's research the best it can be. Specifically, I would also like to thank Justin Mabin who contributed significantly to this work and is a good lab mate and friend; Pooja Gangras, who has truly traveled the valleys and mountaintops of this journey with me; Zhongxia Yi, who is always ready to talk about science and keeps us all excited about our work; and Manu Sanjeev and Caleb Embree, both of whom are great lab mates and contribute to lively discussion.

I want to also thank the members of my committee: Dr. Sharon Amacher for her comments and advice in experimental design; Dr. Paul Herman for talks about career choices and advice, and Dr. Stephen Kolb for his help in shaping my thesis. In addition, Ralf Bundschuh and his student, Robert Patton have spent several hours in discussion regarding my data and best analyses to complete, and I am truly thankful for their help and contribution to my work.

I would also like to acknowledge the second floor of the Bioscience building. In particular, I would like to thank current and past members of the Cole lab: Dr. Skye Bochter, Kara Braunreiter , and Dr. Kanu Wahi as wells as current and past members of the Fisk lab: Jennifer Perkins, Tan Nguyen, and Dr. Joey Marquardt for invaluable discussion and helping learn new techniques.

Vita

May 2009.....Benjamin Logan High School
May 2013.....Mount Vernon Nazarene University
2014 to present.....Graduate Research Associate, Department of
Molecular Genetics, The Ohio State University

Publications

1. **Woodward L**, Mabin J, Gangras P, Singh G (2017) The Exon Junction Complex: A Lifelong Guardian of mRNA Fate. Wiley Interdiscip. Rev. RNA 8.
2. Mabin J*, **Woodward L***, Patton R*, Yi Z, Jia M, Wysocki V, Bundschuh R, Singh G (2018) The exon junction complex undergoes a compositional switch that alters mRNP structure and nonsense-mediated mRNA decay activity. Cell Rep. 25, 2431-2446.e7.
3. **Woodward, L.**, Gangras, P., Singh, G. Identification of Footprints of RNA:Protein Complexes via RNA Immunoprecipitation in Tandem Followed by Sequencing (RIPiT-Seq). J. Vis. Exp. (149), e59913, doi:10.3791/59913 (2019).

Fields of Study

Major Field: Molecular Genetics

Table of Contents

Abstract.....	ii
Dedication.....	vi
Acknowledgments.....	vii
Vita.....	ix
List of Tables.....	xiv
List of Figures.....	xv
Chapter 1 Introduction.....	1
1.1 RNA-binding proteins influence mRNA fate.....	1
1.2 Splicing enhances post-transcriptional steps via deposition of the EJC.....	2
1.3 EJC core components and assembly.....	3
1.4 Peripheral Interacting Proteins.....	8
1.5 EJC canonical binding site.....	10
1.6 Non-canonical EJCs.....	11
1.7 The role of EJCs in mRNP packaging.....	13
1.8 EJC in nonsense-mediated decay.....	15
1.9 EJC removal and recycling.....	18
1.10 EJCs role in pre-mRNA splicing.....	21
1.11 Summary.....	24

Chapter 2	EJC composition changes as mRNPs move from nucleus to cytoplasm	35
2.1	Abstract:	35
2.2	Introduction	35
2.3	Materials and Method	38
2.3.1	Stable cell lines	38
2.3.2	Endogenous and FLAG-tag Immunoprecipitations	38
2.3.3	Western blot imaging and analysis	40
2.3.4	RNA:protein immunoprecipitation in tandem (RIPiT)	40
2.3.5	High-throughput sequencing library preparation	42
2.3.6	Adapter trimming and PCR removal	43
2.3.7	Alignment and removal of multimapping reads	43
2.3.8	Removal of stable RNA mapping reads	43
2.3.9	Differential enrichment analysis	44
2.3.10	Estimation of nuclear versus cytoplasmic levels	44
2.3.11	Ribosome occupancy and mRNA half-life estimates	44
2.3.12	Gene ontology analysis	45
2.3.13	Glycerol gradient fractionation	45
2.4	Results	46
2.4.1	RNPS1 and CASC3 Associate with the EJC Core in a Mutually Exclusive Manner	46
2.4.2	Alternate EJCs are structurally distinct	47
2.4.3	RIPiT-seq reveals that RNPS1 and CASC3 Bind RNA via the EJC Core with Key Distinctions	47
2.4.4	Subcellular mRNP Localization and Nuclear Retention Mechanisms Affect Relative RNPS1 and CASC3 Occupancy	48
2.4.5	Kinetics of Translation and mRNA Decay Impacts the Pool of Alternate EJC-Bound mRNAs	50
2.4.6	Translation of 5'TOP mRNAs is dependent on CASC3	53
2.4.7	The alternate EJC occupancy landscape is also impacted by mRNA decay kinetics	54
2.5	Discussion	55

2.5.1	EJC Composition and mRNP Structure	55
2.5.2	CASC3-EJC and Pre-translation mRNPs	56
Chapter 3	The role of PYM in EJC occupancy.....	77
3.1	Abstract	77
3.2	Introduction.....	78
3.3	Materials and Methods.....	82
3.3.1	Stable cell lines	82
3.3.2	FLAG-tag Immunoprecipitations.....	82
3.3.3	Western Blotting	83
3.3.4	RIPiT.....	84
3.3.5	siRNA knockdowns	84
3.3.6	RT-PCR and qPCR	85
3.3.7	Immunofluorescence.....	86
3.3.8	RNAseq of PYM knockdown.....	87
3.3.9	Preparation of RIPiT RNA samples for deep sequencing	87
3.3.10	Adapter trimming and PCR removal	88
3.3.11	Alignment and removal of multimapping reads	88
3.3.12	Differential enrichment analysis.....	89
3.3.13	GO-term analysis	89
3.4	Results.....	90
3.4.1	EJC-PYM interaction is not necessary for removing EJCs from canonical positions	90
3.4.2	Loss of PYM interaction causes Magoh to bind more at non-canonical positions	91
3.4.3	PYM interacts with the EJC core at canonical and non-canonical positions	93
3.4.4	PYM-eIF4AIII complexes and MagohE117R-EJCs are enriched on single exon transcripts	94
3.4.5	PYM knockdown reduces nascent translation	95
3.4.6	PYM knockdown results in widespread misregulation of many transcripts.	95
3.5	Discussion	96

Chapter 4	Cytoplasmic EJCs are repositioned to the 3'UTR of mRNA and disassembled during translation	134
4.1	Introduction.....	134
4.2	Materials and Methods.....	136
4.2.1	Stable cell lines	136
4.2.2	Endogenous and FLAG-tag Immunoprecipitations	136
4.2.3	Polysome profiling:.....	136
4.2.4	Cell Fractionation.....	137
4.2.5	RNA:protein immunoprecipitation in tandem (RIPiT).....	137
4.2.6	RIPiT High-throughput sequencing library preparation	139
4.2.7	Adaptor trimming and PCR duplicate removal	139
4.2.8	Metaplots of exon stop codons	140
4.3	Results.....	141
4.3.1	EJCs exist in the last exon and 3'UTR	141
4.3.2	EJCs are present in the polysome	141
4.3.3	Translation inhibition reduces the number of EJC footprints in the 3'UTR	142
4.3.4	The number of upstream exons impacts the amount of EJC occupancy in the 3'UTR	143
4.3.5	More EJC occupancy in the 3'UTR of cytoplasmic fraction than nuclear fraction	143
4.3.6	EJCs in the 3'UTR do not destabilize mRNA	144
4.3.7	PYM-eIF4AIII interacts with the EJC in the 3'UTR.....	145
4.4	Discussion	146
Chapter 5	Conclusions and Future Directions	165

List of Tables

Table 1.1 EJC-interacting proteins	29
Table 2 RIPiT library information	171

List of Figures

Figure 1.1 EJC Lifecycle	27
Figure 1.2 Non-canonical EJCs	30
Figure 1.3 Role of EJCs in mRNP packaging.....	31
Figure 1.4 EJCs enhance nonsense-mediated decay (NMD).....	32
Figure 1.5 EJC deposition affects splicing	33
Figure 2.1 CASC3 and RNPS1 bind the EJC in a mutually exclusive manner in multiple cell types	59
Figure 2.2 Alternate EJCs are structurally distinct	61
Figure 2.3 RIPiT schematic	62
Figure 2.4 Tetracycline inductions of FLAG-RNPS1 and FLAG-CASC3 HEK293 cell lines used for RIPiT experiments.....	63
Figure 2.5 Alternate EJCs bind RNA via the EJC	64
Figure 2.6 mRNP subcellular localization correlates with EJC composition.....	65
Figure 2.7 Translation impacts CASC3-EJC occupancy on mRNA	67
Figure 2.8 CASC3 is enriched on transcripts with lower translation efficiency	68
Figure 2.9 CASC3-EJCs are enriched on ribosomal protein mRNAs.....	69
Figure 2.10 CASC3-containing RNPs interacts with SRP protein, SRP68.....	70
Figure 2.11 CASC3 interacts with LARP1	71
Figure 2.12 Serum starvation promotes CASC3-LARP1 interaction.....	73
Figure 2.13 CASC3 depletion upregulates protein level of 5'TOP mRNA-encoded proteins, eIF3A and rps6.....	74
Figure 2.14 mRNAs enriched in RNPS1 over CASC3 have shorter half-life	75
Figure 2.15 EJC remodeling from RNPS1-EJCs to CASC3-EJCs.....	76
Figure 3.1 Schematic depicting two proposed mechanisms of EJC disassembly.....	101
Figure 3.2 FLAG-MagohE117R was expressed at near endogenous levels.....	102
Figure 3.3 A single amino-acid change in Magoh (E117R) disrupts its interaction with PYM.....	103
Figure 3.4 RNA footprints from FLAG-Magoh RIPiTs.....	104
Figure 3.5 Protein fractions from FLAG-Magoh:CASC3 RIPiTs.....	105
Figure 3.6 PCA plots comparing Magoh-CASC3 and MagohE117R-CASC3 RIPiT libraries	106
Figure 3.7 Disruption of ribosome translocation but not PYM-EJC interaction alters EJC occupancy at canonical positions.....	108
Figure 3.8 Disruption of PYM interaction does not significantly affect EJC occupancy at canonical positions.....	109
Figure 3.9 Loss of PYM interaction somewhat affects EJC occupancy on entire transcript	110
Figure 3.10 FLAG-MagohE117R:CASC3 footprints are detected less at the canonical position than FLAG-Magoh:CASC3	111
Figure 3.11 Frequency distribution of non-canonical/canonical ratios of wild-type and mutant EJCs	112

Figure 3.12 MagohE117R-EJCs have a higher non-canonical/canonical ratio than wild-type Magoh EJCs	113
Figure 3.13 Subcellular distribution of Magoh and MagohE117R.....	114
Figure 3.14 PYM effect on EJC-sensitive alternative splicing events.....	116
Figure 3.15 IPO13 interacts with both wild-type Magoh and MagohE117R.....	117
Figure 3.16 FLAG-PYM:eIF4AIII RIPiT protein	118
Figure 3.17 Formaldehyde crosslinking allows capture of FLAG-PYM:CASC3 footprints.....	119
Figure 3.18 Comparison of end of exon distribution of FLAG-PYM:eIF4AIII and FLAG-CASC3:eIF4AIII.....	120
Figure 3.19 Cumulative distribution frequency of enrichment of FLAG-PYM:eIF4AIII in canonical regions, non-canonical regions, single exons, and all genes	121
Figure 3.20 FLAG-PYM:CASC3 and FLAG-MagohE117R:CASC3are both enriched on single exon genes compared to all genes	122
Figure 3.21 Cumulative distribution frequency of single exons genes in FLAG-MagohE117R vs. FLAG-Magoh	123
Figure 3.22 Cumulative distribution frequency of single exons genes in FLAG-PYM vs. FLAG-CASC3	124
Figure 3.23 Integrated Genome Viewer (IGV) shot of EJC density on IRS4	125
Figure 3.24 GO terms enriched in FLAG-PYM.....	126
Figure 3.25 GO-terms enriched in FLAG-MagohE117R include histone genes.....	127
Figure 3.26 PYM knockdown causes reduced translation.....	128
Figure 3.27 PYM knockdown in HEK293 cells results in misregulation of many transcripts.....	129
Figure 3.28 PYM knockdown results in stabilization of many NMD transcripts.	131
Figure 3.29 Schematic of spontaneous EJC assembly in the absence of PYM interaction with Y14:Magoh.....	132
Figure 3.30 Schematic illustrating that reduced translation efficiency upon the loss of PYM interaction results in EJCs remaining bound to RNA longer	133
Figure 4.1 Non-canonical EJCs exist in the last exon and 3'UTR.....	148
Figure 4.2 EJC core proteins co-sediment with polysome fraction.....	150
Figure 4.3 Puromycin-treated EJC core RIPiT fractions.....	152
Figure 4.4 RNA footprints from puromycin treated RIPiTs.....	154
Figure 4.5 RNA from cycloheximide treated RIPiT.....	155
Figure 4.6 FLAG-Magoh:eIF4AIII wild-type, mutant, and cycloheximide-treated RIPiT protein fractions	156
Figure 4.7 EJC core occupancy in 3'UTR is reduced with translation inhibition	157
Figure 4.8 Exon number affects EJC occupancy at stop codon in multiple datasets.....	158
Figure 4.9 Exon number influences the average EJC counts in 3'UTR.....	159
Figure 4.10 Cytoplasmic EJCs are more abundant in 3'UTR than nuclear EJCs	160
Figure 4.11 CASC3-EJCs are more abundant in the 3'UTR than RNPS1-EJCs	161
Figure 4.12 EJC knockdown has minimal effect on non-canonical NMD targets	162
Figure 4.13 PYM footprints exist in the 3'UTR.....	164

Chapter 1 Introduction

1.1 RNA-binding proteins influence mRNA fate

During posttranscriptional stages, many cellular machineries act on mRNA during processes spread over a vast cellular landscape, some even separated by a physical barrier in the form of the nuclear envelope. Coupling between different steps during posttranscriptional stages is achieved primarily through RNA-binding proteins (RBPs). Some RBPs are acquired early during mRNA biogenesis and persist on RNA until subsequent steps. One of the best-studied examples of such proteins is the multi-subunit exon junction complex (EJC), which is imprinted on RNA during pre-mRNA splicing and enhances posttranscriptional events, until the EJC is removed when mRNA is fully engaged with the translation machinery in the cytoplasm (Figure 1.1).

The EJC assembles ~24 nucleotides (nt) upstream of exon–exon junctions during pre-mRNA splicing (Le Hir et al., 2000b, 2000a). Since its discovery, work from numerous laboratories has established that the EJC functions as a molecular coupler between splicing and mRNA export, cytoplasmic mRNA localization, and translation, thus enhancing mRNA expression.

1.2 Splicing enhances post-transcriptional steps via deposition of the EJC

In 1979, two studies reported that in recombinant simian virus 40 strains, intronless mRNA have drastically reduced expression at posttranscriptional level when compared with their intron-containing counterparts (Hamer and Leder, 1979; Khoury et al., 1979). Later, it was discovered in multiple vertebrate model systems that inclusion of a generic intron greatly increases heterologous gene expression (Callis et al., 1987; Le Hir et al., 2003; Palmiter et al., 1991). The biochemical underpinnings of how introns positively influence posttranscriptional gene expression were revealed much later. Comparing expression of matching intron-containing and intronless reporters in *Xenopus* oocyte nuclei revealed that spliced transcripts were exported from the nuclei to the cytoplasm far more efficiently than the unspliced, intronless equivalents (Luo and Reed, 1999). An important insight came with the observation that spliced mRNPs are distinct from unspliced ones, and have slower mobility in native gels (Luo and Reed, 1999). This was the first indication that the process of splicing alters the complement of mRNA-bound proteins.

In the early 1990s, Urlaub et al. discovered that nonsense codons in the internal exons of dihydrofolate reductase (DFHR) mRNA, but not the terminal exon, led to reduced cytoplasmic RNA levels (Urlaub et al., 1989). Subsequently, several other investigators observed that nonsense-mediated decay (NMD) of diverse mRNAs requires splicing of an intron downstream of a stop codon (Belgrader et al., 1994; Carter et al., 1996; Thermann et al., 1998; Zhang et al., 1998). To explain the link between nuclear splicing and cytoplasmic translation, these groups proposed that pre-mRNA splicing

somehow marks the exon junctions in the nucleus. This mark is recognized by the translation machinery in the cytoplasm and discriminates premature termination codons from normal ones (Belgrader et al., 1994; Carter et al., 1996; Thermann et al., 1998; Zhang et al., 1998).

The molecular components of the splicing-dependent mark were revealed using site-specific incorporation of photo-reactive crosslinking groups, several proteins were found to be recruited close to exon–exon junctions in a splicing-dependent manner (Le Hir et al., 2000b). Subsequently, regions 20–24 nt upstream of exon–exon junctions of *in vitro* spliced RNAs showed a striking protection from DNA oligo-directed RNase H cleavage, suggesting the presence of stably bound proteins at this site (Le Hir et al., 2000a). eIF4AIII, a DEAD-box RBP, which clamps the EJC on the RNA is the anchor around which the EJC is assembled (Chan et al., 2004; Ferraiuolo et al., 2004).

1.3 EJC core components and assembly

Before their assembly into an EJC, eIF4AIII and Y14:Magoh heterodimer are independently recruited to the spliceosome. All three proteins are detected in activated (C complex) spliceosomes (Bessonov et al., 2008; Reichert et al., 2002; Zhou et al., 2002). There is no evidence to suggest that EJCs are able to assemble outside of the splicing reaction *in vivo* (Gehring et al., 2009a). Consistent with their association with spliceosomes, these three proteins associate with RNA splicing intermediates, even in the absence of EJC deposition site on the upstream exon (Gehring et al., 2009a; Ideue et al., 2007; Merz et al., 2007; Shibuya et al., 2006). This suggests that EJC factor recruitment

occurs prior to and independent of EJC deposition. The mechanism of recruitment to the spliceosome is not known. One protein that could initially link the EJC core proteins to introns is IBP160, a SF1 RNA helicase present in the spliceosome C complex. IBP160 binds specific positions within introns in a sequence-independent manner (Hirose et al., 2006). IBP160 in the C complex is in contact with all the three EJC core proteins even in the absence of final EJC deposition site (i.e., -24 position), suggesting IBP160 may recruit EJC core protein to introns prior to EJC assembly. However, the molecular details of IBP160's role in EJC factor recruitment remain unknown.

At the heart of protein–protein and protein–RNA interactions of the EJC is eIF4AIII, which belongs to the eIF4A family of DEAD-box proteins, also including translation initiation factors eIF4AI and eIF4AII. eIF4AIII is composed of two RecA domains connected by a flexible linker domain. The RecA domains possess nine conserved motifs characteristic of the DEAD-box proteins, which work together to achieve ATP binding and hydrolysis, and RNA binding. An additional eight patches are conserved in eIF4AIII proteins but absent in eIF4AI (Li et al., 1999). These patches confer EJC-specific eIF4AIII function, as mutations in many of these residues impair EJC assembly and/or NMD. As evident in the EJC core crystal structure, many of these patches engage with Y14 and Magoh. Y14 (also known as Rbm8a) possesses a central RNA recognition motif (RRM) and flanking N- and C-terminal regions rich in basic residues (Bono et al., 2006). Despite its classification as an RRM, this surface of Y14 is primarily active in the protein–protein interaction between Y14 and Magoh within the Y14:Magoh heterodimer. Magoh, itself, is a highly conserved protein with no identifiable

domains. Both Y14 and Magoh are small proteins that exist as a stable heterodimer, and are extensively engaged in protein interactions within the EJC core.

Among the core EJC proteins, the recruitment of eIF4AIII to the spliceosome is most clearly understood. CWC22, an essential splicing protein, recruits eIF4AIII to the spliceosome and thereby bridges the splicing reaction with EJC deposition. CWC22 directly interacts with eIF4AIII in a protein complex distinct from the assembled EJC (Barbosa et al., 2012). CWC22 bears homology to eIF4G, possessing a middle portion of eIF4G (MIF4G) domain. The CWC22–eIF4AIII interaction is analogous to the eIF4G–eIF4A interaction (within the eIF4F complex assembled during translation initiation) (Alexandrov et al., 2012; Barbosa et al., 2012; Buchwald et al., 2013; Steckelberg et al., 2012). The MIF4G domain of CWC22 interacts with each of the two RecA domains of eIF4AIII to hold it in a conformation, such that the RNA-binding residues on the two RecA domains are faced away from each other (Buchwald et al., 2013). This conformation also prevents eIF4AIII's ability to bind ATP, preventing it from clamping onto RNA. Thus, CWC22 may sequester eIF4AIII in an inactive state until the splicing reaction is complete, reinforcing the splicing-dependent nature of EJC deposition. Interestingly, Nom1 is another eIF4G-like protein with MIF4G domain that directly binds eIF4AIII and connects it to ribosomal RNA biogenesis (Alexandrov et al., 2012). Perhaps, eIF4AIII-binding partners that interact with eIF4AIII via the same interaction surface serve to biochemically segregate the protein into its diverse cellular functions, such as ribosome biogenesis (Nom1) and mRNP assembly (CWC22).

Following the recruitment of EJC proteins to the B/C complexes, eIF4AIII, Magoh, and Y14 assemble into the EJC after the first transesterification reaction and before exon ligation (Gehring et al., 2009a). At this time the spliceosome makes extensive contacts with the last 27nt of the 5' exon. Concurrent with exon ligation, contact with regions flanking the EJC deposition site is lost– possibly due to dissociation of spliceosome-associated proteins – leaving behind the assembled EJC core (Reichert et al., 2002). It is possible that proteins bound to these regions around the EJC play a role in EJC assembly. One such candidate is Prp22, an RNA helicase that works very late in the splicing reaction to release ligated exons from the spliceosome (Schwer and Gross, 1998). The loss of helicase function of Prp22 has a dominant negative effect on EJC assembly, suggesting that its function is needed for stable EJC deposition (Zhang and Krainer, 2007).

The crystal structure of the EJC core lends insights into how the complex stably clamps on the RNA in a sequence-independent manner. This is achieved through eIF4AIII's contact with the sugar-phosphate backbone in a characteristic fashion of a DEAD-box protein. Within the assembled core, eIF4AIII residues contact phosphates within the phosphodiester linkages of as many as six nucleotides and at least three contiguous 2'OH groups of the ribose sugar (Andersen et al., 2006; Bono et al., 2006; Mishler et al., 2008). None of the bases are in contact with eIF4AIII but a fourth protein, CASC3, engages in stacking interactions with the base on the 5' end of the binding site.

Y14 and Magoh are essential components of the RNA-bound EJC and are extremely well conserved between species. However, the Y14:Magoh heterodimer alone

has very low affinity for RNA outside of its function in the EJC (Lau et al., 2003). These proteins exist as an extremely stable heterodimer where a hydrophobic region of Magoh binds to an RNA-binding domain (RBD) of Y14 with 1nM binding K_d , implying that a substantial amount of energy would be required to separate or remodel the heterodimer (Lau et al., 2003). Indeed, Y14 and Magoh protein stability is decreased when not in complex with one another (Ma et al., 2019). A point mutation in the RBD of Y14 (L118R) causes a Thrombocytopenia-absent radius syndrome. The Y14 L118R mutation abrogates Y14:Magoh heterodimer formation, and subsequently leads to more rapid protein degradation of Y14 L118R in comparison with its wild-type counterpart (Ma et al., 2019). Likewise, Magoh L136R loses interaction with Y14 and is also rapidly degraded (Ma et al., 2019).

CASC3 (also known as MLN51 or Barenstz) is the least conserved core protein (~52% sequence identity between human and zebrafish sequence when compared with >94% sequence identity between the other three proteins in these organisms). CASC3 contains a highly conserved SELOR (speckle localizer and RNA binding) domain, which is necessary and sufficient for EJC incorporation (nuclear speckle localization, eIF4AIII binding, and RNA binding) (Degot et al., 2004). The crystal structure of the EJC core was obtained with a portion of CASC3 (Bono et al., 2006). Thus, the EJC is often referred to as a tetrameric complex of eIF4AIII, Y14, Magoh, and CASC3, yet many studies cast doubt on categorizing CASC3 as a true core component. Unlike other core proteins CASC3, fails to interact with splicing intermediates and is able to join the EJC after the splicing reaction is complete and is not required for assembly of eIF4AIII, Y14,

and Magoh (Gehring et al., 2009a). CASC3 is also substiochiometric to other EJC components (Singh et al., 2012). Mouse genetics also imply that CASC3 is not an EJC core protein. Extensive characterization of EJC core protein heterozygous phenotypes in mice has shown that eIF4AIII, Magoh, and Y14 all display haploinsufficiency resulting in disruption of neuronal viability and neural progenitor proliferation ultimately causing severe microcephaly (Mao et al., 2015; Miller et al., 2017; Silver et al., 2010). However, CASC3 did not display haploinsufficiency in mouse models; homozygous loss of CASC3 was embryonic lethal and lead to a proportionate decrease in embryo size (Mao et al., 2017). Thus, there are distinctions between EJC core and CASC3 phenotypes in mice. Furthermore, while CASC3 and the other EJC core components shuttle between the nucleus and cytoplasm, CASC3 is primarily localized to the cytoplasm, while the other EJC core proteins are primarily nuclear. Nonetheless, CASC3 assembles into a stable complex with the trimeric core, which is able to withstand >250mM NaCl (data not shown). Taken together, these observations indicate that a more accurate description of the EJC is a trimeric core (eIF4AIII, Y14, and Magoh) that interacts stably, transiently, or dynamically with a myriad of peripherally interacting proteins, which include CASC3.

1.4 Peripheral Interacting Proteins

The sheer number of proteins in the EJC proteome and their wide range of function imply that the EJC is not a static complex (Table 1.1). Rather, the EJC is a dynamic hub whose protein composition may change with EJC position on mRNA, with mRNP subcellular location, and with stages in posttranscriptional gene expression. Such

compositional EJC heterogeneity is supported by distinct subcellular distribution of some peripheral EJC proteins and by significant differences in stoichiometric amounts of the core-associated EJC-interacting proteins (Mabin et al., 2018; Singh et al., 2012). Variable EJC composition may serve to diversify EJC-regulated gene networks and/or enhance efficiency of downstream posttranscriptional events. The EJC is also present in both low molecular weight and high molecular weight complexes (Singh et al., 2012). Such variable hydrodynamic properties of the EJC suggest that the EJC participates in mRNPs with different overall structure and compaction, but the protein composition and significance of these different EJC varieties have yet to be addressed.

SR- and SR-like proteins represent a category of EJC-peripherally interacting proteins which interact with the EJC in a super-stoichiometric manner (Singh et al., 2012). SR proteins are distinguished by serine-arginine (SR)- rich domains and RNA-binding motifs (RRMs). SR-like proteins are active in both constitutive and alternative splicing (Shepard and Hertel, 2009), mRNA export (Huang and Steitz, 2001; Huang et al., 2003; Müller-McNicoll et al., 2016), genome stabilization (Li and Manley, 2005), translation (Sanford et al., 2004), and nonsense-mediated mRNA decay (NMD) (Zhang and Krainer, 2004).

RNPS1 is an SR-like protein that binds the EJC core through its association with EJC-associated ASAP and PSAP complexes (Boehm et al., 2018). Through its interaction with the EJC core, RNPS1 prevents exon skipping by masking spurious splice sites located near canonical EJC binding sites (Boehm et al., 2018). RNPS1 is primarily a nuclear protein but shuttles between the nucleus and cytoplasm.

1.5 EJC canonical binding site

The major occupancy site of the EJC lies ~24 nt upstream of exon–exon junctions. The original observation of the EJC's binding site was elucidated through RNaseH-protection assay of a reporter (Le Hir et al., 2000b). It was widely assumed that splicing of every intron invariably leads to EJC deposition, but it was not until 2012 that two studies used complementary biochemical approaches based on crosslinking immunoprecipitation (CLIP) and RNP immunoprecipitation in tandem (RIPiT) followed by high-throughput sequencing to provide a transcriptome-wide view of EJC deposition in human cell lines (Saulière et al., 2012; Singh et al., 2012). Both studies revealed that the major *in vivo* EJC deposition site lies ~24 nt upstream of exon–exon junctions, and that these canonical EJC (cEJC) sites were devoid of any sequence motifs, indicating sequence-independent deposition.

The EJC's positioning is likely to be dictated by the three dimensional (3D) architecture of the spliceosome, which may restrict access of eIF4AIII to phosphates and 2'OH groups on RNA at the deposition site. Any factors that mask these functional groups at the preferred deposition site, such as binding of competing protein factors or RNA secondary structure, can cause the deposition site to shift upstream or downstream with concurrent drop in degree of EJC occupancy (Mishler et al., 2008). Indeed, RNA secondary structure at the –24 position correlates to decreased EJC occupancy *in vivo* (Saulière et al., 2012; Singh et al., 2012). Both of these studies revealed that *in vivo* EJC occupancy at the –24 nt position is not absolute; approximately 20% of exon–exon

junctions within transcripts are free of EJC binding (Saulière et al., 2012; Singh et al., 2012). Even among the occupied sites, for a given mRNA, the EJC occupancy is variable from exon to exon. In drawing conclusions regarding EJC deposition being variable or invariable at an exon junction, it is important to be mindful that each of these approaches (CLIP-seq and RIPiT-seq) measures only the presence or absence of the EJC at a position within a population of transcripts and thus, the question remains if every splicing event leads to EJC deposition.

1.6 Non-canonical EJCs

A major surprise from in vivo EJC occupancy landscapes was that, in addition to the canonical -24 nt position, the EJC is also detected at non-canonical (nc) positions (Figure 1.2). The fact that ncEJCs were identified both by CLIP-seq (Saulière et al., 2012) and by RIPiT-seq (Singh et al., 2012) clearly establishes them as *bona fide* EJCs. Remarkably, ~40–50% of EJC-associated RNA sequences map to non-canonical sites. Unlike the strict positioning of the cEJC, ncEJCs do not occupy a preferred position with respect to exon 3' ends (or any other gene structure landmark). The discovery of ncEJCs significantly altered our view of EJC occupancy and has major implications for EJC functions, yet we have only limited understanding of their origin and function. Since the original studies investigating endogenous EJC core binding sites, other studies have concentrated on the binding sites of some EJC peripherally interacting proteins (Hauer et al., 2016). Transcriptome-wide investigation of the binding of CASC3 with CLIP-seq revealed that CASC3 binds primarily at the expected -24 nt position (Hauer et

al., 2016). This suggests that, although CASC3 may not be a core component of the EJC, it is recruited to RNA via the assembled EJC and forms a stable complex. On the other hand, RNPS1 binding sites, which were also investigated in the same study, were largely associated with non-canonical positions. This result may be indicative of the different efficiencies of RNPS1 and CASC3 to crosslink to RNA, but suggests still that different EJC peripheral proteins incorporate into EJCs that have different RNA binding patterns.

Two models have been proposed to explain the association of EJCs with non-canonical sites. One interpretation is that ncEJCs are eIF4AIII-containing complexes that assemble with aid from SR proteins. In support of this, both Saulière et al. and Mabin et al. observed that like canonical sites, ncEJC sites frequently have an SRSF1-binding sites in immediately adjoining regions (Mabin et al., 2018; Saulière et al., 2012). Thus, SR proteins or other RBPs that interact with EJC could positively affect its assembly at non-canonical sites. Although in vitro studies showed that some flexibility exists in the choice of EJC site, this model does not explain how spliceosome can deposit ncEJCs as far as >50nt away from canonical site (Mishler et al., 2008). Singh et al. observed that unlike cEJC, ncEJC sites are enriched in sequences recognized by multiple SR proteins (SRSF1, SRSF3, and Tra2a) (Singh et al., 2012). A second interpretation is that ncEJCs are binding sites of EJC-associated RBPs (and not eIF4AIII) that associate with cEJC via protein–protein interactions possibly within a packaged three-dimensional RNP. In support of this model, Singh, et al. reported on the ability of EJCs to multimerize into megadalton-sized structures. Metkar, et al. extended this observation further, developing a method for probing the 3-dimensional conformation of RNP structures by combining

RNase protection and proximity ligation of protected RNA ends (Metkar et al., 2018). This technique revealed that mRNAs are compacted by their associated proteins into linear, rod-like structures (Metkar et al., 2018). Whether these non-canonical EJC binding patterns are associated with different levels of mRNP compaction or large-scale biochemical changes in the EJC's complement of peripheral proteins has yet to be addressed. It is quite possible that these two interpretations are not mutually exclusive, and the ncEJCs are likely explained by a combination of these and possibly other models. Nevertheless, *in vivo* experiments show that there is more to the story of EJC deposition and composition than *in vitro* experiments informed.

In addition to internal exons, non-canonical EJCs are also detected in final exons and within the 3'UTR of mRNA. Our current understanding of EJC deposition on RNA requires a downstream splicing event. Thus, the presence of EJCs in regions devoid of such a downstream splicing event is puzzling. Furthermore, EJCs positioned downstream of a stop codon (i.e. EJCs in the 3'UTR) would be expected to trigger rapid mRNA turnover via NMD (Mabin et al., 2018). How EJCs come to be positioned in 3' stretches of mRNA and how the cell is able to clear these EJCs to prevent aberrant mRNA decay is not known.

1.7 The role of EJCs in mRNP packaging

Another process that coincides with co-transcriptional RNA processing is the incorporation of newly transcribed/processed RNA into mRNPs. Many RBPs such as SR proteins and TREX complex, which are co-transcriptionally recruited to mRNA, are also

known to play a role in RNP packaging. Such a role has also been proposed for the EJC, based on findings that endogenous EJCs exist mainly as multimers within high-molecular-weight complexes (Figure 1.3). In these >2 megadalton size complexes, multimeric EJCs also partner with SR proteins (mainly SRSF1, SRSF3, and SRSF7) to sheath long RNA stretches up to ~100nt, thus packaging the resident RNA into higher order mRNP particles of completely unknown 3D structure (Singh et al., 2012). RNA packaging into RNPs could serve several important purposes. As compared to a naked, stretched out RNA, a packaged RNA can exist in a compact 3D structure more favorable to efficient transport through the crowded cell interior. Estimates suggest that packaged mRNAs are compacted in length ~10-fold (yeast mRNPs) to ~200-fold (Balbiani ring mRNPs) (reviewed in (Singh et al., 2015)), yet the impact of packaging on transport kinetics and gene expression remains unknown.

A more established function of mRNP packaging is its role in maintaining genome integrity. Nascent RNA has a propensity to hybridize to ssDNA within transcription bubbles leading to the formation of R-loops, which are hotspots for recombination and translocation ultimately leading to genomic instability. Co-transcriptional sequestering of RNA into RNPs limits RNA–DNA hybrids and hence R-loop formation. The depletion of SRSF1 and TREX components from cells leads to increased recombination rates and genomic instability (Domínguez-Sánchez et al., 2011; Li and Manley, 2005). Intriguingly, the effects of SRSF1 depletion can be rescued by overexpression of RNPS1, supporting a co-transcriptional role in RNP packaging for EJC proteins (Li and Manley, 2005). Similarly, depletion of EJC core proteins has also been observed to cause genomic

instability and altered chromosome number in mouse neural progenitor cells, yet it is not known if this stems from defects in RNP packaging and increased R-loop formation (Silver et al., 2010).

1.8 EJC in nonsense-mediated decay

One of the most notable roles of the EJC is its function in NMD, a translation-dependent pathway that degrades mRNAs undergoing premature translation termination. According to current models of NMD, in vertebrates, an EJC sufficiently downstream (>50–55 nucleotides) of a termination codon promotes assembly of mRNA decay-promoting complex upon termination of pioneer or early rounds of translation (Karousis et al., 2016; Lykke-Andersen and Jensen, 2015) (Figure 1.4). NMD was originally thought of as an RNA surveillance pathway that limits synthesis of aberrant proteins by rapidly degrading RNAs that have acquired nonsense codons as a result of mutations or processing errors. We now understand that NMD also serves as an important posttranscriptional gene regulation mechanism, and controls between 10 and 25% of the mammalian transcriptome (Chan et al., 2007; Lykke-Andersen and Jensen, 2015). It is important to note that while vertebrate NMD is largely EJC dependent, NMD can still occur without a downstream EJC in all organisms tested. Thus, EJC is more of a stimulator of the NMD pathway and not an essential component.

A close examination of the available evidence from eukaryotes supports a notion that EJC's role in NMD has evolved as a function of increased intron content within genomes. Intron-rich genomes of fungi encode the three EJC core proteins eIF4AIII,

Y14, and Magoh, whereas these proteins are absent in intron-poor *Saccharomyces cerevisiae* genome. However, there exists conflicting evidence for EJC's requirement for NMD among intron-rich fungi. EJC seems dispensable for NMD in the fission yeast but is essential for NMD of 3'UTR intron-containing genes in *Neurospora crassa* (Wen and Brogna, 2010; Zhang and Sachs, 2015). Notably, the EJC independence of NMD in the fission yeast is based on artificial GFP-based reporters whereas in *Neurospora* NMD of natural 3'UTR intron-containing mRNAs is EJC-dependent. Similarly, while reporter gene assays have suggested that NMD in invertebrates (*D. melanogaster*, *Caenorhabditis elegans*) is EJC-independent, more recent studies of natural intron-containing mRNAs have shown that fly NMD can be EJC-dependent (Gatfield et al., 2003; Longman et al., 2007). In *Drosophila* S2 cells, natural 3'UTRs with introns can enhance NMD in an EJC-dependent manner (Saulière et al., 2010). In the same cells, 3'UTR introns were found to be more prevalent in natural NMD targets when compared with nontargets (Hansen et al., 2009). These data are consistent with studies in zebrafish and mammalian cells, where mRNAs with 3'UTR introns are generally EJC-dependent for efficient NMD. Thus, the EJC function in NMD is likely to be more prevalent in the eukaryotic tree than currently appreciated.

Furthermore, the extreme conservation of the EJC proteins throughout eukaryotic evolution could be explained, at least partly, due to their role during NMD to monitor functional versus nonfunctional outcomes of intron splicing, and to govern regulatory switches based on splicing linked NMD. According to the current models of vertebrate NMD, the EJC functions at least at two different steps. First, it aids in assembly of a

complex consisting of Upf1, Upf2, and Upf3 proteins that discriminates a premature termination event from a normal one. The assembly of this complex proceeds as a series of steps initiated by Upf1 recruitment to the terminated ribosome by the release factors eRF3 and eRF1, and also the recruitment of Upf1-regulator, Smg1 kinase. This complex, termed the SURF complex, morphs into a decay-inducing (DECID) complex upon recruitment of Upf2 and Upf3. Upf1 interaction with Upf2 causes a conformational change in Upf1 leading to activation of its helicase function and its phosphorylation by Smg1 (Karousis et al., 2016; Yamashita, 2013). This activated form of Upf1 attracts to mRNP yet another set of factors that promote mRNA degradation either by decapping and deadenylation (Smg5/7) or by endonucleolytic cleavage (Smg6). Additionally, phosphorylated Upf1 prevents new translation initiation by directly interacting with eIF3 to prevent assembly of 80S translation initiation complexes (Isken et al., 2008). When an EJC is present downstream of a terminated ribosome, NMD is much more efficient as the EJC promotes assembly of the DECID complex. In this scenario, the Upf2-UPF3 complex bound to the EJC via Upf3's EBM is situated in a vantage position downstream of a terminated ribosome to activate Upf1. The second function of the EJC in NMD could be to provide additional contact between the Smg6 endonuclease and the NMD bound mRNP (Kashima et al., 2010). It would be expected that Upf3 is removed from the EJC for Smg6 to interact with the same EJC core surface. Some intriguing variations have emerged regarding EJC function during NMD. By artificially tethering EJC proteins downstream of termination codons, it has been noted that different EJC proteins can activate NMD independent of Upf2 or Upf3 proteins (Gehring et al., 2005). RNA-

tethered RNPS1 can induce NMD in a Upf2-dependent but Upf3-independent fashion whereas tethered EJC core proteins can trigger NMD in a Upf2-independent manner. Therefore, unlike the original proposal of strict requirement of Upf1–Upf2–Upf3 complex formation, these data and other subsequent studies have suggested that NMD of endogenous mRNAs can proceed independently of Upf2 or Upf3 (Chan et al., 2007; Gehring et al., 2005). Such an idea of branched NMD pathways is further supported by evidence that cells derived from patients with Upf3b genetic mutations can support NMD (Tarpey et al., 2007). An attractive possibility is that the EJC composition can influence the choice of the NMD branch that is activated on a particular mRNA. However, eIF4AIII, RNPS1, and CASC3 have very similar occupancy on mRNAs targeted to NMD (Hauer et al., 2016; Mabin et al., 2018). Therefore, simple occupancy of different EJC proteins on an mRNA may have little effect on its requirement for Upf2 and Upf3b to undergo NMD. What forms the precise molecular basis of Upf2- and Upf3b-independent NMD branches, and how EJC contributes to this phenomenon, if at all, remains unknown.

1.9 EJC removal and recycling

During the pioneer round of translation, EJCs residing within the internal exons of the ORF are removed by translocating ribosomes (Dostie and Dreyfuss, 2002). Currently, available evidence from human cells suggests a model where active EJC disassembly is coupled to the translating ribosome via PYM, an interaction Partner of Y14 and Magoh. Using its C-terminal region, PYM associates with ribosomes during translation (Bono et

al., 2004). When an elongating ribosome reaches an EJC, the N-terminus of PYM interacts with Y14:Magoh residues adjacent to those used for eIF4AIII interaction (al., 2004; Diem et al., 2007; Gehring et al., 2009b). This leads to the formation of a transient intermediate that destabilizes Y14:Magoh interaction with eIF4AIII allowing for the release of ADP and inorganic phosphate, rendering eIF4AIII in its open conformation (Andersen et al., 2006; Ballut et al., 2005; Bono et al., 2004, 2006). Stable PYM-Y14:Magoh complex formation and steric hindrance prevents Y14:Magoh from re-associating with eIF4AIII. PYM is released from Y14:Magoh following the association of the PYM-Y14:Magoh heterotrimer with Importin-13 (Imp13). This leads to Imp13-mediated reentry of Y14:Magoh into the nucleus in a RanGTP-dependent manner (Bono et al., 2010). Upon EJC disassembly and ATP hydrolysis, eIF4AIII enters its open state. CASC3 has been shown to interact with this apo-form of eIF4AIII, but it is not known how eIF4AIII and CASC3 are separated (Bono et al., 2006). Eventually, eIF4AIII is re-imported into the nucleus to effectively complete the EJC life cycle. The details of eIF4AIII nuclear import have not been carefully detailed. Shibuya, et al. conducted an extensive mutational analysis of eIF4AIII, including a putative N-terminal nuclear localization signal (NLS) ¹⁴RKRLK¹⁹. This analysis concluded that the putative NLS was not required for nuclear localization nor interaction with Y14:Magoh (Shibuya et al., 2006).

New twists to this EJC disassembly model have emerged from work in *Drosophila*, where PYM does not appear to associate with ribosomes, suggesting that, in this system, PYM-mediated EJC disassembly is translation-independent (Ghosh et al., 2014).

Consistent with this, exogenously expressed PYM leads to EJC disassembly in vivo (Gehring et al., 2009b). Thus PYM-mediated EJC disassembly does not require translation, and the translating ribosome is sufficient—at least in *Drosophila*—to remove EJCs from coding stretches of RNA. Thus, there is much more to understand about the biological function of PYM in EJC disassembly and RNA maturation in mammalian cells.

PYM may be important for recycling EJC core components for nascent RNA processing events in the nucleus. Knockdown of PYM results in accumulation of EJC proteins in the cytoplasm (Gehring et al., 2009b). It is possible that PYM is important for the “rescue” EJCs from 3’UTRs and non-coding stretches of RNA. EJC core components presumably need to be recycled from the cytoplasm to the nucleus. This is based on several observations. First, all three core components are known to shuttle between the nucleus and cytoplasm. Second, the Y14:Magoh heterodimer is reimported into the nucleus by Imp13. And finally, the number of exon junctions present in the transcriptome at steady state requires that EJC proteins be reused. The quantity of EJC core proteins present in endogenous cell lysates (eIF4AIII: 12,000, Y14:Magoh: 40,000) are at least an order of magnitude fewer than the number of exon-exon junctions (~400,000) (Gehring et al., 2009b). According to transcriptome-wide binding studies of the EJC, an EJC is deposited upstream of virtually every exon-exon junction. While it follows that EJC proteins must be recycled to ensure proper processing, this has not been investigated.

Knockdown of EJC core components causes changes in splicing, including the exclusion of exons due to recognition of spurious splice sites close to canonical sites of

EJC deposition. When EJCs do not assemble near these spurious splice sites, the spliceosome recognizes the spurious splice sites, and exons are spliced. Whether PYM's cytoplasmic disassembly function is coupled to nuclear EJC-dependent splicing events also an as yet unexplored avenue (Figure 1.5).

1.10 EJCs role in pre-mRNA splicing

The bulk of pre-mRNA splicing takes place co-transcriptionally; therefore, much of EJC deposition must occur on nascent RNAs associated with chromatin (Das et al., 2006; Hicks et al., 2006). It is now well established that transcriptional processes and chromatin structure are intimately coupled to pre-mRNA splicing and mRNP formation (Saldi et al., 2016). SR proteins, which intimately associate with the EJC and form a major mRNP component, are also tightly linked to transcriptional elongation and chromatin architecture via RNA Pol II CTD and histone tails, respectively (Zhong et al., 2009). A similar functional coupling may exist between the EJC and transcriptional processes and/or chromatin signatures. This is supported by a report showing that the EJC modulates alternative splicing by changing RNA Pol II elongation rate (Wang et al., 2014) (Figure 1.5). Depletion of EJC core proteins in HeLa cells, or their inability to assemble into the complex, leads to changes in alternative splicing patterns (Wang et al., 2014). Evidence that altered splicing patterns of some genes in cells depleted of EJC core proteins are partially reversed upon treatment with drugs that interfere with RNA Pol II elongation indicates a functional link between the EJC and the transcription elongation machinery (Wang et al., 2014). It is noteworthy that the three nuclear EJC core

proteins—together with RNPS1—were identified in complexes that purify with CDK12, a cyclin dependent kinase that functions in 3'-regions of transcriptional units and phosphorylates RNA Pol II at serine-2 within the CTD heptad repeats provided some insight into the mechanism by which the EJC can modulate splicing events through interaction with RNA Pol II (Bartkowiak and Greenleaf, 2015). Unlike other CDKs, CDK12 has extended N- and C-terminal arms containing stretches of RS dipeptides, which may serve as protein–protein interaction platforms for SR proteins. Together, these findings changed the view of the connection between EJCs, transcription, and possibly chromatin landscapes. More evidence for EJC function in splicing comes from studies in *D. melanogaster*. Multiple genetic studies have revealed that the EJC can impact pre-mRNA splicing more directly by serving as a splicing factor for neighboring introns. Two independent screens identified EJC core proteins to be necessary for splicing of long introns within the MAPK gene, which contains exceptionally long introns (up to 25kb) (Ashton-Beaucage et al., 2010; Roignant and Treisman, 2010). Both studies extended the EJC's regulation of long intron splicing beyond MAPK gene to find that exons with flanking long introns are skipped in EJC-depleted cells. This supports a role for the EJC in exon definition to facilitate splicing of long introns in flies. Interestingly, EJC depletion in *Drosophila* affects splicing, especially of heterochromatic genes with long introns, such as MAPK, once again indicating a link between chromatin environment and EJC function (Roignant and Treisman, 2010). EJC's impact on splicing in *Drosophila* also impinges on the piRNA biogenesis pathway, which keeps transposons in check in the germline and surrounding somatic tissues. The loss of one of the trimeric EJC core

proteins, Acinus, or RNPS1 leads to intron retention in piwi pre-mRNA, which causes loss of piRNA production and transposon mobilization (Malone et al., 2014). The splicing of the fourth intron in piwi pre-mRNA is especially sensitive to EJC depletion and splicing of flanking introns (Hayashi et al., 2014; Malone et al., 2014). EJC-dependent introns are often the first introns that precede an abnormally long second intron. EJC deposition on neighboring exon junctions can promote efficient recognition of splice sites, perhaps by promoting splicing functions of EJC-associated splicing cofactors SRm160, Acinus, RNPS1, and other SR proteins. All studies from *Drosophila* are in agreement that CASC3 is not required for EJC function in splicing of MAPK, piwi, or any other transcripts (Ashton-Beaucage et al., 2010). This is in contrast to the mammalian cells where CASC3—like the EJC core proteins—is important for splicing events by controlling RNA Pol II transcription rate (Wang et al., 2014). While such differences could be easily attributed to differences in CASC3 function in the two organisms, it is also possible that these different modes of EJC action in splicing depend on whether the EJC associates with CASC3 or not. The latter possibility is supported by findings in human cells that eIF4AIII, Y14, RNPS1, SAP18, and Acinus, but not CASC3, are important for the preferential expression of the longer anti-apoptotic isoform (Bcl-xL) over the smaller pro-apoptotic splice variant (Bcl-xS) from Bcl-x pre-mRNA (Michelle et al., 2012). Boehm, et al. described a mechanism whereby the EJC influences splicing through its interaction with RNPS1 (Boehm and Gehring, 2016). Knockdown of RNPS1 leads to a change in splice-site selection. This occurs through RNPS1's interaction with the EJC core through the PSAP complex. When EJCs are deposited during splicing and

recruit the PSAP/RNPS1 complex, together, these proteins mask spurious splice sites located near the canonical EJC position (Boehm et al., 2018). Similarly, EJC core proteins are important for preventing recursive splicing in humans (Blazquez et al., 2018). The EJC is required for inclusion of recursively spliced exons in the mouse brain, which explains the microcephaly phenotype exhibited in haploinsufficient Magoh mice (Blazquez et al., 2018).

Early *in vitro* work showed that EJC proteins are not necessary for efficient splicing and are thus dispensable for assembly of functional spliceosomes (Zhang and Krainer, 2007). We have come a long way in appreciating the breadth of the EJC's role in splicing, but only very recent studies have provided some insight into the mechanism through which EJC deposition influences splicing events. While it is unlikely that the EJC functions in the catalytic steps of splicing, future work should provide insights into how *in vivo* EJC functions begin soon after its assembly on nascent transcripts—much ahead of its more well-established downstream functions.

1.11 Summary

The EJC assembles ~24 nucleotides (nt) upstream of exon–exon junctions during pre-mRNA splicing. Since its discovery, work from numerous laboratories has established that the EJC functions as a molecular coupler between splicing and mRNA export, cytoplasmic mRNA localization, and translation, thus enhancing mRNA expression. The EJC ensures high fidelity of gene expression by aiding in identification and rapid elimination of nonsense mutation containing mRNAs via nonsense-mediated

mRNA degradation. The EJC serves as a binding platform for a myriad of peripherally interacting proteins. These peripherally interacting proteins influence post-transcriptional steps via their association with the EJC. PYM and the translating ribosome accomplish the removal of EJCs from RNA. Disassembly marks the end of the EJCs influence on mRNA fate. EJC subunits are then reimported to the nucleus and are deposited on nascent RNA and the cycle begins again.

Both PYM and the translating ribosome are sufficient for disassembly; however, several observations suggest the necessity of PYM in this step is somewhat dubious. For example, EJCs are apparently removed from mRNPs in the absence of PYM, as PYM knock-out has no observable impact on cell viability (Paix et al., 2017). Furthermore, PYM is nonessential in *D. melanogaster* (Ghosh et al., 2014). Others have suggested that EJC recycling is required to ensure proper EJC assembly on nascent transcripts, but the significance of EJC recycling and nuclear re-import have not been thoroughly studied (Gehring et al., 2009b).

Recent trends have also shed light on the EJC's role in influencing splice site selection in cis upon the RNAs on which it is deposited. The observation that EJCs are found at non-canonical positions such as the 3'UTR has expanded our view on how EJCs participate in the mRNP structure. The biological significance of these non-canonical EJCs remains unanswered in the field.

Chapter 2 of this thesis shows that the EJC composition changes as mRNPs migrate from the nucleus to the cytoplasm, and that CASC3 interacts with cytoplasmic, pre-translation mRNPs. With translation, CASC3-EJCs are stripped from mRNA.

Chapter 3 defines the role of PYM in the disassembly of EJCs from cytoplasmic mRNPs and identifies EJC disassembly events that require PYM function. Finally, Chapter 4 explains the origin of non-canonical EJCs positioned in the 3'UTR as EJCs displaced from upstream binding sites due to the action of the translating ribosome.

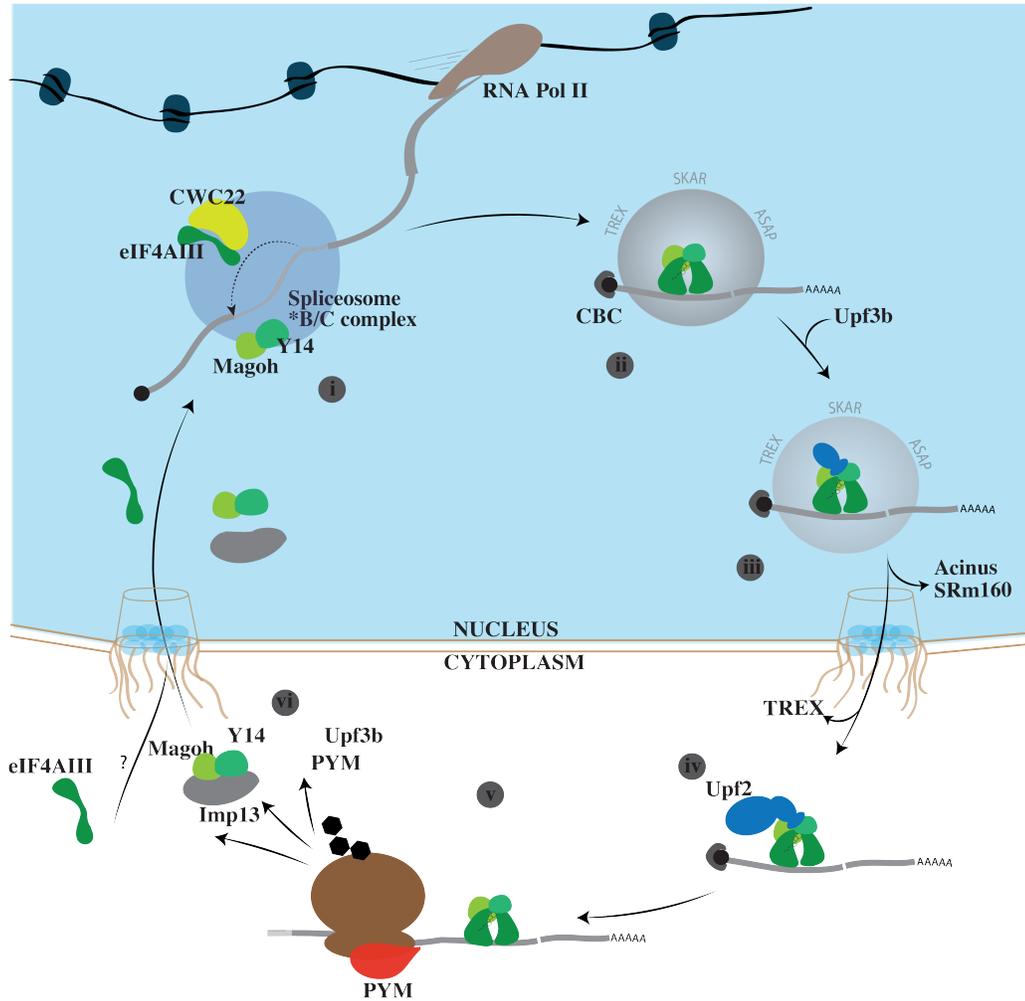


Figure 1.1 EJC Lifecycle

The exon junction complex (EJC) cycle: (i) The EJC subunits are recruited co-transcriptionally to the activated spliceosome. The recruitment of eIF4AIII occurs via interaction with CWC22. (ii) Upon exon ligation and release, the EJC is stably bound to messenger RNA (mRNA) 24 nt upstream of exon–exon junctions. In addition to the core, many peripheral factors are deposited in a splicing-dependent manner (indicated by a gray oval around EJC core). Specific examples of nuclear peripheral factors are indicated: TREX, ASAP, and Upf3b. (iii) Some peripheral factors are removed prior to export, while others remain bound and travel with the EJC into the cytoplasm. (v) 40S ribosome subunit-

Figure 1.1 EJC lifecycle (continued)

associated protein, PYM (red), allows for translation-dependent disassembly of the EJC during the first (pioneer) round of translation. (vi) Following disassembly, the Y14:Magoh heterodimer is reimported to the nucleus via Imp13. The mechanism of eIF4AIII nuclear import are unknown. Recycled subunits can reenter the EJC cycle. Thick gray lines, exons; thin gray line, intron; thick black lines, DNA; black circle, mRNA cap

Figure adapted from Woodward et. Al (Woodward et al., 2017) with permission from Wiley Interdisciplinary Reviews.

Name	Alias	EJC-dependent function	Distribution	Major Domain
eIF4AIII	DDX48	All EJC functions	Shuttles	DEAD-box helicase motifs
Y14	RBM8A tsunagi	All EJC functions	Shuttles	RRM (protein interaction)
Magoh	Mago nashi	All EJC functions	Shuttles	
CASC3	MLN51 Barentsz	Splicing; mRNA localization; translation; NMD	Cytoplasm Perinuclear Nuclear Speckles	SELOR
RNPS1		Splicing; translation; NMD	Shuttles	RRM; SR-rich
SAP18		Splicing	Shuttles	
Acinus		Splicing	Nuclear	
Pinin	PNN	Splicing	Nuclear, desmosome	
PYM	WIBG	Translation; EJC disassembly	Shuttles	eIF2A-like region
Upf3a	RENT3a	Translation; NMD inhibition	Shuttles	RRM; EBM
Upf3b	RENT3b	Translation; NMD	Shuttles	RRM; EBM
SKAR	POLDIP3	Translation; mRNA export	Shuttles	RRM
Aly/REF	Ref2	mRNA export	Shuttles	RRM; SLiM
UAP56	DDX39b	mRNA export		DEAD-box helicase motifs
UIF		mRNA export	Nuclear	UAP56-binding motif (UBM)
URH49	DDX39a	mRNA export	Nuclear	DEAD-box helicase motifs
CWC22		eIF4AIII recruitment to spliceosome	Nuclear	MIF4G domain
SRm160	SRRM1	mRNA export	Nuclear	PWI domain
IBP160	Aquarius	EJC protein recruitment and deposition	Nuclear	

Table 1.1 EJC-interacting proteins

Table information adapted from Woodward et. Al (Woodward et al., 2017) with permission from Wiley Interdisciplinary Reviews.

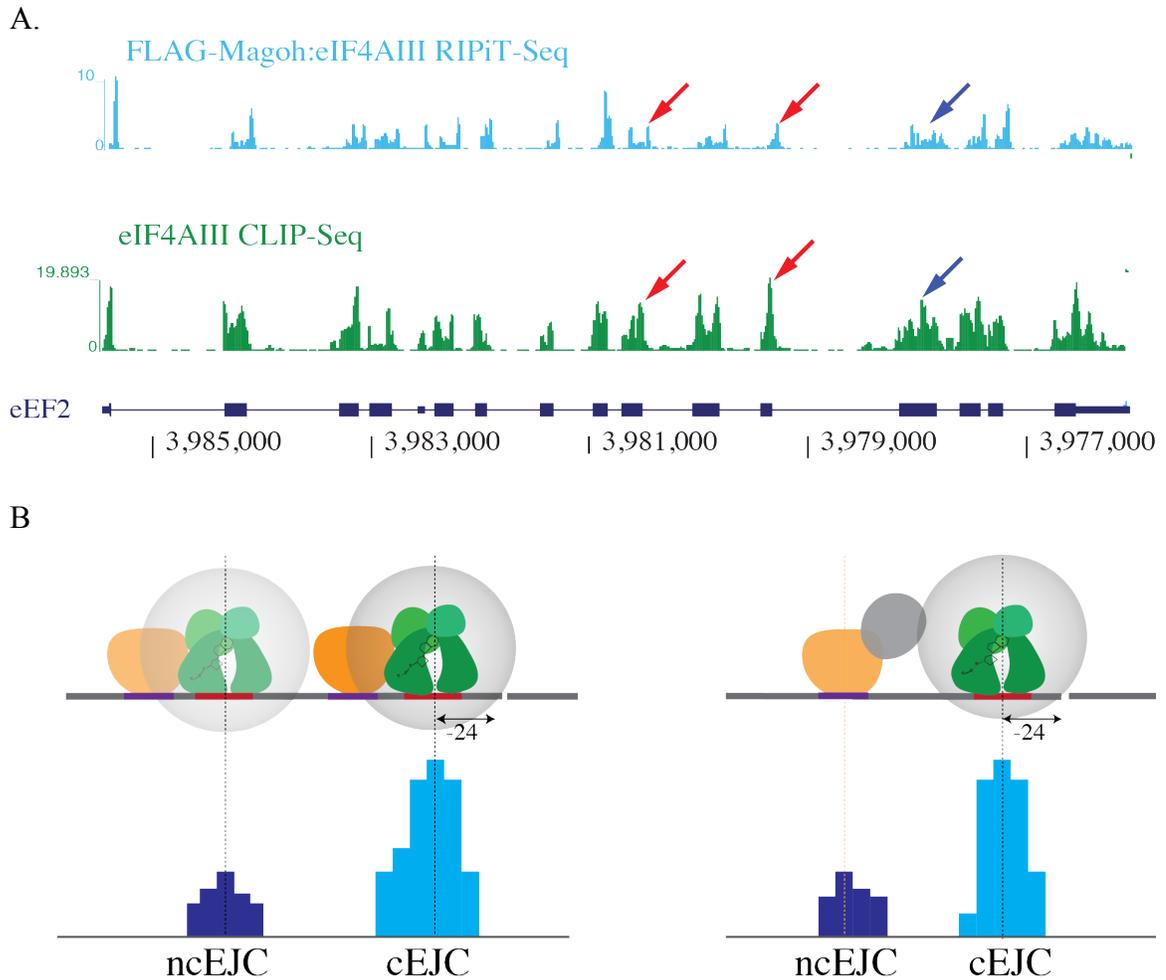


Figure 1.2 Non-canonical EJCs

- A. Genome browser shot of the gene eEF2 and the EJC footprints detected in FLAG-Magoh:eIF4AIII RIPiT-Seq (above) and eIF4AIII CLIP-Seq (below). Red arrows highlight some of the canonical EJC footprints at the \sim -24 nt position. Blue arrows highlight the presence of EJC footprints at non-canonical positions.
- B. Two possible models for origin of non-canonical exon junction complexes (ncEJCs). (left) Both cEJC and ncEJC consist of EJC core factors, and could be stabilized by nearby EJC-interacting RNA-binding proteins (RBPs; e.g., SR proteins; orange shape). (right) Unlike canonical EJCs, which consist of the core factors, the ncEJCs are binding sites for EJC-binding proteins, such as SR proteins. Interactions between cEJCs and ncEJCs may be bridged by other EJC-interacting proteins (gray oval). The site on RNA directly in contact with eIF4AIII is shown in red, whereas sites of EJC-interacting RBPs are in purple. Figure adapted from Woodward et. Al (Woodward et al., 2017) with permission from Wiley Interdisciplinary Reviews.

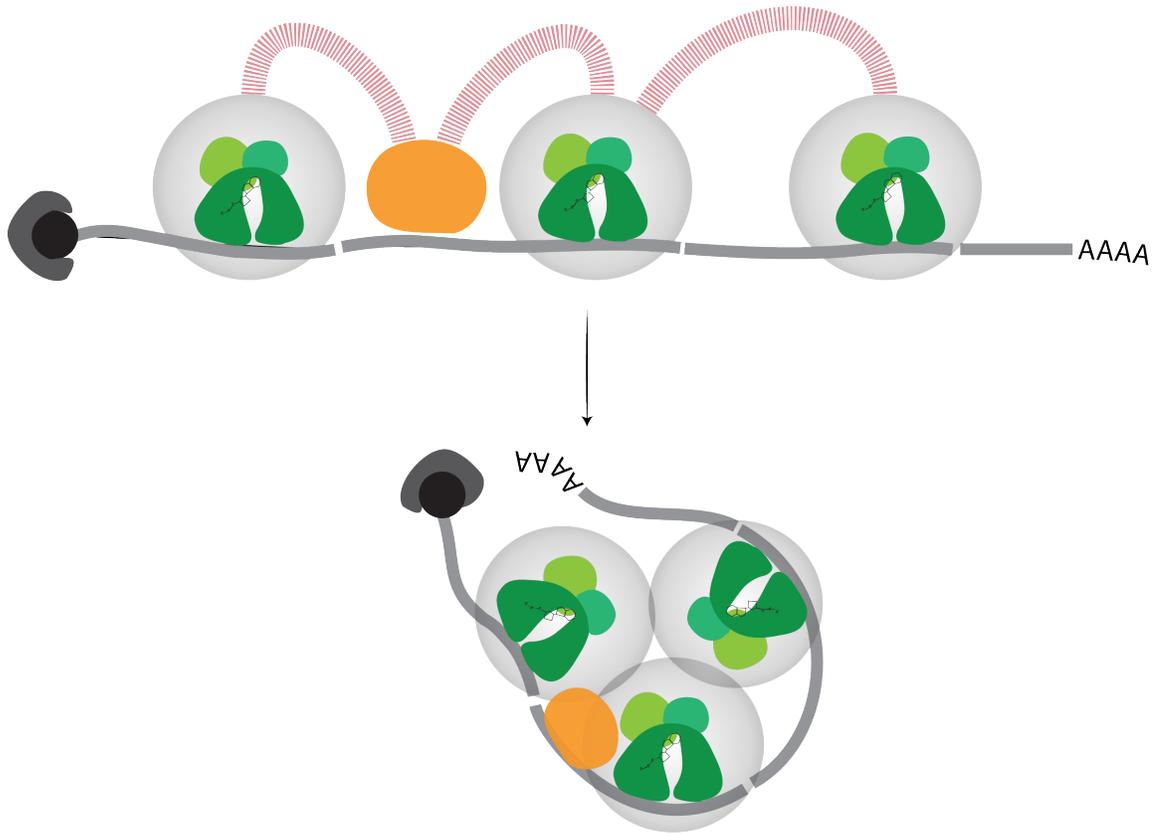


Figure 1.3 Role of EJCs in mRNP packaging

Schematic depicting how EJCs within the same mRNP interact with one another and with other RBPs resulting in a complex and compacted 3-dimensional structured mRNP. Low complexity interactions between EJCs and SR- and SR-like proteins' may result in complex 3-dimensional RNP structures. SR- and SR-like proteins are represented in orange. Low complexity protein-protein interactions are represented in red. Figure adapted from Woodward et. Al (Woodward et al., 2017) with permission from Wiley Interdisciplinary Reviews.

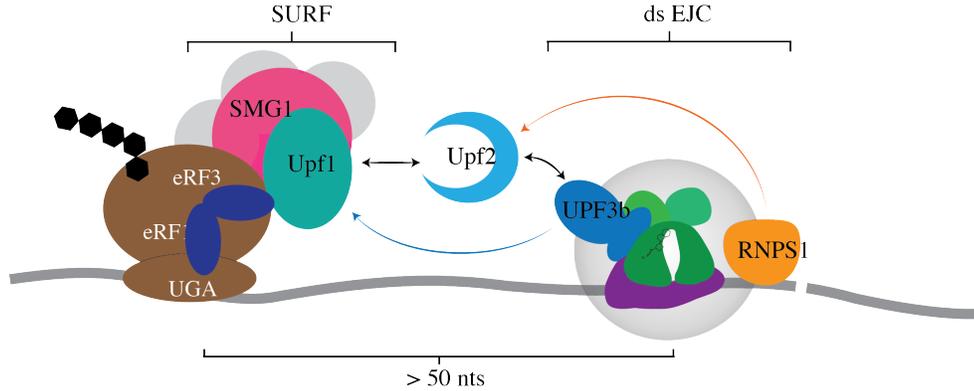


Figure 1.4 EJs enhance nonsense-mediated decay (NMD)

Schematic of an EJC positioned at least 50nt downstream of a termination codon (dsEJC) recruiting NMD machinery. The EJC core interacts directly with Upf3b, which in turn interacts with Upf2. Upf2 bridges the dsEJC to the SURF complex (Upf1 and Smg1), which is recruited to the terminating ribosome (interactions illustrated by solid black arrows). Alternatively, SR-proteins and SR-like (such as RNPS1) may communicate via Upf2 interaction with the SURF complex to initiate NMD (interaction illustrated by orange arrow). Studies have also suggested that Upf3b and Upf2 may also be competent to interact with the SURF complex independently of one another to trigger NMD (Chan et al., 2007; Tarpey et al., 2007).

Figure adapted from Woodward et al. (Woodward et al., 2017) with permission from Wiley Interdisciplinary Reviews.

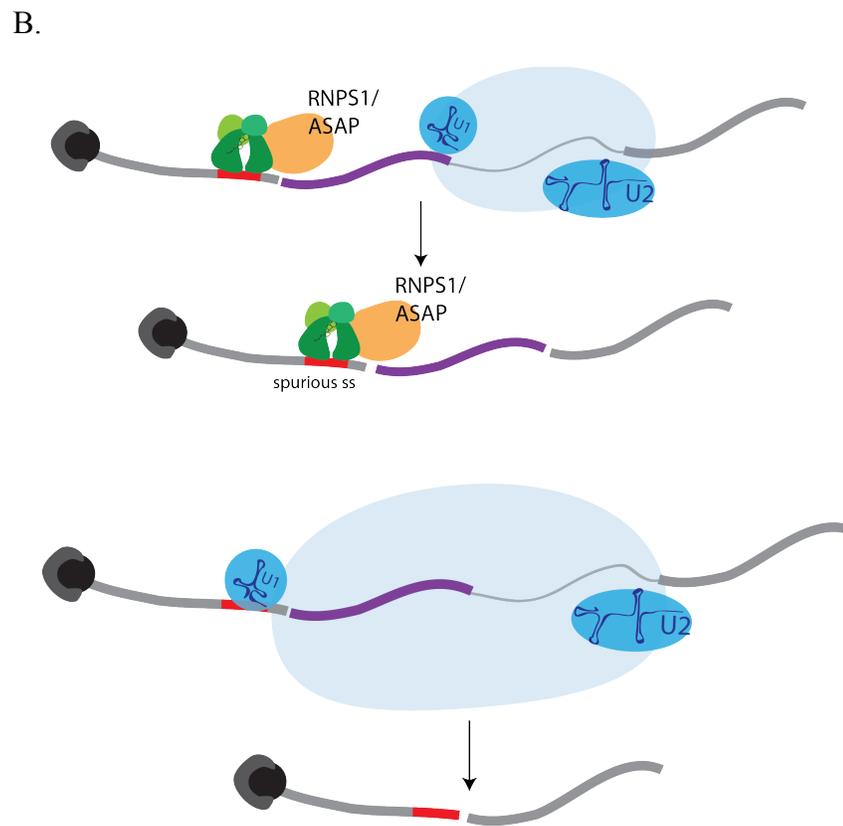
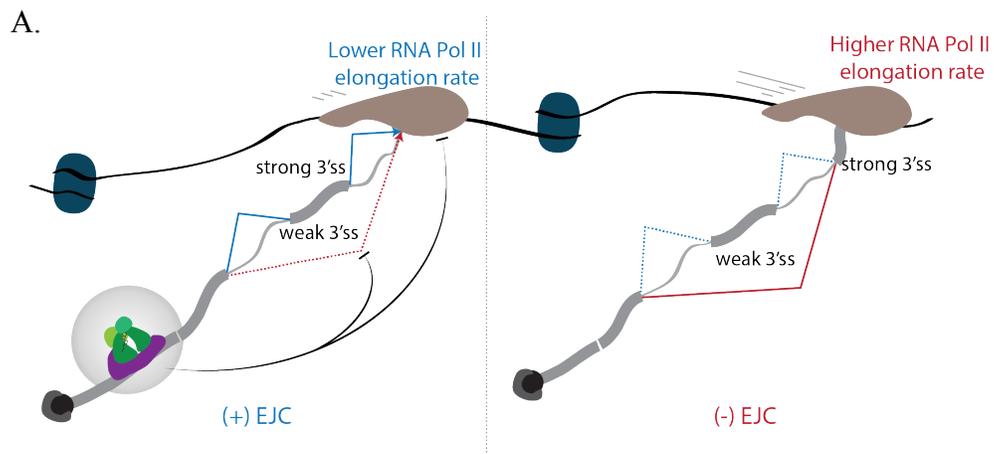


Figure 1.5 EJC deposition affects splicing

Figure 1.5 EJC deposition affects splicing (continued)

A. Schematic illustrating model of how EJC modulates RNA polymerase II (Pol II) elongation rate to control alternative splicing. The presence of EJC (left) slows down RNA Pol II elongation (shown by the inhibitory arrow) to introduce a delay in transcription of downstream strong splice site. This provides more time for splicing to occur at available weak splice sites (exons connected by solid blue lines). In the absence of EJC (right), faster RNA Pol II elongation rate reveals the stronger downstream splice sites more rapidly, leading to exon skipping due to the use of downstream splice site (exons connected by solid red lines) and omission of weaker splice sites (shown by dotted blue lines). Instead of exon skipping, such a scenario can also lead to retention of first intron shown in the schematic if 5'-splice site of intron 2 is used (not shown).

Figure adapted from Woodward et. Al (Woodward et al., 2017) with permission from Wiley Interdisciplinary Reviews.

B. Deposition of EJCs can mask spurious splice sites and prevent their recognition by the splicing machinery (above). Loss of EJC deposition results in derepression of spurious splice sites (below).

Chapter 2 EJC composition changes as mRNPs move from nucleus to cytoplasm¹

2.1 Abstract:

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. CASC3 is enriched on mRNAs that have low translation efficiency or are targeted to the endoplasmic reticulum for translation.

2.2 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 (Hentze et al., 2018) and is

¹ This chapter is based on a published article Mabin et al., 2018 and has contribution from Justin Mabin and Robert Patton.

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). 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; 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 eIF4AIII is locked in place by 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). 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; 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.

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 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 intra-nuclear 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. 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.

2.3 Materials and Method

2.3.1 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 μ g of a pcDNA5-FRT/TO-FLAG construct along with 1.8 μ 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 μ g/ml) and Hygromycin (100 μ 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.

2.3.2 Endogenous and FLAG-tag 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 µg/ml. Following five minute incubation on ice, cell lysates were cleared at 15,000 × g. Complexes were then captured on Protein A/G Dynabeads (Life Technologies) conjugated to IgG, α-eIF4AIII, α-CASC3 or α-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 × with 1 mL IsoWB and eluted in IsoWB supplemented with 125 µg/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 µg/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.

The following primary antibodies were used for endogenous immunoprecipitation and western blotting: rabbit polyclonal anti-CASC3 (Bethyl) 4 µg for immunoprecipitation and 1:250 dilution for western blotting; anti-RNPS1 (Sigma) 4 µg for immunoprecipitation; anti-eIF4AIII (Bethyl) 4 µg for immunoprecipitation and 1:2000 dilution for western blot; anti-hnRNPA1 (Santa Cruz Biotechnology) 1:1000 dilution for western blot; anti-Acinus 1:2000 dilution for western blot; anti-Magoh 1:500 for western blot; anti-Y14 (Bethyl) 1:1000 dilution for western blots; anti-SAP18 (Santa Cruz Biotechnology) 1:500 dilution for western blots.

2.3.3 Western blot imaging and 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.

2.3.4 RNA:protein immunoprecipitation in tandem (RIPiT)

RIPiTs were carried out with and without formaldehyde crosslinking following a detailed protocol available in (Gangras et al., 2018; Singh et al., 2014). Briefly, for native RIPiTs, total extracts from five 15-cm plates were prepared by disrupting cells via sonication using Branson Digital Sonifier-250 at 10% amplitude using a Microtip for a total of 30 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

$\mu\text{g/ml}$ Aprotinin (Sigma, A1153-10MG), 1 $\mu\text{g/ml}$ Leupeptin (Sigma, L9783-5MG), 1 μM Pepstatin (Sigma, L4265-5MG), 1 mM PMSF (Sigma, P7626-5G)] supplemented with 150mM NaCl. Extracts were cleared by centrifugation at $15,000 \times g$ 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 eight 15-cm plates in denaturing lysis buffer. 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 $\mu\text{g/ml}$ FLAG peptide with gentle shaking at 4°C for 2 hr. The recovered elution volume was adjusted to 400 μL 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-eIF4AIII (Bethyl A302-980A, 10 $\mu\text{g/RIPiT}$), anti-CASC3 (Bethyl A302-472A, 8 $\mu\text{g/RIPiT}$), anti-RNPS1 (HPA044014-100UL, 8 $\mu\text{g/RIPiT}$). 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 $\mu\text{g/ml}$ CHX for 3 hr prior to harvesting. CHX

was included at the same concentration in PBS (for washes before lysis) and cell lysis buffers.

2.3.5 High-throughput sequencing library preparation

For RIPiT-seq, RNA extracted from ~80% of RIPiT elution was used to generate strand-specific libraries. For RNA-seq libraries, 5 µg 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 5nt barcode sequence, and ends in GG at the 5' end. Following RT, the extended RT primer is gel purified, circularized using CircLigase (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 100nt read lengths).

2.3.6 Adapter trimming and PCR removal

After demultiplexing, fastq files containing unmapped reads were first trimmed using Cutadapt. A 12nt sequence on read 5' ends consisting of a 5nt random sequence, 5nt 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 TGAATTCTCGGGTGCCAAGG was removed from the 3' end as possible. Any reads less than 20nt in length after trimming were discarded.

2.3.7 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.

2.3.8 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. 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).

2.3.9 Differential enrichment analysis

Differential analysis of exons and transcripts between CASC3 and RNPS1 pull down was conducted with the DESeq2 package in R (Love et al., 2014). 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.

2.3.10 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.

2.3.11 Ribosome occupancy and mRNA half-life estimates

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

2.3.12 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.

2.3.13 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 μ 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 μ L fractions. Proteins were precipitated using TCA and resuspended in 15 μ L of 1 \times SDS loading buffer for analysis on 12% SDS-PAGE followed by western blotting.

2.4 Results

2.4.1 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 eIF4AIII or the substoichiometric EJC proteins RNPS1 and CASC3 from RNase-A-treated HEK293 total cell extracts. As expected, eIF4AIII IP enriches EJC core, as well as all peripheral proteins tested. In contrast, the IPs of substoichiometric factors enrich distinct sets of proteins. CASC3 immunopurified eIF4AIII, Y14, and Magoh but not the peripheral proteins ACIN1 and SAP18 (Figure 2.1; 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 2.1). 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 eIF4AIII but not with each other (Figure 2.1). Thus, in mammalian cells, the EJC core forms mutually exclusive complexes with RNPS1 and CASC3, which we refer to as alternate EJCs.

2.4.2 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 2.2). On the other hand, CASC3 is mainly detected in lower-molecular-weight complexes purified via FLAG-Magoh. Further, FLAG-CASC3 complexes are exclusively comprised of lower-molecular-weight complexes, likely to be EJC monomers. Thus, compositional distinctions between the two alternate EJCs give rise to two structurally distinct complexes.

2.4.3 RIPiT-seq reveals that 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 (Gangras et al., 2018; 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., eIF4AIII, Magoh) and variable (e.g., RNPS1, CASC3) components (Figure 2.3). We carried out RIPiTs from HEK293 cells by expressing FLAG-tagged protein at near endogenous levels (Figure 2.4) 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., 2010; Singh et al., 2012). However, to capture unperturbed, steady-state populations of RNPS1- and CASC3-EJCs, we performed RIPiT-seq without translation inhibition.

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 cell lysis. A strong correlation was observed between crosslinked and uncrosslinked samples (data not shown). All analysis presented below is from two well-correlated biological replicates of formaldehyde crosslinked RIPiT-seq datasets of RNPS1- and CASC3-EJC. 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).

2.4.4 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, the two proteins are often detected on the same sites on RNA, leading to their similar apparent occupancy on individual exons as well as entire transcripts

(Figure 2.5 A and B). These results suggest that the two alternate EJC factors bind to two distinct pools of the same RNAs.

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, 2001) (Figure 2.6A). 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) 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; Figure 2.6C). 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 2.6C). Consistent with CASC3 shuttling between the nucleus and cytoplasm (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. Therefore, while at steady-state, CASC3 binds primarily to cytoplasm-localized RNPs, it does not exclusively bind cytoplasmic RNPs.

2.4.5 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 2.10). 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 2.7). 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 (Figure 2.8A). A search for functionally related genes in the two sets revealed that each contains diverse groups (Figure 2.9A). Under normal conditions, the largest and most significant CASC3-EJC-enriched group encodes signal-peptide bearing secretory/membrane proteins (Figure 2.9A), which has significantly higher ribosome occupancy as compared to all transcripts (data not shown). We reason that, despite their higher ribosome occupancy, the “secretome” (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 et al., 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 2.10). Presumably, the increased CASC3 occupancy on ER-targeted transcripts reflects the prolonged time in which these transcripts exist in the untranslated state, which allows for EJCs greater time to undergo the compositional switch from RNPS1 to CASC3. When we considered only the cytosol-translated transcripts (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 2.8B).

Another functional group enriched in CASC3-EJC under normal conditions comprises the ribosomal protein (RP)-coding mRNAs (Figure 2.9A). Strikingly, transcripts encoding ~50% of all cytosolic ribosome proteins, as well as 13 mitochondrial ribosome subunits, are among this group. Although RP mRNAs are among the most well translated in the cell, a sizeable fraction of RP mRNAs exist in a dormant untranslated state (Geyer et al., 1982; Meyuhas 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 2.9B). 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 2.8). 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.

As reported by Hauer et al., RP mRNAs are depleted of CASC3-EJC upon translation inhibition (Figures 2.9). 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 2.9B) (Hauer et al.,

2016). 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).

2.4.6 Translation of 5' TOP mRNAs is dependent on CASC3

CASC3-EJCs are enriched on RP mRNAs. Many RP mRNAs are part of a distinct class of transcripts characterized by a 5' terminal oligo pyrimidine motif, or 5' TOP (Meyuhas and Kahan, 2015). 5' TOP motifs are located immediately downstream of the m⁷G cap of mRNA and their translation is regulated by the mTOR signaling pathway. An important target of mTOR regulation is LARP1, which binds to the 5' TOP motif directly to repress translation during cellular stress (Fonseca et al., 2018). Since CASC3 is enriched on TOP mRNAs and has been previously shown to regulate translation (Chazal et al., 2013), we sought to investigate the biological role for CASC3 occupancy on 5' TOP motif containing mRNAs. Both CASC3 and LARP1 localize to stress granules (Baguet et al., 2007; Hopkins et al., 2016). Consistently, immunofluorescence showed colocalization of CASC3 and LARP1 in granule-like punctae upon sodium arsenite or cycloheximide treatment (Kedersha and Anderson, 2007) (Figure 2.11 A and B). Additionally, CASC3 interacts with LARP1 in both an RNA-dependent and RNA-independent manner (Figure 2.11C). The mTOR pathway regulates cell proliferation by

sensing environmental cues, thus we reasoned that CASC3's interaction with LARP1 might be sensitive to cellular stress. Indeed, cell stress induced by serum starvation increases LARP1 co-IP with CASC3 1.4-fold (Figure 2.12).

We then reasoned that in binding to TOP mRNAs, CASC3 may function with LARP1 to repress translation. Consistently, depletion of CASC3 by siRNA increases abundance of a TOP mRNA encoded proteins: eIF3A and rps6, by 1.23 and 3.5- fold, respectively (Figure 2.13). Taken together, these data suggest that CASC3 may function in translation of TOP RNAs specifically through interaction with LARP1. However, the mechanism through which CASC3-EJCs act in LARP1 and mTOR-mediated translation of TOP mRNAs awaits further investigation.

2.4.7 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 2.14). 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) are enriched in CASC3-EJC upon translation inhibition (Figure 2.9A) (Schwanhäusser et al., 2011).

2.5 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.

2.5.1 EJC Composition and mRNP Structure

Our findings suggest that, when EJCs first assemble during co-transcriptional splicing, the core complex consisting of eIF4AIII, Y14, and Magoh engages with SR proteins and SR-like factors, including RNPS1. 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. This network of interactions bridges adjacent and distant stretches of mRNA, winding the mRNA up into a higher-order structure, which is characteristic of pre-translation RNPs purified from human cells via the EJC core or RNPS1 (Metkar et al., 2018; Singh et al., 2012). 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., 2013). 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 SR-rich and RNPS1-containing higher-order EJCs persist through much of their nuclear lifetime. 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 (Figure 2.15). 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. The CASC3 bound form of the EJC is likely the main target of translation dependent disassembly, although RNPS1-EJC may also undergo similar disassembly.

2.5.2 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. 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 eIF4AIII, Y14, and Magoh lead to premature neuronal differentiation and apoptosis in the mutant brains (Mao et al., 2015, 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 eIF4AIII 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).

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. 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.

In Chapter 3, this thesis will specifically discuss the role of translation on EJC occupancy and discuss the role of the purported EJC-disassembly factor PYM. CASC3-EJCs are used as a proxy in Chapter 3 for cytoplasmic RNAs.

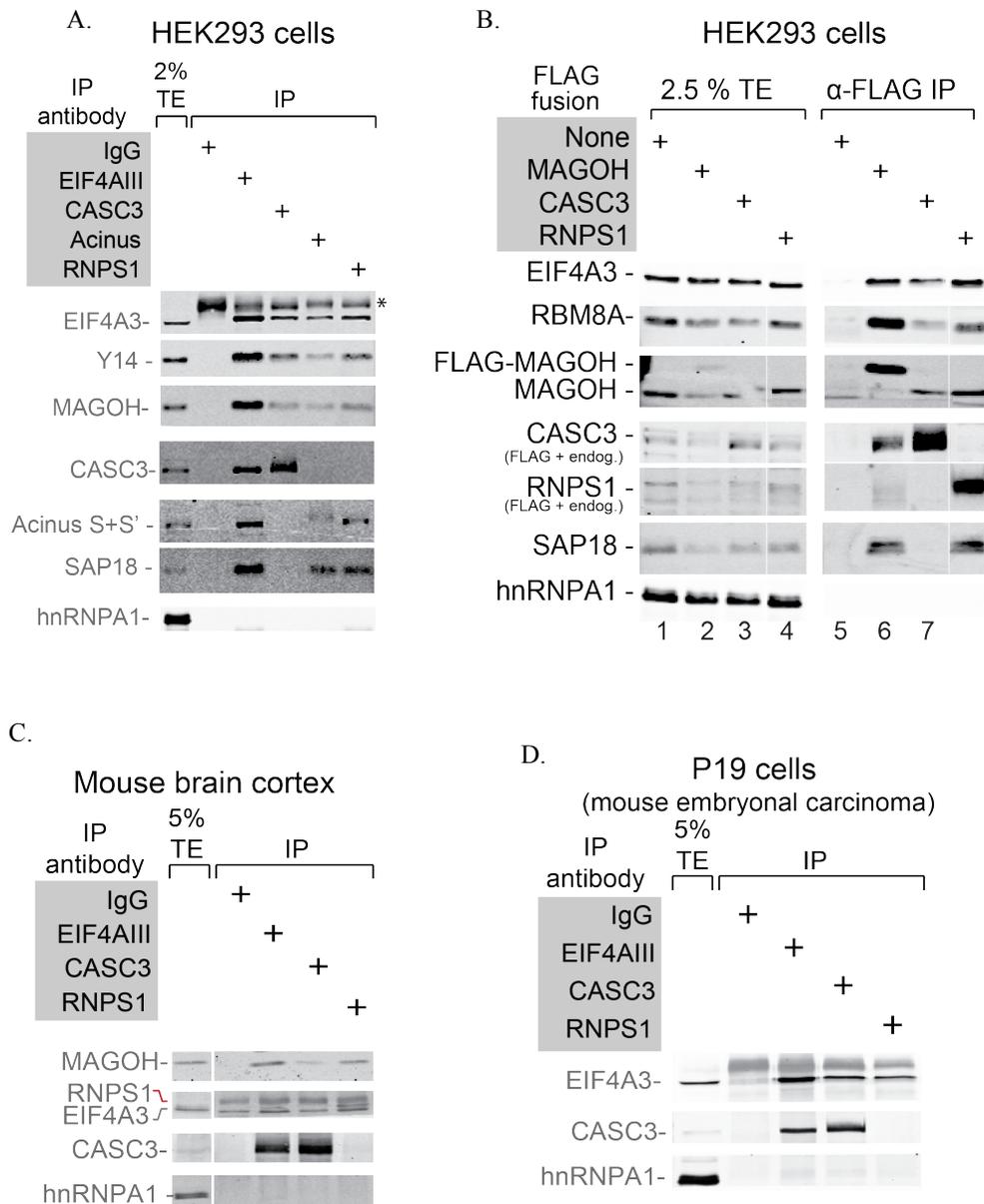


Figure 2.1 CASC3 and RNPS1 bind the EJC in a mutually exclusive manner in multiple cell types

A. Western blots showing proteins on the right in RNase-A-treated total HEK293 cell extract (TE) or in the immunoprecipitates (IP) of the antibodies listed on the top. Asterisk (*) indicates IgG heavy chain.

Figure 2.1 CASC3 and RNPS1 bind the EJC in a mutually exclusive manner in multiple cell types (continued)

B. Western blots showing proteins on the right in RNase-A-treated total HEK293 cell extract (TE) or in the anti-FLAG immunoprecipitates (IP) of HEK293 cell lines expressing the FLAG-tagged protein indicated at the top.

C. Western blots showing proteins on the right in RNase-A-treated total mouse brain cortex extract (TE) or in the immunoprecipitates (IP) of the antibodies listed on the top.

D. Western blots showing proteins on the right in RNase-A-treated total P19 cell extract (TE) or in the immunoprecipitates (IP) of the antibodies listed on the top.

Figure adapted from Mabin, Woodward, and Patton, et al. (Mabin et al., 2018a) with permission from Cell Reports.

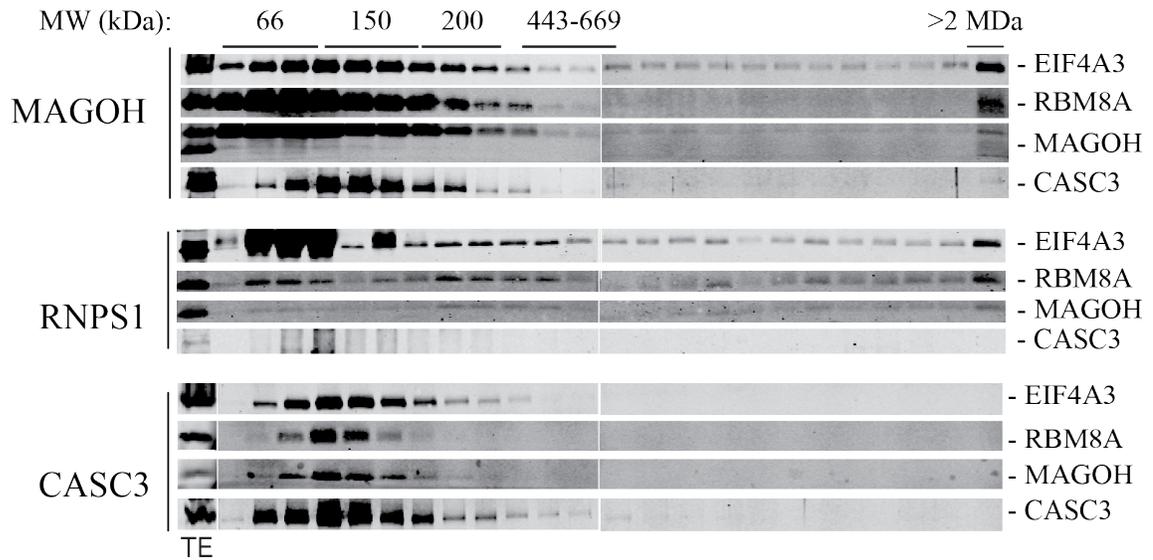


Figure 2.2 Alternate EJC complexes are structurally distinct

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: molecular weight standards.

Experiment done by Justin Mabin.

Figure adapted from Mabin, Woodward, and Patton, et al. (Mabin et al., 2018a) with permission from Cell Reports.

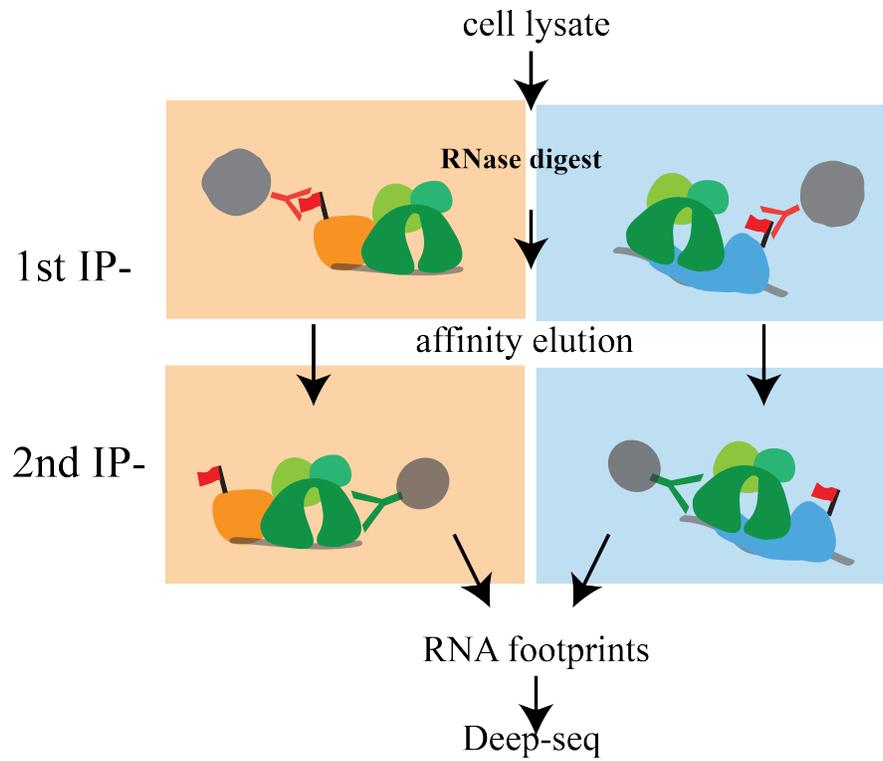


Figure 2.3 RIPiT schematic

Schematic depicting the major steps in RIPiT workflow. Different EJC compositions were purified from HEK293 cell extract by enriching either FLAG-RNPS1 (orange) or FLAG-CASC3 (blue) with two IP steps. In the first step, stably expressed FLAG-tagged CASC3 or RNPS1 were IP'd with anti-FLAG. In the second step, the elution from the first IP was subjected to an anti-eIF4AIII IP to enrich distinct assembled EJC complexes. RNA was purified from elutions of the second IP and prepared for sequencing.

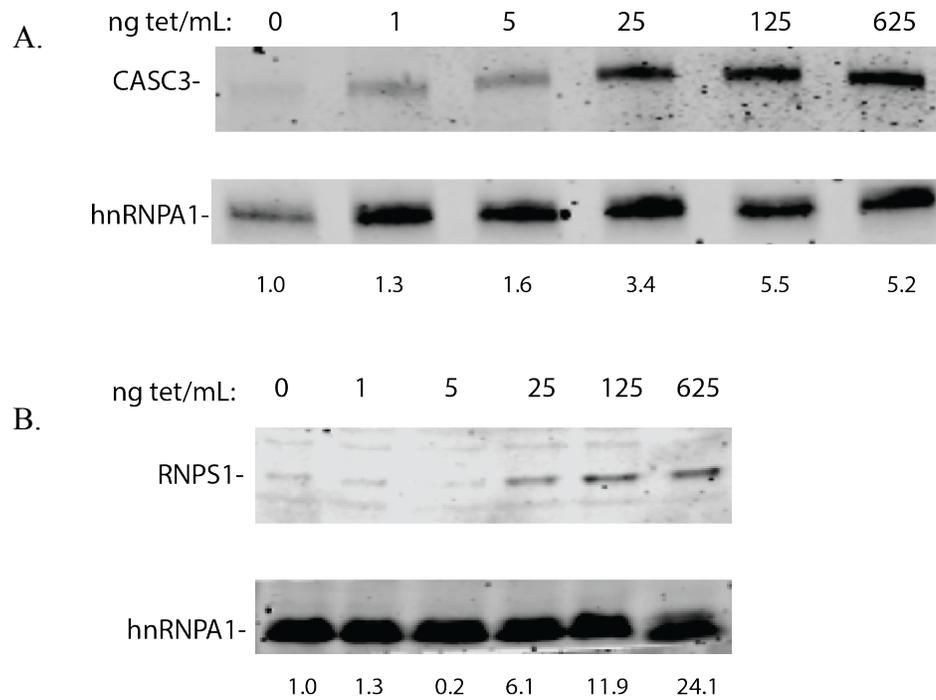


Figure 2.4 Tetracycline inductions of FLAG-RNPS1 and FLAG-CASC3 HEK293 cell

lines used for RIPiT experiments.

A. Western blot depicting tetracycline-induction (above) of FLAG-CASC3. Anti-CASC3 (left) was used to detect endogenous and FLAG-tagged CASC3 levels.

B. Western blot depicting tetracycline-induction (above) of FLAG-RNPS1. Anti-RNPS1 (left) was used to detect endogenous and FLAG-tagged RNPS1 levels.

Figure adapted from Mabin, Woodward, and Patton, et al. (Mabin et al., 2018a) with permission from Cell Reports.

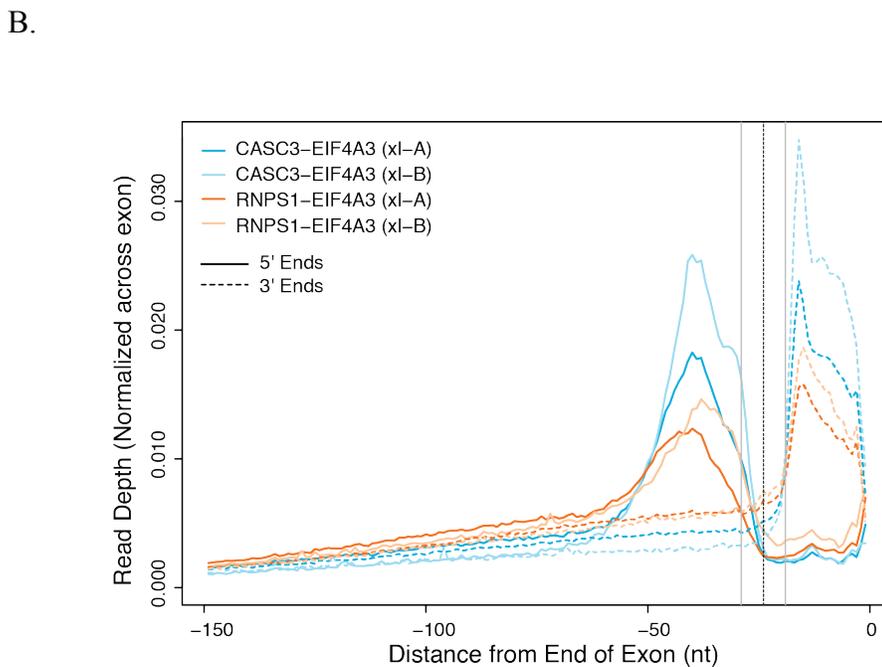
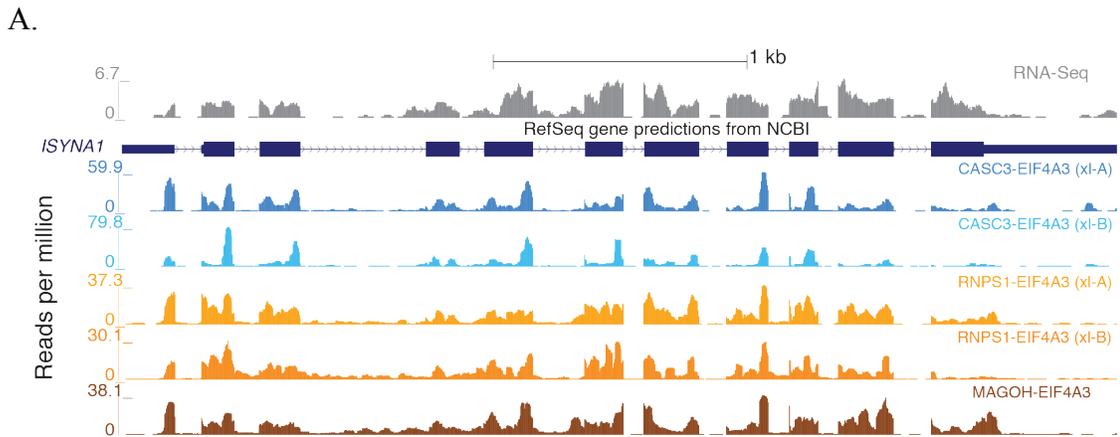
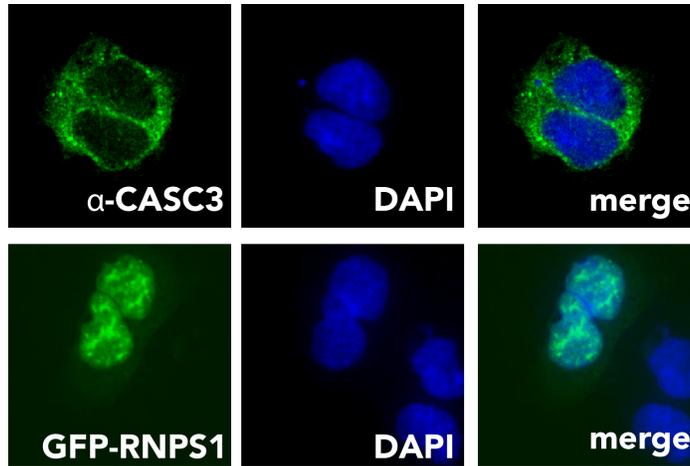


Figure 2.5 Alternate EJCs bind RNA via the EJC

A. 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. B. Meta-exon plots showing read depth in different RIPiT-seq or RNA-seq libraries (indicated in the middle) in the 150nt from the exon 5' (left) or 3' end (right). Vertical black dotted line: canonical EJC position at -24 nt. Figure adapted from Mabin, Woodward, and Patton, et al. (Mabin et al., 2018a) with permission from Cell Reports.

A.



B.

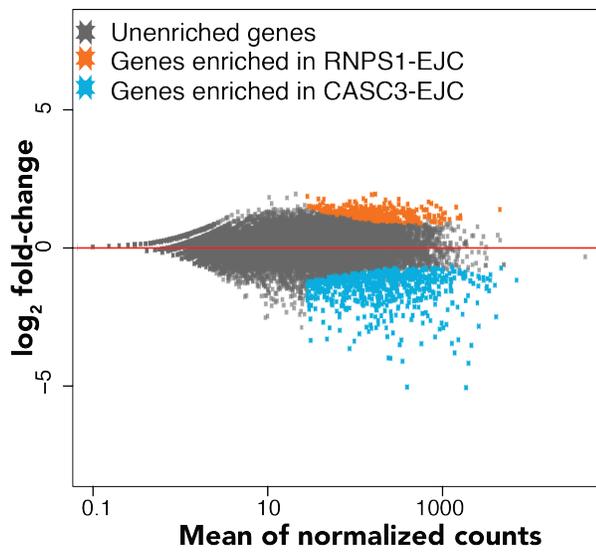


Figure 2.6 mRNP subcellular localization correlates with EJC composition

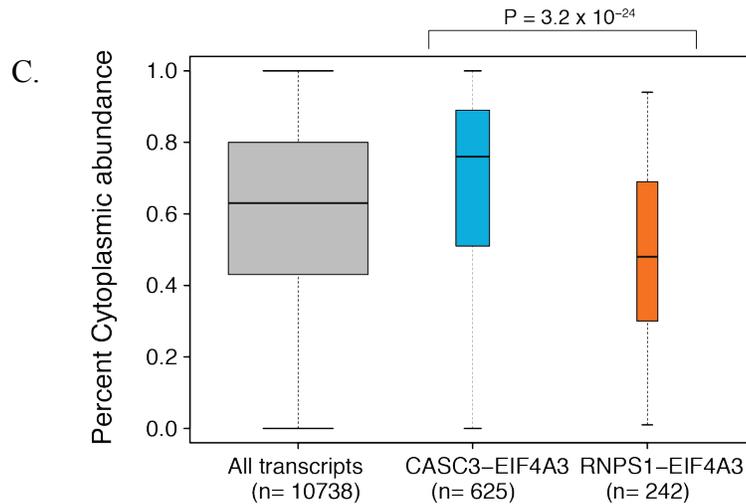


Figure 2.6 mRNP subcellular localization correlates with EJC composition (continued)

A. Immunofluorescence image of CASC3 distribution in HEK293 cells (above) and fluorescence image of GFP-RNPS1 distribution in HEK293 cells (below).

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. Data analyzed by Robert Patton.

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. Data analyzed by Robert Patton.

Figure adapted from Mabin, Woodward, and Patton, et al. (Mabin et al., 2018a) with permission from Cell Reports.

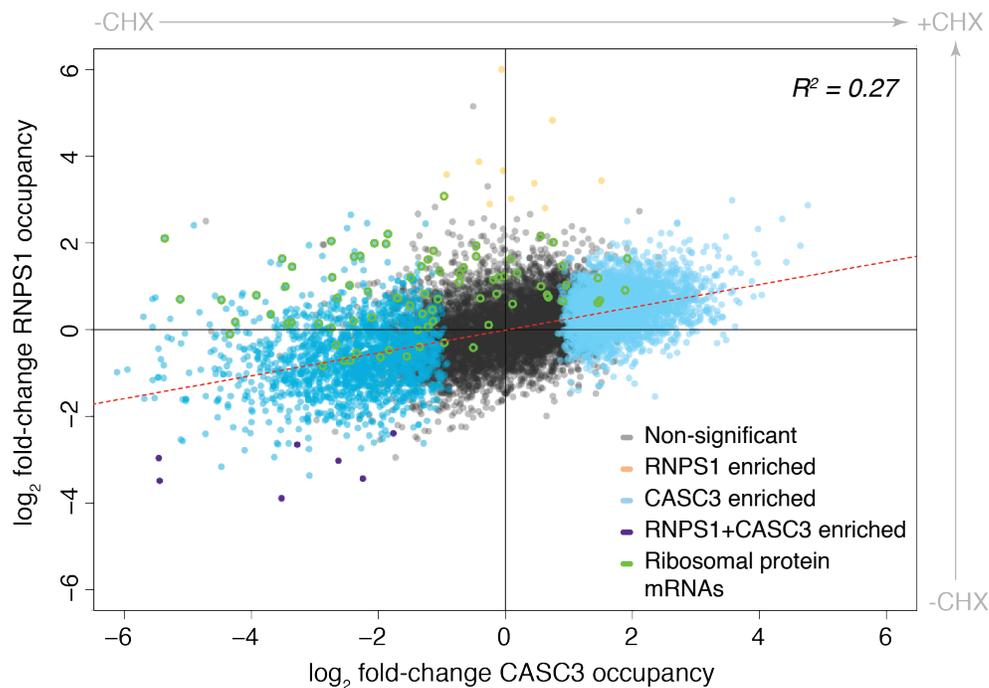


Figure 2.7 Translation impacts CASC3-EJC occupancy on mRNA

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). Figure adapted from Mabin, Woodward, and Patton, et al. (Mabin et al., 2018a) with permission from Cell Reports.

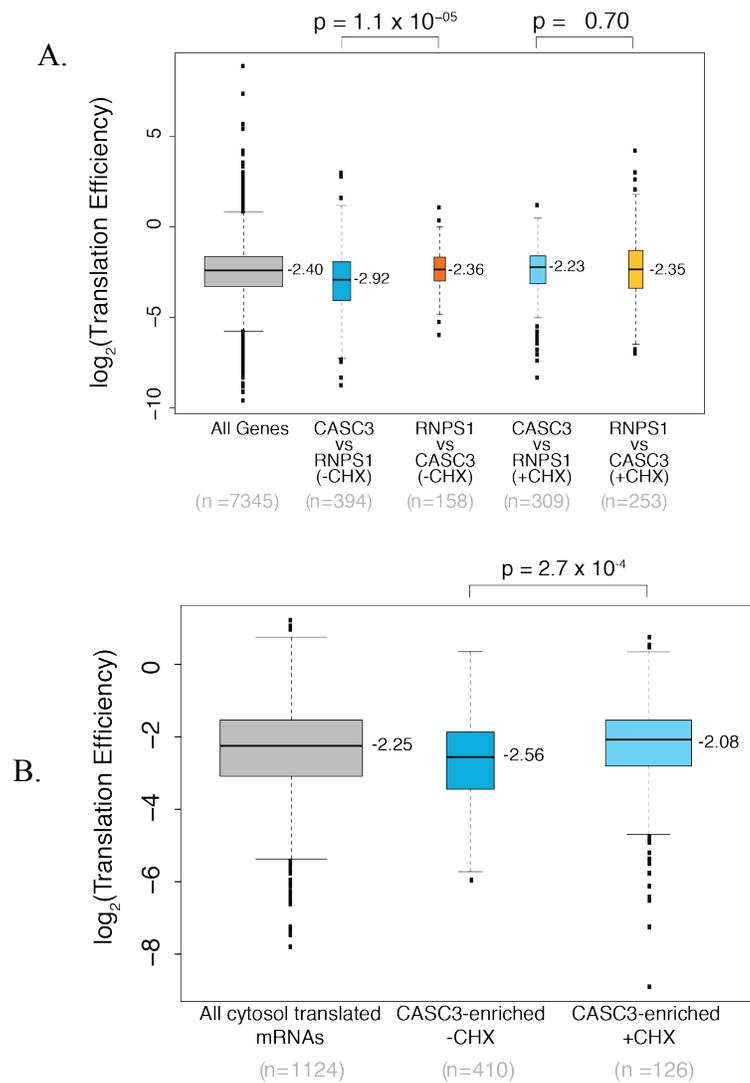


Figure 2.8 CASC3 is enriched on transcripts with lower translation efficiency

A. Boxplots comparing mRNA ribosome occupancy of transcript groups from as indicated on the bottom. Ribosome occupancy data are from (Kiss et al., 2017). Data analyzed by Robert Patton.

B. Boxplot comparing the ribosome occupancy of transcripts enriched in FLAG-CASC3:eIF4AIII RIPiT in untreated versus cycloheximide-treated cells (Kiss et al., 2017).

Data analyzed by Robert Patton.

Figure adapted from Mabin, Woodward, and Patton, et al. (Mabin et al., 2018a) with permission from Cell Reports.

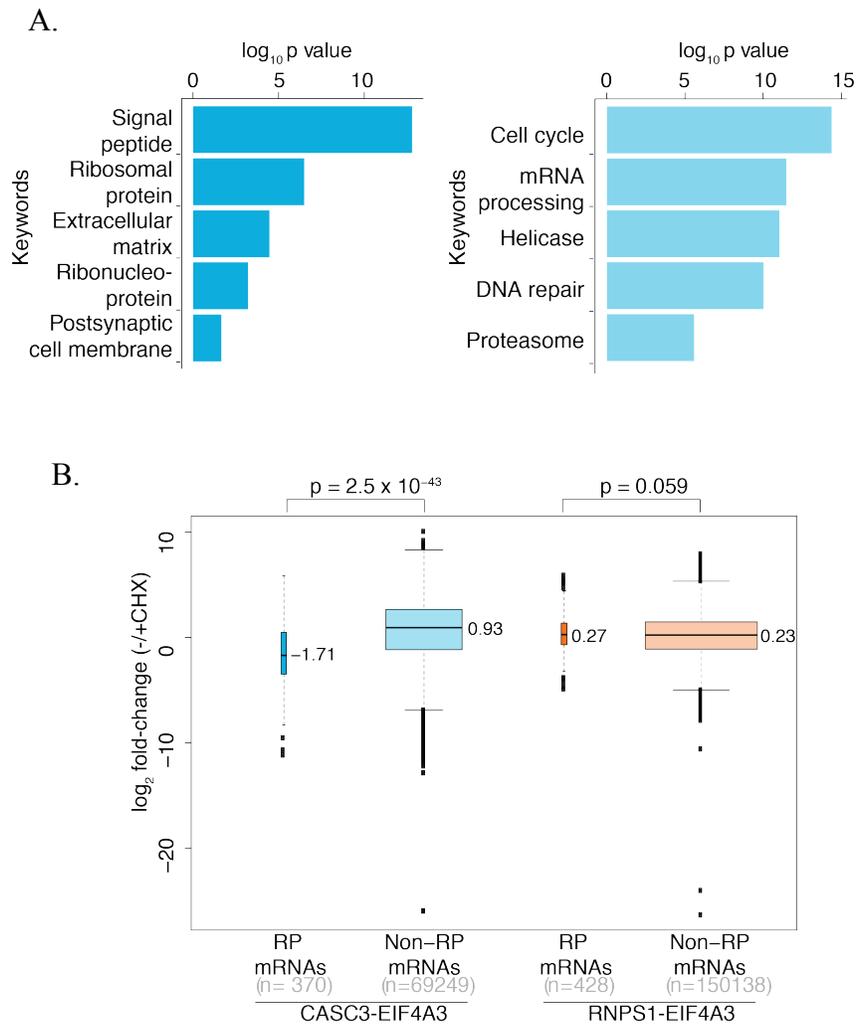


Figure 2.9 CASC3-EJCs are enriched on ribosomal protein mRNAs

A. Go-term analysis of most enriched genes in FLAG-CASC3-EJC RIPiTs with and without cycloheximide treatment.

B. Boxplot comparing CASC3-EJC and RNPS1-EJC occupancy on ribosomal protein-encoding mRNAs in untreated and cycloheximide-treated conditions.

Figure adapted from Mabin, Woodward, and Patton, et al. (Mabin et al., 2018a) with permission from Cell Reports.

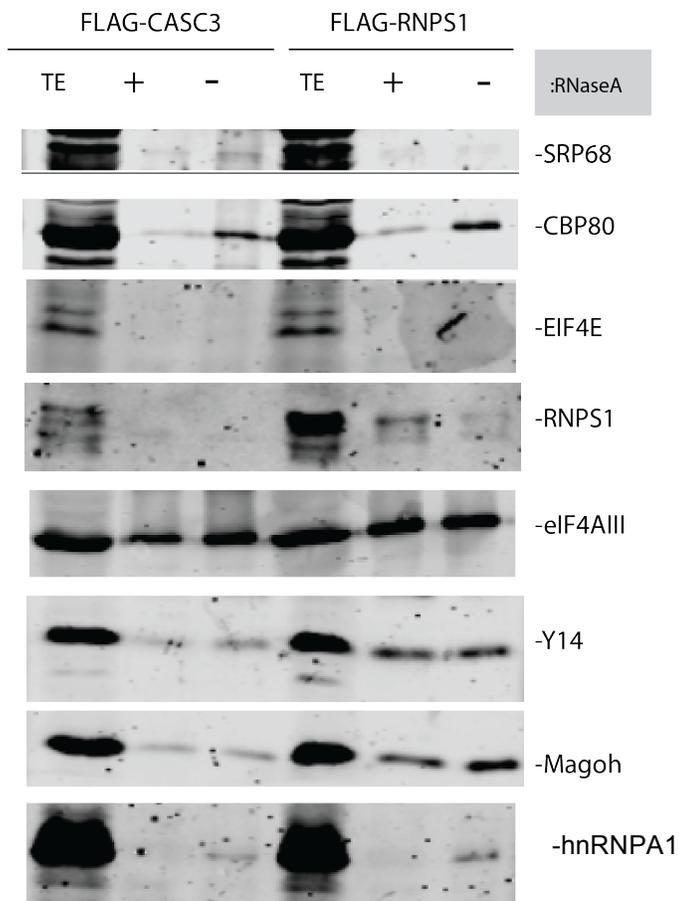


Figure 2.10 CASC3-containing RNPs interacts with SRP protein, SRP68

Western blot of FLAG-CASC3 and FLAG-RNPS1 IPs from HEK293 cells with and without RNase treatment (above). IP'd proteins are at top and proteins detected in IP or Total extract lanes (TE) are on right.

Figure adapted from Mabin, Woodward, and Patton, et al. (Mabin et al., 2018a) with permission from Cell Reports.

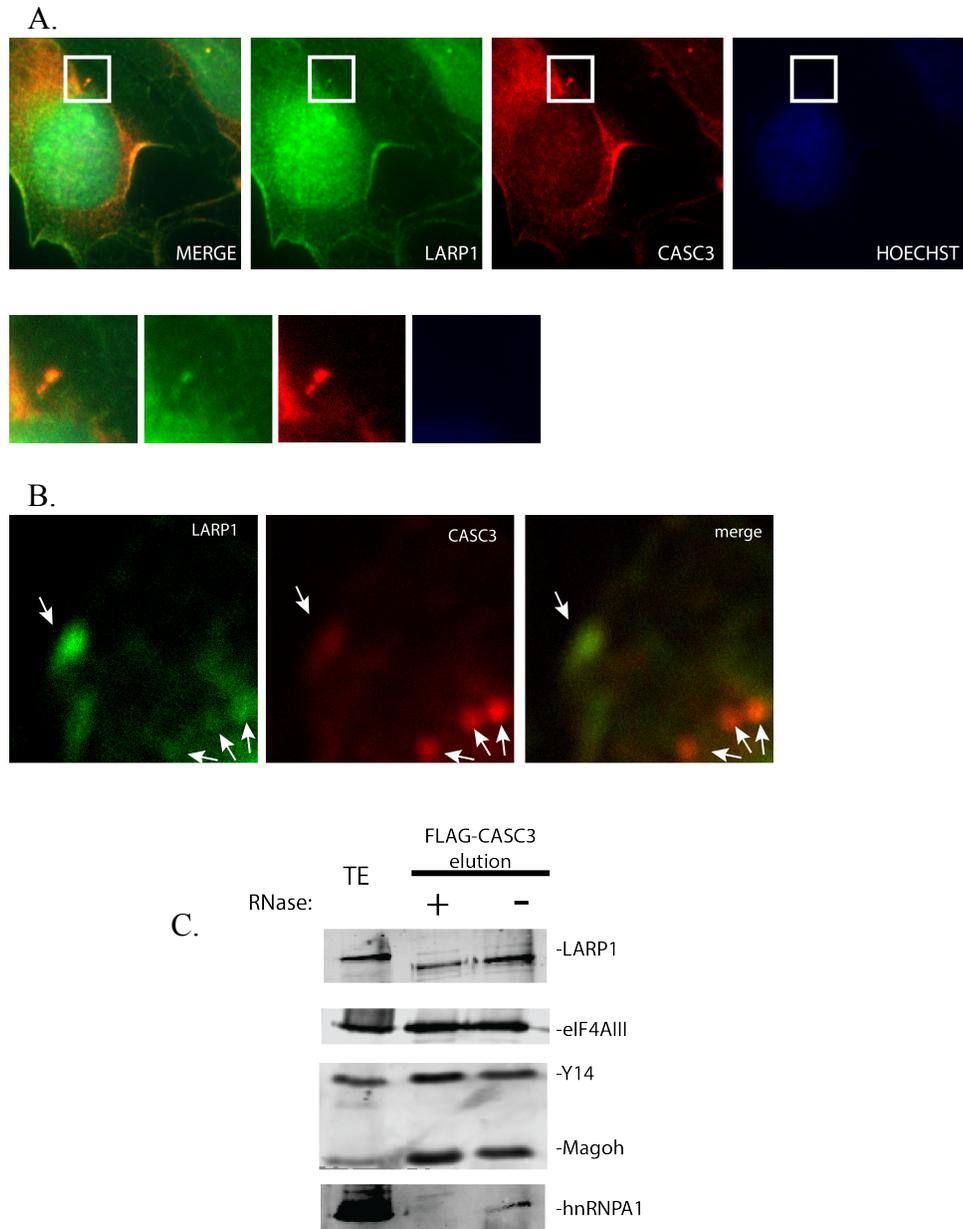


Figure 2.11 CASC3 interacts with LARP1

A. Immunofluorescence of CASC3 and LARP1 in HeLaSW cells upon 3hrs cycloheximide treatment to induce stress granules. White box highlights puncta depicting colocalization of CASC3 and LARP1 and area within white box is shown zoomed-in (below).

Figure 2.11 CASC3 interacts with LARP1 (continued)

B. Immunofluorescence of CASC3 and LARP1 in HeLaSW cells upon treatment with 200 μ M sodium arsenite to induce stress granules for 1 hour. Punctae with colocalization are highlighted with white arrows.

C. Western blot depicting co-IP of LARP1 with FLAG-CASC3 with and without RNaseA treatment. LARP1 interaction is detected in eluate in an RNA-dependent manner.

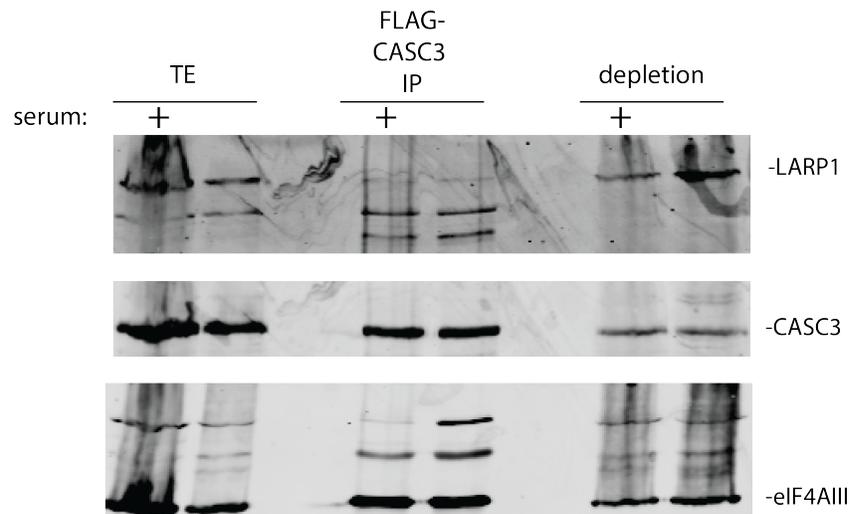


Figure 2.12 Serum starvation promotes CASC3-LARP1 interaction

Western blot depicting LARP1 co-IP with FLAG-CASC3 from cells which were incubated in 0.5% serum for 24 hours or cells which were kept under normal conditions. Levels of LARP1 co-IP were normalized to co-IP of eIF4AIII.

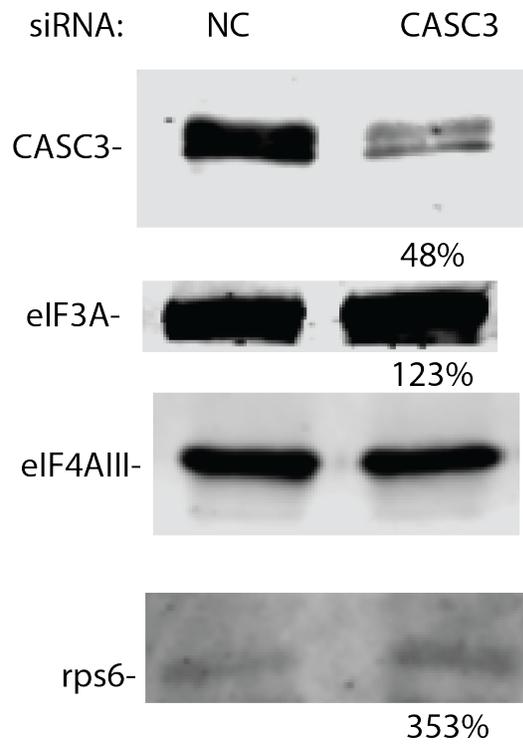


Figure 2.13 CASC3 depletion upregulates protein level of 5'TOP mRNA-encoded proteins, eIF3A and rps6

Western blot depicting detected protein levels (right) of HEK293 cell lysates from cells transfected with control or CASC3 siRNA (top).

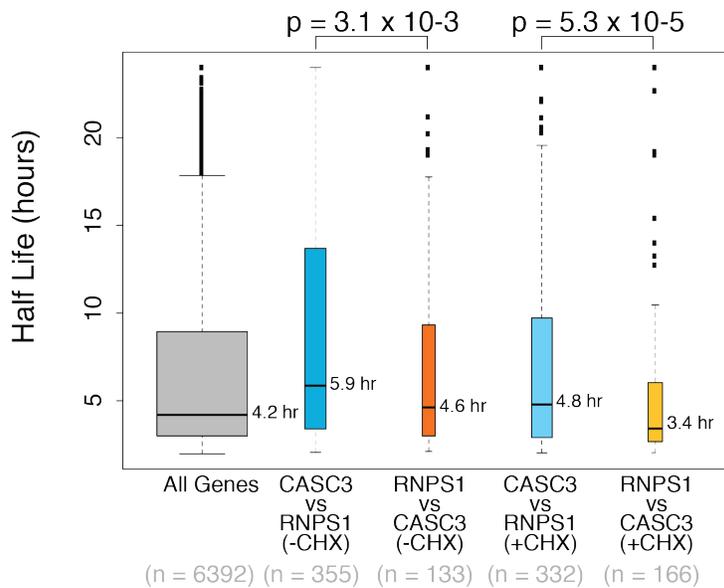


Figure 2.14 mRNAs enriched in RNPS1 over CASC3 have shorter half-life

Boxplots comparing mRNA half-life of transcript groups from as indicated on the bottom. mRNA half-life data are from (Tani et al., 2012).

Data analyzed by Robert Patton

Figure adapted from Mabin, Woodward, and Patton, et al. (Mabin et al., 2018a) with permission from Cell Reports.

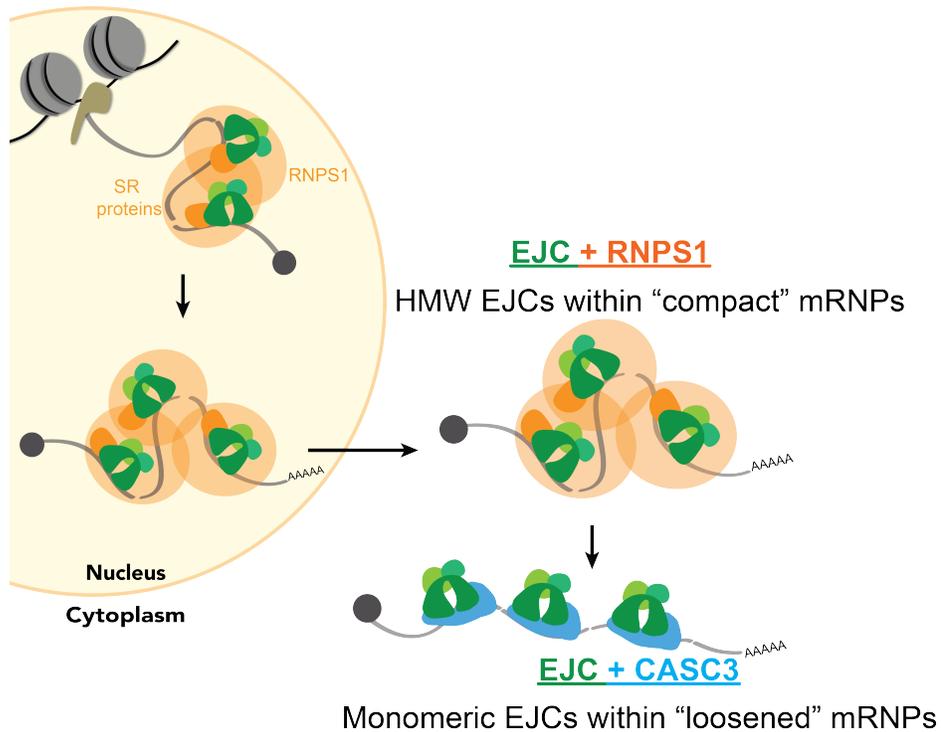


Figure 2.15 EJC remodeling from RNPS1-EJCs to CASC3-EJCs

Schematic illustrating a model for EJC remodeling from RNPS1-containing to CASC3-containing EJCs. Early EJCs in the nucleus are primarily composed of RNPS1 and SR-protein-associated EJCs (orange) with a compact mRNP structure. Eventually, these mRNPs are exported to the cytoplasm. Likely after export, the RNPS1-EJCs undergo a compositional change to CASC3-EJCs (blue).
 Figure adapted from Mabin, Woodward, and Patton, et al. (Mabin et al., 2018a) with permission from Cell Reports.

Chapter 3 The role of PYM in EJC occupancy

3.1 Abstract

Both *in vitro* experiments and transcriptome-wide footprinting analysis of the EJC reveal it is assembled 24 nt upstream of exon-exon junctions during splicing; however, transcriptome-wide binding studies have revealed that the EJC footprints exist at non-canonical positions, or areas distant from the -24 positions. As yet, very little is known about why EJCs exist at these sites or how they are removed.

It is thought that PYM's primary function is to recruit the pioneer ribosome to untranslated mRNPs and disassembles EJCs at their site of deposition (-24 nt position). According to current models, PYM binds to the Y14:Magoh heterodimer, interrupting the heterodimer's interaction with eIF4AIII ultimately resulting in co-translational EJC disassembly. However, there are unresolved issues with this model. Primarily, PYM's EJC disassembly function is not required for translation to occur, suggesting that the processivity of ribosome is sufficient to displace EJCs from RNA (Ghosh et al., 2014; Paix et al., 2017).

Here, we investigated the effects of interaction with PYM and translation on EJC occupancy on mRNA and show that PYM is largely inconsequential for EJC removal from canonical positions. However, PYM prevents EJC binding at non-canonical

positions, including intronless mRNAs. We propose two models to explain PYM's function as it relates to non-canonical EJCs.

3.2 Introduction

Beginning with transcription of nascent mRNAs and through their translation, transcripts exist as ribonucleoprotein complexes (RNPs). The protein complement of mRNPs controls stability and fate of a message until it is either removed or the mRNA is destroyed. The EJC is an integral part of spliced mRNPs.

Upon splicing, the EJC is deposited 24 nt upstream of exon-exon junctions at the so-called canonical site (Singh et al., 2012). The EJC is composed of a trimeric core, containing eIF4AIII and heterodimer of Y14 and Magoh. The very stable EJC core binds a more dynamic set of peripherally interacting partners and remains stably associated with RNA until the EJC is removed during translation (Diem et al., 2007). Notably, transcriptome-wide binding studies of the EJC consistently show EJC occupancy outside of the canonical (-24 nt) position (Hauer et al., 2016; Mabin et al., 2018; Saulière et al., 2012; Singh et al., 2012). These are considered non-canonical binding sites.

The removal of EJC proteins from RNA is presumed essential, as the EJC components are recycled, and EJCs must be removed to accommodate ribosomal decoding of the transcript (Gehring et al., 2009b). PYM has long been regarded as the EJC disassembly factor. Bono et al. solved the crystal structure of the PYM-Y14-Magoh trimer in 2004. The N-terminal of PYM interacts with a composite surface of the Y14:Magoh heterodimer, participating in many electrostatic interactions with Magoh and hydrophobic interactions with Y14 (Bono et al., 2004). Whether in isolation, or bound to

PYM or eIF4AIII, the Y14:Magoh heterodimer maintains a very rigid structure, implying that Y14:Magoh serves as a sort of scaffold for the protein-protein interactions it participates in (Bono et al., 2004, 2006; Diem et al., 2007; Lau et al., 2003). Consistently, PYM does not interact with either Y14 or Magoh in isolation (Bono et al., 2004). Bono et al. also observed that the interaction of Y14:Magoh in the assembled EJC would not accommodate PYM interaction due to steric clashing between eIF4AIII and PYM and first postulated that PYM might be a cytosolic EJC disassembly factor (Bono et al., 2004). Consistently, PYM does not stably interact with the other EJC core protein, eIF4AIII (Diem et al., 2007; Gehring et al., 2009b; Ghosh et al., 2014). Following studies focused more on the PYM's EJC disassembly function. In 2009, Gehring et al. reported that exogenous overexpression of PYM resulted in reduction of EJC interaction with reporter RNA, solidifying PYM's reputation as an EJC disassembly factor (Gehring et al., 2009b).

PYM also interacts with the small ribosomal subunit/pre-initiation complex, suggesting a role for PYM in coupling translation and splicing (Diem et al., 2007). Diem et al. proposed a model where pioneer ribosomes were recruited to untranslated mRNA via interaction with PYM (Diem et al., 2007). This mechanism for coupling EJC deposition to enhanced translation is analogous to a similar mechanism proposed by Chazal et al. where CASC3's interacts with eIF3A recruits ribosomes to untranslated mRNA (Chazal et al., 2013). Similarly, the recruitment of translation machinery via SKAR's interaction with S6K1 results in increased translation through activation of the mTOR pathway when cell conditions are conducive to protein synthesis (Ma et al., 2008).

S6K1 associates with and phosphorylates components of the pre-initiation complex, thus increasing protein synthesis (Peterson and Sabatini, 2005). Recruitment of translation machinery to newly processed mRNA through peripheral EJC proteins—PYM, CASC3, or SKAR—would prioritize recruitment of translation machinery to untranslated mRNAs and enhance the pioneer round of translation, thereby explaining the enhanced translation of spliced mRNAs (Braddock et al., 1994; Callis et al., 1987; Carter et al., 1996; Lu and Cullen, 2003; Matsumoto et al., 1998; Moore and Proudfoot, 2009; Nott et al., 2004). However, it is noteworthy that while PYM depletion results in modestly reduced translational efficiency of a luciferase reporter, PYM's EJC disassembly function is not required for translation to occur (Diem et al., 2007; Ghosh et al., 2014; Paix et al., 2017).

According to the current model, PYM joins the pioneer ribosome and disassembles EJCs at their site of deposition (-24 nt position); however, several paradoxical observations related to this model persist. First, PYM levels are far substoichiometric to assembled ribosomes or EJCs, and thus it is unlikely that PYM is responsible for removal of the majority of EJCs from RNA (Gehring et al., 2009b). Further, PYM is not essential for viability in *Drosophila* or mammalian cells (Ghosh et al., 2014; Paix et al., 2017). PYM null mutants in *Drosophila* are viable as is a PYM knockout mammalian cell line (Ghosh et al., 2014; Paix et al., 2017). Further, *Drosophila* PYM null mutants exhibit no obvious phenotype. By comparison, loss of EJC core proteins is embryonic lethal in all organisms investigated and display haploinsufficiency in mammalian systems (Mao et al., 2015, 2016, 2017; McMahon et al., 2016; Miller et

al., 2017; Silver et al., 2010). Even 48hr siRNA knockdown of EJC core components in mammalian cell lines is lethal (my unpublished results).

Taken together, these observations suggest that the processivity of ribosome is sufficient to displace EJCs from the CDS. This is not in agreement with the current model of EJC disassembly by the so-called EJC disassembly factor, PYM. Although these incongruous observations regarding PYM's role in EJC disassembly exist, there has not been a comprehensive study of PYM's role in EJC occupancy on RNA. PYM's interaction with Y14:Magoh is conserved in *Drosophila* and mammalian systems, so the suggestion that PYM is not necessary for EJC disassembly begs the question: what is PYM's function in the EJC lifecycle?

Here, we investigate the relationships between EJC occupancy, PYM-mediated disassembly and translation and show that PYM interacts with EJCs both at canonical and non-canonical positions along exons. Surprisingly, loss of interaction between the EJC and PYM is largely inconsequential in terms of EJC occupancy with some noticeable exceptions. Loss of PYM interaction causes a greater accumulation of EJCs on single exon mRNAs and in non-canonical stretches of exons. Consistently, PYM interacts with EJC core protein, eIF4AIII more in non-canonical regions compared to sites occupied by assembled EJC core proteins. Taken together, PYM's primary function is not—as has been reported—to disassemble EJCs from canonical positions during translation, but instead, PYM may prevent aberrant splicing-independent association of EJCs in non-canonical positions.

3.3 Materials and Methods

3.3.1 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 μg of a pcDNA5-FRT/TO-FLAG construct along with 1.8 μg 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 $\mu\text{g}/\text{ml}$) and Hygromycin (100 $\mu\text{g}/\text{ml}$) 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.

3.3.2 FLAG-tag Immunoprecipitations

For immunoprecipitation of FLAG-tagged proteins, stable HEK293 cells expressing FLAG-tagged EJC protein were lysed 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

$\mu\text{g/mL}$. Following five minute incubation on ice, cell lysates were cleared at $15,000 \times g$. Lysates were then incubated with anti-FLAG M2-agarose beads (Sigma) for 1-2 hr at 4°C , washed 6 times with 1 mL Iso Wash Buffer (IsoWB) [20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% NP-40] and eluted in IsoWB supplemented with $125 \mu\text{g/mL}$ FLAG-peptide with gentle shaking for 2 hr at 4°C . Total cell extracts from mammalian cells prepared in an isotonic buffer supplemented with $125 \mu\text{g/ml}$ RNase A were incubated at 4°C for 2 hours 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.

3.3.3 Western Blotting

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.

The following primary antibodies were used for western blotting: anti-PYM (Novus Biologicals) 1:500; anti-CASC3 (Bethyl) 1:250; anti-eIF4AIII (Bethyl) 1:2000; anti-hnRNPA1 (Santa Cruz Biotechnology) 1:1000; anti-Magoh 1:500; anti-Y14 (Bethyl) 1:1000 dilution; anti-IPO13 (Proteintech) 1:500 dilution; anti-Myc 1:1000; anti-FLAG 1:1000 (Sigma).

3.3.4 RIPiT

PYM-eIF4AIII RIPiTs were carried out as described previously, but with the addition of formaldehyde (to a final concentration of 0.1%) to stabilize transient protein interactions (Woodward et al., 2019). For formaldehyde crosslinked RIPiTs, total extracts were prepared from six 15-cm plates. Magoh wild-type and E117R:CASC3 RIPiTs, total extracts from four 15-cm plates were prepared as described previously RIPiTs to enrich EJC footprints upon cycloheximide (CHX) treatment were carried out as described previously except that cells were incubated with 100 µg/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.

Following RNA extraction, RNA samples were prepared for deep sequencing as described previously (Gangras et al., 2018; Singh et al., 2014; Woodward et al., 2019). Data was aligned using HISAT and differential occupancy at the gene level was determined using DESeq2. Canonical EJC positions were defined as the last 50nt of first and internal exons of at least 100nt in length in multi-exon genes.

3.3.5 siRNA knockdowns

HEK293 cells were reverse transfected with siRNA using RNAiMax transfection reagent. The following siRNAs were used: PYM1 (Dharmacon, D-014868-03-0002) PYM1 (Dharmacon, D-14868-04-0002), eIF4AIII_187 (CGAGCAAUCAAGCAGAUCAUU), UPF1_1879

(AAGAUGCAGUCCGCUCCAUU), AllStars Negative Control (QIAGEN, SI03650318). Transfection mixes were prepared by diluting 1.5pmol siRNA in 200 μ L of OMEM and then adding 1.7 μ L of RNAiMax. Transfection mix was incubated for 10-15 minutes at room temperature and 200 μ L of the transfection mix was added to each well. HEK293 cells trypsinized and resuspended in serum-free DMEM to a density of 1.5×10^5 cells/mL, and 1mL of cell dilution was added to the transfection mix in each well. For each knockdown condition, 4 wells were transfected. Cells were incubated for 48 hours. Cells were then resuspended in PBS from wells and pooled. One-tenth of the pooled cells were separated for assay of protein knockdown via western blot. Cells were pelleted at 400xg for 5 minutes at 4°C and cells were resuspended in 500 μ L Trizol. RNA was extracted according to manufacturers instructions, DNaseI digested for 10 minutes at 37°C, extracted with phenol:chloroform isosamyl alcohol (pH 4.5), and precipitated with ethanol. For RNA sequencing, RNA concentration and quality were measured by RNA-HS bioanalyzer. Only RNA samples with a RNA integrity number (RIN) of at least 9.0 were used for library preparation.

3.3.6 RT-PCR and qPCR

Reverse transcription reaction was prepared by mixing 2000ng of total RNA, 1 μ L of 10mM dNTPs, 1 μ L of Oligo d(T)16, RNase-free ddH₂O to 13 μ L. Above solution was incubated at 65°C for 10 minutes. Then, 0.5 μ L of SuperscriptIII, 1 μ L RNaseIn, 4 μ L 10 x First Strand Buffer, and 1 μ L 100 mM DTT. Reverse transcription reaction was

then incubated at 50°C for 1 hour. RNaseH digest was performed by adding 1 µL to the reaction mix and incubating at 37°C for 1 hour.

For RT-PCR, 0.5% of cDNA was used as template for PCR reaction in 2x Sybr green PCR master mix with 2.5µM of each primer. For each reaction, 24 cycles was completed, and PCR products were resolved on a 1.0% agarose gel and imaged.

For RT-qPCR, 0.1% of cDNA was used as a template for each qPCR reaction. Reactions were prepared in a total volume of 10µL with 2x iScript master mix.

3.3.7 Immunofluorescence

For immunofluorescence, 0.4mL of HEK293 at a concentration of $\sim 8 \times 10^4$ cells/mL were seeded into chamber well slides, treated for tissue culture (Ibidi). Cells were allowed to grow from 24-48 hours. In experiments where FLAG-tagged proteins were overexpressed, tetracycline was added to the media at a concentration of 125ng/mL 18 hours prior to fixing the cells. Media was removed from cells, and cells were washed once with PBS. Then, cells were fixed with 0.5mL of 4% paraformaldehyde for 5 minutes at room temperature. Paraformaldehyde was washed off with PBS 4 times and then cells were permeabilized in 0.1% Triton-X-100 for 5 minutes at room temperature. Cells were incubated with blocking buffer for 1 hour and then incubated with primary antibody dilutions in blocking buffer overnight. Primary antibody was removed with 4 0.5mL washes of PBS. Samples were then incubates with 0.5mL of a 1:1000 dilution of fluorophore-conjugated antibody against either rabbit or mouse and Hoechst (for DNA staining) for 1 hour at room temperature. Secondary antibody was washed away with 4

washes of PBS, mounting media and a cover slip were Images were acquired using a spinning-disk confocal imaging system that consists of a spinning-disk scan head and a microscope (model IX81, Olympus, Tokyo, Japan) equipped with a 100×/1.40 NA oil objective. For excitation, diode lasers 488nm for GFP and 561nm for RFP and 461nm were used. Images are of a single Z-slice.

3.3.8 RNAseq of PYM knockdown

PYM knockdown and negative control RNA samples were obtained as described in section 3.3.5. The Lexogen Corall kit was used to deplete rRNA, ligate adapter sequences, reverse transcribe, and amplify RNA as per manufacturer's instructions.

3.3.9 Preparation of RIPiT RNA samples for deep sequencing

RNA fragments were 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 5nt barcode sequence, and ends in GG at the 5' end. Following RT, the extended RT primer is gel purified, circularized using CircLigase (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 100nt read lengths).

3.3.10 Adapter trimming and PCR removal

After demultiplexing, fastq files containing unmapped reads were first trimmed using Cutadapt. A 12nt sequence on read 5' ends consisting of a 5nt random sequence, 5nt 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 TGGAATTCTCGGGTGCCAAGG was removed from the 3' end as possible.

3.3.11 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.

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. The primary reference transcriptome used in all post-alignment analysis was obtained from Ensembl and only transcripts considered were those annotated as principal isoforms (P1) by the APPRIS database (Rodriguez et al., 2013).

3.3.12 Differential enrichment analysis

Differential analysis of exons and transcripts between FLAG-PYM:eIF4AIII, FLAG-CASC3:eIF4AIII, FLAG-Magoh:CASC3, and FLAG-MagohE117R:CASC3 pull downs were conducted with the DESeq2 package in R (Love et al., 2014). 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.

3.3.13 GO-term analysis

DAVID gene ontology tool (Huang et al., 2009) was used to compare the set of genes predicted by DESeq2 analysis to be significantly enriched ($p \leq 0.3$ and fold change of >1.5) in FLAG-MagohE117R:CASC3 versus FLAG-Magoh:CASC3 and FLAG-PYM:eIF4AIII versus FLAG-CASC3:eIF4AIII against the DAVID tool default background list. Select non-redundant categories with lowest p value are reported.

3.4 Results

3.4.1 EJC-PYM interaction is not necessary for removing EJCs from canonical positions

Our previous work showed that inhibition of translation using translation elongation inhibitor cycloheximide (CHX) has a dramatic effect on the occupancy of CASC3-containing cytoplasmic EJCs at canonical positions (Mabin et al., 2018). These observations suggested that ribosome's translocation activity plays a key role in removal of EJCs from its deposition sites. As EJC disassembly is also reported to depend on PYM activity, we decided to disrupt either ribosome translocation or PYM-EJC interaction to directly compare the effect of these two activities on EJC occupancy in human cells (Figure 3.1). To assess the effect of ribosome translocation on EJC occupancy, we treated HEK293 cells stably expressing FLAG-Magoh with CHX for 3 hours before cell lysis, or with methanol (vehicle for CHX) as a control, and determined occupancy of FLAG-Magoh and CASC3 containing complexes via RIPiT-Seq. To assess the role of PYM-EJC interaction on EJC occupancy, we generated HEK293 cell lines that stably expressed at near endogenous levels FLAG-tagged Magoh-E117R, which lacks PYM-EJC interaction but retains EJC assembly and NMD activity (Bono et al., 2004; Gehring et al., 2009) (Figure 3.2), (Figure 3.3). To quantify in vivo occupancy of PYM-interaction deficient EJCs, we quantified in vivo footprints of FLAG-Magoh-E117R:CASC3 complexes via RIPiT-Seq from normally translating cells (Figures 3.4 and 3.5). In all RIPiTs, CASC3

was the target of the second IP in order to specifically enrich cytoplasmic EJCs, which are likely to be the most susceptible to both PYM and translation-dependent disassembly. Upon alignment and analysis, principle component analysis (PCA) showed that replicates of the conditions clustered with one another (Figure 3.6). As we reported previously, total canonical EJC occupancy of several protein-coding genes shows a significant difference in CHX treated versus untreated cells suggesting that ribosome translocation impacts EJC occupancy. In comparison, loss of PYM interaction was largely inconsequential in terms of EJC occupancy at canonical positions (Figures 3.7 and 3.8). Only one gene displayed a significant difference in canonical site occupancy between EJCs containing Magoh wild-type and MagohE117R (Figure 3.7). Thus, at least for canonical EJCs, PYM interaction is dispensable, and the major determinate affecting EJC occupancy is translation.

3.4.2 Loss of PYM interaction causes Magoh to bind more at non-canonical positions

When EJC occupancy was compared similarly for entire transcripts (instead of only canonical positions) several transcripts were found to have a significant change in EJC occupancy upon loss of PYM interaction (Figure 3.9). This suggests perhaps that the disassembly function of PYM has greater impact on non-canonical regions. In support of this, although both wild type Magoh and MagohE117R are enriched at the canonical position (Figure 3.10), loss of PYM interaction is accompanied by a reduced occupancy of Magoh on the canonical positions and an increase in the amount of MagohE117R bound to non-canonical positions (Figure 3.10) In agreement, ratio of signal observed for

wild-type and mutant (E117R) EJCs in non-canonical regions over canonical regions illustrates that EJCs devoid of PYM interaction are bound more to non-canonical stretches of RNA than wild-type EJCs (Figure 3.11). Comparison of the Non-canonical/Canonical ratios of MagohE117R vs. wild-type Magoh shows that the majority of genes have a value >1.0 , indicating that MagohE117R binds more to non-canonical regions of a gene than its wild-type counterpart (figure 3.12). Taken together, these data suggests that loss of PYM-EJC interaction leads to increased occupancy of EJCs at non-canonical positions.

To rule out the possibility that loss of PYM interaction may result in changes in subcellular distribution of EJC subunits and impair proper splicing-dependent deposition of EJCs, we compared the subcellular localization of MagohE117R and wild-type Magoh and found that Magoh localizes primarily to the nucleus, with no obvious change in localization between wild-type and mutant proteins (Figure 3.13). Further, comparison of subcellular localization of Y14 shows no obvious change between cells expressing FLAG-Magoh and FLAG-MagohE117R (Figure 3.13). Consistently, we also find that several EJC-dependent splicing events are unperturbed by depletion of PYM and/or overexpression of MagohE117R (Boehm et al., 2018) (Figure 3.14). Perhaps this is because MagohE117R is just as capable as wild-type Magoh to be reimported to the nucleus by importin 13 (IPO13) (Figure 3.15). Taken together, these results imply that the function of PYM in preserving proper EJC positioning occurs outside of splicing-dependent deposition. Thus, loss of PYM interaction through knockdown or FLAG-MagohE117R expression does not-- as previously suggested-- so impair EJC subunit

recycling that EJC-dependent splicing events are impaired (Gehring, 2009). As splicing-dependent EJC deposition is not impaired by loss of PYM interaction, perhaps one possibility is that the mutant EJCs are accumulating in non-canonical positions sometime after splicing is complete.

3.4.3 PYM interacts with the EJC core at canonical and non-canonical positions

If PYM is more important for limiting EJC assembly at non-canonical positions, it is expected that PYM itself will be detected at such sites. To directly test this possibility, the actual locations of PYM-EJC interactions on RNA were determined. To this end, RNP complexes containing both eIF4AIII and PYM were purified via RIPiT and RNA footprints were sequenced (Figures 3.16 and 3.17). It is notable that PYM binds exclusively to the Magoh:Y14 heterodimer and prevents Magoh:Y14 interaction with eIF4AIII. eIF4AIII is not a reported interaction partner with PYM. To stabilize this transient interaction between PYM and intact EJCs, proteins were crosslinked with formaldehyde prior to cell lysis and the RIPiT procedure. Upon formaldehyde crosslinking, RNPs containing both eIF4AIII and FLAG-PYM were recovered (Figures 3.16 and 3.17). In agreement with the hypothesis that PYM may be more consequential for EJC positioning at non-canonical positions, metagene exon plots comparing EJC core (FLAG-CASC3:eIF4AIII) and PYM-EJC (FLAG-PYM:eIF4AIII) reveal that PYM-EJC footprints are less abundant at the canonical position (Figure 3.18). Consistently, PYM-EJC interactions are significantly enriched at non-canonical positions compared to the EJC core itself (Figure 3.19).

3.4.4 PYM-eIF4AIII complexes and MagohE117R-EJCs are enriched on single exon transcripts

If PYM was required for EJC disassembly, one would expect MagohE117R-containing EJCs to be enriched on the same mRNAs on which PYM acts to disassemble EJCs. Comparison of FLAG-PYM:eIF4AIII versus FLAG-CASC3:eIF4AIII and FLAG-MagohE117R:CASC3 versus FLAG-Magoh:CASC3 DESeq2 \log_2 FoldChanges revealed a weak correlation between enrichment for genes in PYM and MagohE117R (Figure 3.20). However, single exon mRNAs were enriched in FLAG-MagohE117R:CASC3 compared to all transcripts. (Figures 3.20 and 3.21). Single exon genes are also enriched in FLAG-PYM:eIF4AIII compared to all transcripts (Figures 3.20 and 3.22). This is consistent with the hypothesis that PYM interaction is important in non-canonical regions, as single exon mRNAs are non-canonical by definition. In accordance with the model for splicing-dependent deposition of the EJC onto mRNA, intronless mRNAs are not expected to display EJC enrichment exceeding background levels. Nonetheless, EJC footprints are consistently detected across several EJC RIPiT (Figure 3.23) and CLIP datasets (data not shown) (Saulière et al., 2012). Interestingly, gene ontology (GO) term analysis of genes upregulated in FLAG-PYM-eIF4AIII vs. CASC3-eIF4AIII RIPiT datasets revealed a significant enrichment of genes encoding histone mRNAs (Figure 3.24). Consistently, histone genes are also enriched in MagohE117R (Figure 3.25).

3.4.5 PYM knockdown reduces nascent translation

Kaposi Sarcoma-associated Herpes Virus (KSHV) encodes a viral protein that recruits PYM to intronless viral transcripts. This results in the enhanced translation of viral transcripts and increased viral replication. We reasoned that perhaps PYM enhances translation of endogenous single exon mRNAs in a similar manner. To test this, a puromycin incorporation assay was used to measure levels of nascent translation upon PYM knockdown. In agreement with our hypothesis, PYM knockdown resulted in a 25% reduction in nascent peptide synthesis (Figure 3.26). However, this does not address the role of PYM in translation enhancement of single exon transcripts, specifically.

3.4.6 PYM knockdown results in widespread misregulation of many transcripts.

One explanation for the increased occupancy EJC on non-canonical stretches of mRNA is that EJCs may assemble outside of the splicing reaction in the cytoplasm more in the absence of PYM interaction. PYM is primarily localized to the cytoplasm at steady state, where it interacts with the Y14:Magoh heterodimer and prevents its binding to eIF4AIII. If interaction with PYM is lost, such as is the case in the MagohE117R mutant EJCs, then these EJCs may be more inclined to assemble on mRNA in the cytoplasm. An anticipated consequence of this would be widespread misregulation of transcripts, as EJCs bind to stretches of mRNA and destabilize transcripts. Indeed, PYM knockdown in HEK293 cells reveals significant misregulation of many transcripts (Figure 3.27). Surprisingly PYM knockdown results in upregulation of many NMD substrates (Figure

3.28). This, however, reveals a parallel between loss of PYM interaction and flaviviral infection. Several flaviviruses target PYM and interrupt its EJC-interaction; concurrent with flavivirus infection, NMD is suppressed. Thus, interruption of PYM-EJC interaction may be the mechanism through which flaviviruses repress NMD within infected cells. Further, this suggests that PYM has an, as yet, unappreciated role in maintaining proper surveillance of NMD substrates in uninfected cells. Perhaps spontaneous assembly of EJCs in non-canonical triggers aberrant decay of transcripts without a premature stop codon, simultaneously titrating NMD machinery away from transcripts bearing a PTC.

3.5 Discussion

The field's current model of PYM function is its role in EJC disassembly (Bono et al., 2004; Gehring et al., 2009; Ghosh et al., 2014). This model suggests that PYM is required for removing EJCs from mRNA such that translation may proceed unimpeded and EJC subunits may be recycled back into the nucleus. However, the basis for PYM's function in EJC disassembly is based on exogenous overexpression and in vitro experiments that may not completely capture in vivo function (Gehring et al., 2009).

Indeed only two instances of in vivo PYM function have been reported. In *D. melanogaster*, PYM is not required, as null mutants survive and have no obvious phenotype, implying that even in the absence of PYM, EJCs are removed from coding sequences (Ghosh et al., 2014). Despite PYM's near ubiquitous expression in different human cell types (with the exception of blood cells), in HEK293 human cells, knockout of the PYM locus had no effect on cell viability (Paix et al., 2017). It is possible that

PYM may exhibit cell specific or tissue specific phenotypes. Another possibility is that PYM is required for cell viability only under specific conditions, such as stress or at key points of development; however, neither of these possibilities have been investigated.

Sequencing of EJC footprints that have no PYM interaction revealed that PYM interaction is nonessential for in vivo removal of EJCs from endogenous transcripts (Figures 3.7 and 3.9). Furthermore, my data suggest that the disassembly function of PYM is more noticeable at non-canonical regions of mRNA (Figures 3.9, 3.10, 3.11, and 3.12).

In agreement with PYM being nonessential for EJC disassembly and re-import to the nucleus, PYM depletion in human cells did not impair nuclear EJC-mediated splicing events. However, this does not definitively rule out that PYM is important for EJC recycling, as PYM depletion may simply not result in a dramatic enough reduction of nuclear EJC subunits to impact splicing. In all, the necessity of PYM in EJC disassembly and recycling is dubious; however a new function of PYM in reducing EJC's binding spontaneously to non-canonical regions is emerging from my work.

Loss of PYM interaction results in a global increase in EJC binding to non-canonical stretches of mRNA, while having little impact on EJC disassembly from canonical regions. These non-canonical regions encompass intronless mRNA, including those encoding histone proteins. Further, PYM interaction with the EJC is similarly enriched at histone genes, suggesting a role for PYM interaction specifically with EJCs deposited at non-canonical regions. To address the question of the origin of non-canonical EJCs as well as PYM's relevance, we propose that EJCs can assemble outside

of the splicing reaction (as EJC footprints are so consistently detected on intronless mRNA). Spontaneous EJC assembly on RNA may explain some of the high percentage (~40-50%) of non-canonical reads in EJC footprinting datasets. As MagohE117R is enriched at non-canonical regions, PYM interaction is likely important for preventing spontaneous EJC assembly (Figure 3.29).

There are some interesting parallels between loss of PYM interaction and flaviviral infection. PYM knockdown resulted in overall misregulation of many transcripts, including significant upregulation of NMD substrates (Figures 3.27, and 3.28). Additionally, loss of PYM-EJC interaction results in increased EJC binding to non-canonical stretches of mRNA similar to the increased EJC binding to flaviviral RNA observed by Li, et al. (Li et al., 2019). Flaviviral RNA does not undergo nuclear processing, so any EJC binding must be the result of EJC assembly outside of the splicing reaction. In agreement with loss of PYM interaction with the EJC promoting pro-viral conditions in the cell, FLAG-MagohE117R:CASC3 footprints are enriched on mRNAs with GO terms associated with viral transcription and viral lifecycle (Figure 3.25).

EJCs are assembled at the canonical -24 nt position during splicing. The actual positioning of the EJC is completely sequence independent; instead, the -24 nt position is most probably dictated by geometric constraints of the spliceosome during EJC recruitment to the RNA. *In vivo* assembly of EJCs outside of the splicing reaction has not been reported or investigated. This is likely, in part, because such an event represents a very small percentage of all EJC assembly events. However, EJCs have been successfully assembled *in vitro* outside of the splicing reaction (Bono et al., 2006). Given

a high enough local concentration of EJC subunits and single stranded RNA, there is no known reason that should render spontaneous EJC assembly improbable *in vivo*. Future investigation will determine if MagohE117R is more adept at splicing-independent assembly of the EJC and if splicing-independent assembly contributes substantially to the population of non-canonical EJCs.

Other models have been proposed to explain the origin of EJCs at non-canonical positions. For example, EJC subunits may be recruited to non-canonical regions by SR-proteins; many non-canonical peaks are enriched at or adjacent to regions bound by SR-proteins (Rodriguez et al., 2013). A second interpretation is that ncEJCs are binding sites of EJC-associated RBPs (and not eIF4AIII) that associate with cEJC via protein–protein interactions possibly within a packaged three-dimensional RNP. In support of this model, Singh, et al. reported on the ability of EJCs to multimerize into megadalton-sized structures, and Metkar, et al. revealed mRNAs are compacted by their associated proteins into linear, rod-like structures (Metkar et al., 2018). While putative explanations exist, no model to explain the origin of non-canonical EJCs has been tested. Further, neither of these models addresses EJC footprints on unspliced transcripts.

PYM has previously been implicated in processing of a specific class of unspliced mRNAs. Kaposi's sarcoma-associated herpes virus (KSHV) depends on expression of several intronless mRNAs which encode viral proteins (Boyne et al., 2010). KSHV encodes a protein, ORF57, that interacts directly with components of the 48S pre-initiation complex (Boyne et al., 2010). Depletion of PYM results in the reduction of ORF57 interaction with the pre-initiation complex and reduced translation of intronless

KHSV mRNA (Boyne et al., 2010). Authors proposed that interaction between ORF57 and PYM resulted in efficient recruitment of translation machinery to the viral mRNAs.

This work presents evidence that PYM may, have a broader role in the translation of intronless mRNAs. One explanation for the accumulation of MagohE117R on intronless mRNAs could be that they are inefficiently translated as a consequence of reduced PYM interaction (Figure 3.30). In support of this, PYM knockdown causes a global reduction in translation (Figure 3.27). A caveat to this hypothesis is that MagohE117R is co-expressed with the endogenous Magoh, allowing for both normal and mutant EJC's to assemble on intronless mRNAs. To more clearly address PYM's function in intronless mRNA translation, the translational status of intronless mRNAs upon loss of PYM must be investigated.

In sum, the work presented in this chapter has lent insight into the role of PYM within the EJC lifecycle and expanded the conversation regarding the origin of non-canonical EJC's. Future work will investigate the biological significance of PYM's role in reducing non-canonical EJC's and the significance of EJC occupancy on unspliced mRNAs.

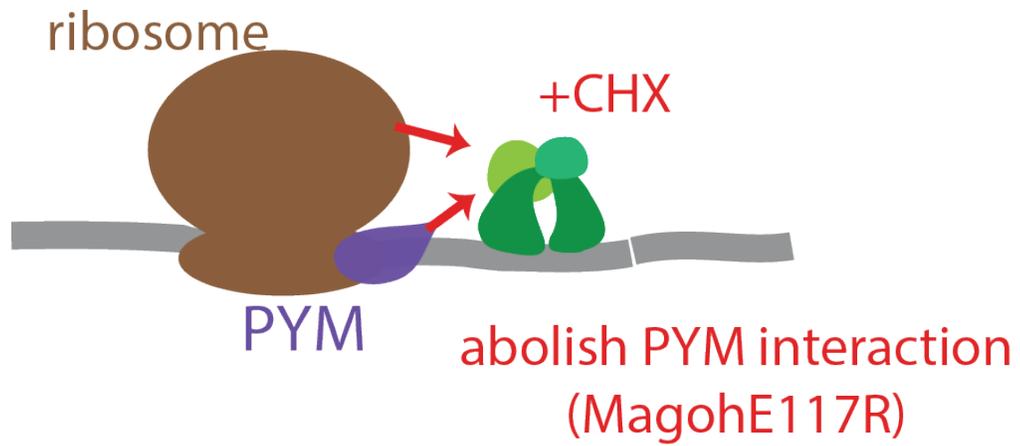


Figure 3.1 Schematic depicting two proposed mechanisms of EJC disassembly

EJC can be disassembled via translating ribosome or ribosome-associated PYM. Action of the ribosome can be impeded by treatment of cells with cycloheximide (CHX) while PYM's EJC-related functions are impeded in EJCs containing MagohE117R.

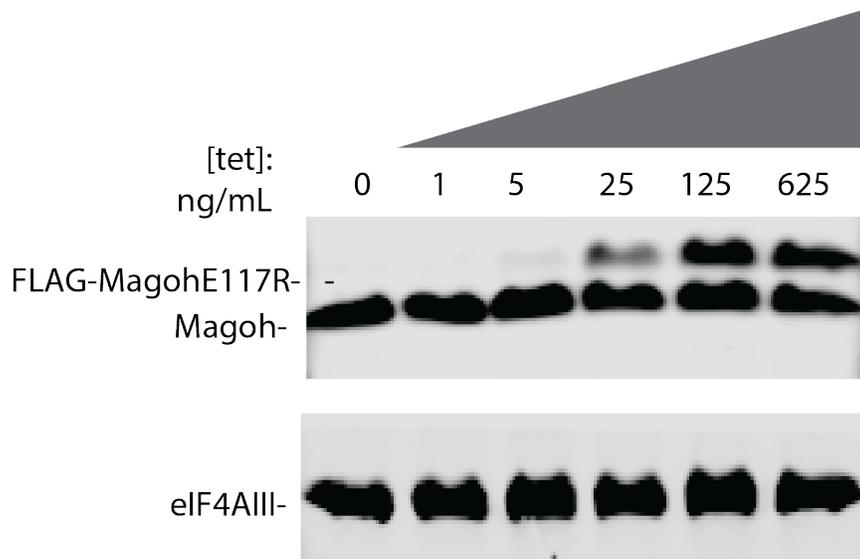


Figure 3.2 FLAG-MagohE117R was expressed at near endogenous levels

Western depicting protein levels (left) in cell extracts upon tetracycline induction (top) of FLAG-MagohE117R. Western was probed with anti-Magoh, which recognizes both endogenous and FLAG-tagged Magoh.

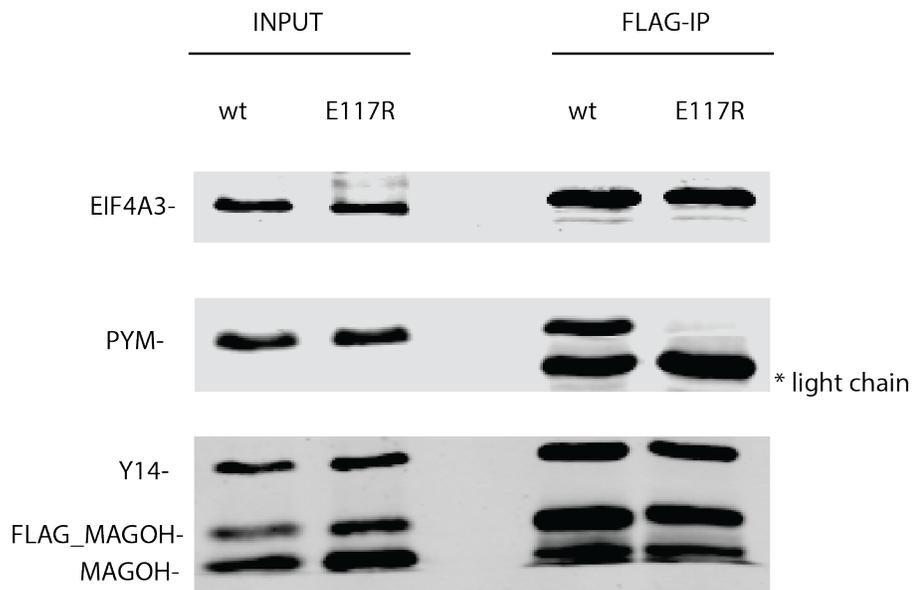


Figure 3.3 A single amino-acid change in Magoh (E117R) disrupts its interaction with PYM

Western blot depicting proteins (left) in total extract (input) and in FLAG-IP of FLAG-Magoh wild-type (wt) and FLAG-MagohE117R (E117R). eIF4AIII is used as a control for co-IP efficiency.

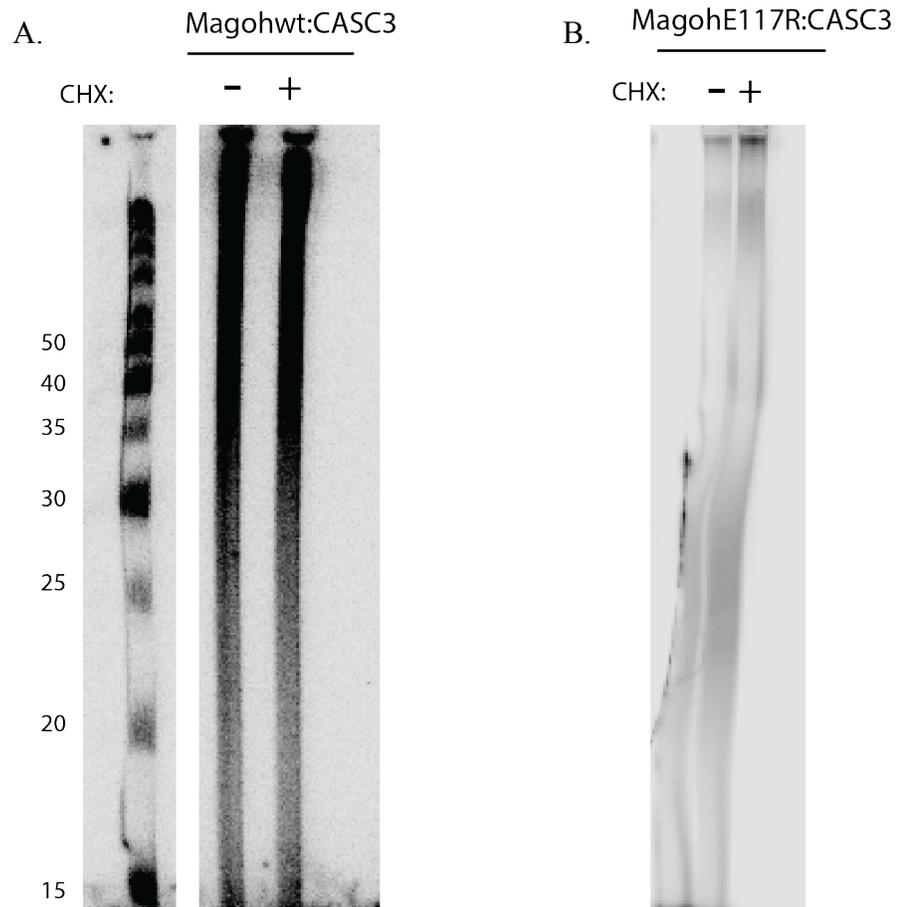


Figure 3.4 RNA footprints from FLAG-Magoh RIPiTs

- A. Autoradiograph of end-labeled RNA footprints from FLAG-Magoh wild-type RIPiTs with and without cycloheximide (CHX) (above).
- B. Autoradiograph of end-labeled RNA footprints from FLAG-MagohE117R RIPiTs with and without cycloheximide (CHX) (above).

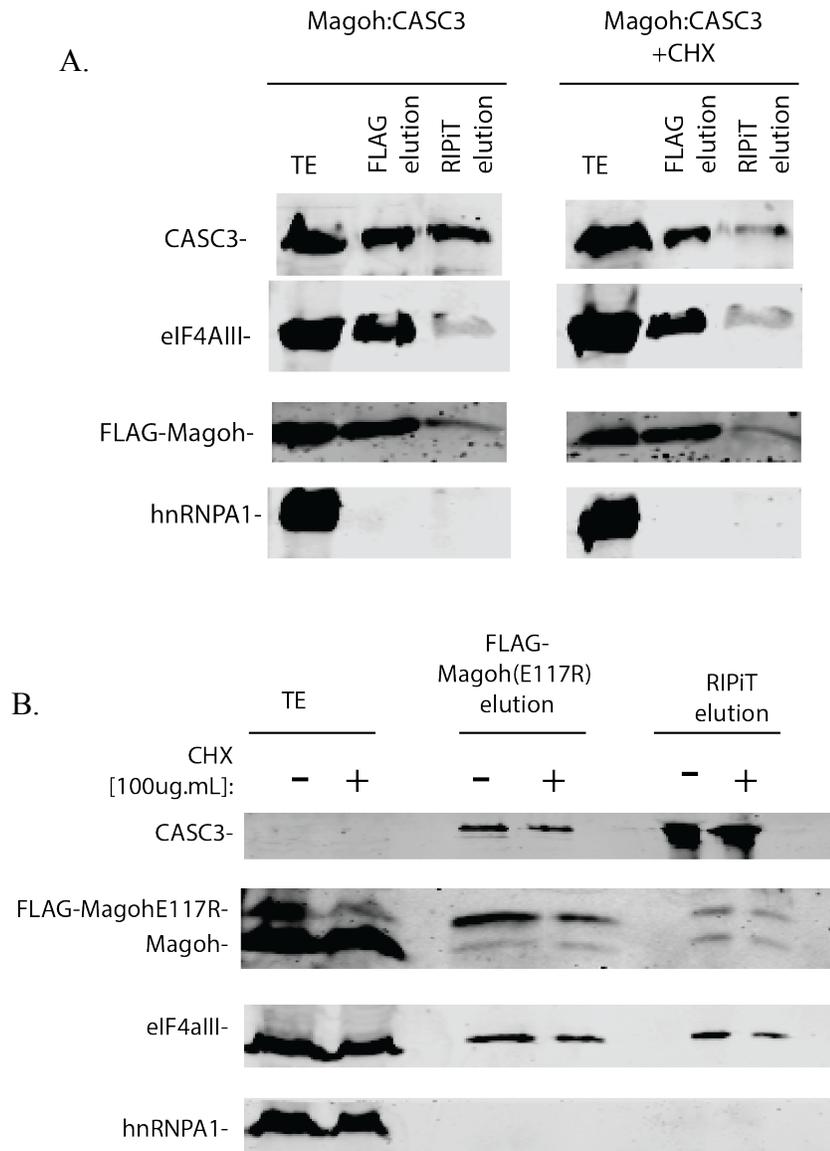


Figure 3.5 Protein fractions from FLAG-Magoh:CASC3 RIPiTs

A. Western blot of RIPiT fractions of FLAG-Magoh:CASC3 with and without CHX treatment. Total extract (TE), elution from FLAG-IP (FLAG elution), and elution from second IP (RIPiT elution) are indicated at top and proteins detected on the left.

B. Western blot of RIPiT fractions of FLAG-MagohE117R:CASC3 with and without CHX treatment. Total extract (TE), elution from FLAG-IP (FLAG elution), and elution from second IP (RIPiT elution) are indicated at top and proteins detected on the left.

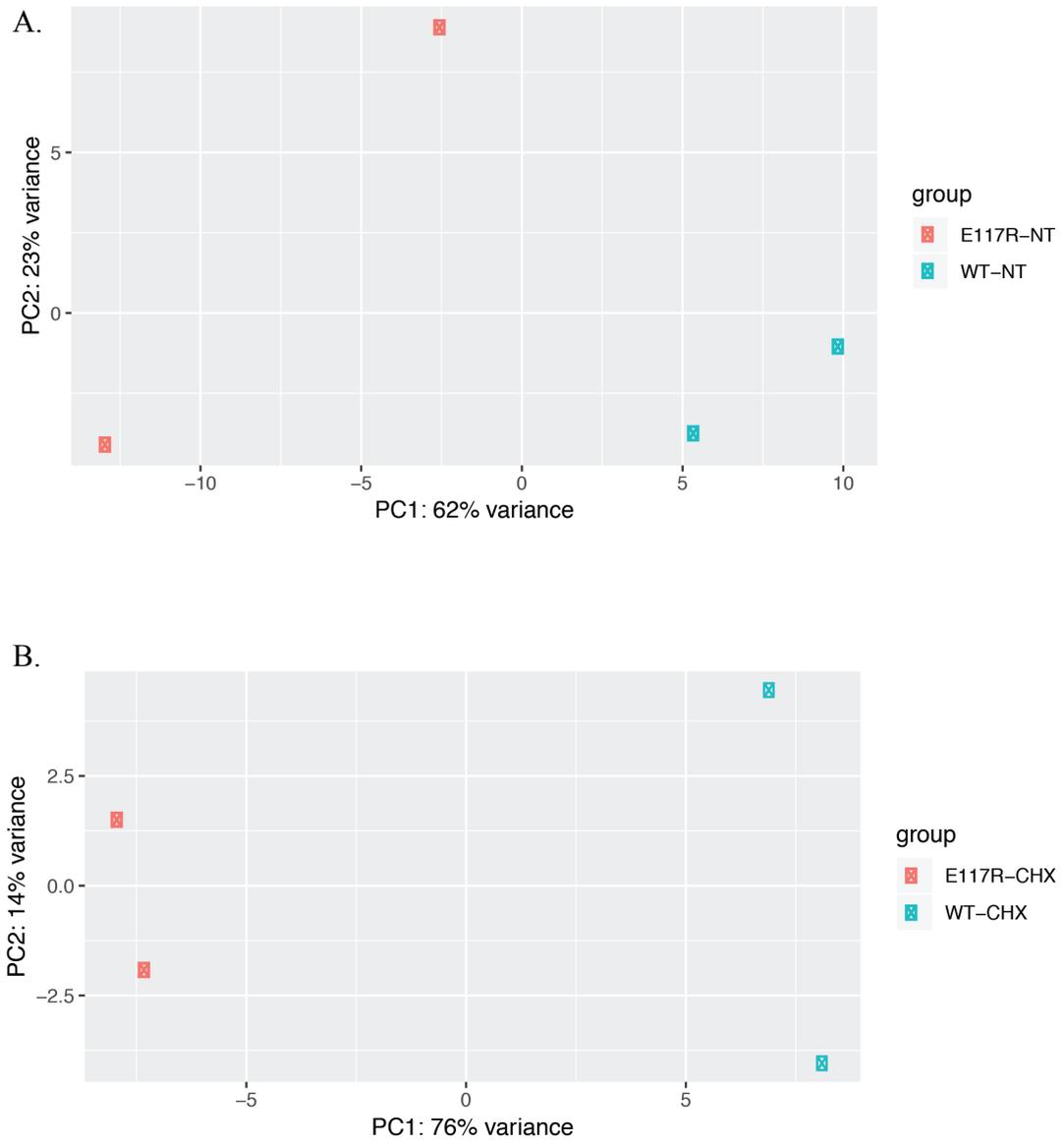


Figure 3.6 PCA plots comparing Magoh-CASC3 and MagohE117R-CASC3 RIPiT libraries

Figure 3.6 PCA plots comparing Magoh-CASC3 and MagohE117R-CASC3 RIPiT libraries (continued)

A. Principle component analysis (PCA) plots comparing libraries without cycloheximide treatment (NT) from FLAG-Magoh:CASC3 (WT-NT) to FLAG-MagohE117R:CASC3 (E117R-NT).

B. Principle component analysis (PCA) plots comparing libraries with cycloheximide treatment (CHX) from FLAG-Magoh:CASC3 (WT-CHX) to FLAG-MagohE117R:CASC3 (E117R-CHX).

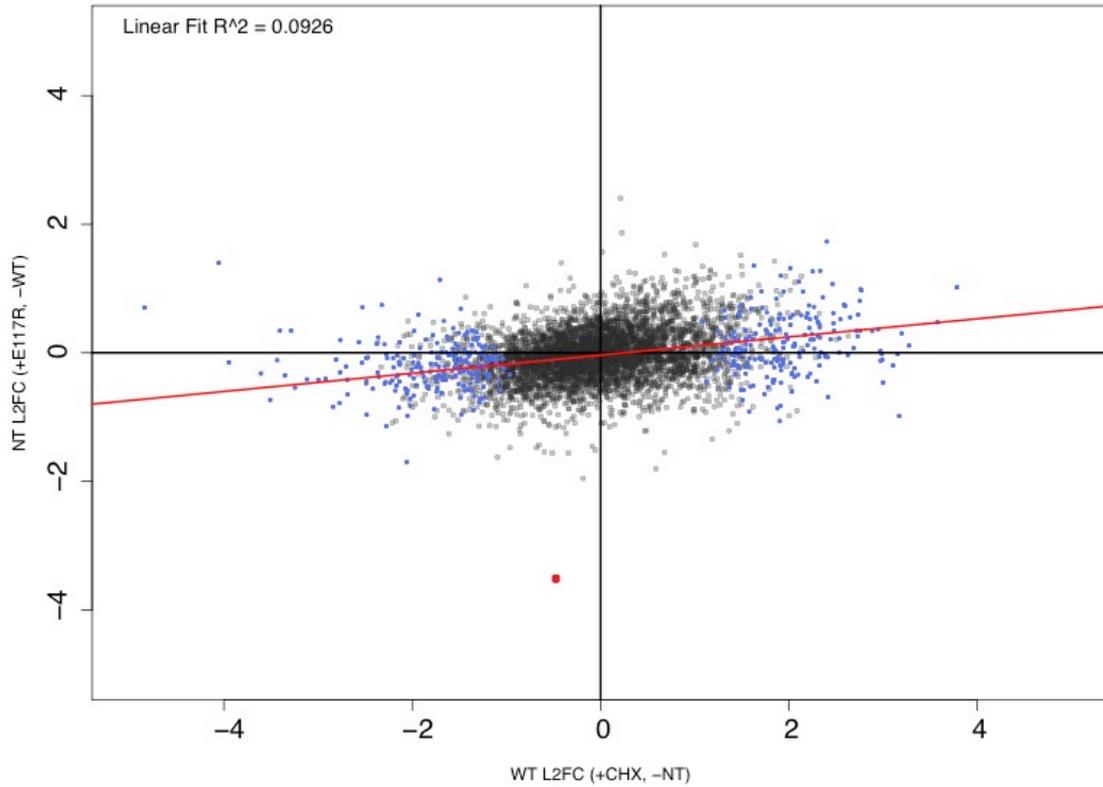


Figure 3.7 Disruption of ribosome translocation but not PYM-EJC interaction alters EJC occupancy at canonical positions

Scatter plot illustrating the effect of loss of PYM interaction and CHX treatment in only canonical regions per gene. Enrichment (\log_2 Fold Change (L2FC)) of MagohE117R-EJCs (+E117R) versus wild-type Magoh-EJCs (-WT) is plotted along the Y-axis. Enrichment of wild-type Magoh on genes in the presence (+CHX) versus absence (-NT) of CHX is plotted on the X-axis. Significantly enriched genes between -NT and +CHX are depicted in blue. Significantly enriched genes between +E117R and -WT are depicted in red.

Data analyzed by Robert Patton.

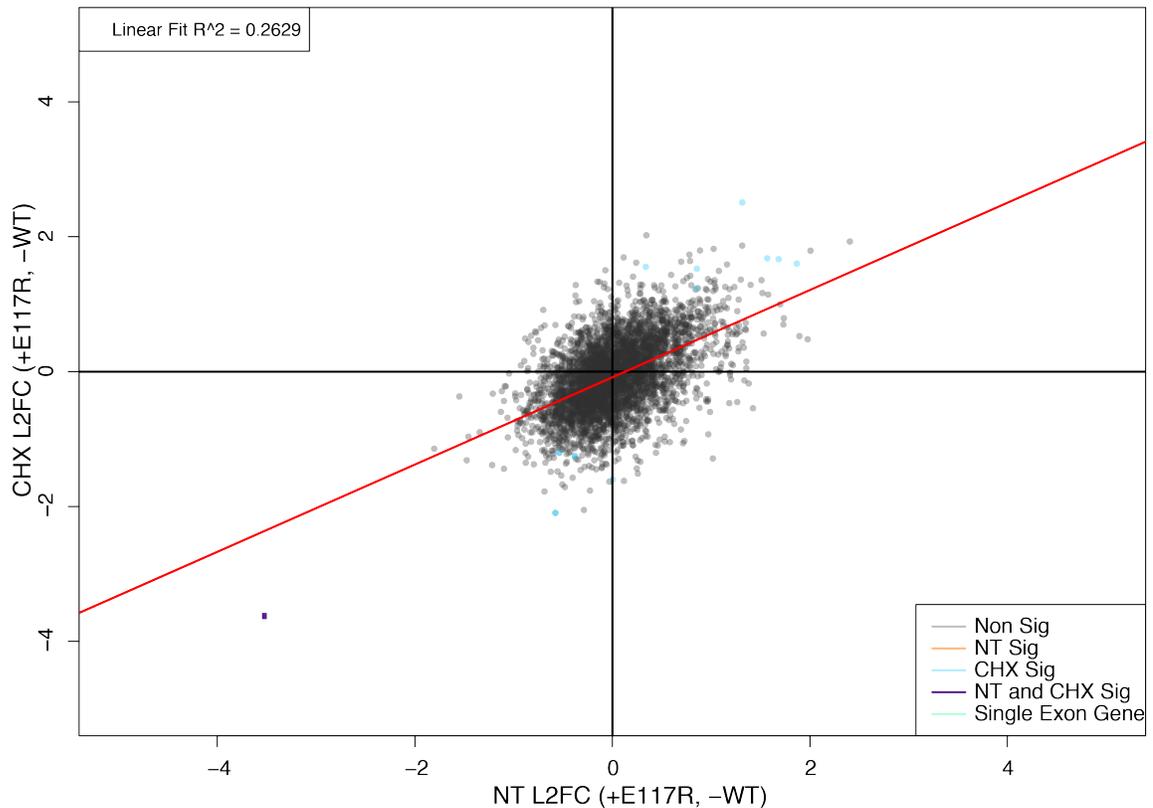


Figure 3.8 Disruption of PYM interaction does not significantly affect EJC occupancy at canonical positions

Scatter plot illustrating the similarity between $-NT$ and $+E117R$ in only canonical regions per gene. Enrichment (\log_2 Fold Change (L2FC)) of CHX-treated MagohE117R-EJCs ($+E117R$) versus wild-type Magoh-EJCs ($-WT$) is plotted along the Y-axis. Enrichment of untreated MagohE117R-EJCs ($+E117R$) versus wild-type Magoh-EJCs ($-WT$) is plotted along the X-axis. Significantly enriched genes between $-NT$ and $+CHX$ are depicted in orange. Significantly enriched genes between $+E117R$ and $-WT$ are depicted in blue. Single exon genes are highlighted with green outline. Data analyzed by Robert Patton.

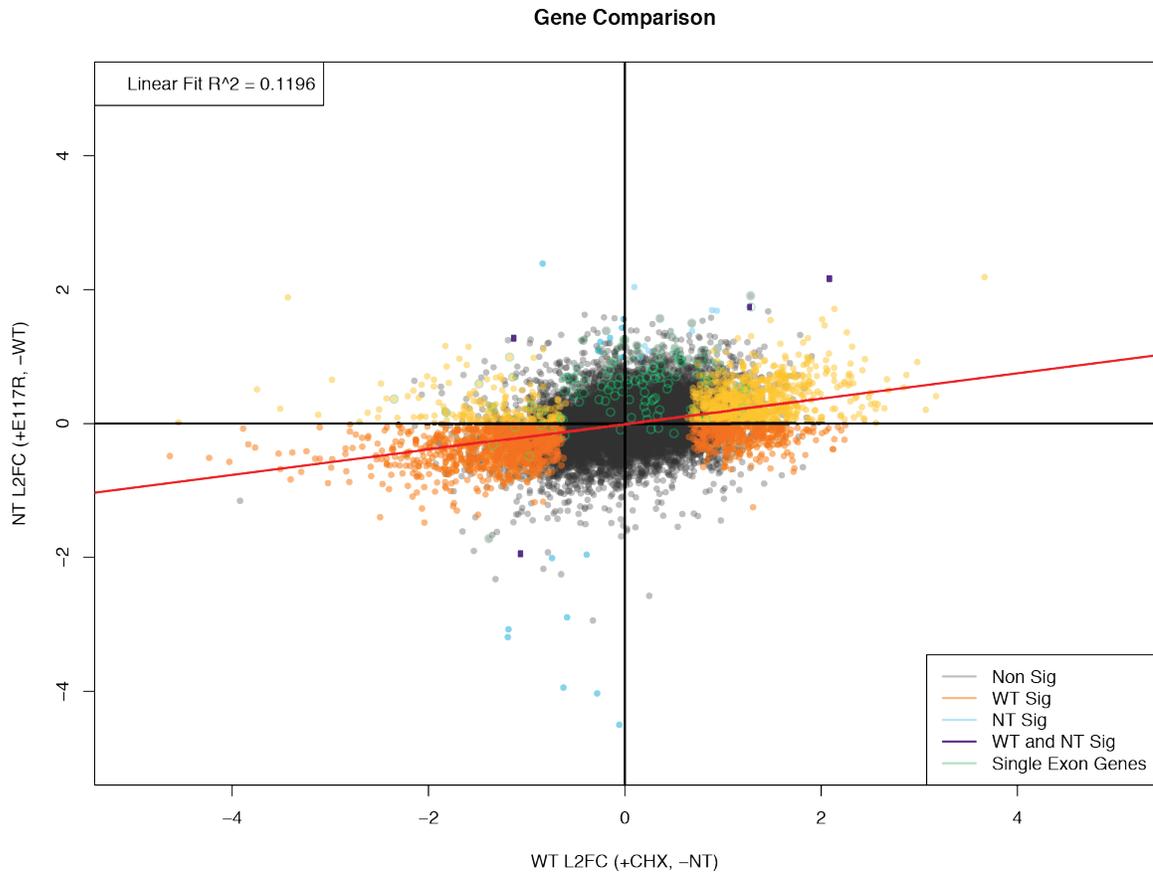


Figure 3.9 Loss of PYM interaction somewhat affects EJC occupancy on entire transcript

Scatter plot illustrating the effect of EJC occupancy upon loss of PYM interaction and CHX treatment per gene. Enrichment (\log_2 Fold Change (L2FC)) of MagohE117R-EJCs (+E117R) versus wild-type Magoh-EJCs (-WT) is plotted along the Y-axis. Enrichment of wild-type Magoh on genes in the presence (+CHX) versus absence (-NT) of CHX is plotted on the X-axis. Significantly enriched genes between -NT and +CHX are depicted in orange. Significantly enriched genes between +E117R and -WT are depicted in blue. Single exon genes are highlighted in green outline.

Data analyzed by Robert Patton.

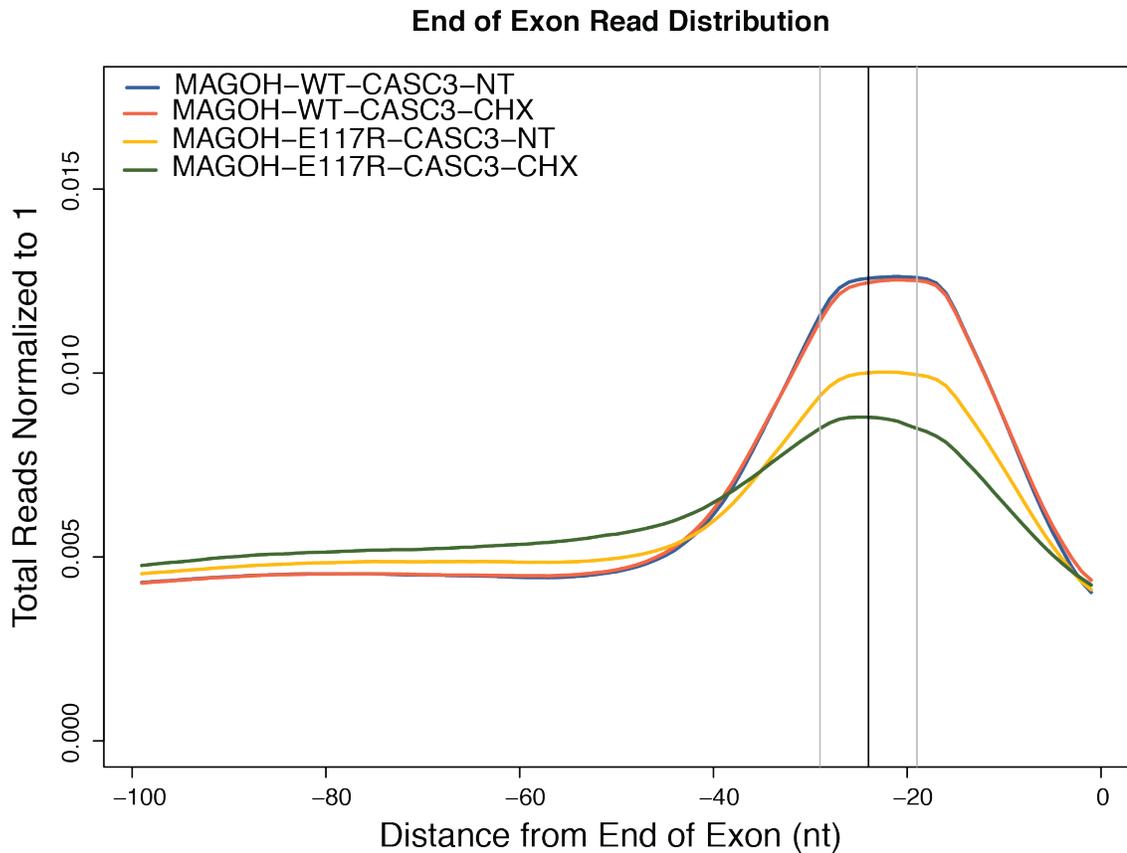


Figure 3.10 FLAG-MagohE117R:CASC3 footprints are detected less at the canonical position than FLAG-Magoh:CASC3

End of exon distribution of wild-type FLAG-Magoh:CASC3 and FLAG-MagohE117R:CASC3 RIPiTs. Distributions are each a composite of two replicate RIPiT libraries. Both cycloheximide treated (CHX) and untreated (NT) distributions are shown. Data analyzed by Robert Patton.

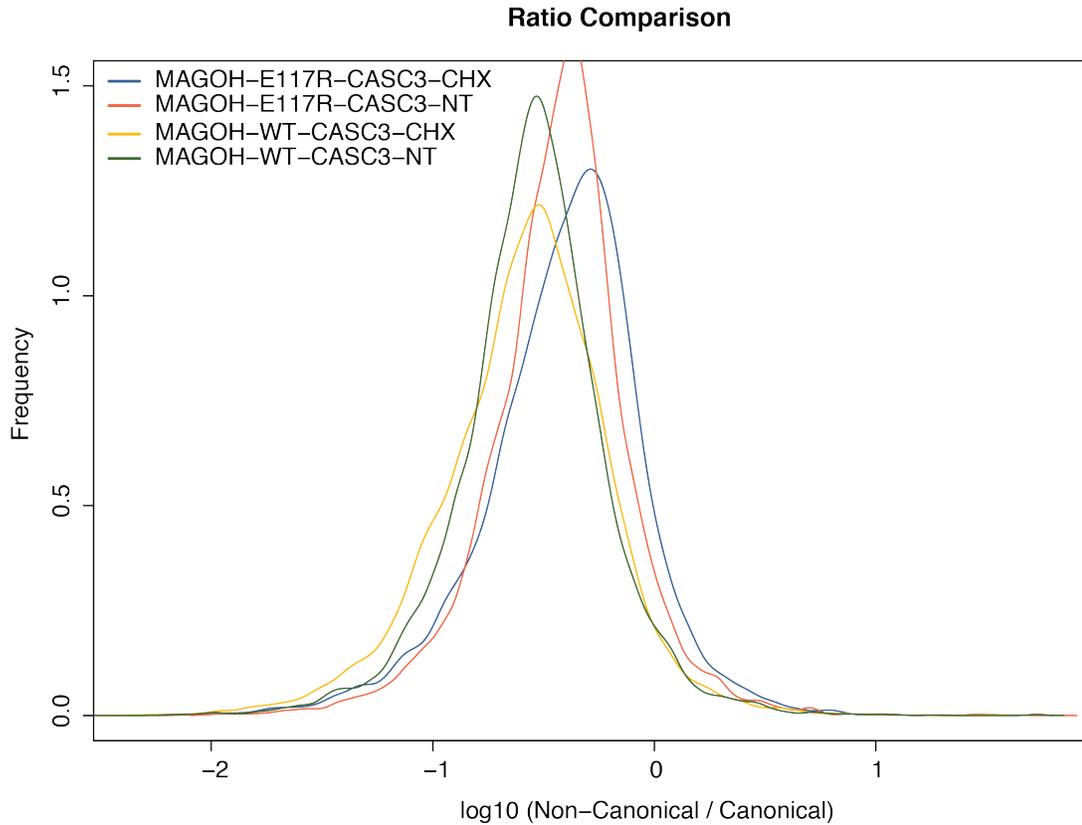


Figure 3.11 Frequency distribution of non-canonical/canonical ratios of wild-type and mutant EJC

Frequency distribution of non-canonical RPKM/canonical RPKM in FLAG-Magoh:CASC3 and FLAG-MagohE117R:CASC3 footprints either with cycloheximide treatment (CHX) or untreated (NT). Distributions are each a composite of two replicate RIPiT libraries. The ratio of non-canonical to canonical footprint RPKM was calculated per gene. FLAG-MagohE117R:CASC3 RIPiT libraries depict a higher median than FLAG-Magoh:CASC3 RIPiT libraries.

Data analyzed by Robert Patton.

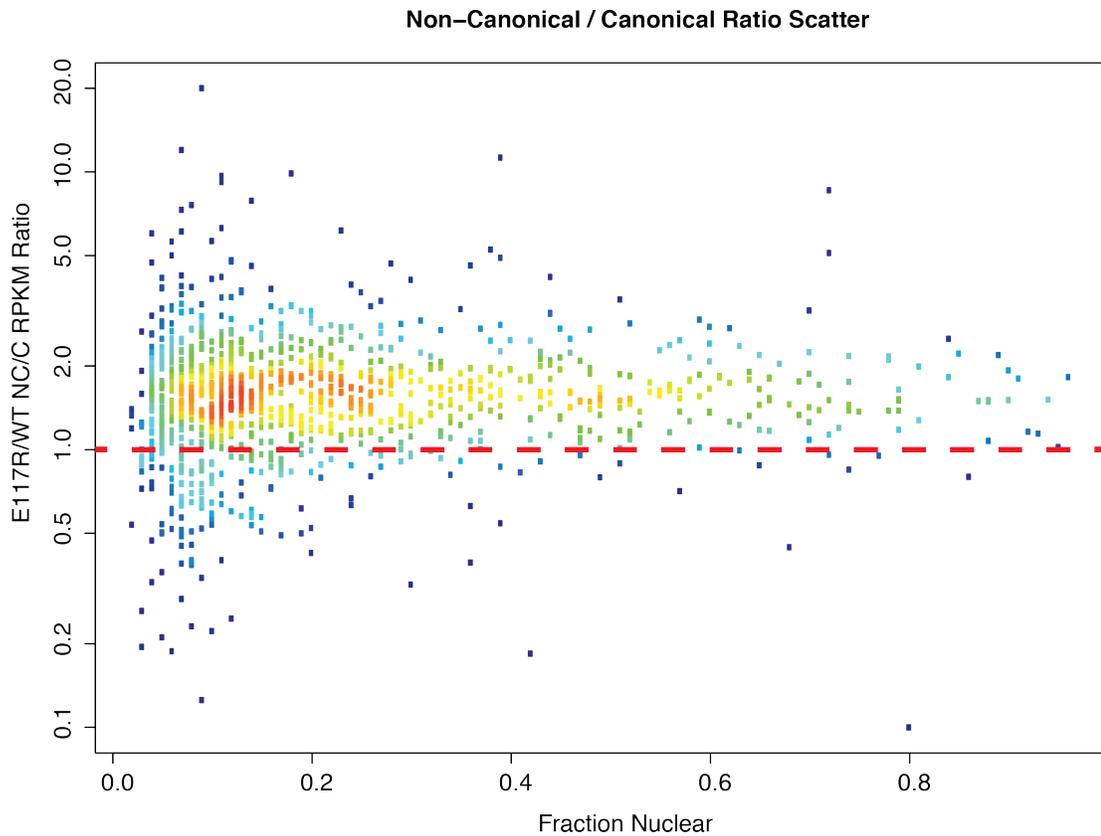


Figure 3.12 MagohE117R-EJCs have a higher non-canonical/canonical ratio than wild-type Magoh EJCs

RPKM ratios (Y-axis) per gene versus the fraction nuclear of each gene (X-axis). There is no observable relationship regarding non-canonical/canonical ratio vs. the fraction of mRNAs that are localized to the nucleus (X-axis), but the majority of genes display a E117R/WT NC/C ratio of >1.0 indicating that MagohE117R has a greater non-canonical/canonical ratio than wild-type. Data analyzed by Robert Patton.

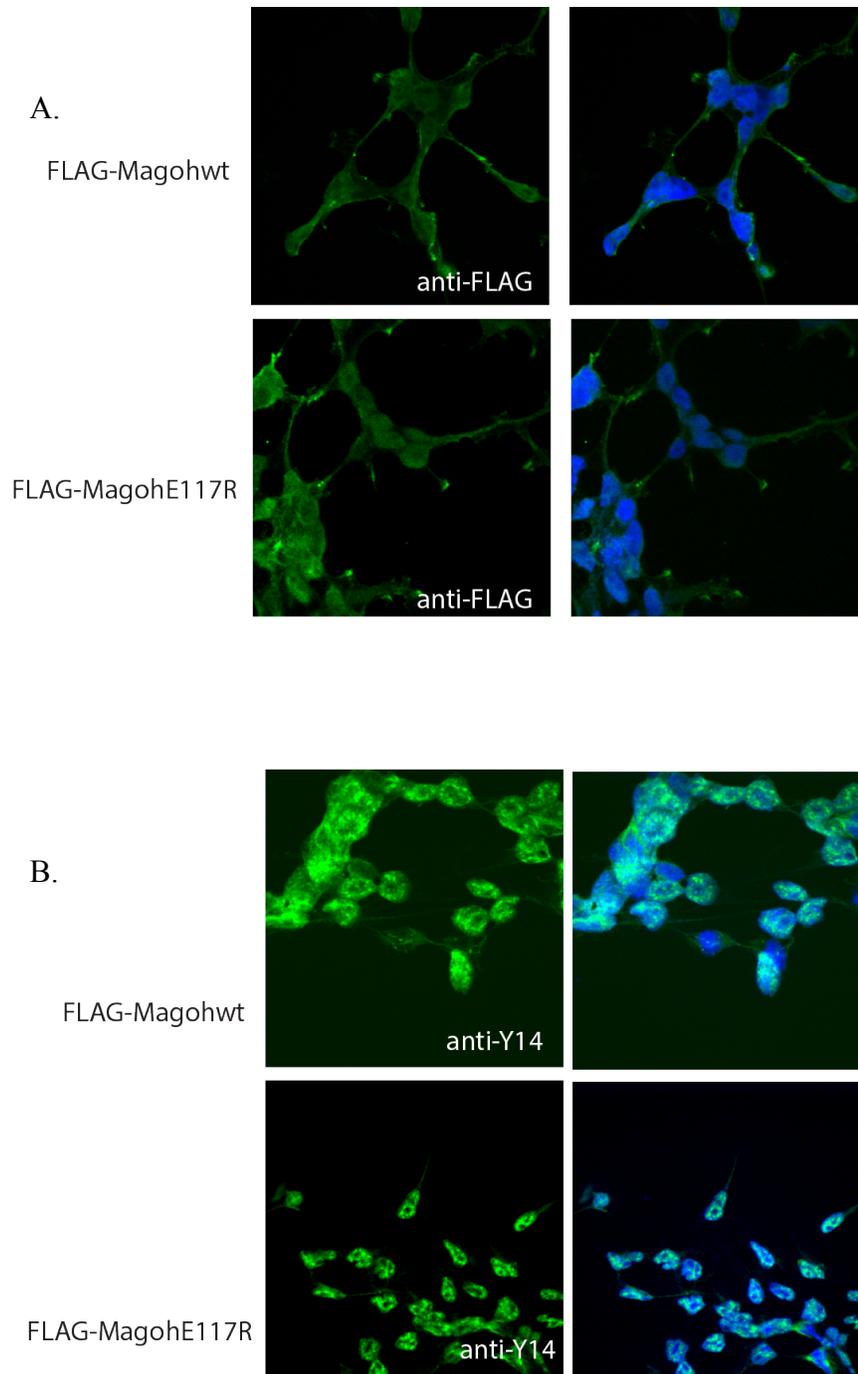


Figure 3.13 Subcellular distribution of Magoh and MagohE117R

Figure 3.13 Subcellular distribution of Magoh and MagohE117R (continued)

A. Immunofluorescence image of HEK293 cells against FLAG-tagged Magoh wild-type and E117R using antibody against FLAG show no obvious change in subcellular distribution of Magoh wild-type and MagohE117R.

B. Immunofluorescence image of HEK293 cells using antibody against Y14 in HEK293 cells overexpressing FLAG-tagged Magoh wild-type or E117R show no obvious change in subcellular distribution of Y14.

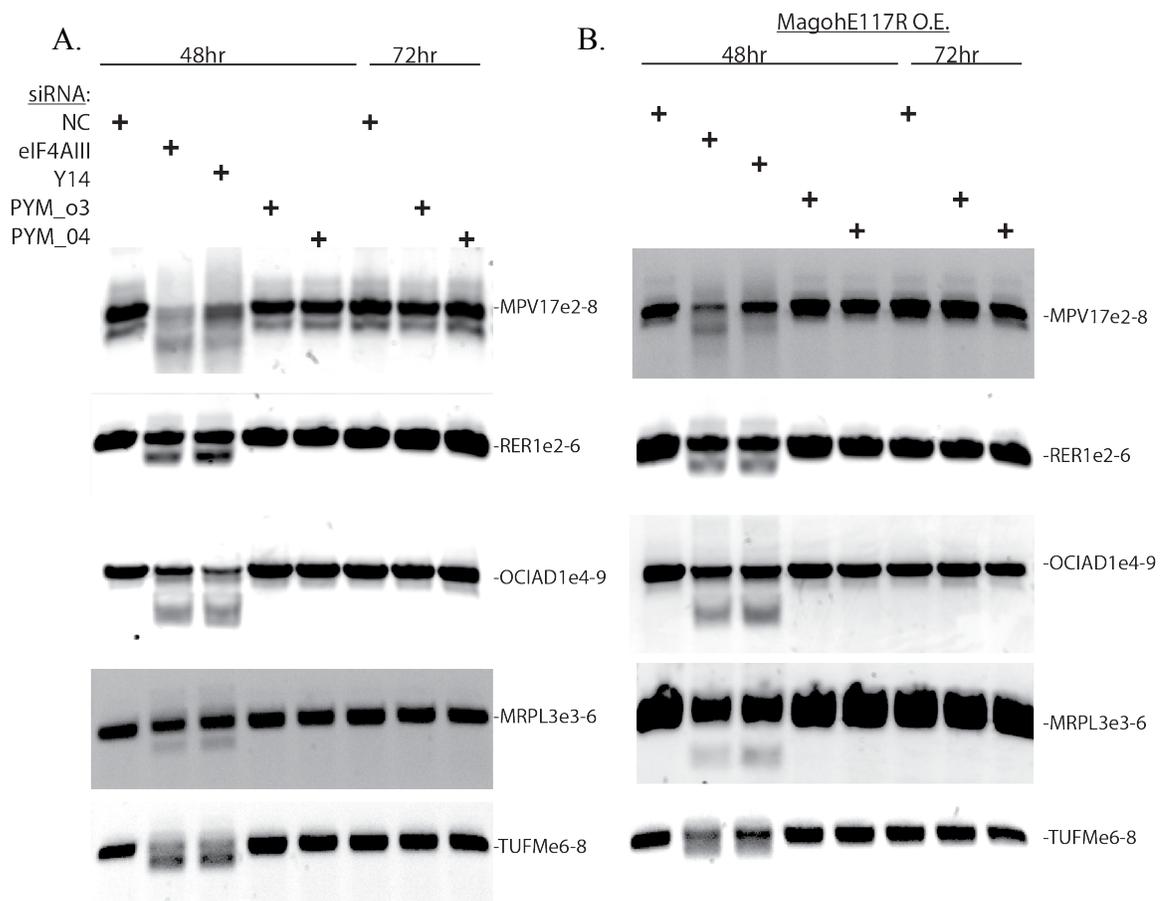


Figure 3.14 PYM effect on EJC-sensitive alternative splicing events

A. HEK293 cells were treated with eIF4AIII, Y14, or two different PYM siRNAs for either 48 or 72 hours as shown above each lane. RT-PCR of from RNA of each condition revealed that in no instance did PYM knockdown cause a change in EJC-dependent splicing.

B. The same was done as in (A.) in FLAG-MagohE117R cell lines with tetracycline-induced overexpression of MagohE117R. RT-PCR of from RNA of each condition revealed that in no instance did PYM knockdown cause a change in EJC-dependent splicing.

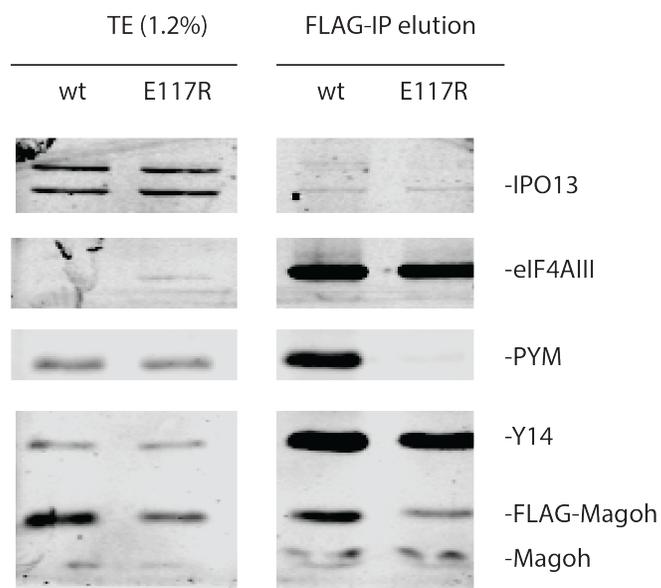


Figure 3.15 IPO13 interacts with both wild-type Magoh and MagohE117R

Western blot depicting proteins present (right) in FLAG-IP and total extract (TE) of HEK293 cells expressing either FLAG-Magoh (wt) or FLAG-MagohE117R (E117R).

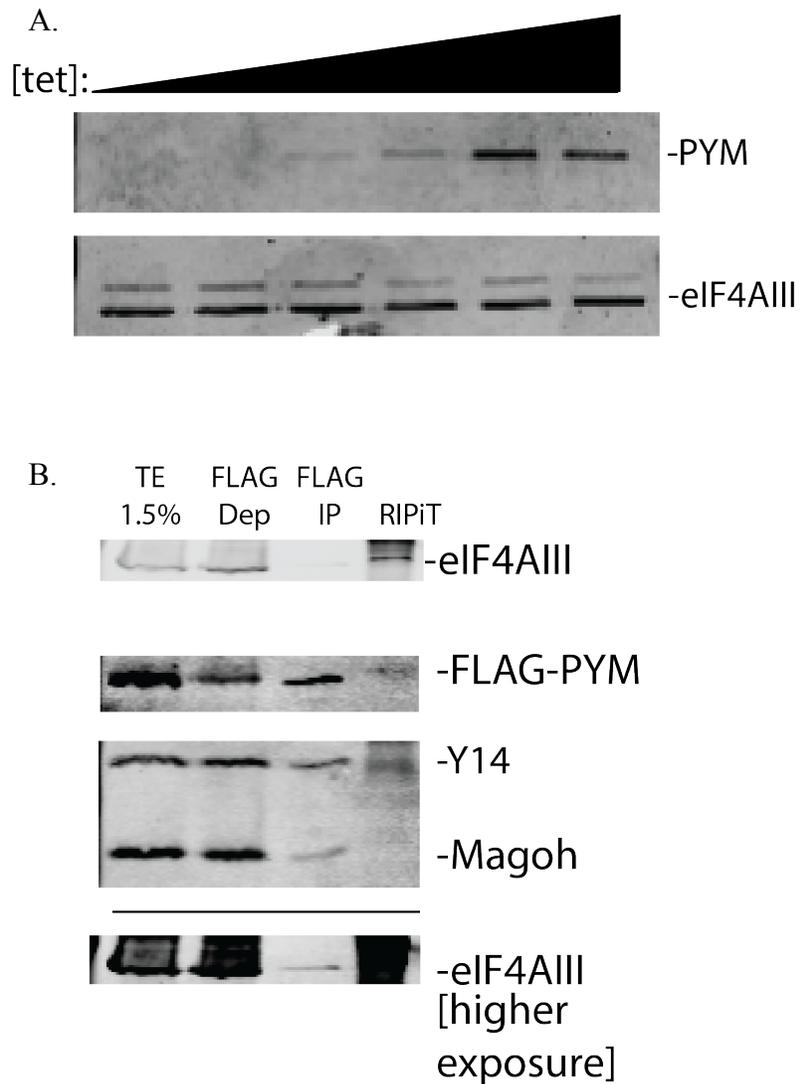


Figure 3.16 FLAG-PYM:eIF4AIII RIPiT protein

A. Western blot depicting proteins (right) in extracts of HEK293 cells stably expressing FLAG-PYM upon tetracycline induction (top)

B. Western blot depicting proteins present (right) in FLAG-PYM-eIF4AIII RIPiT total extract (TE), FLAG-IP unbound fraction (FLAG Dep), first IP elution (FLAG IP), and RIPiT elution (RIPiT).

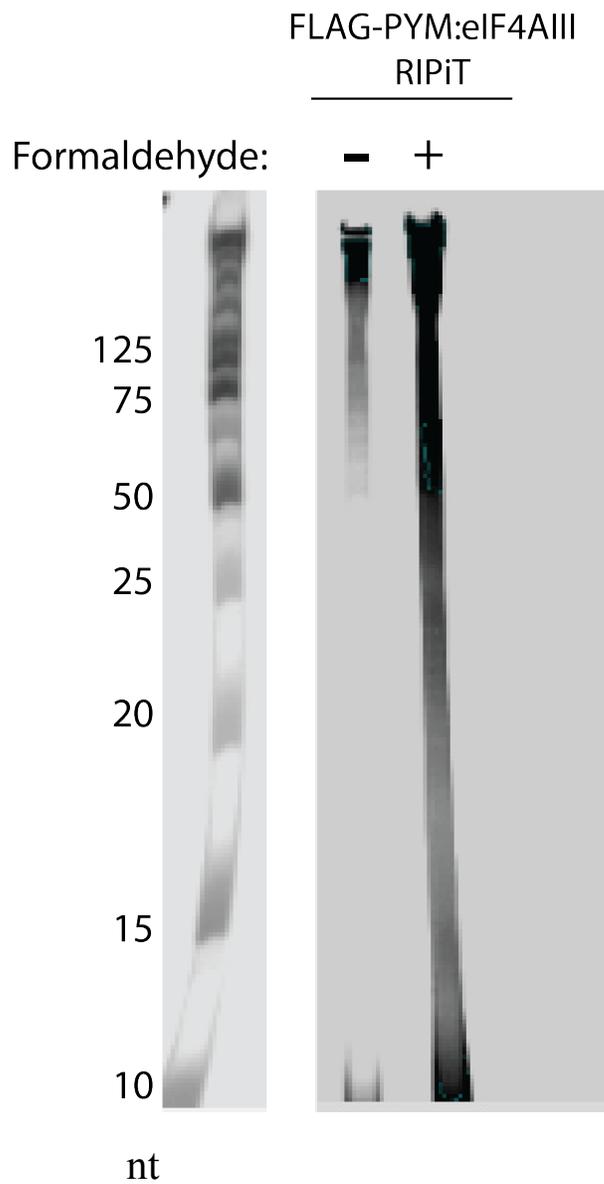


Figure 3.17 Formaldehyde crosslinking allows capture of FLAG-PYM:CASC3 footprints

Autoradiograph of end-labeled RNA footprints of FLAG-PYM:eIF4AIII with and without formaldehyde crosslinking (top). Nucleotide marker on left shows size of RNA footprints captured.

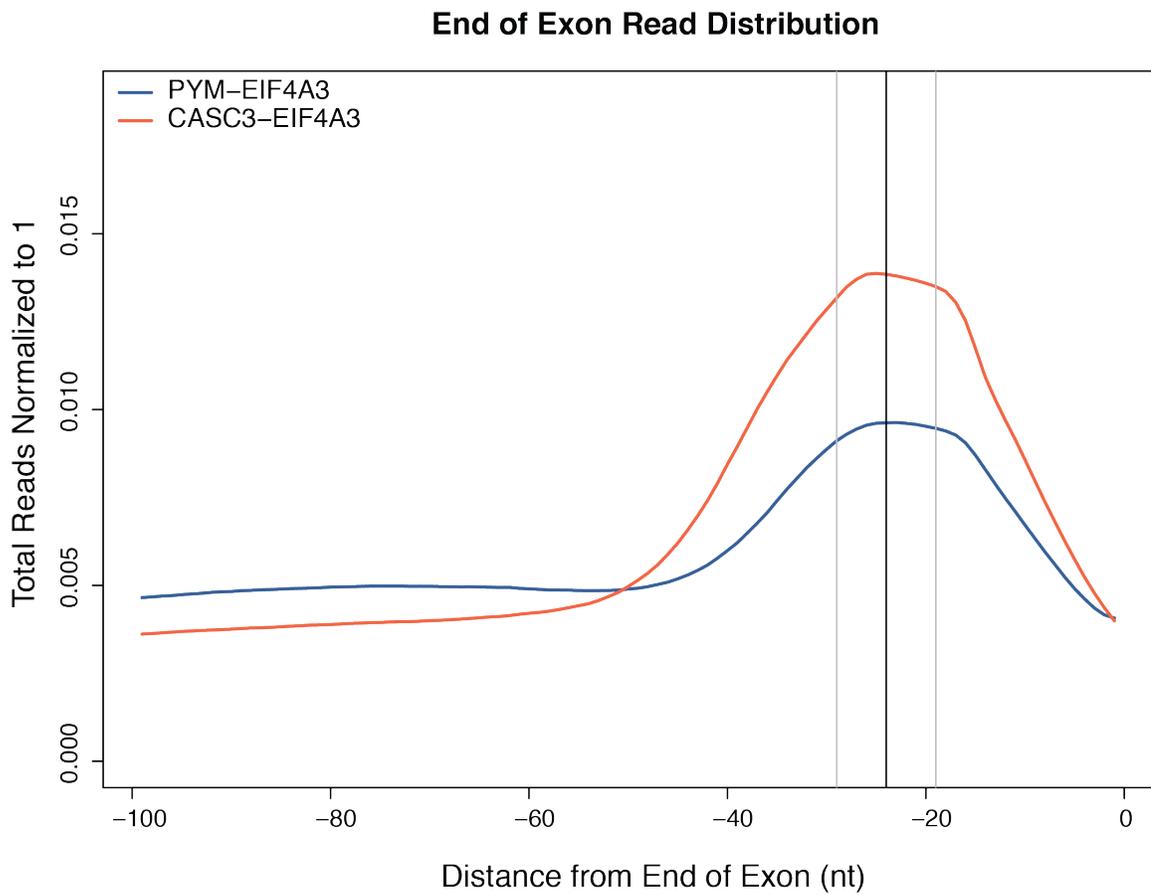


Figure 3.18 Comparison of end of exon distribution of FLAG-PYM:eIF4AIII and FLAG-CASC3:eIF4AIII

End of exon distribution of wild-type FLAG-PYM:eIF4AIII and FLAG-CASC3:eIF4AIII RIPiTs. Distributions are each a composite of two replicate RIPiT libraries. Data analyzed by Robert Patton.

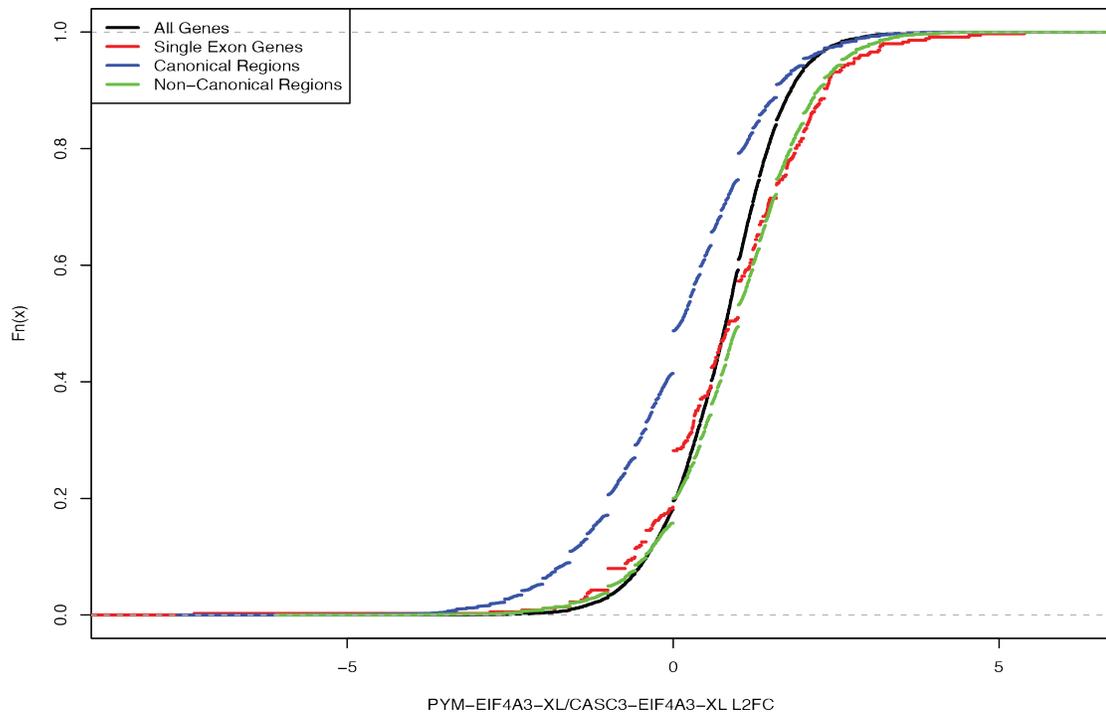


Figure 3.19 Cumulative distribution frequency of enrichment of FLAG-PYM:eIF4AIII in canonical regions, non-canonical regions, single exons, and all genes

DESeq2 log₂ Fold Changes were determined for each gene in canonical regions, non-canonical regions, single exon genes, and all genes. Frequencies of log₂ Fold Changes were then plotted for each class.
Data analyzed by Robert Patton.

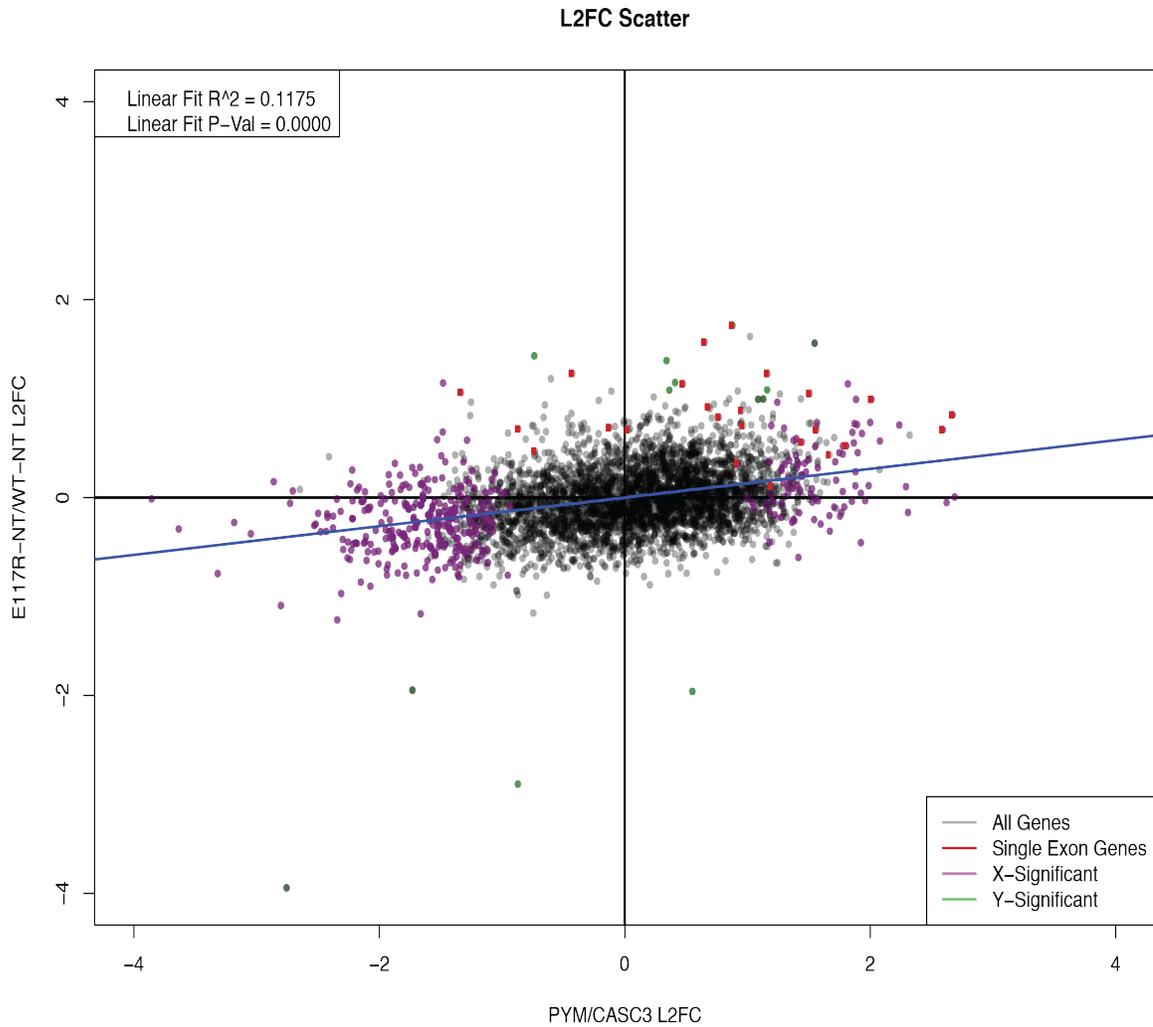


Figure 3.20 FLAG-PYM:CASC3 and FLAG-MagohE117R:CASC3 are both enriched on single exon genes compared to all genes

Scatter plot illustrating the effect correlation between FLAG-PYM:CASC3 and FLAG-MagohE117R enrichment on entire genes. Enrichment (\log_2 Fold Change (L2FC)) of MagohE117R-EJCs versus wild-type Magoh-EJCs with no cycloheximide treatment (NT) is plotted along the Y-axis. Enrichment of PYM-EJCs versus CASC3-EJCs is plotted on the X-axis. Significantly enriched genes between FLAG-MagohE117R and FLAG-Magoh are depicted in green. Significantly enriched genes between FLAG-PYM and FLAG-CASC3 are depicted in purple. Single exon genes are highlighted in red.

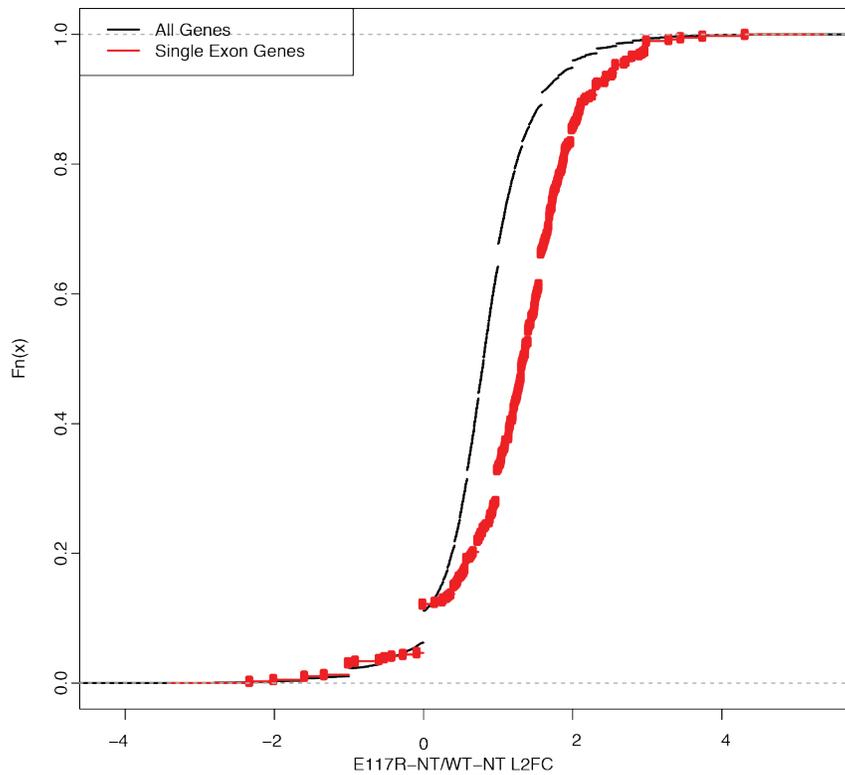


Figure 3.21 Cumulative distribution frequency of single exons genes in FLAG-MagohE117R vs. FLAG-Magoh

DESeq2 log₂ Fold Changes were determined for single exon genes and all genes. Frequencies of log₂ Fold Changes were then plotted for each class. Data analyzed by Robert Patton.

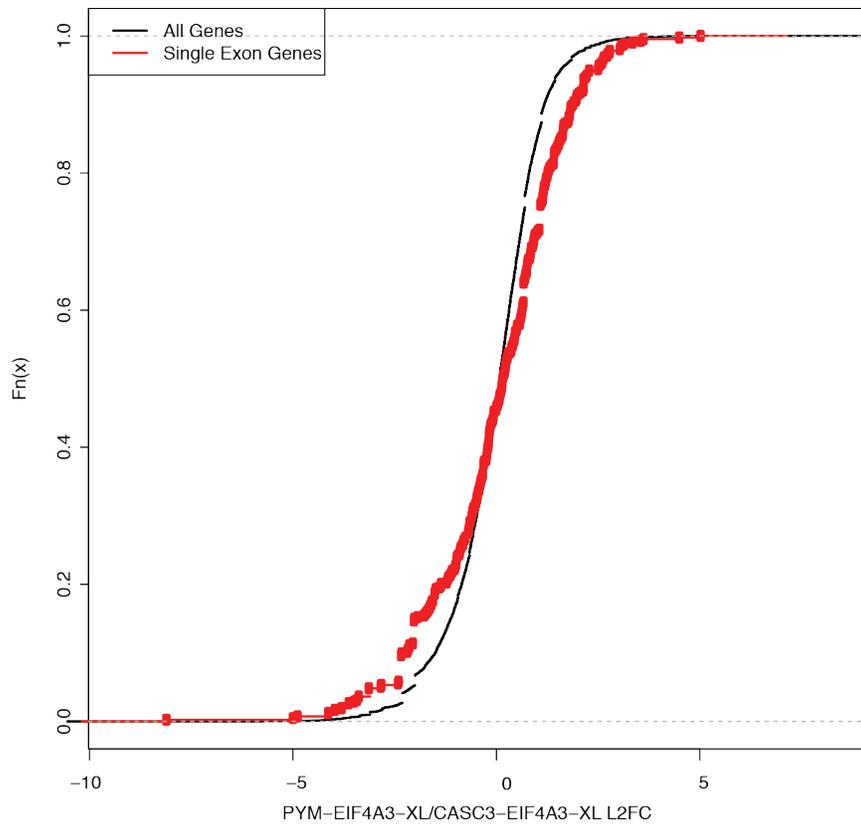


Figure 3.22 Cumulative distribution frequency of single exons genes in FLAG-PYM vs. FLAG-CASC3

DESeq2 log₂ Fold Changes were determined for single exon genes and all genes. Frequencies of log₂ Fold Changes were then plotted for each class.

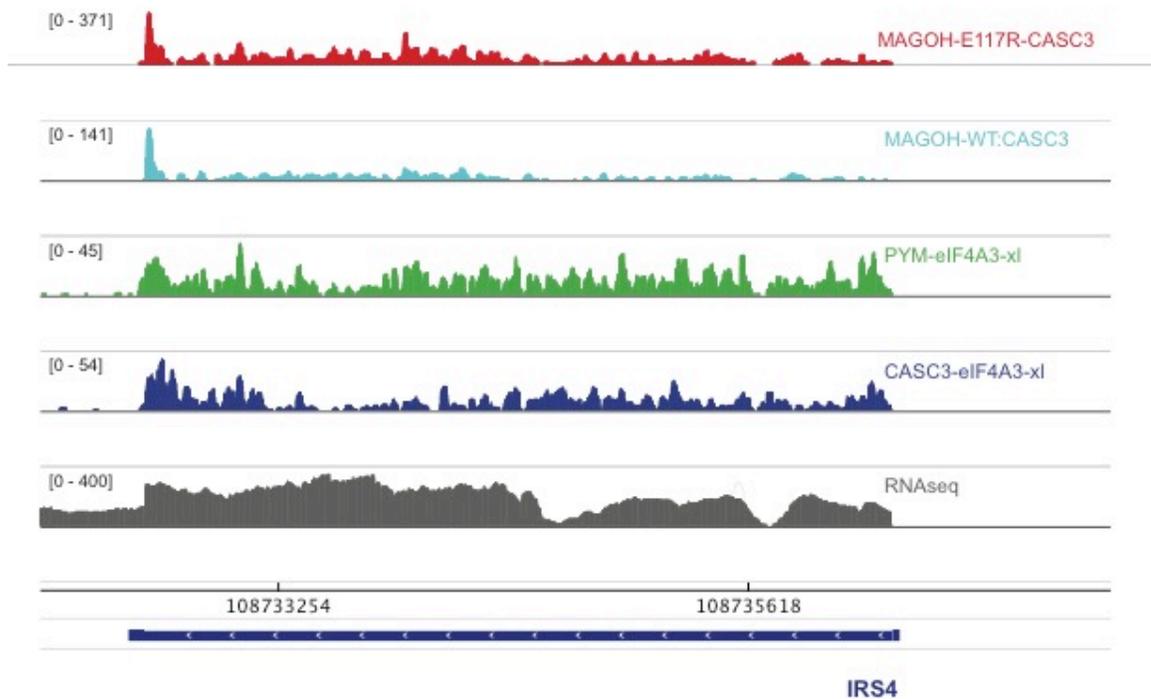


Figure 3.23 Integrated Genome Viewer (IGV) shot of EJC density on IRS4

IGV browser shot of EJC footprints in various RIPiT libraries. Each track represents coverage from one of two replicates per condition on single exon gene, IRS4. HEK293 RNAseq coverage is included for reference. Axis scale for each library is shown in brackets (left) and chromosomal coordinates are listed (below).

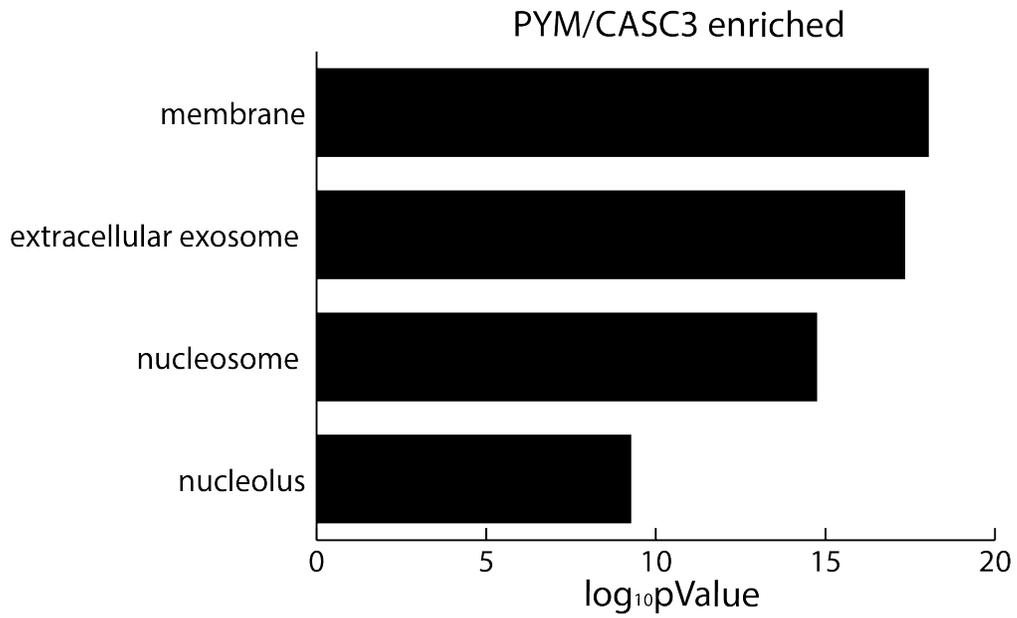


Figure 3.24 GO terms enriched in FLAG-PYM

Select, non-redundant GO terms most significantly represented among genes enriched in FLAG-PYM. Enriched genes had a p-value of <0.3 and a fold change of > 1.5 . 2240 genes were used as an input for DAVID.

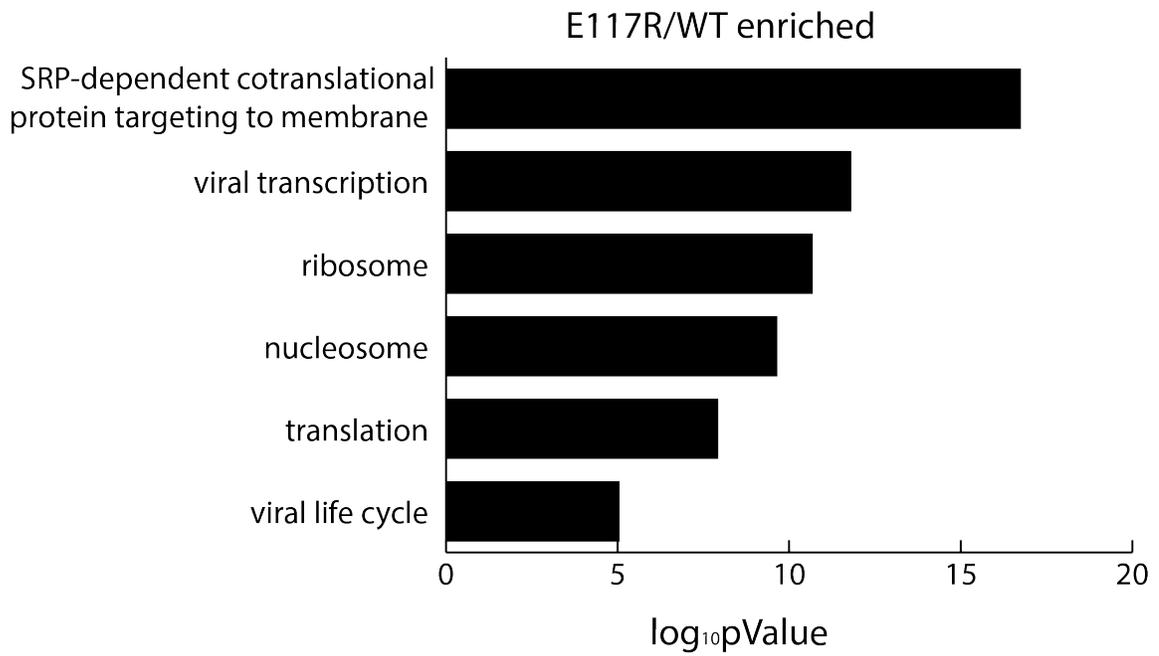


Figure 3.25 GO-terms enriched in FLAG-MagohE117R include histone genes

Select, non-redundant GO-Terms most significantly represented among genes enriched in FLAG-MagohE117R. Enriched genes had a p-value of <0.3 and a fold change of > 1.5. 1213 genes were used as an input to DAVID.

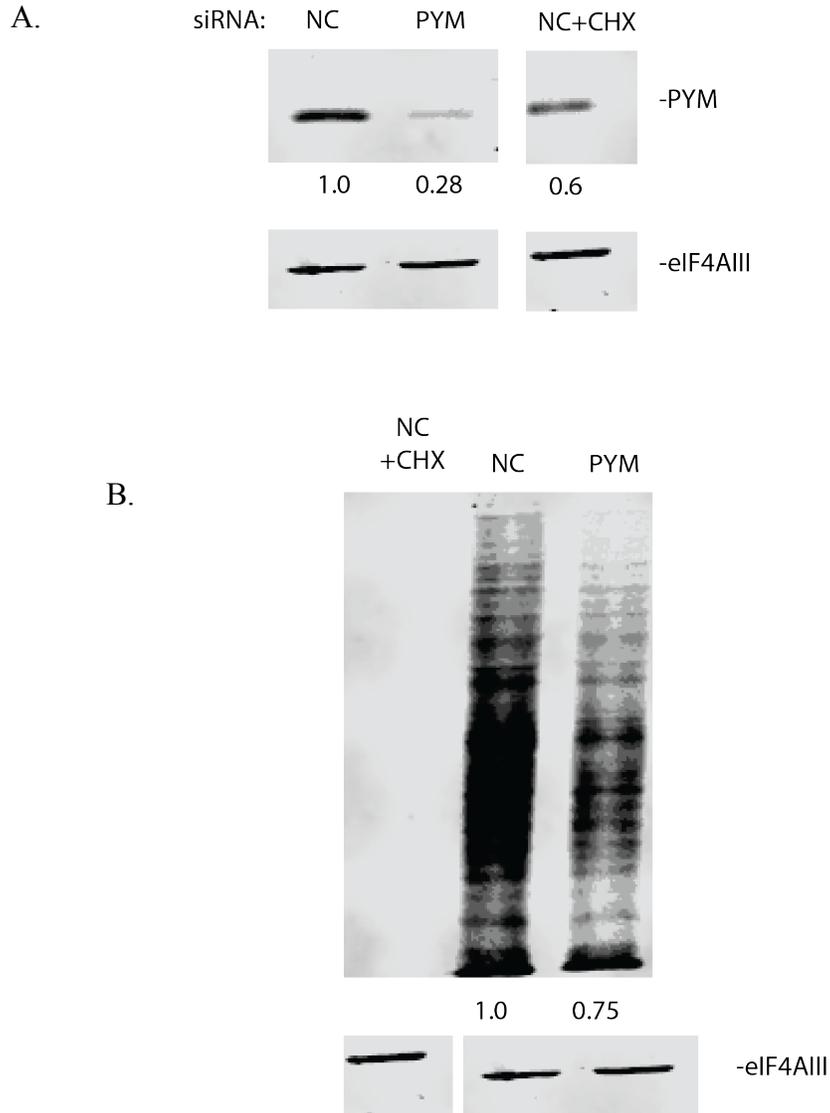


Figure 3.26 PYM knockdown causes reduced translation

A. Western blot depicting extent of knockdown of PYM following siRNA transfection.
 B. Western blot illustrating effect of PYM knockdown on nascent translation by measuring puromycin incorporation. Puromycin incorporation was normalized against eIF4AIII protein levels from the same samples. Cycloheximide treatment served as a positive control for inhibition of nascent translation.

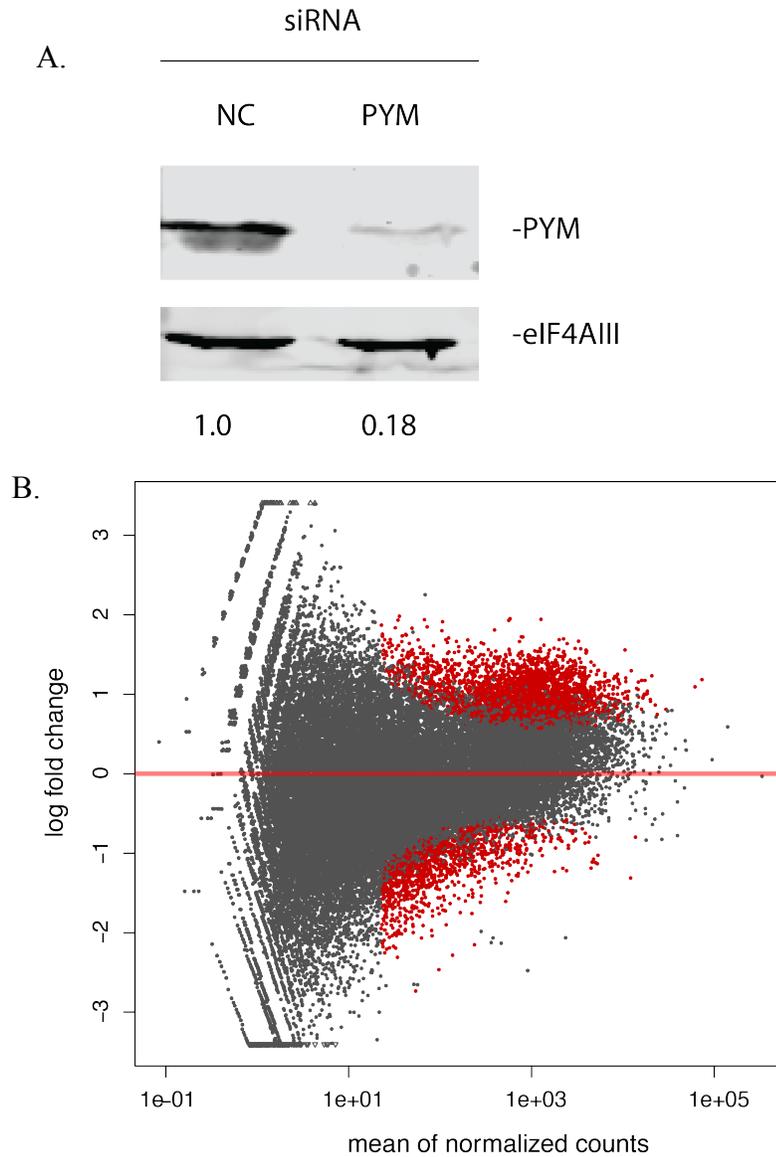


Figure 3.27 PYM knockdown in HEK293 cells results in misregulation of many transcripts

A. Proteins detected in HEK293 cells (right) transfected with either NC (negative control) or PYM siRNA (above). Percent knockdown was calculated by normalizing PYM protein level to eIF4AIII.

Figure 3.27 PYM knockdown in HEK293 cells results in misregulation of many transcripts (continued)

B. MA-plot generated using DESeq2 (Love et al., 2014) comparing enrichment transcripts in PYM knockdown and negative control RNAseq. Results are based on triplicate experiments of each condition. Genes significantly more abundant in either the negative control or PYM knockdown are highlighted in red. These transcripts are show a \log_2 Fold Change of at least ± 1 and $p \leq 0.1$.

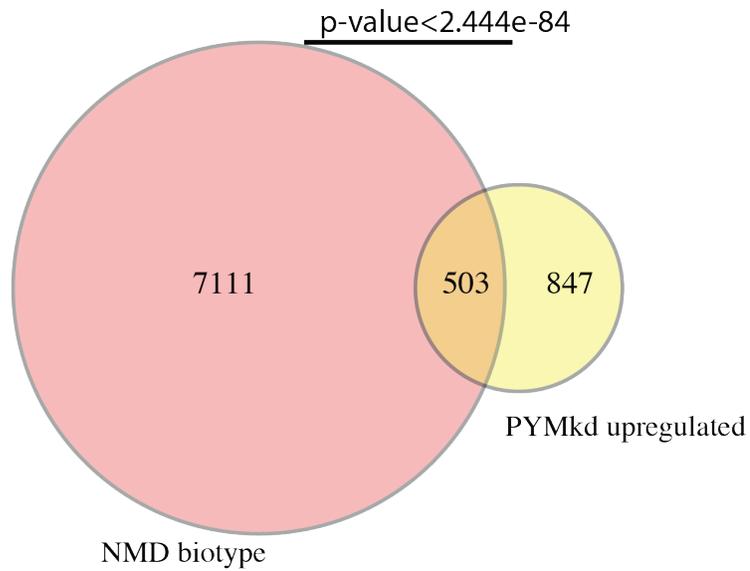


Figure 3.28 PYM knockdown results in stabilization of many NMD transcripts.

PYM knockdown resulted in the upregulation of 1350 transcripts relative to the negative control (see figure 3.27). Of these 1350 transcripts, 503 overlapped with a list of Ensembl annotated NMD substrates. An exact Fisher test based on hypergeometric distribution showed that the overlap between these two classes of transcripts is significant.

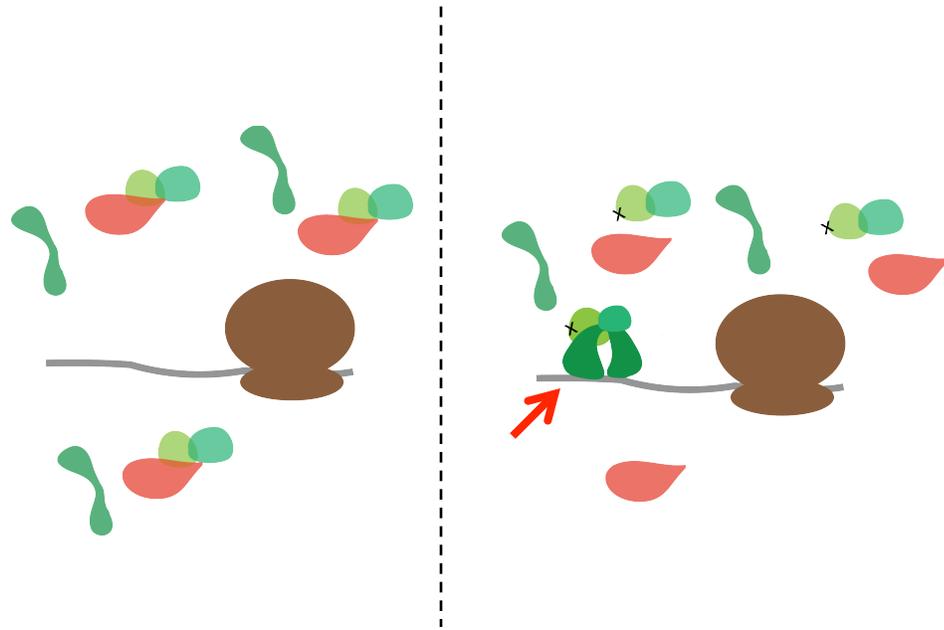


Figure 3.29 Schematic of spontaneous EJC assembly in the absence of PYM interaction with Y14:Magoh

On the left, PYM (red) sequesters Y14:Magoh to block EJC re-assembly with eIF4AIII (shown in open conformation) after EJCs are removed from RNA by the ribosome (brown). On the right, Y14:Magoh heterodimers containing MagohE117R (mutation indicated by “x”) do not interact with PYM and are thus able to interact with eIF4AIII and assemble into EJCs on non-canonical regions of RNA (red arrow).

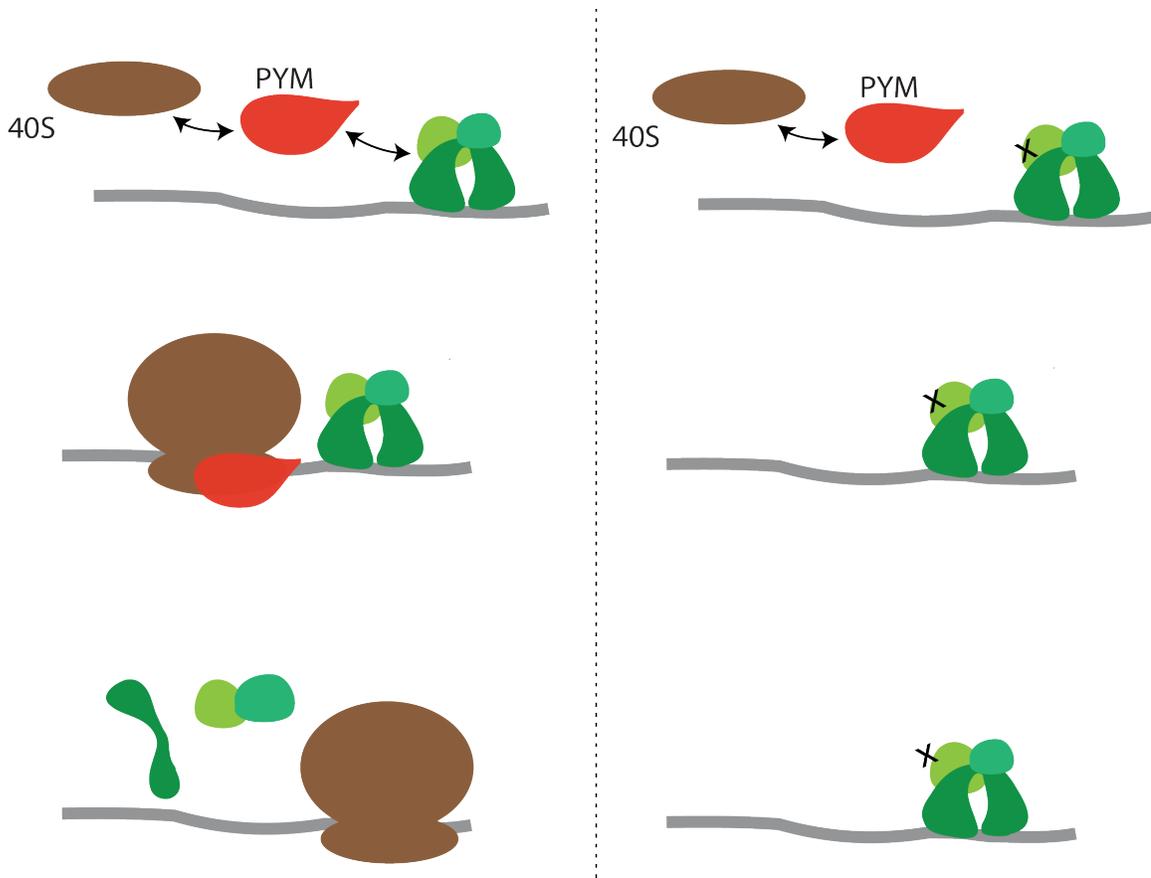


Figure 3.30 Schematic illustrating that reduced translation efficiency upon the loss of PYM interaction results in EJCs remaining bound to RNA longer

On the right, PYM recruits translation machinery to mRNAs bound by wild-type EJCs. Upon translation, EJCs are rapidly removed from the RNA. On the left, PYM is incapable of recruiting translation machinery (via interaction with the 40S subunit of the ribosome) to mRNAs bound by EJC mutants containing MagohE117R. Thus, mRNAs bound by mutant EJCs persist longer in the untranslated state, and mutant EJCs remain bound to the mRNA longer than their wild-type counterparts.

Chapter 4 Cytoplasmic EJCs are repositioned to the 3'UTR of mRNA and disassembled during translation

4.1 Introduction

The trimeric EJC core, consisting of eIF4AIII, Y14, and Magoh is deposited onto mRNAs 24 nt upstream of exon-exon junctions during the splicing reaction. The EJC core associates with many different peripherally interacting proteins as it traverses through the nucleoplasm and out into the cytoplasm. Both the complement of peripheral EJC proteins and the overall mRNP structure change as the mRNP makes this journey. For example, RNPS1 binds to more nuclear EJCs participating in a multimeric mRNP structure, while CASC3 binds primarily to cytoplasmic EJCs participating in a monomeric structure (Mabin et al., 2018).

Eventually, the mRNA engages with the pre-initiation complex. Upon recognition of the start codon and assembly of the remainder of the translation machinery, the ribosome begins elongation and processively translocates along the stretch of the coding sequence. Current model states that PYM (partner of Y14 and Magoh) interacts with the ribosome via its C terminus (Bono et al., 2004; Diem et al., 2007). PYM stably interacts with the Y14:Magoh heterodimer thus interrupting the heterodimer's binding interface with eIF4AIII. This allows for the ATP bound by eIF4AIII to be hydrolyzed and eIF4AIII enters an open conformation while Y14:Magoh remains sequestered by PYM.

According to current models, PYM associates with translating ribosomes and mediates the co-translational removal of EJCs from their sites of deposition (Diem et al., 2007; Dostie and Dreyfuss, 2002; Gehring et al., 2009b). However, there is contradictory evidence to this model. For example, in *Drosophila*, PYM does not interact with the ribosome (Ghosh et al., 2014). In mammalian cells, overexpression of PYM causes increased EJC disassembly. Thus, it appears that the functions of PYM and the ribosome may not be totally coupled. The function of PYM in EJC removal from endogenous RNA is incompletely understood.

Mapping of transcriptome-wide binding sites of the EJC have shown that the EJC binds RNA not only at the expected canonical position (-24 nt), but also at other non-canonical positions. Non-canonical EJC binding events occur within internal exons (at regions distal from the -24 position) as well as in the last exon and 3'UTR of mRNA. Notably, we observe a similar phenomenon in EJC CLIP-seq datasets. As the EJC is deposited upstream of exon-exon boundaries, it remains unknown how the EJC comes to be located in the 3'UTR or last exon of mRNA.

Here, we show that the non-canonical EJCs in the 3'UTR of mRNAs represent EJCs that are displaced from their original deposition sites and moved to the 3' end of the RNA by translating ribosomes.

4.2 Materials and Methods

4.2.1 Stable cell lines

As described previously in Section 2.3.1.

4.2.2 Endogenous and FLAG-tag Immunoprecipitations

As described previously in Section 2.3.2.

4.2.3 Polysome profiling:

HEK293 cell lysates were prepared by lysing $\sim 6 \times 10^6$ cells (1 10cm plate 80% confluent) growing at log phase in 0.5mL of ice-cold polysome lysis buffer [50mM Tris-HCl pH 7.5, 10mM KCl, 10mM MgCl₂, 150mM NaCl, 2mM DTT, 0.5 mM PMSF, 1mM Na-orthovanadate, 1 × Sigma protease inhibitor cocktail, 200μg/mL cycloheximide, 0.2% IGEPAL]. This gentle hypotonic lysis allows nuclear membrane to remain intact. Following ten minute incubation on ice, cell lysates were cleared at 15,000 × g. Lysates were snap frozen in liquid N₂ and stored at -80°C. 11mL 10-50% sucrose gradients were prepared and left to equilibrate overnight at 4°C. 400μL of cell lysate was dispensed onto gradient. Gradients were then centrifuged at 35,000rpm for 3 hours in a Sorvall-TH641 rotor at 4°C. Gradient fractions were collected at 500uL intervals with continuous measurement of UV absorbance at 254nm. Proteins were precipitated from fractions with TCA and presence of proteins in each fraction was detected via western blot.

4.2.4 Cell Fractionation

HEK-293 cells expressing a FLAG-tagged protein of interest were lysed from eight 15-cm plates in RSB-150 (10mM Tris-HCl pH 7.5, 100mM NaCl, 5mM MgCl₂, 2mM DTT, 40µg/mL Digitonin, 100 µg/mL cycloheximide, 1mM PMSF, Aprotinin (Sigma, A1153-10MG), 1 µg/ml Leupeptin (Sigma, L9783-5MG), 1 µM Pepstatin (Sigma, L4265-5MG)). Cytoplasmic fraction was separated from insoluble nuclear fraction at 2,000 xg for 10 minutes at 4°C. The insoluble pellet was resuspended in RSB-150(+0.5% Triton-X-100) and incubated on ice for 10 minutes then sonicated 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). Both the cytoplasmic and nuclear lysates were cleared by centrifugation at 15,000 x g for 10 minutes at 4°C. The supernatant was collected for each and used as an input for FLAG-IPs as below.

4.2.5 RNA:protein immunoprecipitation in tandem (RIPiT)

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 µg/ml Aprotinin (Sigma, A1153-10MG), 1 µg/ml Leupeptin (Sigma, L9783-5MG), 1 µM Pepstatin (Sigma, L4265-5MG), 1 mM PMSF (Sigma, P7626-5G)] supplemented with

150mM NaCl. Extracts were cleared by centrifugation at $15,000 \times g$ 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 $\mu\text{g/ml}$ FLAG peptide with gentle shaking at 4°C for 2 hr. The recovered elution volume was adjusted to 400 μL 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-eIF4AIII (Bethyl A302-980A, 10 $\mu\text{g/RIPiT}$). 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 $\mu\text{g/ml}$ CHX for 3 hr prior to harvesting. CHX was included at the same concentration in PBS (for washes before lysis) and cell lysis buffers.

4.2.6 RIPiT High-throughput sequencing library preparation

For RIPiT-seq, RNA extracted from ~80% of RIPiT elution was used to generate strand-specific libraries. For RNA-seq libraries, 5 µg 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 5nt barcode sequence, and ends in GG at the 5' end. Following RT, the extended RT primer is gel purified, circularized using CircLigase (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 100nt read lengths).

4.2.7 Adaptor trimming and PCR duplicate removal

After demultiplexing, fastq files containing unmapped reads were first trimmed using Cutadapt. A 12nt sequence on read 5' ends consisting of a 5nt random sequence, 5nt 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 TGAATTCTCGGGTGCCAAGG was removed from the 3' end as possible. Any reads less than 20nt 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.

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 (Saulière et al., 2012; Singh et al., 2012), and any reads overlapping by more than 50% were removed. Reads aligned to chrM (mitochondrial) were also counted and removed.

Human reference transcriptome

4.2.8 Metaplots of exon stop codons

Coding gene or exon coordinates were obtained from UCSC table browser for GENCODE v.29 track and knownCanonical table. These genes were filtered by CCDS ID to remove all but canonical coding transcripts. Coordinates of 50 or 100nt on either side of the stop codon were determined. All mRNAs with an intron within 50nt of the stop codon were removed from analysis using Bedtools intersect. Stop codons

coordinates were defined by taking the single nucleotide interval representing the final nucleotide of the CDS. Metaplots of EJC coverage were generated using NGSplot or DeepTools.

4.3 Results

4.3.1 EJCs exist in the last exon and 3'UTR

In 2012, two groups reported on the transcriptome-wide occupancy of the EJC using different experimental approaches (Saulière et al., 2012; Singh et al., 2012). Saulière et al. performed eIF4AIII CLIP-seq, while Singh et al. piloted a novel approach, RIPiT-seq. The findings of the two groups displayed very good agreement. In both studies, approximately 50% of peaks were localized to -24 nucleotides upstream of exon-exon junctions. The remaining peaks were localized outside of this region, and are referred to as non-canonical peaks. Included among these non-canonical regions are peaks in the last exon and 3'UTR of mRNAs (Figure 4.1).

4.3.2 EJCs are present in the polysome

According to current models, EJCs remain bound to RNA until the pioneer round of translation (Dostie and Dreyfuss, 2002; Lejeune et al., 2002; Maquat et al., 2010). Pioneer-round and pre-pioneer round mRNPs are characterized by association with the nuclear cap-binding protein, CBP80. CBP80 is exchanged for the cytoplasmic cap-binding protein, EIF4E, after nuclear export (Lejeune et al., 2002). EJC proteins interact

with CBP80-bound mRNPs, but not EIF4E-bound mRNPs (Lejeune et al., 2002). This suggests that EJCs are removed during early rounds of translation. In agreement, *in vitro* translation has been shown to separate Y14 from reporter mRNA (Dostie and Dreyfuss, 2002). However, my data shows presence of EJC core proteins in the polysome containing fractions. This association is sensitive to EDTA treatment, suggesting that the presence of EJCs in higher fractions is indeed dependent on *bona fide* intact polysomes (Figure 4.2). The association of EJC subunits with the polysome is inconsistent with the removal of EJCs during the pioneer round of translation and suggests that EJCs may remain intact on actively translating RNA.

4.3.3 Translation inhibition reduces the number of EJC footprints in the 3'UTR

As EJCs are present in the 3'UTR and associate with actively translating mRNA, I reasoned that EJCs placement in the 3'UTR may result from ribosome activity in the ORF. To test this, EJC core footprints were purified via RIPiT of FLAG-Magoh:eIF4AIII-containing complexes (Figures 4.3, 4.4, 4.5, and 4.6). Upon treatment with translation inhibitors, the amount of EJC occupancy in the 3' end of the CDS and into the 3'UTR is reduced (Figure 4.7). If EJCs are more abundant in the 3'UTR as a result of translation, perhaps the ribosome is actively pushing EJCs downstream as it decodes the mRNA.

4.3.4 The number of upstream exons impacts the amount of EJC occupancy in the 3'UTR

If EJCs are, in fact being displaced the 3'UTR, it stands to reason that the amplitude of coverage in a given transcripts 3'UTR would depend on the number of upstream splicing events. To test this, 3'UTR EJC occupancy was compared to exon number per transcript in RIPiT libraries of the EJC core (FLAG:Magoh:eIF4AIII). Transcripts were split into classes based on their exon count, and the amount of EJC occupancy near the 3'UTR was determined for each class. The exon number for each class was chosen to maintain a similar number of genes in each category and ensure similar median expression of each class. As expected, transcript classes with a higher number of exons had more reads in the area near the stop codon compared to transcripts with fewer exons (Figures 4.8 and 4.9). The same pattern was observed in an eIF4AIII CLIP-seq dataset prepared by Saulière et al.; however, there is no relationship between exon count and read density at the 3'UTR in RNAseq datasets. This suggests that upstream EJC deposition events influence the amount of EJC footprints seen in the 3'UTR.

4.3.5 More EJC occupancy in the 3'UTR of cytoplasmic fraction than nuclear fraction

The mRNPs in the nuclear compartment are not expected to be subject to any translation-specific effects. Thus we reasoned that 3'UTR EJC footprints from nuclear and cytoplasmic EJCs should differ. To test this, cell extract were separated into nuclear

and cytoplasmic fractions and conducted RIPiT-seq from these separated fractions. Metagene plots of stop codon coordinates show less nuclear EJC occupancy in the 3'UTR (Figure 4.10). To test this a different way, 3'UTR occupancy of alternate EJCs representing primarily nuclear or cytoplasmic mRNPs were compared. Chapter 2 was devoted to the characterization of CASC3 and RNPS1 containing EJCs. RNPS1 is largely representative of nuclear EJCs, while CASC3 is enriched on mRNPs in the cytoplasm. Consistent with the observations that cytoplasmic EJC cores were more abundant in the 3'UTR than nuclear EJC cores, CASC3-EJCs are more abundant in the 3'UTR than RNPS1-containing EJCs (Figure 4.11).

4.3.6 EJCs in the 3'UTR do not destabilize mRNA

Traditionally, EJC proteins in the 3'UTR are a hallmark of NMD substrates. This is because the presence of a downstream EJC characterizes upstream translation termination events as premature. Terminating ribosomes at premature termination codons recruit components of the SURF complex and the mRNA is targeted for NMD. However, in the last several years, a key observation of NMD substrates has emerged. As much as 10%-30% of the transcriptome is sensitive to levels of key NMD proteins even without the presence of a characteristic NMD-inducing feature, which are referred to here as non-canonical NMD targets (Hug et al., 2016). Transcriptome-wide studies of key NMD components, Upf1 and Smg6, in mammalian cells have identified targets of the NMD pathway (Schmidt et al., 2015; Tani et al., 2012). A high confidence panel of non-canonical NMD targets was identified by compiling genes that are both targets of Smg6

cleavage and Upf1-mediated decay. While this panel was made up of high-confidence NMD targets, they did not contain any obvious NMD-inducing feature, such as a 3'UTR intron. As is the case for nearly all transcripts, each of the transcripts in the panel displayed non-canonical EJC footprints in their 3'UTR. To test if the presence of non-canonical EJCs present in the 3'UTR could destabilizing these transcripts by recruiting NMD machinery, EJC core protein, eIF4AIII was depleted by siRNA transfection. ~60% depletion of eIF4AIII protein (data not shown) resulted in little or no change in the abundance of the non-canonical NMD substrates (Figure 3.12). Thus, the EJC signal present in the 3'UTR does not appear to have any significant destabilizing effect on the mRNAs tested.

4.3.7 PYM-eIF4AIII interacts with the EJC in the 3'UTR

It is possible that EJCs present in the 3'UTR do not have a destabilizing effect on mRNA, because the interaction between the EJC and 3'UTR stretches of mRNA is transient. This would imply that the region downstream of the stop codon is a site of active EJC disassembly. In order to investigate sites of EJC disassembly in Chapter 3, I conducted RIPiT-seq of the formaldehyde-stabilized interaction between the EJC disassembly factor, PYM, and the EJC core protein, eIF4AIII. Metagene plots aligned around the stop codon of mRNA show PYM:eIF4AIII interactions are detected at the stop codon and into the 3'UTR (Figure 3.13). This is consistent with the hypothesis that EJCs are actively disassembled at the 3'UTR, and thus prevented from destabilized bound transcripts via NMD.

4.4 Discussion

Since the discovery of the EJC in the early 2000's, we have come to a much greater understanding of how EJCs are deposited on mRNA and influence mRNA fate. EJC deposition during splicing upstream of exon-exon junctions was firmly established by several *in vitro* studies. The advent of deep sequencing technology and the ability to determine RBP footprints not only affirmed prior models of EJC properties and function, but also revealed further unknowns in the EJCs function and lifecycle.

Namely, the high percentage of EJC footprints in non-canonical stretches of RNA was unexpected and as yet, unexplained. Among these non-canonical stretches, EJC footprints are observed in single exon genes (discussed in Chapter 3) and in the final exons of spliced RNAs.

The results of these analyses indicate that EJC density in the 3'UTR is influenced by active translation and depends on the number of exons as well as the subcellular location of the mRNP to which the EJC is bound. Perhaps the action of the translating ribosome allows EJCs to be disassembled from the canonical position and they subsequently reassemble in 3' stretches. Alternatively, the action of the ribosome may physically displace assembled EJCs to the 3'UTR. A similar phenomenon is observed in for cohesin and single-stranded binding proteins (SSB); these are DNA binding proteins which are pushed to intergenic regions as a result of active transcription by RNAPolIII (cohesin) and processive translocase (SSB) (Glynn et al., 2004; Lengronne et al., 2004; Sokoloski et al., 2016). The EJC can be removed from RNA *in vitro* by DExH RNA helicases Ded1 and NPH-II, but whether these helicases are pushing the EJC downstream

or disassembling EJCs on contact has not been addressed (Fairman et al., 2004; Jankowsky and Bowers, 2006)

EJCs deposited in the 3' UTR due to splicing result in nonsense-mediated mRNA decay; however, mRNA's subject to NMD with EJC signal in the 3'UTR are not stabilized in response to EJC knockdown. Thus, the presence of non-canonical EJCs in the 3'UTR does not result in the destabilization of mRNAs. Perhaps this can be explained by PYM's action in the 3' UTR to prevent extrasplicing assembly of EJCs, as PYM-EJC footprints are detected in 3'UTRs.

In conclusion, the observation that EJC are ubiquitous in the 3'UTR of transcripts in numerous datasets is a novel characteristic of the EJC. The effect of translation, exon count, and subcellular location on 3'UTR EJC occupancy support a model where EJCs are displaced to 3' stretches in a translation-dependent manner. It remains to be determined if these 3' UTR, non-canonical EJCs arise from physical pushing of assembled EJCs to the 3'UTR or reassembly events. Translation-dependent EJC disassembly may result in a high local concentration of EJC subunits, allowing for spontaneous assembly of EJCs outside of the ORF.

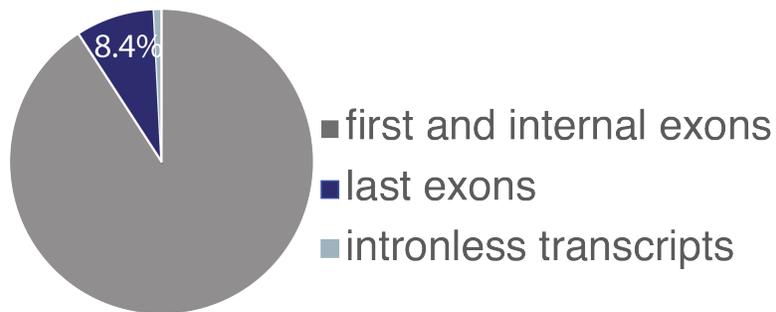
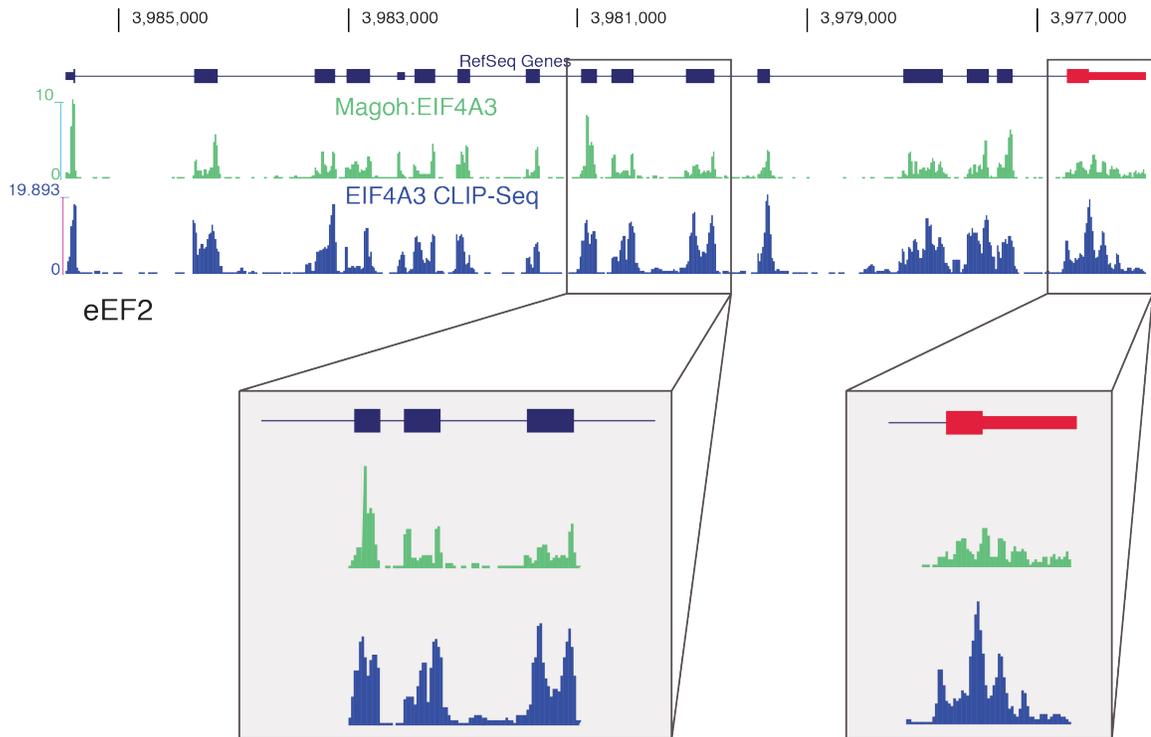


Figure 4.1 Non-canonical EJCs exist in the last exon and 3'UTR

A. Genome browser shot of eEF2 show EJC footprint density along gene from FLAG-Magoh:eIF4A3 RIPiT-seq and eIF4A3 CLIP-seq (Saulière et al., 2012). Thick blue boxes represent exons; the red boxes highlight the last exon and 3'UTR; thin lines represent introns.

Figure 4.1 Non-canonical EJCs exist in the last exon and 3'UTR (continued)

B. Pie-chart illustrating genic locations of EJC peaks from RIPiT-seq analysis. RIPiT samples and peak calling were conducted by Singh et al. (Singh et al., 2012).

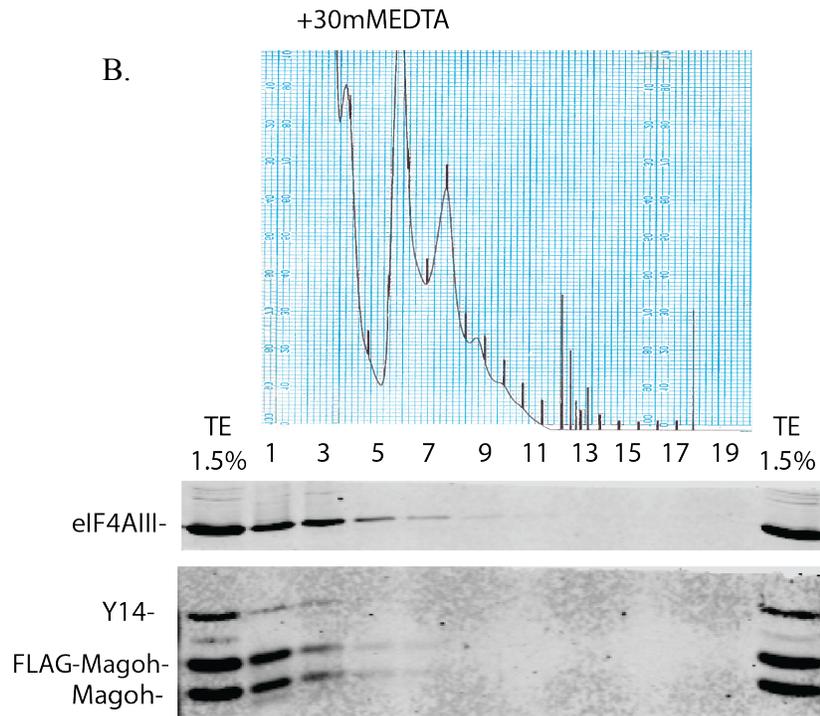
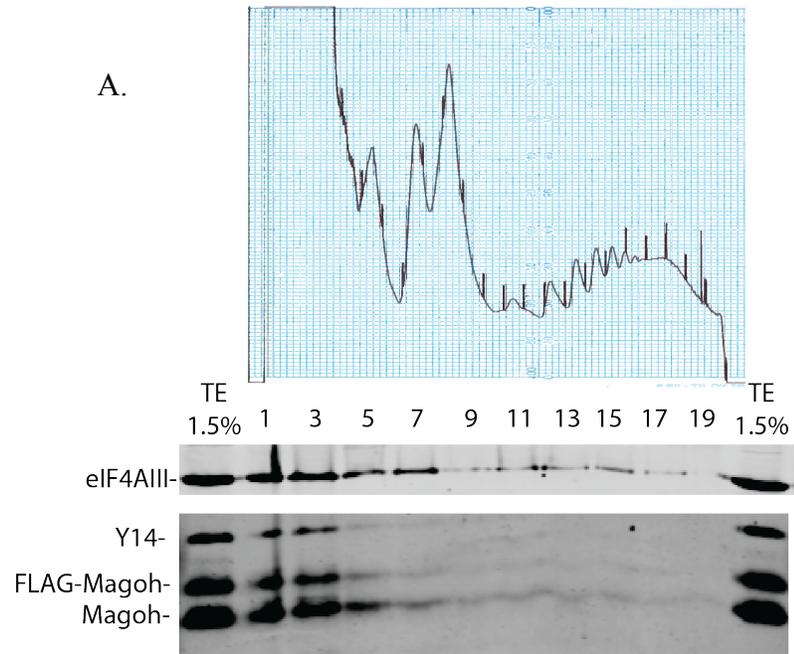


Figure 4.2 EJC core proteins co-sediment with polysome fraction

Figure 4.4.2 (continued)

A. Western blot of proteins (left) precipitated from sucrose gradient resolved fractions (above) of HEK293 cell lysate. Polysome-containing fractions (fractions 9-19) were identified by monitoring continuous UV absorbance at 254nm.

B. Western blot of proteins (left) precipitated from sucrose gradient resolved fractions (above) of HEK293 cell lysate subject with addition of 30mM EDTA to dissociate polysomes. Polysome-containing fractions (fractions 9-19) were identified by monitoring continuous UV absorbance at 254nm.

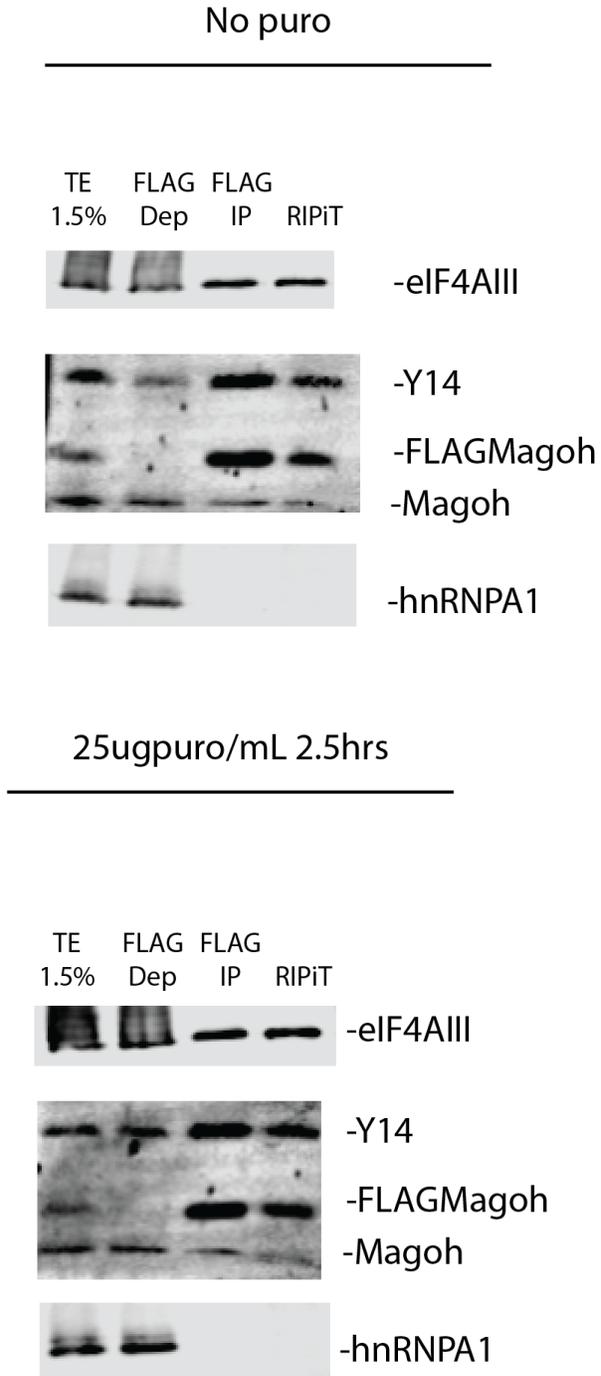


Figure 4.3 Puromycin-treated EJC core RIPiT fractions

Figure 4.3 Puromycin-treated EJC core RIPiT fractions (continued)

A. Western blots depicting proteins (right) present in fractions taken from total extract (TE), unbound fraction in FLAG IP (FLAG Dep), elution from first IP (FLAG IP), and elution from RIPiT (RIPiT).

B. Western blots depicting proteins (right) present in fractions taken from total extract (TE), unbound fraction in FLAG IP (FLAG Dep), elution from first IP (FLAG IP), and elution from RIPiT (RIPiT) from HEK293 cells treated with puromycin for 2.5 hours at a concentration of 2.5 μ g/mL.

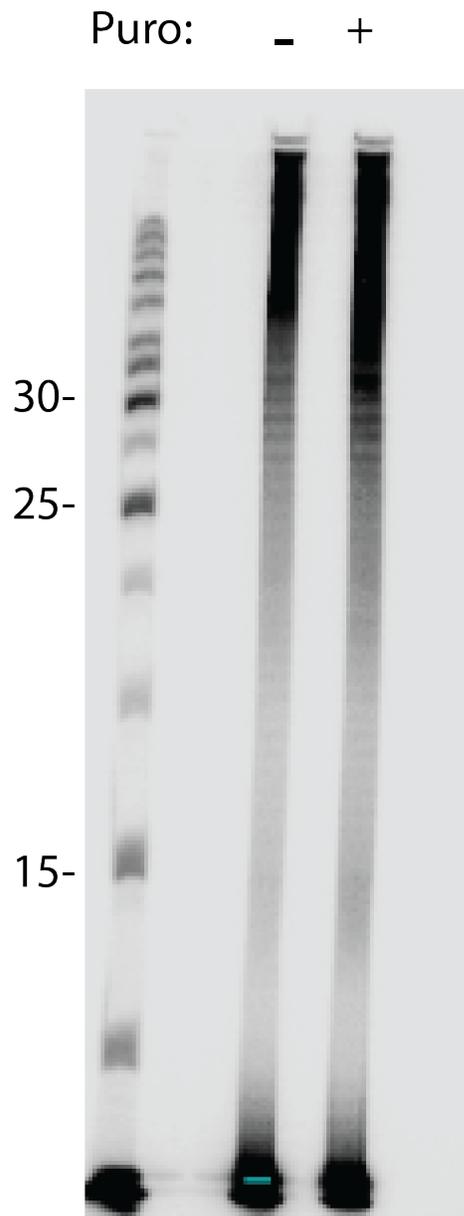


Figure 4.4 RNA footprints from puromycin treated RIPiTs

Autoradiogram of end-labeled RNA extracted from FLAG-Magoh:eIF4AIII RIPiT with and without puromycin treatment (above). Nucleotide ladder (left) illustrates length in of RNA footprints.

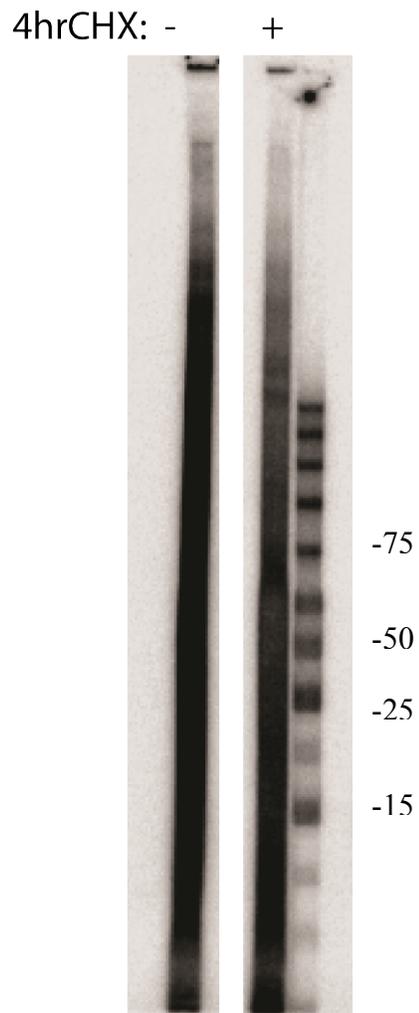


Figure 4.5 RNA from cycloheximide treated RIPiT

Autoradiogram of end-labeled RNA extracted from FLAG-Magoh:eIF4AIII RIPiT with and without cycloheximide treatment. Nucleotide ladder (right) illustrates the size of the RNA footprints.

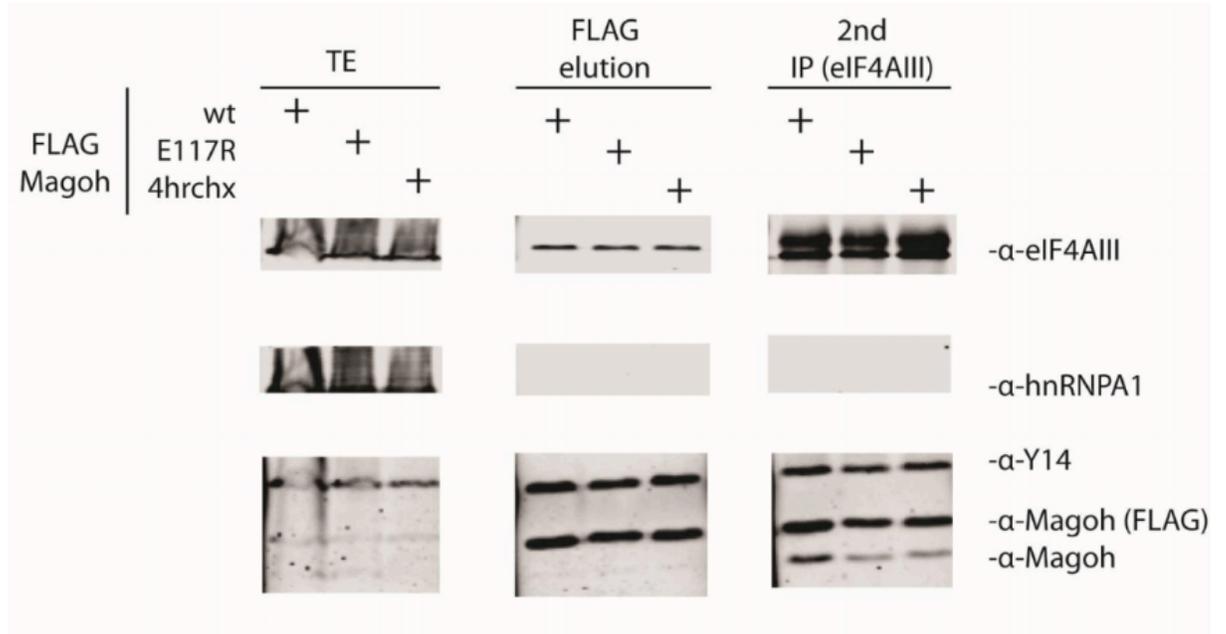


Figure 4.6 FLAG-Magoh:eIF4AIII wild-type, mutant, and cycloheximide-treated RIPiT protein fractions

Western blots depicting proteins (right) present in fractions taken from total extract (TE), elution from first IP (FLAG elution), and elution from RIPiT (2nd IP) (top) from HEK293 cells expressing FLAG-Magoh (wt) or FLAG-MagohE117R (E117R) with and without cycloheximide treatment (4hrchx).

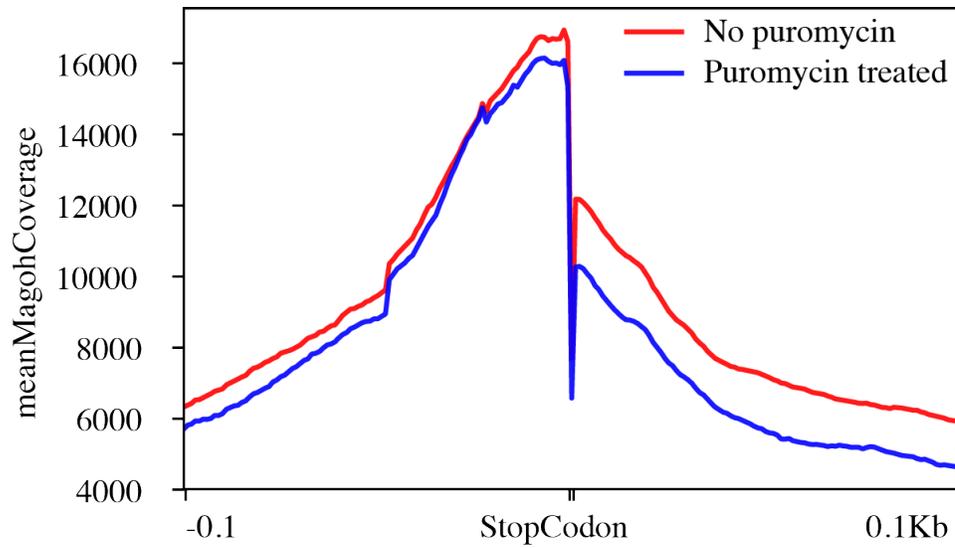


Figure 4.7 EJC core occupancy in 3'UTR is reduced with translation inhibition

Stop codon metagene plots of FLAG-Magoh:eIF4AIII coverage with or without puromycin treatment. Plot generated with DeepTools suite(Ramírez et al., 2014).

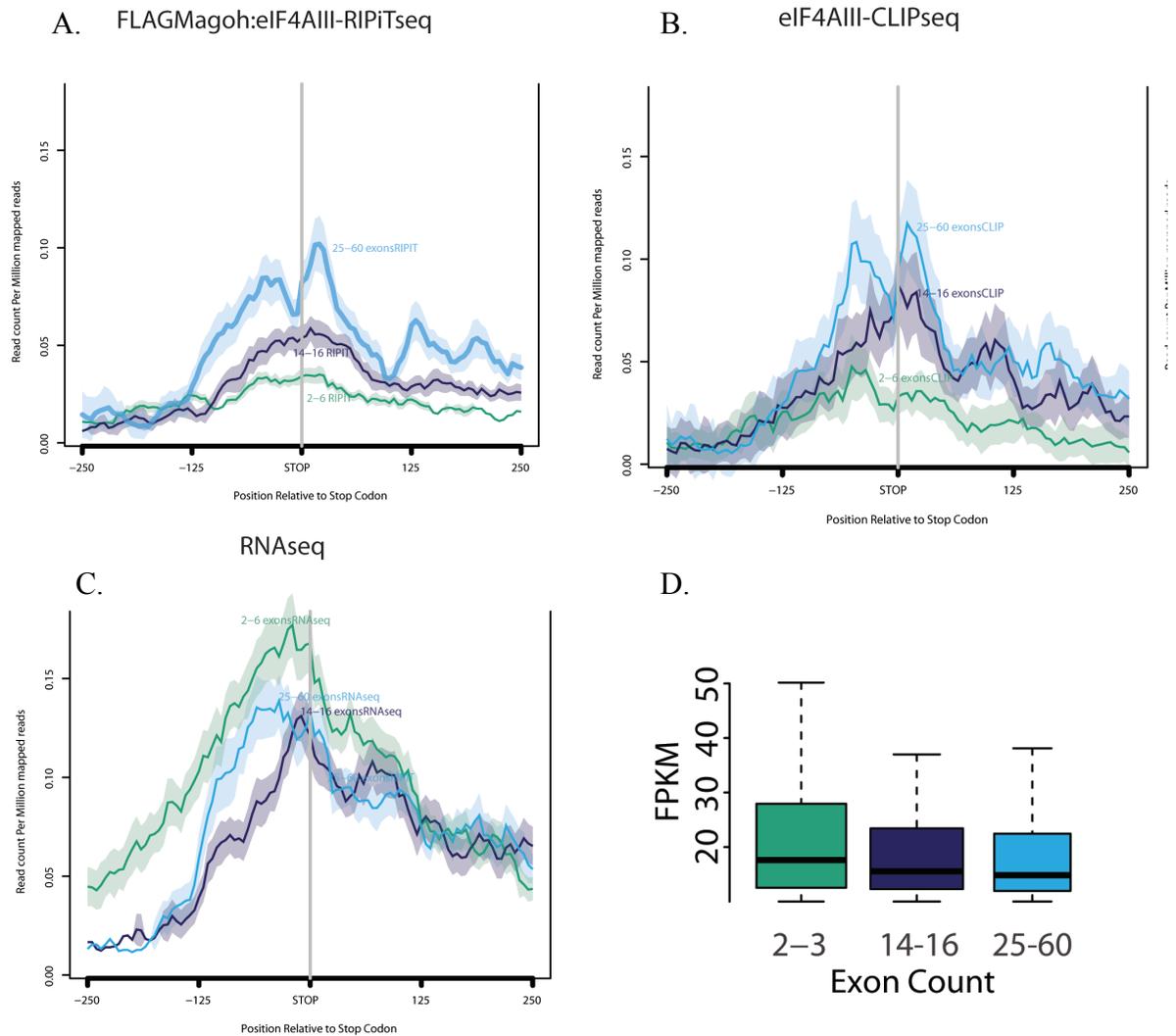


Figure 4.8 Exon number affects EJC occupancy at stop codon in multiple datasets

Stop codon metagene plot of protein coding genes compares FLAG-Magoh:eIF4AIII (A) eIF4AIII-CLIP (B) (Saulière et al., 2012) and RNA-seq (C) read density on mRNAs with 2-3 (n=7153), 14-16 (n=1455), and 25-60 (n=3608) exons. Boxplot (D) shows the FPKM spread of mRNAs in each category. Plots generated with ngs.plot package (Shen et al., 2014).

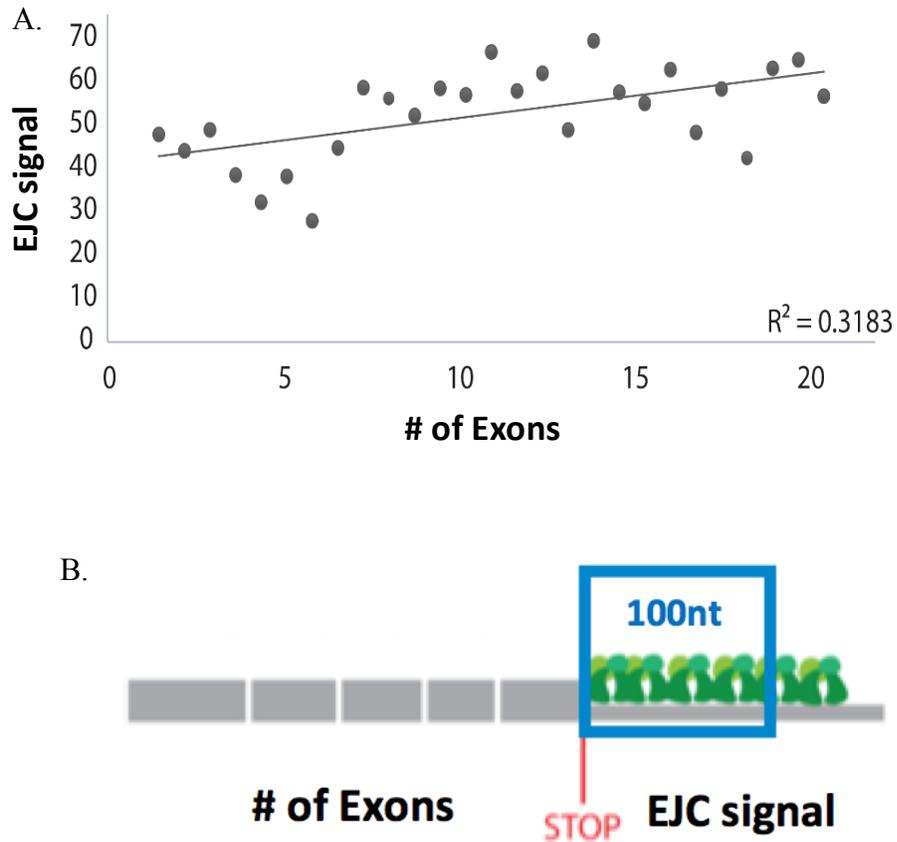


Figure 4.9 Exon number influences the average EJC counts in 3'UTR

- A. Scatter plot of number of exons in mRNAs versus FLAG-Magoh:eIF4AIII RIPiT average counts in 100nt region downstream of stop codon for all genes with FPKM >10.
 B. Schematic explaining the location of EJC footprint in scatter plot in (A.).

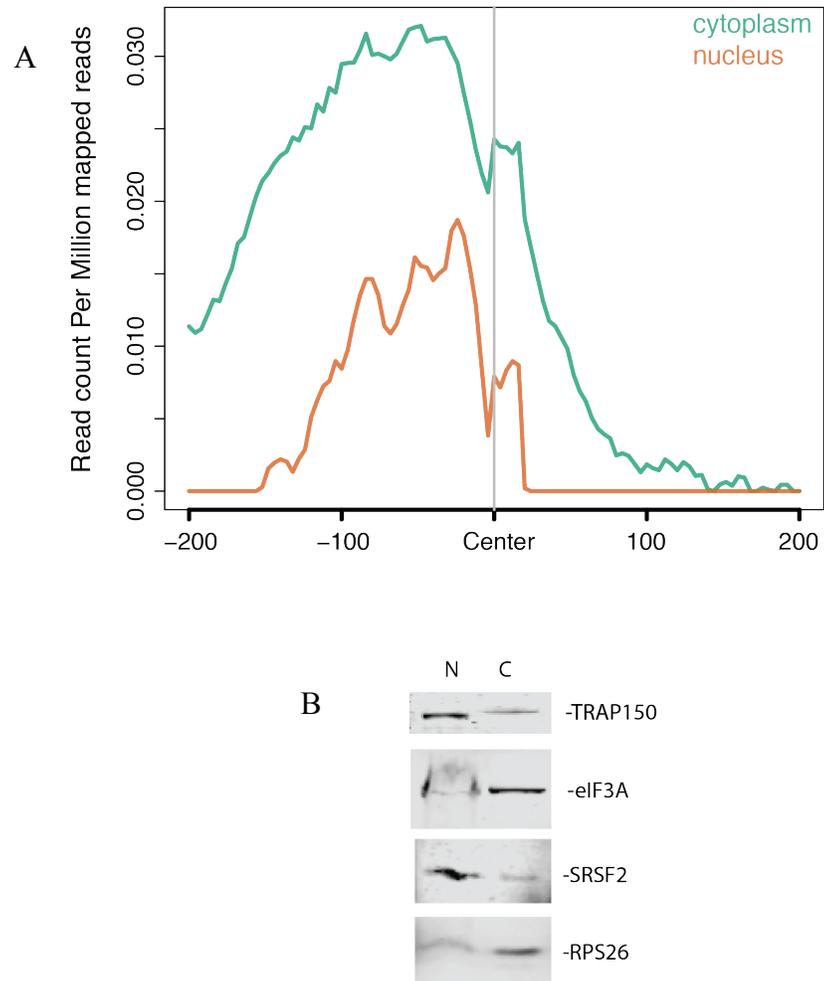


Figure 4.10 Cytoplasmic EJC are more abundant in 3'UTR than nuclear EJC

A. Stop codon metagene plot of FLAG:Magoh:eIF4AIII RIPiT read density from either nuclear or cytoplasmic extracts generated with ngs.plot (Shen et al., 2014).

B. Western blot showing protein present (right) in the nuclear (N) and cytoplasmic f© fractions of HEK293 cells from which the RIPiTs in (A.) were conducted. Nuclear markers: TRAP150, SRSF. Cytoplasmic markers: eIF3A, RPS26.

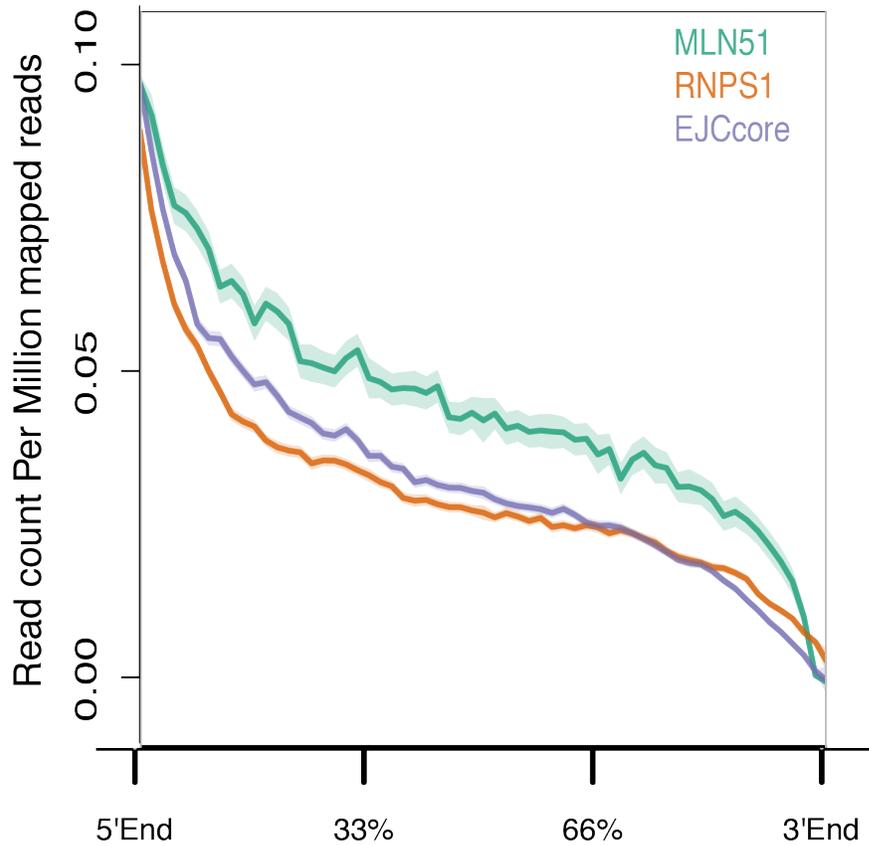


Figure 4.11 CASC3-EJCs are more abundant in the 3'UTR than RNPS1-EJCs

3'UTR metagene plot of mRNAs comparing read density of FLAG-CASC3:eIF4AIII, FLAG-RNPS1:eIF4AIII, and FLAG-Magoh:eIF4AIII (core) RIPiT-seq. Plot generated with ngs.plot (Shen et al., 2014).

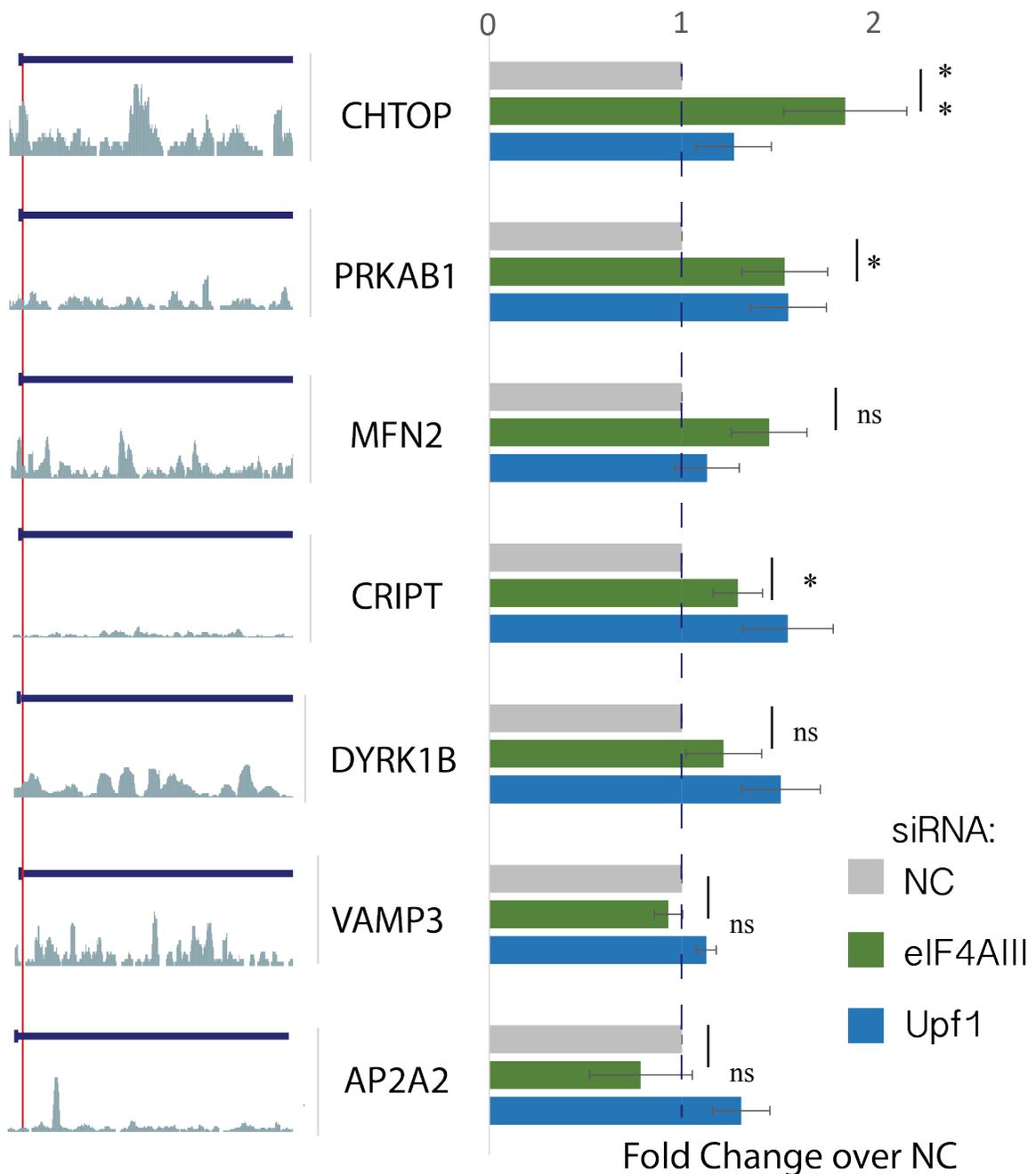


Figure 4.12 EJC knockdown has minimal effect on non-canonical NMD targets

Genes sensitive to Upf1 and Smg6 depletion with EJC signal in 3'UTRs were assayed for abundance after siRNA knockdown of eIF4AIII. Red vertical line (left) is stop codon of genes, and the blue horizontal lines are the 3'UTRs with FLAG-Magoh:eIF4AIII coverage represented in gray below. Fold changes were calculated using the Delta-Delta Cq method. Fold change values are depicted for control(NC), eIF4AIII, and Upf1

Figure 4.12 EJC knockdown has minimal effect on non-canonical NMD targets
(continued)

knockdowns to the right. eIF4AIII kd $\geq 60\%$ (data not shown) and Upf1 kd $\geq 80\%$. Knockdowns and RT-qPCR fold change values and standard deviation calculated from experiments conducted in triplicate with Beta Actin mRNA serving as an internal control. p-values determined with student T-test. n.s = p-value > 0.1 ; * =p-value < 0.1 ; **=pvalue < 0.05 .

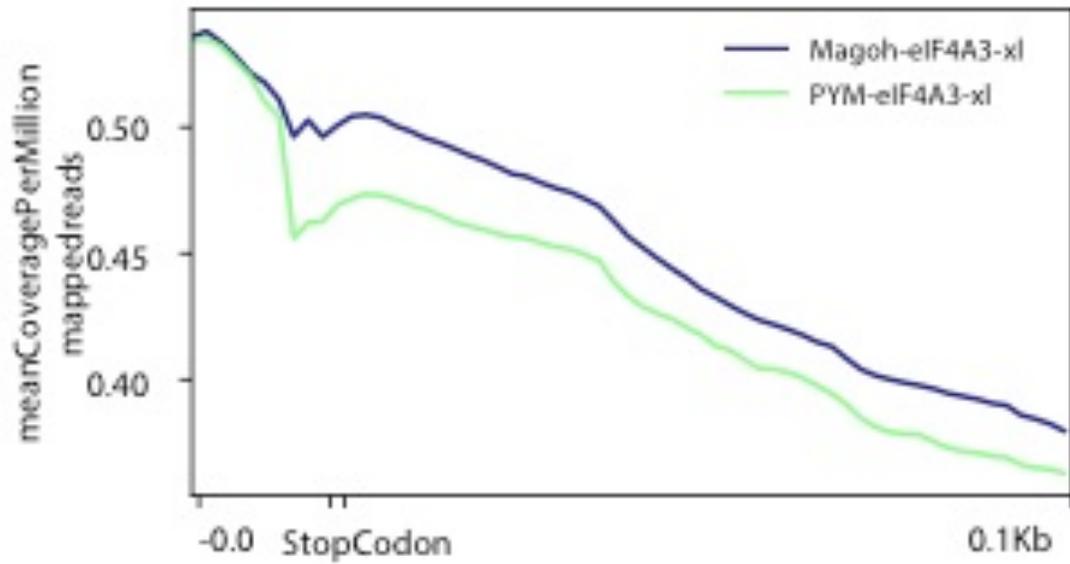


Figure 4.13 PYM footprints exist in the 3'UTR

Metagene plot of 3'UTRs comparing read density of FLAG-Magohwt:eIF4AIII and FLAG-PYM:eIF4AIII. Plot was generated using DeepTools package (Ramírez et al., 2014).

Chapter 5 Conclusions and Future Directions

Forty years ago, two studies reported that intronless mRNA have drastically reduced expression at posttranscriptional level when compared with their intron-containing counterparts (Hamer and Leder, 1979; Khoury et al., 1979). Since then, the influence of splicing on RNA fate has been an ever-developing theme in the realm of post-transcriptional gene regulation. It was reasoned that a splicing-dependent mark must be left on mRNAs to couple splicing to downstream processes, such as nuclear export, translation, and nonsense-mediated mRNA decay (Thermann et al., 1998; Zhang et al., 1998). The molecular components of the splicing-dependent mark were revealed to be a complex of several proteins recruited 24 nt upstream of exon–exon junctions (Le Hir et al., 2000b, 2000a).

At the heart of protein–protein and protein–RNA interactions of the EJC is eIF4AIII (Chan et al., 2004; Ferraiuolo et al., 2004; Li et al., 1999). In the context of the EJC, eIF4AIII is locked onto the RNA via its interaction with an extremely stable heterodimer of Y14 and Magoh (Bono et al., 2006). These three proteins form the stable core of the EJC that, in turn serves as binding platform for a myriad of other EJC peripheral proteins. Indeed, the EJC binds a proteome of at least 50 distinct proteins (Table1) (Singh et al., 2012; Woodward et al., 2017). It is unlikely that the EJC binds so many proteins at one time, especially as some of the peripheral proteins have distinctly

different subcellular localizations than others. Thus, the EJC has been assumed to be dynamic in nature. The complement of peripheral proteins bound to the EJC is even more dynamic than previously appreciated.

While the stable trimeric core remains constant, the composition of the peripheral complement of EJC proteins varies with subcellular localization, translational status, and mRNA half-life. Chapter 2 of this thesis discussed the discovery of two distinct and mutually exclusive EJCs: RNPS1-EJCs and CASC3-EJCs. RNPS1-EJCs interact with SR- and SR-like proteins in large, multimeric RNP structures, while CASC3-EJCs interact with more cytoplasmic mRNPs in monomeric structures. CASC3-EJCs are much more sensitive to translational status of mRNAs than RNPS1-EJCs.

While it was generally not surprising that the composition of the EJC can vary, the dramatic binary nature of the EJC with respect to CASC3 and RNPS1 binding was unexpected. As yet, there are only two unique EJC compositions that have been characterized. Future studies will likely uncover other compositionally and functionally distinct EJCs.

One possibility is that EJCs of certain compositions may trigger NMD via different mechanisms or target distinct mRNA. This has already been observed in the case of CASC3- and RNPS1-EJCs which initiate NMD with different efficiencies (Mabin et al., 2018). Distinct branches of NMD have been observed by others, who have proposed that Upf3b and Upf2 are capable of triggering NMD independent of one another (Chan et al., 2007). It is possible that EJC composition may influence differential recruitment of either Upf2 or Upf3b, but this hypothesis remains to be tested.

Overall, distinct compositions of CASC3 and RNPS1 EJCs do not bind to differential targets, with some notable exceptions. Namely, CASC3-EJCs are enriched on 5' TOP mRNAs, including ribosomal protein mRNAs. 5' TOP mRNA translational status is regulated by mTOR signaling pathway via LARP1-binding to the 5' TOP RNA. Data presented in Chapter 2 suggest a mechanistic link between CASC3-EJCs and LARP1-mediated translational regulation of 5' TOP mRNA. Regulation of 5' TOP mRNAs through interactions between CASC3-EJCs and the mTOR pathway may cause translational repression of 5' TOP mRNAs. As 5' TOP mRNAs encompass many ribosomal protein-encoding transcripts, this would provide rapid translational silencing of pre-translation 5' TOP mRNPs in response to unfavorable conditions for protein synthesis. These mRNPs would remain preserved in an untranslated state until favorable conditions are restored. While this is an attractive model with preliminary results to support it, the mechanism and scope CASC3's role in mTOR signaling and 5' TOP mRNA regulation awaits further study.

Regardless of the composition of the EJC, its role in the regulation of mRNA fate is thought to conclude upon translation, when the EJC is removed from mRNA. PYM interacts stably with the Y14:Magoh heterodimer along the same interface required for eIF4AIII binding and also interacts with the small subunit of the ribosome (Bono et al., 2004). *In vitro*, PYM is able to reduced EJC binding to a RNA reporter in a concentration dependent manner (Gehring et al., 2009b). Thus, the current model in the field is that PYM bridges the EJC to the translation machinery and disassembles EJCs (Diem et al., 2007; Gehring et al., 2009b). Work in Chapter 3 shows that PYM is largely dispensable

for removal of EJC from canonical regions. In contrast, loss of PYM interaction results in an accumulation of EJCs in non-canonical regions, including single exon genes. Whether PYM promotes disassembly of non-canonical EJCs or prevents spontaneous reassembly of EJC subunits remains to be determined. In either case, EJC assembly outside of the splicing reaction is certainly a new frontier to be investigated. The potential of extra-splicing EJC assembly to influence post-transcriptional regulation—especially of intronless transcripts— should be an area of further research.

PYM's interaction with the EJC is well conserved, yet the biological function of PYM in the cell is unclear. While *in vitro* studies suggest that PYM is an EJC disassembly factor, there is much evidence to suggest that PYM's role in EJC disassembly *in vivo* is dispensable. PYM is not needed for proper translation nor EJC removal, so what is PYM doing, especially in the context of the EJC's interaction with RNPs?

Other's have demonstrated that flaviviruses target PYM-Y14:Magoh interaction and subsequently, Y14 binding to flaviviral transcripts increases as part of promotion of viral replication (Li et al., 2019). Interestingly, flaviviral RNAs do not undergo nuclear processing, so EJC assembly on these transcripts presumably occurs in the cytoplasm or a subcellular compartment apart from the nucleus. Possibly, the abrogation of PYM-Y14:Magoh interaction may promote aberrant EJC assembly on viral transcripts,. Consequently, cytoplasmic EJC assembly on viral RNA may promote viral RNA processing. Future work will focus on the role of PYM's interaction with the EJC to

prevent EJC assembly at non-canonical positions on both endogenous and viral RNA in the cytoplasm

In Chapter 4, another class of non-canonical RNA was examined. 3' UTRs have received much attention for their role in dictating RNA stability (Amrani et al., 2004; Hogg and Goff, 2010; Kurosaki and Maquat, 2013; Zünd et al., 2013). The field is still working toward a greater understanding of the relationship between 3'UTR sequence, structure, and interacting RBPs on influencing mRNA stability and fate. A splicing event in the 3'UTR, resulting in deposition of an EJC downstream of a stop codon is one of the most robust triggers of NMD.

Surprisingly, EJC footprints are observed on nearly all 3'UTRs of expressed mRNAs in deep-sequencing experiments (Hauer et al., 2016; Mabin et al., 2018; Saulière et al., 2012; Singh et al., 2012). Nonetheless, these EJCs have been largely ignored, and no studies have forayed into exploring the origin or biological significance of non-canonical EJCs.

This work observed that non-canonical EJC footprints in the 3'UTR depend on active translation. Further, the number of exons, and therefore the number of splicing events upstream of the 3' UTR causes an increase in the density of EJC footprints in the 3'UTR. This phenomenon is not unique to one experimental procedure, as similar relationships are also observed in eIF4AIII CLIP-seq (Saulière et al., 2012). The subcellular location of the mRNA also affects the density of 3' UTR reads. Primarily cytoplasmic EJCs accumulate more in the 3'UTR than the primarily nuclear EJCs.

How do these EJCs make it to the 3'UTR? Perhaps they are moved during translation from upstream canonical and non-canonical positions to the 3' end. Alternatively, upstream EJCs may disassemble during translation, causing an increased local concentration of EJC subunits, which may result in spontaneous reassembly of EJCs in the UTR as discussed in Chapter 3. It is also unclear how mRNAs with EJCs in the 3'UTR evade NMD. Perhaps there is a difference in peripheral protein composition of 3' UTR EJCs that renders them incapable of triggering NMD. Also possible, EJC interaction with the 3'UTR may be too transient to result in NMD.

The work in this thesis has revealed novel complexities of the EJCs composition, mRNA occupancy, and lifecycle. However with the questions answered here, many new avenues of investigation remain open, ensuring there is still much to learn regarding the ever-increasing complexity of the EJC.

Appendix

RIPiT	treatment	Total Reads	Mapped	(%)	Multi-Mapped	(%)	Final Count
FLAGMagoh:eIF4AIII	4hrs CHX	1335229	1088767	81.54	91406	8.4	327987
FLAG-Magoh:eIF4AIII	NT	5830708	5439447	93.29	515132	9.47	2638065
FLAG-Magoh:eIF4AIII	Nuclear fraction	27155282	17940893	66.1	8164132	45.5	
FLAG-Magoh:eIF4AIII	Cytoplasmic fraction	25977484	17163805	66.1	11141009	64.9	
FLAG-Magoh:eIF4A3_I	NT	19721973	18649907	94.56	1517744	8.14	8395123
FLAG-Magoh:eIF4A3_II	NT	12610630	11943296	94.71	652081	5.46	3499793
FLAG-Magoh:eIF4A3_I	2.5hrs puromycin	11190763	10593531	94.66	721067	6.81	3813187
FLAG-Magoh:eIF4A3_II	2.5hrs puromycin	5592871	5042997	90.17	288934	5.73	1245889
FLAG-CASC3:eIF4A3_I	3hrs CHX HCHO-XL	15795694	15084794	95.50	684318	4.54	9772454
FLAG-CASC3:eIF4A3_II	3hrs CHX HCHO-XL	13937900	13413713	96.24	523380	3.90	7962962
FLAG-PYM:eIF4A3_I	HCHO-XL	22250704	21357141	95.98	745563	3.49	1153362
FLAG-PYM:eIF4A3_II	HCHO-XL	11382168	10766820	94.59	642844	5.97	2216562
FLAG-RNPS1:eIF4A3_I	3hrs CHX HCHO-XL	27687914	26580410	96.00	2258708	8.50	10669549
FLAG-RNPS1:eIF4A3_II	3hrs CHX HCHO-XL	2532100	2274851	89.84	95329	4.19	614313
FLAG-Magoh-E117R:CASC3 I	3hrs CHX	7170648	6039349	84.22	2921568	40.74	2582424
FLAG-Magoh-E117R:CASC3 II	3hrs CHX	6781202	5453716	80.42	2617126	38.59	2177486
FLAG-Magoh-E117R:CASC3 I		29911638	25685235	85.87	14207522	47.50	9430163
FLAG-Magoh-E117R:CASC3 II		9587564	7867429	82.06	4114174	42.91	2710571
FLAG-Magoh:CASC3 I	3hrs CHX	7700873	5985774	77.73	2239434	29.08	2879747
FLAG-Magoh:CASC3 II	3hrs CHX	7708251	6485020	84.13	2409786	31.26	3403925
FLAG-Magoh:CASC3 I		7127668	5861343	82.23	2246066	31.51	2938442
FLAG-Magoh:CASC3 II		8907801	7481441	83.99	2815441	31.61	3944601

Table 2 RIPiT library information

“

Bibliography

- Adivarahan, S., Livingston, N., Nicholson, B., Rahman, S., Wu, B., Rissland, O.S., and Zenklusen, D. (2018). Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. *Mol. Cell* *72*, 727-738.e5.
- Alexandrov, A., Colognori, D., Shu, M.-D., and Steitz, J.A. (2012). Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. *Proc. Natl. Acad. Sci.* *109*, 21313–21318.
- Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A *faux* 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* *432*, 112–118.
- Andersen, C.B.F., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L.P., Pedersen, J.S., Séraphin, B., Hir, H.L., and Andersen, G.R. (2006). Structure of the Exon Junction Core Complex with a Trapped DEAD-Box ATPase Bound to RNA. *Science* *313*, 1968–1972.
- Ashton-Beaucage, D., Udell, C.M., Lavoie, H., Baril, C., Lefrançois, M., Chagnon, P., Gendron, P., Caron-Lizotte, O., Bonneil, É., Thibault, P., et al. (2010). The Exon Junction Complex Controls the Splicing of *mapk* and Other Long Intron-Containing Transcripts in *Drosophila*. *Cell* *143*, 251–262.
- Baguet, A., Degot, S., Cougot, N., Bertrand, E., Chenard, M.-P., Wendling, C., Kessler, P., Hir, H.L., Rio, M.-C., and Tomasetto, C. (2007). The exon-junction-complex-component metastatic lymph node 51 functions in stress-granule assembly. *J. Cell Sci.* *120*, 2774–2784.
- 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.
- 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.
- Bartkowiak, B., and Greenleaf, A.L. (2015). Expression, Purification, and Identification of Associated Proteins of the Full-length hCDK12/CyclinK Complex. *J. Biol. Chem.* *290*, 1786–1795.
- Belgrader, P., Cheng, J., Zhou, X., Stephenson, L.S., and Maquat, L.E. (1994). Mammalian nonsense codons can be cis effectors of nuclear mRNA half-life. *Mol. Cell. Biol.* *14*, 8219–8228.

- Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Lührmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. *Nature* *452*, 846–850.
- Blazquez, L., Emmett, W., Faraway, R., Pineda, J.M.B., Bajew, S., Gohr, A., Haberman, N., Sibley, C.R., Bradley, R.K., Irimia, M., et al. (2018). Exon Junction Complex Shapes the Transcriptome by Repressing Recursive Splicing. *Mol. Cell* *72*, 496-509.e9.
- Boehm, V., and Gehring, N.H. (2016). Exon Junction Complexes: Supervising the Gene Expression Assembly Line. *Trends Genet.* *32*, 724–735.
- Boehm, V., Britto-Borges, T., Steckelberg, A.-L., Singh, K.K., Gerbracht, J.V., Gueney, E., Blazquez, L., Altmüller, J., Dieterich, C., and Gehring, N.H. (2018). Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity. *Mol. Cell* *72*, 482-495.e7.
- Bono, F., Ebert, J., Unterholzner, L., Güttler, T., Izaurralde, E., and Conti, E. (2004). Molecular insights into the interaction of PYM with the Mago–Y14 core of the exon junction complex. *EMBO Rep.* *5*, 304–310.
- 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.
- Bono, F., Cook, A.G., Grünwald, M., Ebert, J., and Conti, E. (2010). Nuclear Import Mechanism of the EJC Component Mago-Y14 Revealed by Structural Studies of Importin 13. *Mol. Cell* *37*, 211–222.
- Boyne, J.R., Jackson, B.R., Taylor, A., Macnab, S.A., and Whitehouse, A. (2010). Kaposi’s sarcoma-associated herpesvirus ORF57 protein interacts with PYM to enhance translation of viral intronless mRNAs. *EMBO J.* *29*, 1851–1864.
- Braddock, M., Muckenthaler, M., White, M.R.H., Thorburn, A.M., Sommerville, J., Kingsman, A.J., and Kingsman, S.M. (1994). Intron-less RNA injected into the nucleus of *Xenopus* oocytes accesses a regulated translation control pathway. *Nucleic Acids Res.* *22*, 5255–5264.
- Buchwald, G., Schüssler, S., Basquin, C., Le Hir, H., and Conti, E. (2013). Crystal structure of the human eIF4AIII-CWC22 complex shows how a DEAD-box protein is inhibited by a MIF4G domain. *Proc. Natl. Acad. Sci. U. S. A.* *110*, E4611-4618.
- Callis, J., Fromm, M., and Walbot, V. (1987). Introns increase gene expression in cultured maize cells. *Genes Dev.* *1*, 1183–1200.
- Carter, M.S., Li, S., and Wilkinson, M.F. (1996). A splicing-dependent regulatory mechanism that detects translation signals. *EMBO J.* *15*, 5965–5975.

- Chan, C.C., Dostie, J., DIEM, M.D., FENG, W., MANN, M., RAPPILBER, J., and DREYFUSS, G. (2004). eIF4A3 is a novel component of the exon junction complex. *RNA* *10*, 200–209.
- Chan, W.-K., Huang, L., Gudikote, J.P., Chang, Y.-F., Imam, J.S., MacLean, J.A., and Wilkinson, M.F. (2007). An alternative branch of the nonsense-mediated decay pathway. *EMBO J.* *26*, 1820–1830.
- Chazal, P.-E., Dagueuet, E., Wendling, C., Ulryck, N., Tomasetto, C., Sargueil, B., and Hir, H.L. (2013). EJC core component MLN51 interacts with eIF3 and activates translation. *Proc. Natl. Acad. Sci.* *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.
- Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M., and Reed, R. (2006). Functional coupling of RNAP II transcription to spliceosome assembly. *Genes Dev.* *20*, 1100–1109.
- 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.
- Degot, S., Hir, H.L., Alpy, F., Kedinger, V., Stoll, I., Wendling, C., Seraphin, B., Rio, M.-C., and Tomasetto, C. (2004). Association of the Breast Cancer Protein MLN51 with the Exon Junction Complex via Its Speckle Localizer and RNA Binding Module. *J. Biol. Chem.* *279*, 33702–33715.
- Diem, M.D., Chan, C.C., Younis, I., and Dreyfuss, G. (2007). PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat. Struct. Mol. Biol.* *14*, 1173–1179.
- Domínguez-Sánchez, M.S., Barroso, S., Gómez-González, B., Luna, R., and Aguilera, A. (2011). Genome Instability and Transcription Elongation Impairment in Human Cells Depleted of THO/TREX. *PLoS Genet.* *7*.
- 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*.

- van Eeden, F.J.M., Palacios, I.M., Petronczki, M., Weston, M.J.D., 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*, 511–524.
- Fairman, M.E., Maroney, P.A., Wang, W., Bowers, H.A., Gollnick, P., Nilsen, T.W., and Jankowsky, E. (2004). Protein displacement by DExH/D “RNA helicases” without duplex unwinding. *Science* *304*, 730–734.
- Ferraiuolo, M.A., Lee, C.-S., Ler, L.W., Hsu, J.L., Costa-Mattioli, M., Luo, M.-J., Reed, R., and Sonenberg, N. (2004). A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 4118–4123.
- Fonseca, B.D., Lahr, R.M., Damgaard, C.K., Alain, T., and Berman, A.J. (2018). LARP1 on TOP of ribosome production. *Wiley Interdiscip. Rev. RNA* e1480.
- 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. In *Argonaute Proteins: Methods and Protocols*, K. Okamura, and K. Nakanishi, eds. (New York, NY: Springer New York), pp. 1–28.
- Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P., and Izaurralde, E. (2003). Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways. *EMBO J.* *22*, 3960–3970.
- 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., Lamprinaki, 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*.
- Gehring, N.H., Lamprinaki, 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.
- Ghosh, S., Obrdlik, A., Marchand, V., and Ephrussi, A. (2014). The EJC binding and dissociating activity of PYM is regulated in *Drosophila*. *PLoS Genet.* *10*, e1004455.

- Glynn, E.F., Megee, P.C., Yu, H.-G., Mistrot, C., Unal, E., Koshland, D.E., DeRisi, J.L., and Gerton, J.L. (2004). Genome-Wide Mapping of the Cohesin Complex in the Yeast *Saccharomyces cerevisiae*. *PLoS Biol.* *2*.
- Hamer, D.H., and Leder, P. (1979). Splicing and the formation of stable RNA. *Cell* *18*, 1299–1302.
- 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., et al. (2009). Genome-wide identification of alternative splice forms down-regulated by nonsense-mediated mRNA decay in *Drosophila*. *PLoS Genet.* *5*, e1000525.
- 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-Horowicz, 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.
- Hicks, M.J., Yang, C.-R., Kotlajich, M.V., and Hertel, K.J. (2006). Linking splicing to Pol II transcription stabilizes pre-mRNAs and influences splicing patterns. *PLoS Biol.* *4*, e147.
- Hir, H.L., 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.
- Hirose, T., Ideue, T., Nagai, M., Hagiwara, M., Shu, M.-D., and Steitz, J.A. (2006). A Spliceosomal Intron Binding Protein, IBP160, Links Position-Dependent Assembly of Intron-Encoded Box C/D snoRNP to Pre-mRNA Splicing. *Mol. Cell* *23*, 673–684.
- Hogg, J.R., and Goff, S.P. (2010). Upf1 Senses 3'UTR Length to Potentiate mRNA Decay. *Cell* *143*, 379–389.
- Hopkins, T.G., Mura, M., Al-Ashtal, H.A., Lahr, R.M., Abd-Latip, N., Sweeney, K., Lu, H., Weir, J., El-Bahrawy, M., Steel, J.H., et al. (2016). The RNA-binding protein LARP1 is a post-transcriptional regulator of survival and tumorigenesis in ovarian cancer. *Nucleic Acids Res.* *44*, 1227–1246.

- Huang, Y., and Steitz, J.A. (2001). Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol. Cell* 7, 899–905.
- Huang, D.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.
- Huang, Y., Gattoni, R., Stévenin, J., and Steitz, J.A. (2003). SR Splicing Factors Serve as Adapter Proteins for TAP-Dependent mRNA Export. *Mol. Cell* 11, 837–843.
- Hug, N., Longman, D., and Cáceres, J.F. (2016). Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.* 44, 1483–1495.
- Ideue, T., Sasaki, Y.T.F., 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.
- Isken, O., Kim, Y.K., Hosoda, N., Mayeur, G.L., Hershey, J.W.B., and Maquat, L.E. (2008). Upf1 Phosphorylation Triggers Translational Repression during Nonsense-Mediated mRNA Decay. *Cell* 133, 314–327.
- 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.
- Jankowsky, E., and Bowers, H. (2006). Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Res.* 34, 4181–4188.
- Karousis, E.D., Nasif, S., and Mühlemann, O. (2016). Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdiscip. Rev. RNA n/a-n/a*.
- Kashima, I., Jonas, S., Jayachandran, U., Buchwald, G., Conti, E., Lupas, A.N., and Izaurralde, E. (2010). SMG6 interacts with the exon junction complex via two conserved EJC-binding motifs (EBMs) required for nonsense-mediated mRNA decay. *Genes Dev.* 24, 2440–2450.
- Kedersha, N., and Anderson, P. (2007). Mammalian Stress Granules and Processing Bodies. B.-M. in *Enzymology*, ed. (Academic Press), pp. 61–81.
- Khoury, G., Gruss, P., Dhar, R., and Lai, C.J. (1979). Processing and expression of early SV40 mRNA: a role for RNA conformation in splicing. *Cell* 18, 85–92.
- Kiss, D.L., Baez, W., Huebner, K., Bundschuh, R., and Schoenberg, D.R. (2017a). Impact of FHIT loss on the translation of cancer-associated mRNAs. *Mol. Cancer* 16, 179.
- Kiss, D.L., Baez, W., Huebner, K., Bundschuh, R., and Schoenberg, D.R. (2017b). Impact of FHIT loss on the translation of cancer-associated mRNAs. *Mol. Cancer* 16.

- Kurosaki, T., and Maquat, L.E. (2013). Rules that govern UPF1 binding to mRNA 3' UTRs. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 3357–3362.
- Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-Regulated Binding of RNA Polymerase II to Fibrous Polymers of Low-Complexity Domains. *Cell* *155*, 1049–1060.
- Lau, C.-K., Diem, M.D., Dreyfuss, G., and Van Duyne, G.D. (2003). Structure of the Y14-Magoh Core of the Exon Junction Complex. *Curr. Biol.* *13*, 933–941.
- 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., Nott, A., and Moore, M.J. (2003). How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci.* *28*, 215–220.
- 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.
- Lejeune, F., Ishigaki, Y., Li, X., and Maquat, L.E. (2002). The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J.* *21*, 3536–3545.
- Lengronne, A., Katou, Y., Yokobayashi, S., Mori, S., Kelly, G., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* *430*, 573–578.
- Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* *122*, 365–378.
- Li, M., Johnson, J.R., Truong, B., Kim, G., Weinbren, N., Dittmar, M., Shah, P.S., Dollen, J.V., Newton, B.W., Jang, G.M., et al. (2019). Identification of antiviral roles for the exon–junction complex and nonsense-mediated decay in flaviviral infection. *Nat. Microbiol.* *1*.
- Li, Q., Imataka, H., Morino, S., Rogers, G.W., Richter-Cook, N.J., Merrick, W.C., and Sonenberg, N. (1999). Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. *Mol. Cell. Biol.* *19*, 7336–7346.

- Longman, D., Plasterk, R.H.A., Johnstone, I.L., and Cáceres, J.F. (2007). Mechanistic insights and identification of two novel factors in the *C. elegans* NMD pathway. *Genes Dev.* *21*, 1075–1085.
- 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.
- Lu, S., and Cullen, B.R. (2003). Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. *RNA* *9*, 618–630.
- Luo, M., and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 14937–14942.
- Lykke-Andersen, J. (2001). Communication of the Position of Exon-Exon Junctions to the mRNA Surveillance Machinery by the Protein RNPS1. *Science* *293*, 1836–1839.
- Lykke-Andersen, S., and Jensen, T.H. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* *16*, 665–677.
- Ma, Q., Tatsuno, T., Nakamura, Y., and Ishigaki, Y. (2019). The stability of Magoh and Y14 depends on their heterodimer formation and nuclear localization. *Biochem. Biophys. Res. Commun.* *511*, 631–636.
- 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.
- Mabin, J.W., Woodward, L.A., Patton, R.D., Yi, Z., Jia, M., Wysocki, V.H., Bundschuh, R., and Singh, G. (2018a). The Exon Junction Complex Undergoes a Compositional Switch that Alters mRNP Structure and Nonsense-Mediated mRNA Decay Activity. *Cell Rep.* *25*, 2431-2446.e7.
- Macchi, P., Kroening, S., Palacios, I.M., Baldassa, S., Grunewald, B., Ambrosino, C., Goetze, B., Lupas, A., Johnston, D.S., 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., Sachidanandam, 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., Pilaz, L.-J., McMahon, J.J., Golzio, C., Wu, D., Shi, L., Katsanis, N., and Silver, D.L. (2015). Rbm8a haploinsufficiency disrupts embryonic cortical development resulting in microcephaly. *J. Neurosci. Off. J. Soc. Neurosci.* *35*, 7003–7018.

- 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 N. Y. N* *23*, 23–31.
- Maquat, L.E., Tarn, W.-Y., and Isken, O. (2010). The Pioneer Round of Translation: Features and Functions. *Cell* *142*, 368–374.
- Matsumoto, K., Wassarman, K.M., and Wolffe, A.P. (1998). Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. *EMBO J.* *17*, 2107–2121.
- 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. Off. J. Int. Soc. Dev. Neurosci.* *55*, 117–123.
- 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.
- 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* *72*, 715-726.e3.
- Meyuhas, O., and Kahan, T. (2015). The race to decipher the top secrets of TOP mRNAs. *Biochim. Biophys. Acta* *1849*, 801–811.
- 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.
- Miller, E.E., Kobayashi, G.S., Musso, C.M., Allen, M., Ishiy, F.A.A., de Caires, L.C., Goulart, E., Griesi-Oliveira, K., Zechi-Ceide, R.M., Richieri-Costa, A., et al. (2017). EIF4A3 deficient human iPSCs and mouse models demonstrate neural crest defects that underlie Richieri-Costa-Pereira syndrome. *Hum. Mol. Genet.* *26*, 2177–2191.
- 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 N. Y. N* *14*, 2657–2670.

- Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* *136*, 688–700.
- Müller-McNicoll, M., Botti, V., de Jesus Domingues, A.M., Brandl, H., Schwich, O.D., Steiner, M.C., Curk, T., Poser, I., Zarnack, K., and Neugebauer, K.M. (2016). SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev.* *30*, 553–566.
- 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.
- 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.
- Paix, A., Folkmann, A., Goldman, D.H., Kulaga, H., Grzelak, M.J., Rasoloson, D., Paidemarry, S., Green, R., Reed, R.R., and Seydoux, G. (2017). Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proc. Natl. Acad. Sci.* *114*, E10745–E10754.
- Palmiter, R.D., Sandgren, E.P., Avarbock, M.R., Allen, D.D., and Brinster, R.L. (1991). Heterologous introns can enhance expression of transgenes in mice. *Proc. Natl. Acad. Sci. U. S. A.* *88*, 478–482.
- 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.
- Peterson, T.R., and Sabatini, D.M. (2005). eIF3: a connectTOR of S6K1 to the translation preinitiation complex. *Mol. Cell* *20*, 655–657.
- Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* *42*, W187–W191.
- 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.
- Rodriguez, J.M., Maietta, P., Ezkurdia, I., Pietrelli, A., Wesselink, J.-J., Lopez, G., Valencia, A., and Tress, M.L. (2013). APPRIS: annotation of principal and alternative splice isoforms. *Nucleic Acids Res.* *41*, D110-117.

- 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.
- Saldi, T., Cortazar, M.A., Sheridan, R.M., and Bentley, D.L. (2016). Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J. Mol. Biol.* *428*, 2623–2635.
- Sanford, J.R., Gray, N.K., Beckmann, K., and Cáceres, J.F. (2004). A novel role for shuttling SR proteins in mRNA translation. *Genes Dev.* *18*, 755–768.
- 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.
- Saulière, J., Murigneux, V., Wang, Z., Marquet, E., Barbosa, I., Le Tonquèze, O., Audic, Y., Paillard, L., Roest Crollius, 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.
- Schmidt, S.A., Foley, P.L., Jeong, D.-H., Rymarquis, L.A., Doyle, F., Tenenbaum, S.A., Belasco, J.G., and Green, P.J. (2015). Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of endogenous NMD targets in human cells. *Nucleic Acids Res.* *43*, 309–323.
- 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.
- Schwer, B., and Gross, C.H. (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J.* *17*, 2086–2094.
- Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* *15*, 284.
- Shepard, P.J., and Hertel, K.J. (2009). The SR protein family. *Genome Biol.* *10*, 242.
- Shibuya, T., Tange, T.Ø., Stroupe, M.E., and Moore, M.J. (2006). Mutational analysis of human eIF4AIII identifies regions necessary for exon junction complex formation and nonsense-mediated mRNA decay. *RNA* *12*, 360–374.
- 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., et al. (2010). The exon junction

- complex component Magoh controls brain size by regulating neural stem cell division. *Nat. Neurosci.* *13*, 551–558.
- 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.
- Sokoloski, J.E., Kozlov, A.G., Galletto, R., and Lohman, T.M. (2016). Chemo-mechanical pushing of proteins along single-stranded DNA. *Proc. Natl. Acad. Sci. U. S. A.* *113*, 6194–6199.
- 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.
- Tani, H., Imamachi, N., Salam, K.A., Mizutani, R., Ijiri, K., Irie, T., Yada, T., Suzuki, Y., and Akimitsu, N. (2012). Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. *RNA Biol.* *9*, 1370–1379.
- Tarpey, P.S., Raymond, F.L., Nguyen, L.S., Rodriguez, J., Hackett, A., Vandeleur, L., Smith, R., Shoubridge, C., Ekins, S., Stevens, C., et al. (2007). Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. *Nat. Genet.* *39*, 1127–1133.
- Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeyer, C., Hentze, M.W., and Kulozik, A.E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J.* *17*, 3484–3494.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* *25*, 1105–1111.
- Urlaub, G., Mitchell, P.J., Ciudad, C.J., and Chasin, L.A. (1989). Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. *Mol. Cell. Biol.* *9*, 2868–2880.
- Walter, P., Ibrahimi, I., and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to in-vitro-assembled polysomes synthesizing secretory protein. *J. Cell Biol.* *91*, 545–550.
- Wang, Z., Murigneux, V., and Le Hir, H. (2014). Transcriptome-wide modulation of splicing by the exon junction complex. *Genome Biol.* *15*.

- Wen, J., and Brogna, S. (2010). Splicing-dependent NMD does not require the EJC in *Schizosaccharomyces pombe*. *EMBO J.* *29*, 1537–1551.
- Woodward, L., Gangras, P., and Singh, G. (2019). Identification of Footprints of RNA:Protein Complexes via RNA Immunoprecipitation in Tandem Followed by Sequencing (RIPiT-Seq). *J. Vis. Exp. JoVE*.
- 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*.
- Yamashita, A. (2013). Role of SMG-1-mediated Upf1 phosphorylation in mammalian nonsense-mediated mRNA decay. *Genes Cells Devoted Mol. Cell. Mech.* *18*, 161–175.
- Zhang, Y., and Sachs, M.S. (2015). Control of mRNA Stability in Fungi by NMD, EJC and CBC Factors Through 3'UTR Introns. *Genetics* *200*, 1133–1148.
- 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. U. S. A.* *104*, 11574–11579.
- Zhang, J., Sun, X., Qian, Y., LaDuca, J.P., and Maquat, L.E. (1998). At Least One Intron Is Required for the Nonsense-Mediated Decay of Triosephosphate Isomerase mRNA: a Possible Link between Nuclear Splicing and Cytoplasmic Translation. *Mol. Cell. Biol.* *18*, 5272–5283.
- 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., Ding, J.-H., Adams, J.A., Ghosh, G., and Fu, X.-D. (2009). Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones. *Genes Dev.* *23*, 482–495.
- Zhou, Z., and Fu, X.-D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* *122*, 191–207.
- Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* *419*, 182–185.
- Zünd, D., Gruber, A.R., Zavolan, M., and Mühlemann, O. (2013). Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3' UTRs. *Nat. Struct. Mol. Biol.* *20*, 936–943.

