Proteomic and transcriptomic characterization of RNPS1:

A sequence-independent regulator of mRNA fate

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Lena Pia Schlautmann

Veröffentlichung: Köln, 2023

Erster Gutachter:	Prof. Dr. Niels H. Gehring
Zweiter Gutachter:	Prof. Dr. Kay Hofmann
Tag der mündlichen Prüfung:	11.01.2023

1 CONTENTS

2 List of Abbreviations

DNA	Desoxyribonucleic acid
RNA	Ribonucleic acid
mRNA	messenger RNA
bp	base pair
nt	nucleotide(s)
3’ UTR	3’ untranslated region
RBP	RNA-binding protein
EJC	Exon junction complex
EIF4A3	Eukaryotic translation initiation factor 4A3
RBM8A	RNA binding motif protein 8A
MAGOH	Mago nashi homolog
ASAP	Apoptosis and splicing-associated protein (complex)
PSAP	Named after ASAP, P for Pinin
RNPS1	RNA-binding protein with serine-rich domain 1
SAP18	Sin3-associated protein of 18 kDa
ACIN1	Acinus, apoptotic chromatin inducer in the nucleus
PNN	Pinin
UPF1, UPF2, UPF3A/B	Up-frameshift1, 2, 3A/B
SMG6, SMG7	Suppressor with morphogenetic effect on genitalia
A3SS	Alternative 3’ splice site
A5SS	Alternative 5’ splice site
IR	Intron retention
ES	Exon skipping

EI	Exon inclusion
MXE	Mutually exclusive exons
MAPK	MAP-kinase
PIWI	P-element induced wimpy testis
RER1	Retention in endoplasmic reticulum sorting receptor 1
FDPS	Farnesyl diphosphate synthase
TAF15	TATA-box binding protein associated factor 15
INTS2, INTS3	Integrator complex subunit 2, 3
RFX5	Regulatory factor X5
FLAG	Protein tag
MS2	Protein tag, MS2 bacteriophage coat protein
V5	Protein tag
GST	Glutathione S-transferase
emGFP	Emerald green fluorescent protein
LC MS/MS	Liquid chromatographie mass spectrometry
IP	Immunoprecipitation
dIF	Delta Isoform Fraction
dPSI	Delta percent spliced in
log2 FC	Log2 foldchange
Padj.	Adjusted P-value

3 ABSTRACT

Due to the genome complexity, the processes of gene expression require tight regulation especially in mammalian cells. Many regulatory proteins have the dedicated purpose to ensure the production of correct mature mRNAs or to otherwise degrade faulty transcripts. One key factor in these processes is the exon junction complex (EJC), which is deposited on the mRNA during the splicing process in a sequence-independent manner. This enables the EJC to orchestrate a variety of co- and posttranscriptional processes, including alternative splicing, mRNA export and nonsense-mediated mRNA decay (NMD). To fulfill this wide range of regulatory functions the EJC serves as a binding platform for various regulatory proteins and complexes, including the ASAP- and PSAP-component RNPS1. Previously, RNPS1 was reported to act on alternative splicing regulation as well as NMD. Here, transcriptome-wide analyses were combined with interactome studies to further enlighten the role RNPS1 plays in these processes. Differential gene expression and differential transcript usage analyses revealed that RNPS1 mildly influences NMD of specific targets rather than being a globally essential NMD factor. However, alternative splicing analyses confirmed that RNPS1 is an important regulator of various types of alternative splicing. Mechanistically, intron retention reporters revealed that RNPS1 positioned at a downstream splice junction can activate splicing of an upstream intron and thereby prevent intron retention. RNPS1 normally requires the assembly of the ASAP or PSAP complex to be recruited to the EJC. However, individual depletion of the ASAP/PSAP components ACIN1 or PNN affected splicing only mildly, suggesting that the complexes might be able to function redundantly. Moreover, the alternative splicing analyses and further knockdown and rescue experiments indicated that, contrary to previous hypotheses, not only the RNPS1 RRM, but also its C-terminus and S domain are involved in splicing regulation. Investigation of the interactome of different RNPS1 deletion mutants revealed that RNPS1 can interact with a large variety of splicing-related factors and that these interactions were reduced or completely abolished, when one or more of its domains are deleted.

Thus, a picture emerges in which RNPS1 promotes correct splicing by gathering varying splicing competent or enhancing complexes on the mRNA by making use of the distinct binding capacities of its domains.

4 ZUSAMMENFASSUNG

Aufgrund der hohen Komplexität des Genoms unterliegen die Prozesse der Genexpression strikter Regulation, insbesondere in Säugetierzellen. Viele regulatorische Proteine haben die Funktion, die Produktion korrekter, reifer mRNAs sicherzustellen oder anderenfalls fehlerhafte Transkripte abzubauen. Ein Schlüsselfaktor während dieser Prozesse ist der Exon Junction Complex (EJC), der während des Spleißvorgangs in einer sequenz-unabhängigen Weise auf der mRNA platziert wird. Dies ermöglicht es dem EJC, eine Vielzahl von co- und posttranskriptionalen Prozessen zu koordinieren, darunter alternatives Spleißen, mRNA-Export und mRNA-Abbau über den Nonsense-mediated mRNA decay (NMD). Um dieses große Spektrum von regulatorischen Funktionen erfüllen zu können, fungiert der EJC als Binde-Plattform für diverse regulatorische Proteine und Komplexe, inklusive die Komponente des ASAP und PSAP Komplex RNPS1. Zuvor wurde dokumentiert, dass RNPS1 sowohl alternatives Spleißen regulieren kann als auch eine Funktion im NMD-Prozess hat. Die Analysen der differentiellen Genexpression (DGE) sowie der differentiellen Transkript-Verwendung (DTU) zeigten, dass RNPS1 eher einen milden Einfluss auf den Abbau spezifischer Transkripte durch den NMD hat, als dass es ein globaler oder essenzieller NMD-Faktor ist. Allerdings haben die Analysen des alternativen Spleißens bestätigt, dass RNPS1 ein wichtiger Regulator verschiedener Typen von alternativem Spleißen ist. Auf mechanistischer Ebene wurde mit Reportern für Intron Retention gezeigt, dass die Positionierung von RNPS1 an einer nachfolgenden Exon-Exon Verbindungsstelle das Spleißen eines vorangehenden Introns aktivieren kann und somit die Retention dieses Introns verhindert. Normalerweise benötigt RNPS1 die Interaktion mit den anderen Komponenten des ASAP- oder PSAP-Komplexes. Die einzelne Depletion der ASAP/PSAP-Komponenten ACIN1 oder PNN hatte allerdings geringen Einfluss auf alternatives Spleißen, was darauf hindeutet, dass die zwei Komplexe möglicherweise redundant fungieren können. Darüber hinaus deuten die Analysen zum alternativen Spleißen darauf hin, dass, entgegen den vorherigen Hypothesen, nicht nur die RNPS1 RRM Domäne, sondern auch der C-Terminus und die S-Domäne an der Regulation von Spleißvorgängen beteiligt sind. Interaktom-Analysen zeigten, dass RNPS1 mit einer großen Bandbreite weiterer Spleißregulatoren interagiert. In verschiedenen RNPS1 Deletions-mutanten interagierten diese Spleißfaktoren schwächer oder gar nicht mehr mit RNPS1.

Letztendlich entsteht dadurch ein Bild, nachdem RNPS1 korrektes Spleißen unterstützt, indem es unter Zuhilfenahme der verschiedenen Bindungskapazitäten seiner Domänen unterschiedliche Komplexe zur Verstärkung oder zur Aktivierung von Spleißen an die mRNA bindet.

5 INTRODUCTION

In eukaryotic cells, most processes require the presence of specified proteins. For the synthesis of those proteins, the genetic information that is encoded in the DNA in the nucleus needs to be transcribed into messenger RNA (mRNA) by the RNA polymerase II (PolII) (Figure 1). The resulting pre-mRNA undergoes multiple processing steps to produce a mature mRNA. These processing steps include the addition of a m7G cap to the 5' end of the mRNA, the removal of non-coding sequences (introns) and the polyadenylation of the 3' end (reviewed in (Moore and Proudfoot 2009)). After these processing steps, the mature mRNA is exported to the cytoplasm through the nuclear pore. Ribosomes can then associate with the mRNA and produce a polypeptide chain by decoding the mRNA codons into the corresponding amino acids. Although this process of gene expression might seem simple and straightforward at first, it is much more complex on the second glance. For each step, different molecular machines are required that can be composed of hundreds of different proteins and oftentimes include also specialized RNAs. Therefore, a multitude of control mechanisms is deployed by the cell to prevent the generation of faulty, potentially harmful proteins. In this thesis, the main focus lies on the investigation of mRNA processing, more precisely on the ability of the cell to correctly detect and remove non-coding sequences and thereby to maintain an intact transcriptome.

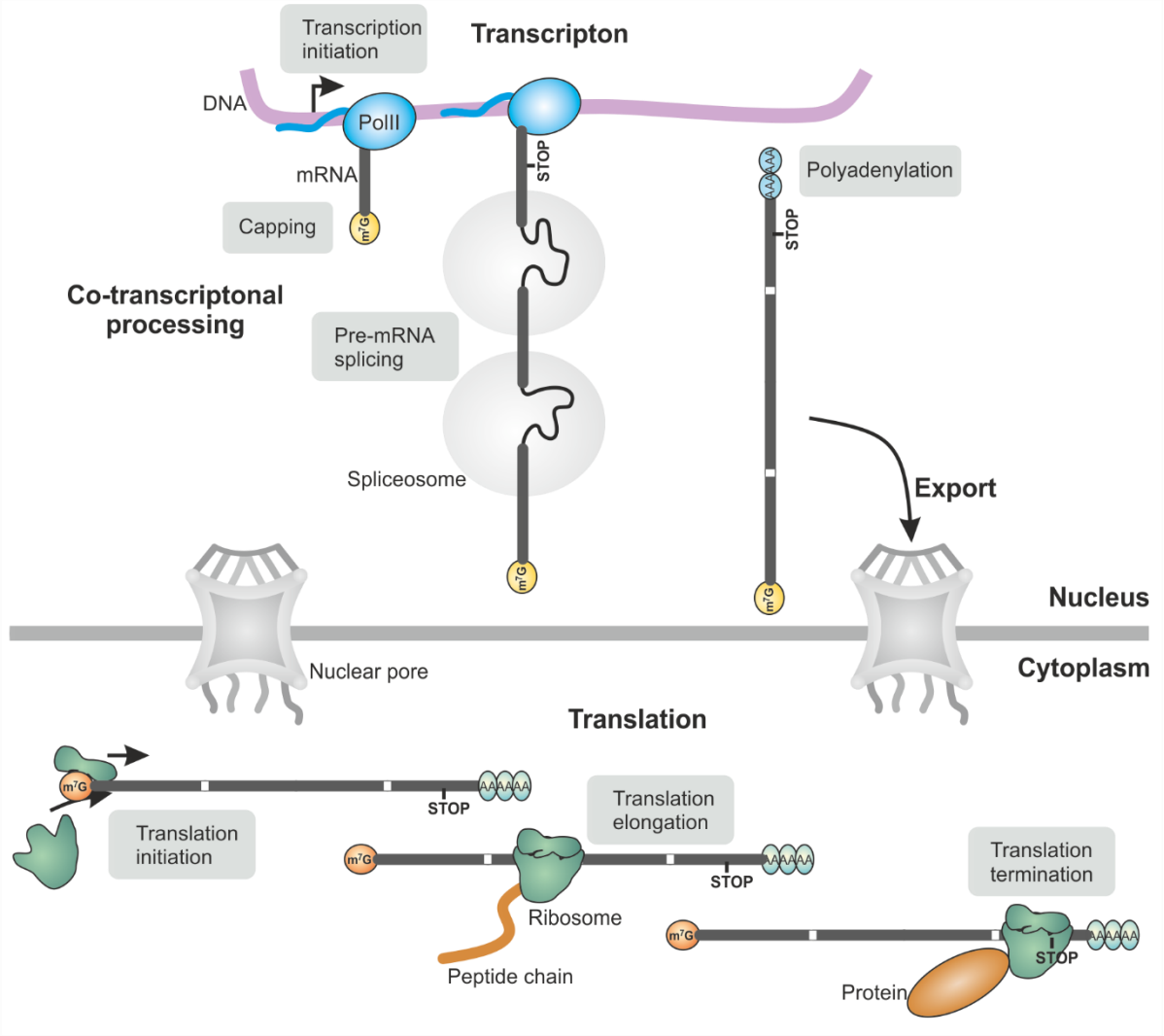


Figure 1: A schematic overview of the processes of gene expression. *Gene expression starts with the transcription of the DNA by PolII. The resulting mRNA is co-transcriptionally spliced and after polyadenylation exported into the cytoplasm. Translation of the mRNA into a polypeptide chain is mediated by the ribosomes that leave the mRNA when they arrive at the stop codon. The bold gray lines depict exons, the thinner black lines depict introns in the mRNA.*

5.1 Gene expression is regulated on the level of mRNA

As mentioned above, the process of gene expression consists of many steps involving different molecular machines, which require tight supervision to prevent mistakes. Mammals have developed a complicated regulatory system for gene expression that, especially on the level of mRNA, attracted more and more attention among the scientific community in the past decades. And although many studies investigate this topic, because of its high complexity there remain countless details unknown.

One decisive regulatory process during gene expression is the control of pre-mRNA splicing in the nucleus. During the splicing process, the spliceosome catalyzes the excision of intronic

sequences. The remaining coding sequences, the exons, are re-ligated and the mRNA is further processed (e.g. polyadenylated). At first, the composition of eukaryotic genes with its intronic and exonic sequences seems to be unnecessarily complicated. But the presence of introns and the necessity of splicing not only enables the production of multiple transcript- and protein-isoforms from a single gene but also holds the opportunity to finetune gene expression (Jo and Choi 2015). This allows to increase the complexity of an organism tremendously while only slightly increasing the size of the genome (Maniatis and Tasic 2002).

The splicing process is performed by the spliceosome, which is a vast machinery consisting of approximately 100 proteins and five small nuclear ribonucleoproteins (snRNPs) that is dynamically restructured throughout the splicing process (Figure 2, (Wahl, Will et al. 2009, Wilkinson, Charenton et al. 2020)). Normally, formation of the spliceosome is initiated by the recognition of the 5' splice site (GU) by the U1 snRNP. The U2 snRNP binding to the branchpoint (A) is facilitated by recognition of the 3' splice site (AG) by U2 small nuclear RNA auxiliary factor 1 and 2 (U2AF1, U2AF2). When the U4/U6.U5 tri-snRNP joins U1 and U2, all spliceosomal subunits are present on the mRNA. After this point, U1 followed by U4 leave the spliceosome. The remaining subunits hold the two exons in close proximity. The 5' splice site is cleaved and forms a connection to the branchpoint instead. Now, the 3' splice site is cleaved, and the exons are ligated and form an exon-exon junction. The lariat-forming intron leaves the ready-spliced mRNA with the spliceosome (reviewed in (Wilkinson, Charenton et al. 2020).

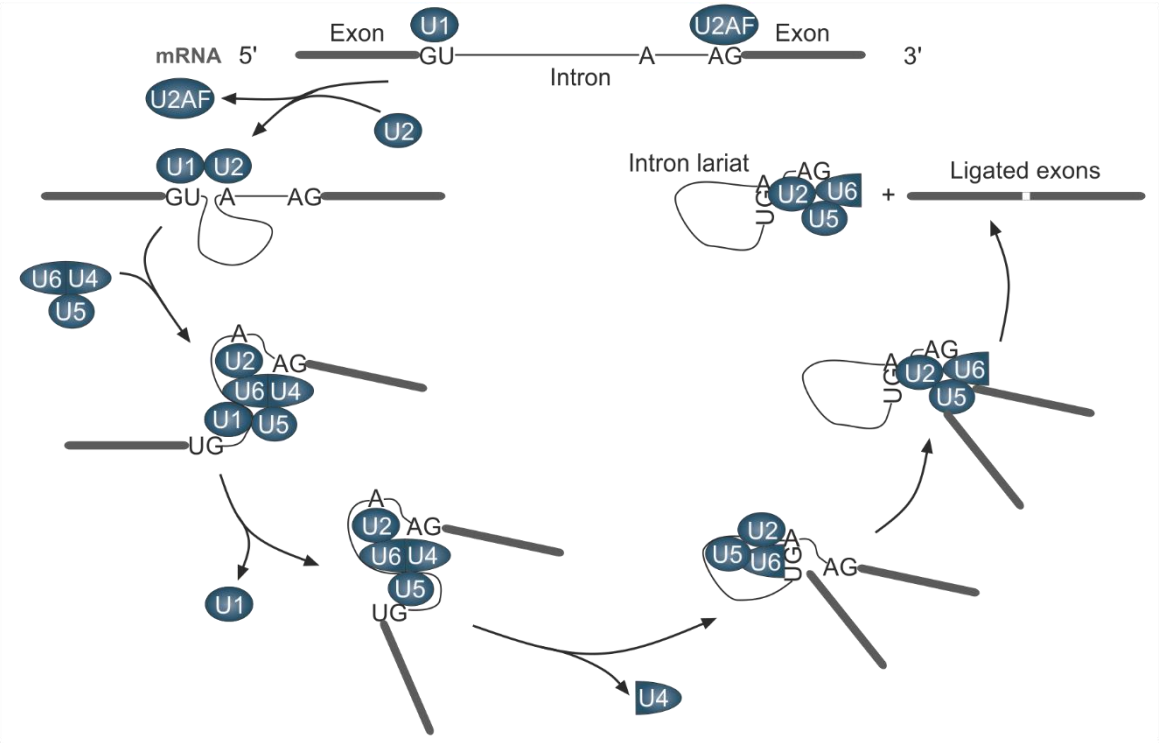


Figure 2: The spliceosome is dynamically rearranged during the splicing of an intron. *The different steps of the splicing process are shown counterclockwise. The spliceosomal compounds are depicted in dark blue, the 5' and 3' splice sites and branchpoint are highlighted.*

This schematic summary of the splicing process only covers the rearrangements of the snRNPs, while many other spliceosomal proteins are left out for simplicity. Considering all intermediate steps and all spliceosomal proteins, splicing is a process of high complexity and its detailed description would be out of scope for this thesis. The spliceosome has the delicate task to discriminate between genuine splice sites and so-called cryptic splice sites. A cryptic splice site resembles the consensus sequence of a genuine splice site but is not used under normal conditions. As the sequences of the splice sites in the human genome are very heterogenous, except for the well-conserved GU and AG motifs, the spliceosome requires some guidance to correctly identify the intended splice sites. This task is fulfilled by many different RNA-binding proteins (RBP) that bind to specific sequences in the exons or introns of the pre-mRNA. These *trans*-acting factors are generally called splicing factors, although this group is very diverse. Splicing factors bind to *cis* elements that are sequences in the pre-mRNA itself. *Cis* elements that improve intron and exon definition and therefore splicing are called exonic or intronic splicing enhancers (ESE/ISE), depending on their position in the mRNA (De Conti, Baralle et al. 2013, Kornblihtt, Schor et al. 2013). ESEs are often bound by splicing factors that belong to the serine/arginine-rich protein (SR protein) family, which have an "exonization" effect and

therefore repress the removal of the bound sequences (Figure 3). SR proteins are a defined group of highly conserved proteins that carry an arginine/serine rich domain (RS domain) and at least one RNA recognition motif (RRM) ((Roth, Zahler et al. 1991, Zahler, Lane et al. 1992) reviewed in (Shepard and Hertel 2009)).

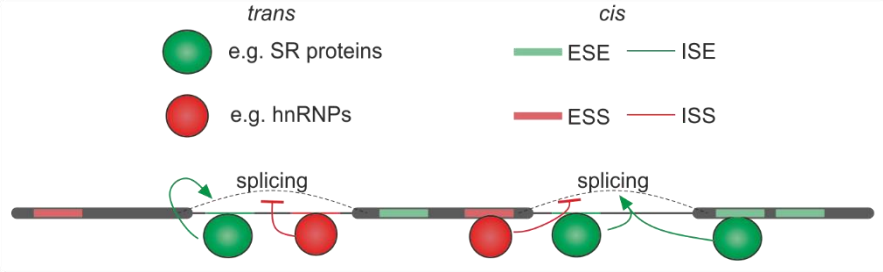


Figure 3: Cis sequence elements can be bound by trans factors to influence splicing. *Both introns and exons carry sequence motifs that can be bound by specific splicing factors that can either enhance (ESE/ISE) or repress splicing (ESS/ISS). Thick gray lines depict exons, thinner black lines depict introns. Colored lines depict the indicated sequence motifs. Spheres indicate the trans-acting factors.*

On the other hand, exonic and intronic splicing silencers (ESS/ISS) can inhibit splicing when they are bound by specific factors. In case of ESS, these factors include proteins of the splicing inhibitory family of heterogenous nuclear ribonucleoproteins (hnRNPs) that have an "intronization" effect and promote removal of the bound sequences. Accordingly, some of these hnRNPs can bind to ISE motifs to enhance splicing of surrounding exons (Wang, Ma et al. 2012). Overall, both types of enhancing or silencing sequence elements and splicing factors contribute to the correct identification of to-be-spliced introns by the spliceosome. Still, these protein families require specific sequence elements to bind to the mRNA which might not always be given. In these cases, a sequence-independent mechanism for splicing regulation would be required. This is provided by the deposition of the trimeric exon junction complex (EJC).

5.2 Composition and functions of the EJC

The EJC is a special protein complex since it is deposited onto the mRNA in a splicing- rather than sequence-dependent manner. It binds 24 nt upstream of the emerging exon-exon junction and consists of the three core components EIF4A3, RBM8A and MAGOH (Figure 4A) (Le Hir, Izaurralde et al. 2000, Sauliere, Murigneux et al. 2012). During splicing, the spliceosomal component CWC22 recruits the DEAD-box helicase eukaryotic translation initiation factor 4A3 (EIF4A3) to the spliceosome (Alexandrov, Colognori et al. 2012, Barbosa,

Haque et al. 2012, Steckelberg, Boehm et al. 2012, Steckelberg, Altmueller et al. 2015). According to a recent study, also CWC27 is involved in this recruitment, but proposedly leaves the spliceosome when the heterodimer formed by RNA binding motif protein 8A (RBM8A) and mago nashi homolog (MAGOH/MAGOHB) joins EIF4A3 (Busetto, Barbosa et al. 2020). EIF4A3 achieves sequence-independent binding by interacting in its ATP-bound state with the phosphate-ribose backbone of the mRNA (Andersen, Ballut et al. 2006, Bono, Ebert et al. 2006). The interaction surface is formed by the two EIF4A3 RecA domains (Ballut, Marchadier et al. 2005). Upon ATP hydrolysis, EIF4A3 undergoes a conformational change and leaves the mRNA. Stable binding to the mRNA is achieved by interaction of EIF4A3 with the RBM8A-MAGOH heterodimer, which prevent EIF4A3 from executing the ATP hydrolysis. In this state, binding of the EJC to mRNA is very stable and is therefore maintained during most of the subsequent steps in the mRNA lifetime. Thus, the EJC can influence multiple steps during gene expression, ranging from splicing over mRNA export and translation to the decay of faulty mRNAs via the nonsense-mediated mRNA decay (NMD) (Figure 4B, (Kim, Kataoka et al. 2001, Wang, Murigneux et al. 2014, Fukumura, Wakabayashi et al. 2016, Gromadzka, Steckelberg et al. 2016, Viphakone, Sudbery et al. 2019), reviewed in (Schlautmann and Gehring 2020)).

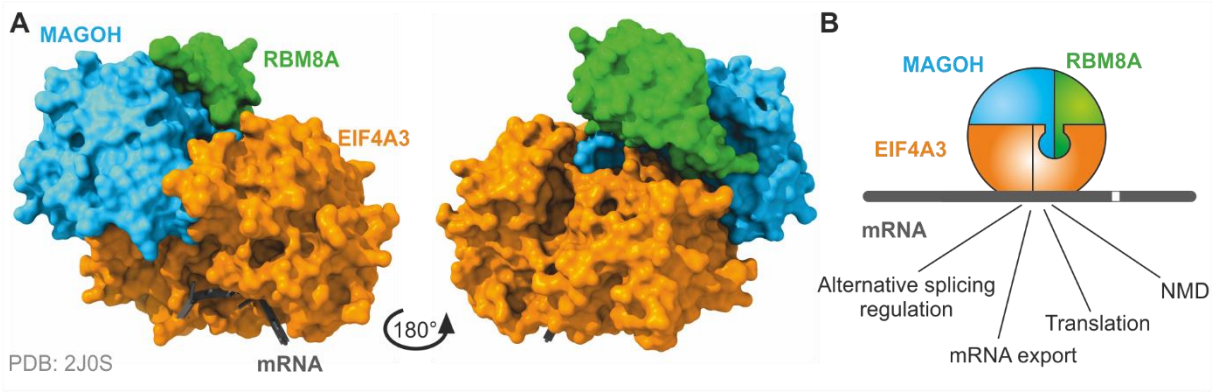


Figure 4: Structure of the trimeric EJC core complex. *(A) Structure of the mRNA-bound EJC is accessible at PDB with the identifier 2J0S (Bono, Ebert et al. 2006). (B) Schematic depiction of the EJC with its associated functions.*

5.2.1 Regulation of alternative splicing by the EJC

One of the central functions of the EJC is the regulation of splicing. Although EJC deposition is only completed after splicing, it can still influence the splicing of other introns in the transcript. Already in 2010 it was shown in *Drosophila* that EJCs can influence splicing of the map kinase gene (Ashton-Beaucage, Udell et al. 2010, Roignant and Treisman 2010). Here, removal of EJC components induced exon skipping, which led to downregulation of the functional protein.

Moreover, the fourth intron of the PIWI mRNA in *Drosophila* was retained in the absence of the EJC (Hayashi, Handler et al. 2014, Malone, Mestdagh et al. 2014). For PIWI intron 4 splicing, splicing of the subsequent intron and EJC deposition were required, suggesting that EJCs can also influence splicing of upstream introns. In *Xenopus*, depletion of EIF4A3 likewise resulted in complete or partial retention of two distinct introns (Haremaki and Weinstein 2012). A few years later, a splicing regulatory function of the EJC was also recorded in mammals (Wang, Murigneux et al. 2014, Fukumura, Wakabayashi et al. 2016). These studies used human cell lines depleted of EJC core factors and analyzed them using RNA sequencing. As a result, exon skipping (ES), exon inclusion (EI), intron retention (IR) and alternative 3' or 5' splice sites (A3SS/A5SS) were more frequently found in the absence of the EJC. Ever since these discoveries, a lot of research focused on the details of splicing regulation by the EJC. Two studies published in 2018 for instance demonstrate that the EJC prevents re-splicing by masking reconstituted splice sites (Blazquez, Emmett et al. 2018, Boehm, Britto-Borges et al. 2018). Thus, the EJC marks already spliced regions to avoid that these areas are spliced again. It can therefore be described as a molecular marker for splicing which prevents the loss of transcriptomic sequences that are essential for proper gene expression.

5.2.2 The EJC induces mRNA degradation via the NMD pathway

In addition to its splicing regulatory functions, the EJC plays an important role during the initiation of mRNA degradation via the NMD pathway. This pathway is responsible for the cytoplasmic degradation of mRNAs that carry a so-called premature translation termination codon (PTC). Normally, stop codons are located in the last exon and no splicing occurs behind this point. PTCs are usually not located in the last exon but otherwise resemble normal termination codons. There are several causes for the occurrence of a PTC. These causes can for instance be the presence of a nonsense-mutation in the DNA, the wrong incorporation of nucleotides by RNA PolII or frame shift caused by alternative splicing. The degradation of such PTC-bearing transcripts is important to prevent the production of C-terminally truncated proteins. These can be harmful for the cell as they can accumulate or even have dominant negative functions. Additionally, the NMD machinery degrades various regular transcripts to finetune gene expression (Kishor, Fritz et al. 2019).

During normal translation termination, the eukaryotic release factor 1 (eRF1) enters the ribosome where it recognizes the stop codon (Figure 5A, (Dever and Green 2012)). Together

with the eukaryotic release factor 3 (eRF3), eRF1 induces the release of the ribosomal subunits and the nascent peptide chain. The principles of aberrant translation termination at PTCs are similar, but the presence of downstream EJCs can lead to the recruitment of the NMD machinery. Normally, all EJCs are located upstream of the stop codon and are therefore removed from the mRNA by translating ribosomes. If an mRNA contains a PTC, the ribosome(s) would stop at this point and EJCs that were deposited at downstream splice junctions would not be removed. When one or more EJCs remain on an mRNA that is actively translated, these EJCs can lead to NMD induction. Importantly, the distance between the PTC and a downstream exon-exon junction must be at least 50 nucleotides (50 nt rule) to induce EJC-dependent NMD (Nagy and Maquat 1998, Zhang, Sun et al. 1998). This distance is required to prevent collision of the ribosome with the EJC, which would result in EJC removal and therefore abolish its NMD-activating potential.

To induce NMD, the EJC serves as an anchoring point for some NMD factors, which eventually leads to the recruitment of the NMD core machinery (Figure 5B). For this purpose, some NMD factors, like up-frameshift 3B (UPF3B), contain an EJC-binding motif (EBM) (Gehring, Neu-Yilik et al. 2003, Kashima, Jonas et al. 2010). This sequence motif binds to a composite surface that is presented from all three EJC core proteins together (Buchwald, Ebert et al. 2010). A long-used NMD model proposed that UPF3B binds to UPF2 and that these two proteins bridge the EJC to UPF1 (Lykke-Andersen, Shu et al. 2000, Kim, Kataoka et al. 2001). Also, in case of UPF3B depletion, its homolog UPF3A can replace it and maintain NMD functionality (Wallmeroth, Lackmann et al. 2022).

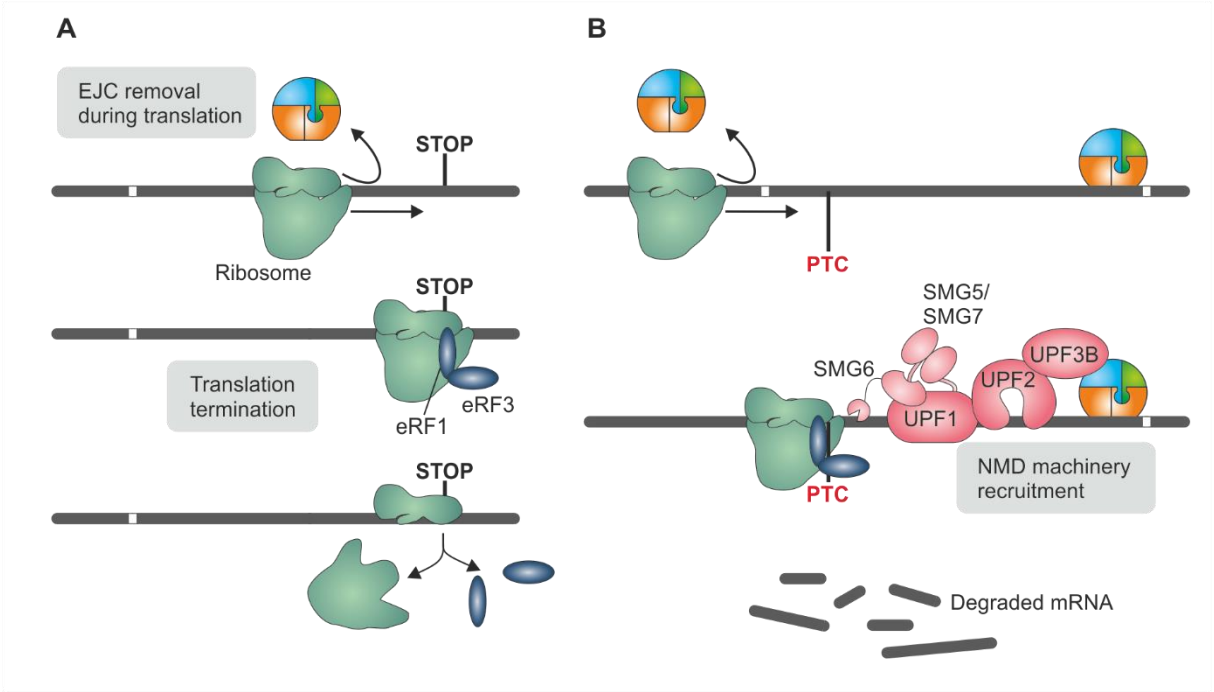


Figure 5: The basic model of normal translation termination and EJC-dependent NMD. *(A) The translating ribosome removes EJCs from the mRNA and stalls at the stop codon. Binding of eRF1 and eRF3 induces dissociation of the ribosome. (B) To induce the degradation of a PTC-containing mRNA, downstream located EJCs interact with the NMD factor UPF3B, which, together with UPF2, bridges the interaction between the EJC and UPF1. The heterodimer SMG5/7 and SMG6 bind to the phosphorylated N- and C-terminus of UPF1.*

The abundant protein UPF1, which is considered to be the central NMD factor, binds non-specifically to all mRNAs and can interact with the eRF1 and eRF3 that are bound to the ribosome during translation termination (Czaplinski, Ruiz-Echevarria et al. 1998). The interactions between the EJC and the UPF proteins induce the phosphorylation of the N- and C-terminal regions of UPF1 (Kashima, Yamashita et al. 2006). Thereby, UPF1 can be bound by the suppressors with morphogenetic effect on genitalia 5, 6 and 7 (SMG5, SMG6 and SMG7) (Okada-Katsuhata, Yamashita et al. 2012). In older NMD models, the heterodimer formed by SMG5 and SMG7 and the endonuclease SMG6 function in separate, redundant downstream pathways (Colombo, Karousis et al. 2017). SMG5 and SMG7 in that case promote deadenylation by recruiting the CCR4-NOT complex (Loh, Jonas et al. 2013). Endonucleolytic cleavage is mediated by SMG6 and leaves two mRNA fragments that can be degraded by exonucleolytic decay (Huntzinger, Kashima et al. 2008) (Gatfield and Izaurralde 2004, Eberle, Lykke-Andersen et al. 2009).

Recent research now suggests a different model, in which both pathways need to be intact to drive degradation of the mRNA (Boehm, Kueckelmann et al. 2021). To this end, the presence

of either SMG5 or SMG7 is sufficient to trigger NMD, since they can act redundantly in enabling the endonucleolytic cleavage by SMG6.

Apart from the induction by PTCs, NMD can also be triggered on transcripts that contain an upstream open reading frames (uORF) or a long 3' UTR. In case of a long 3' UTR, NMD is induced by a too long distance between the ribosome and the poly(A)-binding protein (PABPC1) (Amrani, Ganesan et al. 2004). If an mRNA contains a uORF, the ribosome terminates prematurely at the stop codon of this uORF and splicing and EJC deposition in the regular ORF induce the degradation via NMD (Somers, Pöyry et al. 2013).

5.3 The EJC auxiliary complexes ASAP and PSAP

To fulfill the variety of functions of which some are explained above, the trimeric EJC core alone would not be sufficient, but rather requires the interaction with other protein factors. Because of its splicing-dependent deposition, the EJC can recruit proteins independently of the mRNA sequence to the exon-exon junction. Therefore, it serves as a binding platform to link other gene expression regulatory proteins to the mRNA. In several cases, these so-called peripheral EJC components form functional sub-complexes using the EJC as the anchor on the transcript. Two important EJC-auxiliary complexes are the apoptosis and splicing–associated protein (ASAP) complex and the PSAP complex (Schwerk, Prasad et al. 2003, Murachelli, Ebert et al. 2012), which both are implicated to regulate alternative splicing.

5.3.1 Composition of the ASAP and PSAP complexes

The ASAP complex consists of RNA-binding protein with serine-rich domain 1 (RNPS1), Sin3-associated protein of 18 kDa (SAP18) and apoptotic chromatin condensation inducer in the nucleus (ACINUS, ACIN1) (Schwerk, Prasad et al. 2003). The PSAP complex on the other hand contains both RNPS1 and SAP18, but PININ (PNN) instead of ACIN1 and is termed PSAP for its similarity to the ASAP complex (Murachelli, Ebert et al. 2012).

Interestingly, all individual components of the complexes, except for RNPS1, were at first discovered for other functions than splicing regulation. ACIN1 for example, as the name suggests, was identified as a mediator of chromatin condensation in case of apoptosis (Sahara, Aoto et al. 1999). Three different ACIN1 protein isoforms are produced from the ACIN1 gene which are termed ACIN1 L, ACIN1 S and ACIN1 S'. Different from the two shorter isoforms, the

long N-terminus of ACIN1 L contains an additional SAF-A/B, Acinus and PIAS (SAP) domain (Aravind and Koonin 2000, Rodor, Pan et al. 2016). All three ACIN1 isoforms were shown to be cleaved by Caspase 3 during the onset of apoptosis. After cleavage, a shortened ACIN1 version remains that mediates the condensation of chromatin (Sahara, Aoto et al. 1999). This cleaved ACIN1 part was identified as an RRM, for its similarity to the RRM in the *Drosophila* splicing regulator sex-lethal (Sxl). Shortly thereafter, ACIN1 also appeared in a screen for proteins containing arginine/serine-rich domains (RS domains) (Boucher, Ouzounis et al. 2001). In metazoans, RS domains are frequently found in regulators of splicing, especially members of the SR protein family. Strikingly, ACIN1 not only contains an RS domain but also an RRM, which are both characteristics of SR proteins. Therefore, ACIN1 is defined as an SR-like or SR-related protein (Boucher, Ouzounis et al. 2001). SR-like proteins differ from SR proteins as they can lack the RRM, have a distinct domain structure and are not recognized by the SR protein-specific antibody (reviewed in (Shepard and Hertel 2009)). The hypothesis of ACIN1 having splicing-regulatory potential was further supported by the finding that it was co-purified with the EJC (Tange, Shibuya et al. 2005).

Like ACIN1, also SAP18 was first identified in a different context. First, it was found to be an 18 kDa interactor of the mammalian transcriptional repressor Sin3 (mSin3) and termed accordingly (Zhang, Iratni et al. 1997). In a complex with mSin3 and the histone deacetylases 1 and 2 (HDAC1, HDAC2), SAP18 repressed transcription.

The PSAP component PNN was first identified to be localized within mature desmosomes of epithelial cells (Ouyang and Sugrue 1992, Ouyang and Sugrue 1996). However, the association with desmosomes was soon questioned by a study that found PNN to be localized in the nucleus (Brandner, Reidenbach et al. 1997). Later, PNN was claimed to modulate splicing, which was supported by its interaction with the splicing activator RNPS1 (Wang, Lou et al. 2002, Li, Lin et al. 2003). Still, the tested alternative splicing events were independent of its interaction with RNPS1, suggesting that PNN might have some splicing regulatory activity on its own (Wang, Lou et al. 2002).

Different from the other ASAP/PSAP components, RNPS1 was already initially identified as an activator of splicing (Mayeda, Badolato et al. 1999). Its identification as an SR-like protein matches this finding (Boucher, Ouzounis et al. 2001). In addition to a central RRM, RNPS1

contains a C-terminal RS or arginine-serine/proline-rich (RS/P) domain and an N-terminal serine-rich (S) domain (Mayeda, Badolato et al. 1999, Sakashita, Tatsumi et al. 2004).

In their study in 2012, Murachelli et al. extensively investigated the formation of the ASAP complex and thereby revealed the existence of the PSAP complex (Murachelli, Ebert et al. 2012). They identified the interaction surfaces of the ASAP components and solved the crystal structure of the minimal ASAP complex using the *Drosophila* ACIN1 homolog and the mouse SAP18 homolog (Figure 6). Interestingly, in RNPS1 the RRM is not required for interaction with the RNA but is essential for the assembly of the ASAP or PSAP complexes. In ACIN1, the conserved motif that was found to be responsible for ASAP assembly was accordingly termed RNPS1-SAP18-binding (RSB) motif. SAP18 interacts with both proteins via its ubiquitin-like fold, which was already previously suggested to mediate protein-protein interaction (Zhang, Iratni et al. 1997, McCallum, Bazan et al. 2006). While RNPS1 and ACIN1 can form a binary complex, SAP18 only interacts with both proteins together (Tange, Shibuya et al. 2005, Murachelli, Ebert et al. 2012). In PNN, a motif was found that resembled the RSB motif of ACIN1. Accordingly, also PNN interacts with RNPS1 and SAP18 and can form the PSAP complex (named after the ASAP complex). Both complexes were also shown to be mutually exclusive. Different from ACIN1, PNN was shown to be able to form a dimer with SAP18 in the absence of RNPS1 (Costa, Canudas et al. 2006, Murachelli, Ebert et al. 2012).

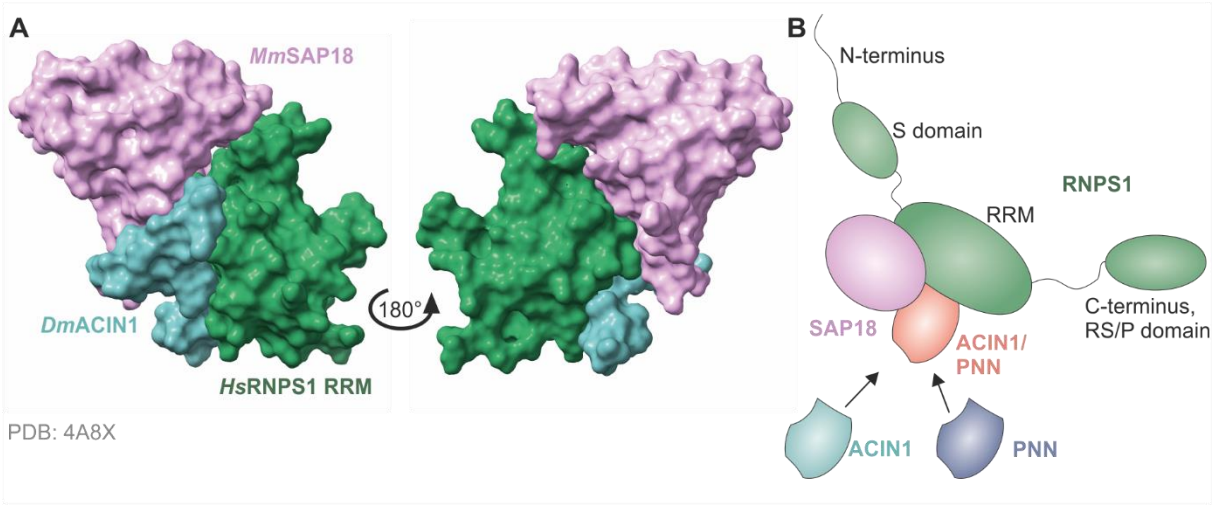


Figure 6: The structure of the minimal ASAP complex. *(A) The different protein components of the minimal ASAP complex are shown in different colors. This crystal structure can be accessed at PDB with the identifier 4A8X (Murachelli, Ebert et al. 2012) (B) The schematic depiction of the ASAP or PSAP complex with the different RNPS1 domains will be used in all following models.*

A few years ago, the interaction of the ASAP complex with the EJC was shown to be mediated by ACIN1 (Wang, Ballut et al. 2018). For the PSAP complex, PNN is thought to be the link to

the EJC since neither SAP18 nor RNPS1 can interact with the EJC by itself. Yet the exact mechanisms of when and how these complexes are assembled and recruited to the EJC remains unknown.

5.3.2 The ASAP and PSAP complex expand the functional repertoire of the EJC

In association with the EJC, both ASAP and PSAP components were shown to regulate alternative splicing. The PIWI intron 4 splicing in *Drosophila*, that required the presence of the EJC, is also dependent on ACIN1 and RNPS1 (Figure 7, (Hayashi, Handler et al. 2014, Malone, Mestdagh et al. 2014)). To enable PIWI intron 4 splicing, intron 5 needed to be spliced first, suggesting that EJC deposition and ASAP recruitment enhance splicing of the otherwise retained intron. One year later, ACIN1 was shown to be involved in splicing of the retinoic acid response (RAR) gene (Wang, Soprano et al. 2015). Interestingly, here ACIN1 increased splicing of an intron that contained a weak 5' splice site while RNPS1 repressed this activity. This suggests that there might be a complex-internal control of the individual splicing regulatory activities.

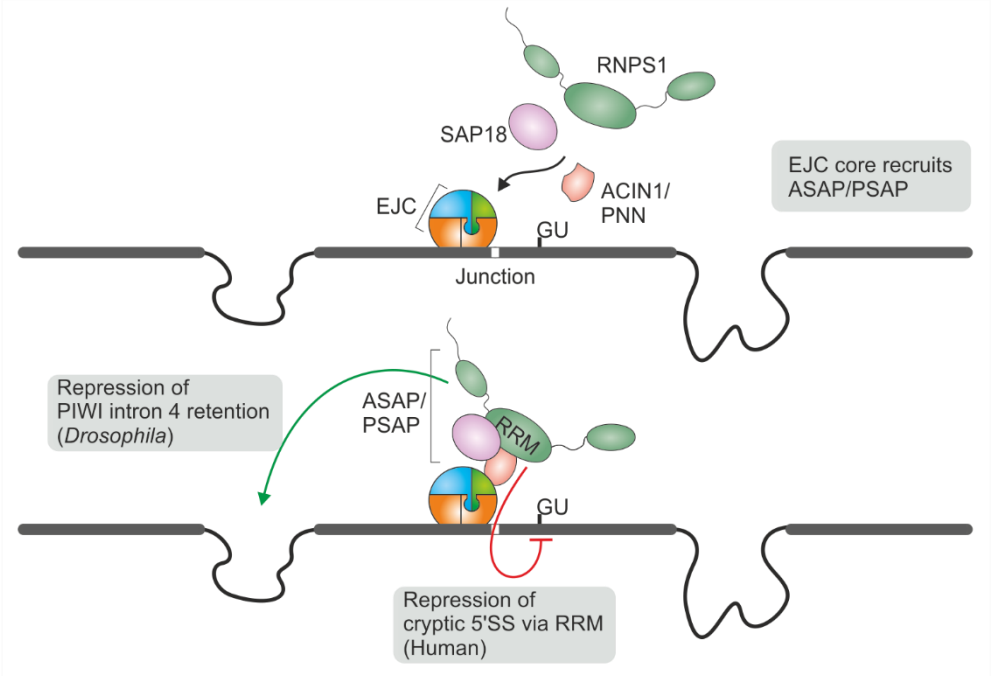


Figure 7: The ASAP/PSAP complexes can regulate alternative splicing in flies and human. *The EJC recruits RNPS1 to the exon-exon junction as a part of the ASAP or PSAP complex. In Drosophila, RNPS1 prevents retention of PIWI intron 4, in human it represses the usage of cryptic 5' splice sites.*

Another example of EJC and RNPS1-dependent splicing is the alternative splicing of MAPK in *Drosophila* (Ashton-Beaucage, Udell et al. 2010, Roignant and Treisman 2010). RNPS1

depletion led to exon skipping in this gene, as the depletion of the EJC factors did. Together with the EJC and PNN, RNPS1 prevents the usage of cryptic 5' splice sites (Figure 7, (Blazquez, Emmett et al. 2018, Boehm, Britto-Borges et al. 2018)). Different from IR in *Drosophila*, this splicing regulatory activity depended on the splicing of the upstream intron. While the EJC represses the usage of reconstituted splice sites by direct masking of splicing relevant sequence elements, the RNPS1-dependent mechanism is not fully understood but requires the RNPS1 RRM (Boehm, Britto-Borges et al. 2018).

Interestingly, these findings assign distinct splicing regulatory activities to ACIN1 and PNN. This was supported by the findings of a recent study, which showed that PNN or PSAP-dependent alternative splicing events were not rescued by ACIN1 (Wang, Ballut et al. 2018). Therefore, although the complexes only differ in one component, ASAP and PSAP probably do not function redundantly.

It furthermore appears that RNPS1 is involved in all of these ASAP/PSAP regulated alternative splicing events, while a dependency on SAP18 was not confirmed. This not only suggests that RNPS1 is the effector molecule in both complexes but also that they might be active without SAP18.

In addition to splicing regulation, RNPS1 is supposed to enable another EJC-dependent gene expression step. Already around the time of the ASAP discovery, RNPS1 was shown to induce NMD of reporter mRNAs (Lykke-Andersen, Shu et al. 2001, Gehring, Kunz et al. 2005). In these experiments, RNPS1 was artificially brought to the 3' UTR of the reporter by tethering. Later, overexpression of RNPS1 was found to increase NMD of a reporter while depletion of RNPS1 led to increased expression of potentially RNPS1-dependent endogenous NMD targets (Viegas, Gehring et al. 2007, Mabin, Woodward et al. 2018). Interestingly, the EJC was found to constitute mutually exclusive NMD-activating complexes with either RNPS1 or with cancer susceptibility candidate gene 3 (CASC3), which was also shown to be required for NMD of many endogenous targets (Mabin, Woodward et al. 2018, Gerbracht, Boehm et al. 2020). Compared to RNPS1-associated EJCs, which mainly localize in the nucleus, CASC3-associated EJCs are mainly present in the cytoplasm.

5.4 Relevance of the EJC and associated factors for human diseases

Since splicing and NMD are very basic steps in mRNA processing and quality control, it can be assumed, that their dysregulation has severe physiological effects. Indeed, an impairment in splicing is often found in various diseases, including cancer (Wang and Cooper 2007, Pedrotti and Cooper 2014). Moreover, in many disease-causing mutations a PTC is formed that leads to the degradation of the affected mRNA via the NMD pathway (Miller and Pearce 2014).

Considering that both pathways can be regulated by the EJC, it is logical that disruption of EJC functionality also induces several diseases. Indeed, the EJC core components EIF3A3 and RBM8A are listed as common essential in cancer dependency map (DepMap: CRISPR screens show that cells constantly depleted of these proteins would not be viable). MAGOH is also listed as common essential, although it can be substituted by its homolog MAGOHB. Many of the diseases that result from dysregulation of EJC core components cause defects in neural development (McMahon, Miller et al. 2016). These include microencephaly caused by haploinsufficiency of either RBM8A or MAGOH (Silver, Watkins-Chow et al. 2010, Mao, Pilaz et al. 2015). In patients with intellectual disabilities in general, copy number variations of EIF4A3 and RNPS1 were frequently found (Nguyen, Kim et al. 2013).

In addition to these neurological disorders, involvement of the EJC and RNPS1 was also detected for several other diseases. Impaired RBM8A expression for instance was proven to be one of the main causes of the thrombocytopenia with absent radius syndrome (TAR syndrome) (Albers, Paul et al. 2012). The Richieri-Costa-Pereira syndrome, an autosomal-recessive acrofacial dysostosis, was found to be linked to a reduced expression of EIF4A3 (Favaro, Alvizi et al. 2014). In mice with ischemic stroke, RNPS1 was upregulated in the brain and reduced RNPS1 expression was detected in mice with adenocarcinoma or dysplasia of the salivary gland (Mäkitie, Reis et al. 2005, Zhang, Guo et al. 2020). Even a small disruption of RNPS1 by introducing a single amino acid mutation that reduces its activity led to defects in hematopoiesis in homozygous mice (Zhong, Choi et al. 2022).

The above-mentioned diseases, which only display a few examples, demonstrate that the integrity of the EJC and RNPS1 is crucial for the overall health of mammalian cells. To be able to understand and maybe even cure or prevent these types of diseases one day, it is important to increase the knowledge of the basic, underlying mechanisms. Therefore, studying the

mechanistic details of EJC and RNPS1 functions is of great interest to expand the understanding of their role in the bigger picture of gene expression regulation.

6 AIMS OF THIS THESIS

The key regulatory steps of mammalian gene expression have been known for a long time and studied extensively. Multiple of these steps rely on the presence of the EJC, a trimeric protein complex, that is deposited during splicing at the exon-exon junction and that accompanies an mRNA throughout most of its lifetime. To fulfill a diversity of functions the EJC is equipped with changing sets of auxiliary factors that enable specific regulations. One of these auxiliary factors is the ASAP and PSAP component RNPS1, which supposedly has both splicing and NMD regulatory functions. However, the involvement of RNPS1 in these processes is not fully understood yet and requires deeper investigation.

In this thesis, specifically three aspects of RNPS1's function as an EJC auxiliary complex will be deciphered.

1. Previously, overexpression of RNPS1 was shown to induce NMD of reporter mRNAs (Viegas, Gehring et al. 2007), artificial binding of RNPS1 to reporters via tethering reduced reporter expression (Lykke-Andersen, Shu et al. 2001, Gehring, Kunz et al. 2005), and multiple endogenous targets were upregulated upon RNPS1 depletion (Mabin, Woodward et al. 2018), suggesting that they are usually targeted by RNPS1-dependent NMD. Here, it will be determined using high-throughput RNA sequencing if RNPS1 is indeed involved in NMD. Furthermore, it will be examined whether the presence of RNPS1 is critical for functional NMD in general or whether RNPS1 regulates NMD of specific target mRNAs.
2. Over the years, RNPS1 was shown to have multiple splicing regulatory functions, including the prevention of intron retention and exon skipping in *Drosophila* (Ashton-Beaucage, Udell et al. 2010, Roignant and Treisman 2010, Hayashi, Handler et al. 2014, Malone, Mestdagh et al. 2014) and the suppression of cryptic 5' splice sites in human cells (Blazquez, Emmett et al. 2018, Boehm, Britto-Borges et al. 2018). For the repression of cryptic 5' splice sites, the RNPS1 RRM was shown to be essential. The RRM is a central domain of RNPS1 that mediates the interaction with the other components of the ASAP/PSAP complexes. In this thesis, high-throughput RNA sequencing will be used to investigate the impact of RNPS1 on all types of alternative splicing and whether the RNPS1 RRM is the general splicing regulatory domain.

Additionally, a mechanism to prevent intron retention in human cells by RNPS1 deposition is examined using minigene reporters.

3. To fulfill its various functions, the EJC core interacts with a multitude of auxiliary proteins. Since RNPS1 is also associated with multiple functions, it proposedly also interacts with other proteins to be capable of all these tasks. Here, the composition of the RNPS1 interactome will be investigated by mass spectrometry analysis. This should give insight into the functions that are mainly regulated by RNPS1 and how this regulation takes place.

7 MATERIAL AND METHODS

The following material and methods section is in parts identical to and otherwise complemented from (Schlautmann, Lackmann et al. 2022).

7.1 Cell Culture

All experiments were performed using one of the following cultured human cell lines: Flp-In-T-REx-293 (HEK/293; human, female, embryonic kidney, epithelial; Thermo Fisher Scientific, RRID:CVCL_U427), HeLa Flp-In-T-REx (HFT; human, female, cervix; Elena Dobrikova and Matthias Gromeier, Duke University Medical Center) and HeLa Tet-Off (HTO; human, female, cervix; Clontech, RRID:CVCL_V352). The cells were cultured in high-glucose GlutaMAX DMEM from Gibco supplemented with 1x Penicillin Streptomycin and 9% fetal bovine serum (Gibco) in a humidified incubator at 37°C and 5% CO_2.

7.2 Transfection and Generation of stable cell lines

7.2.1 Plasmids and Cloning

To generate stable RNPS1 and EIF4A3 rescue, and RNPS1 proximity labeling cell lines, inserts were first cloned into pCI-neo (Promega, Cat# E1841) with N-terminal FLAG-emGFP-tag, FLAG-TurboID-tag (Branon, Bosch et al. 2018) or MYC-UltraID-tag (Zhao, Bitsch et al. 2021) using XhoI and NotI (both New England Biolabs). For RNPS1, pre-existing inserts generated from human cDNA were used. For EIF4A3, a codon-optimized synthetic gene was used (Mr. Gene). After cloning into pCI-neo, the tagged inserts were subcloned into PB-CuO-MCS-BGH-EF1-CymR-Puro (modified from System Biosciences) using NheI (New England Biolabs) and NotI.

RFX5 reporter constructs were generated from human gDNA or cDNA using Q5 polymerase (New England Biolabs) and cloned into pcDNA5/FRT/TO (Thermo Fisher Scientific; Cat# V652020) with an N-terminal FLAG-Tag using again XhoI and NotI. For single-intron reporters, mutagenesis was performed with Q5 polymerase. The RFX5 tethering reporter was ordered from IDT as a gBlock.

GST (as control), ASAP/PSAP and EJC tethering constructs were cloned into pCI-neo that contained an N-terminal MS2-V5-tag using XhoI and NotI.

Plasmids that were used in this work are either listed in Table 1 or in the Supplementary Table 1 from (Schlautmann, Lackmann et al. 2022).

7.2.2 Transfections and stable cell lines

Both stable and transient transfections were performed using the standard Calcium-phosphate transfection method. 24 h before transfection, 2.5–3 × 10^5 cells were seeded in 6-well plates. For PB-CuO plasmids (PiggyBac System), 1 µg of PiggyBac construct was transfected together with 0.8 µg of the Super PiggyBac Transposase expressing vector into 293 or HTO cells. For pcDNA5 plasmids (Flip-in System), 1.5 µg of pcDNA5 construct were co-transfected with 1.5 µg Flippase expression vector (pOG44) into HFT cells.

After 48 h the transfected cells were transferred to 10 cm dishes and selected with either 2 µg ml^{-1} puromycin (InvivoGen, for PiggyBac) or with 100 µg ml^{-1} hygromycin (InvivoGen, for pcDNA5). Expression of the PiggyBac constructs was induced using 30 µg ml^{-1} cumate and expression of the pcDNA5 constructs was induced using 1 µg/ml doxycycline.

For transient transfections, pcDNA5 reporter plasmids were co-transfected with pCI-neo plasmids expressing the desired tethering constructs into HTO cells. In HTO cells, pcDNA5 expression did not require induction since the cells do not contain a Tet-repressor.

7.3 siRNA mediated Knockdown

For siRNA mediated knockdown (KD), cells were seeded at a density of 2-3 x 10^5 in 6-well plates. The cells were reverse transfected according to the manufacturer's protocol using 2.5 µl Lipofectamine RNAiMAX and a total of 60 pmol of siRNA(s). The siRNAs used for the indicated gene knockdowns are listed Table 2 or in in the Supplementary Table 1 from (Schlautmann, Lackmann et al. 2022).

7.4 Endpoint and quantitative RT-PCR

For endpoint (RT-PCR) or quantitative PCR (qRT-PCR), RNA was extracted and reverse-transcribed as follows. The cells were treated with KDs, KD rescues or overexpression of reporter and tethering constructs. Subsequently, RNA was extracted using either peqGOLD TriFast (VWR Peqlab) or RNA-Solv Reagent (Omega Bio-Tek) according to the manufacturer's instructions for TriFast. Instead of adding 200 µl chloroform to induce phase separation, 150 µl

1-bromo-3-chloropropane were used. After a single wash with 75% EtOH, RNA was eluted in 20 µl RNase-free water. For reverse transcription, 0.5-1 µg of RNA, GoScript Reverse Transcriptase from Promega and 10 µM VNN-$(dT)_{20}$ primer were set up in a total reaction volume of 20 µl and incubated as described in the manufacturer's protocol. All depicted RT-PCRs were performed using MyTaq™ Red Mix (Bioline/BIOCAT) according to the manufacturer's instructions and quantified using the Image Lab software (Bio-Rad, version 6.0.1). The RT-qPCRs were performed using the GoTaq qPCR Master Mix (Promega). Which primer was used for which PCR is indicated in the primer list in the Supplementary Table 1 from (Schlautmann, Lackmann et al. 2022).

7.5 RNA sequencing

RNA sequencing was performed using three biological replicates of either 293 or HTO cells with the indicated KDs and rescues. The rescues were induced 24 h or, in case of EIF4A3, 5 h after the siRNA mediated KD. Cells were harvested either four days after KD or 48 h (EJC core set). RNAs of the EJC core set were harvested earlier since cells tend to die usually around three days after EJC core KDs. Other sets were incubated longer to maximize the effects of the KDs. Cells were lysed in either TriFast (VWR Peqlab; HTO) or RNA Solv Reagent (Omega Bio-Tek, HEK) followed by RNA extraction with DIRECTzol Miniprep Kit (Zymo Research) according to the manufacturer's instructions. Sample concentrations were measured on a nanodrop device and adjusted to 50-200 ng/µl, then samples were handed to the sequencing facility, where library preparation and sequencing were performed as described below (Altmüller, CCG). Spike-In Control Mixes (SIRV Set1 SKU: 025.03, Lexogen) were added to the samples according to the Supplementary Table 1 from (Schlautmann, Lackmann et al. 2022). The Spike-Ins were used only for evaluation of the performance but not for the later analyses. The cDNA libraries were prepared using 1 µg total RNA and the TruSeq Stranded Total RNA kit (Illumina). For removal of cytoplasmic and mitochondrial ribosomal RNAs, biotinylated, target-specific oligos and the Ribo-Zero Gold Human/Mouse/Rat kit were used. Afterwards, the RNA was purified, then fragmented and cleaved. Reverse transcription with random primers was used to synthesize the first cDNA strand, while the second strand was synthesized by DNA Polymerase I and RNA was digested using RNase H. Prior to adapter ligation, a single 'A' base was added to the ends of the now double-stranded cDNA. For library finalization, the cDNA was purified and PCR amplified. The resulting library was then validated and quantified on the

Agilent TapeStation and afterwards equimolar amounts of library were pooled. Quantification of the pool was performed using the Peqlab KAPA Library Quantification Kit 587 and the Applied Biosystems 7900HT Sequence Detection System. Final sequencing was performed on an Illumina NovaSeq6000 sequencing instrument with an PE100 protocol. The raw data were analyzed as described in the following paragraphs.

7.6 Bioinformatic analyses of RNA sequencing data

For bioinformatic analysis, the RNA sequencing reads were first mapped to the human genome (version 38, GENCODE release 33 transcript annotations (Frankish, Diekhans et al. 2019), supplemented with SIRVomeERCCome annotations from Lexogen obtained from https://www. lexogen.com/sirvs/download/). Mapping was performed using STAR read aligner (version 2.7.3a, (Dobin, Davis et al. 2013)) and estimates for transcript abundance were calculated using Salmon (version 1.3.0, (Patro, Duggal et al. 2017)) with a decoy-aware transcriptome. Most of the primary bioinformatic analyses were performed by Dr. Volker Boehm.

7.6.1 Differential gene expression analysis

Transcript abundances as computed by Salmon were imported to R using tximport (Soneson, Love et al. 2015) followed by filtering for genes with 10 or more counts in each sample. The differential gene expression (DGE) was then analyzed using the DESeq2 (Love, Huber et al. 2014) R package (version 1.34.0). P-values were calculated by DESeq2 using a two-sided Wald test and corrected for multiple testing using the Benjamini-Hochberg method. As significance thresholds a $|\log2FoldChange| > 1$ and adjusted P-value (Padj) < 0.05 were chosen. The results of the DGE analysis can be found in the Supplementary Table 2 from (Schlautmann, Lackmann et al. 2022).

7.6.2 Differential transcript usage

To compute the differential transcript usage (DTU), the IsoformSwitchAnalyzeR (ISAR, version 1.16.0) and the DEXSeq method (Robinson and Oshlack 2010, Anders, Reyes et al. 2012, Ritchie, Phipson et al. 2015, Soneson, Love et al. 2015, Vitting-Seerup and Sandelin 2017, Vitting-Seerup and Sandelin 2019). Here, a delta isoform fraction $|dIF| > 0.1$ and adjusted P-value (isoform switch q value) < 0.001 were used as significance thresholds. To determine the

PTC status of transcript isoforms with annotated open reading frame, ISAR was used combined with the 50 nucleotide (nt) rule of NMD (Nagy and Maquat 1998, Zhang, Sun et al. 1998). The results of the DTU analysis can be found in the Supplementary Table 3 from (Schlautmann, Lackmann et al. 2022).

7.6.3 Differential splicing analysis

For differential splicing analysis multiple tools were used. First, LeafCutter (version 0.2.9) (Li, Knowles et al. 2018) was used to identify alternative 3'/5' splice sites (A3SS/A5SS), exon skipping (ES) and exon inclusion (EI). Therefore, the significance thresholds |deltaPSI| > 0.1 and Padj < 0.001 were applied. Second, intron retention (IR) was computed with IRFinder (version 1.2.6, (Middleton, Gao et al. 2017)) in FastQ mode and differential IR was calculated using DESeq2 with the significance thresholds |log2FoldChange| > 1 and Padj < 0.001. Lastly, rMATS (version 4.1.1, (Shen, Park et al. 2014)) with novel splice site detection and the significance thresholds |deltaPSI| > 0.2 and Padj < 0.01 was used to identify alternative splicing classes. This was followed by analysis using maser (version 1.8.0). All cutoffs stated above were set as defaults but may differ on individual plots as indicated. The results of the splicing analyses can be found in the Supplementary tables 4, 5 and 6 from (Schlautmann, Lackmann et al. 2022).

7.7 Label-free Mass Spectrometry

7.7.1 Co-Immunoprecipitation

For Co-Immunoprecipitation, stable cell lines were seeded at a density of 1.5×10^6 cells per 10 cm dish and expression of FLAG-emGFP control or FLAG-emGFP tagged RNPS1 was induced with 30 µg ml^{-1} cumate. After 72 h, cells were lysed in 600 µl Buffer E (20 mM HEPES–KOH (pH 7.9), 100 mM KCl, 10% glycerol, 1 mM DTT, Protease Inhibitor) in the presence of 1 µg ml^{-1} RNase A. After sonication using the Bandelin Sonopuls mini20 with 10 pulses (2.5 mm tip, 1 s pulse, 50% amplitude), protein concentration of the lysates was determined using the Pierce Detergent Compatible Bradford Assay Reagent according to the manufacturer's protocol (Thermo Fisher Scientific). Protein concentrations were adjusted and 1 mg in 500 µl were loaded onto Anti-FLAG M2 Magnetic Beads (Sigma Aldrich). After a 2 h overhead shaking at 4°C, the beads were washed four times with mild wash buffer (20 mM HEPES-KOH (pH 7.9), 137 mM NaCl, 2 mM MgCl2, 0.2% Triton X-100, 0.1% NP-40), for 3 min at 4°C with overhead

shaking. Proteins were eluted from the beads using 2 × 21.5 µl (42.5 µl total) of a 200 mg ml^{-1} dilution of FLAG peptides (Sigma) in 1× TBS. If the samples were used for label-free mass spectrometry, 1 volume of 5% SDS in PBS was added, followed by reduction with DTT and alkylation with CAA (final concentrations 5 and 55 mM, respectively).

7.7.2 Proximity labeling

As described above for co-immunoprecipitation, for proximity labeling 1.5 × 10^6 stable cells were seeded per 10 cm dish. Background biotinylation was suppressed by changing the medium to medium containing dialyzed FBS instead of non-dialyzed FBS (Gibco; A3382001) after 72 h. Expression of TurboID or UltraID-tagged RNPS1 constructs was induced as described above using 30 µg ml^{-1} cumate. 24 h after medium change and expression induction, 50 µM biotin was added to the cells for 10 min to biotinylate proteins proximal to the Turbo/UltraID-tagged RNPS1. The cells were then washed two times with PBS and scraped in 1 ml ice-cold PBS, followed by 5 min centrifugation at 100 x g and 4°C. The samples were resuspended in 200 µl phospho-RIPA buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% IGEPAL CA 630, 0.5% deoxycholate, 0.1% SDS, 1 µg/ml RNase A; supplemented with 1 tablet of PhosSTOP (Roche), 100 µl EDTA-free HALT Protease & Phosphatase Inhibitor (Thermo Fisher Scientific) per 10 ml buffer) with subsequent sonication and Bradford protein concentration measurement as described above (7.7.1). 100 µg of protein in 50 µl were mixed with SDS-sample buffer and stored as input samples.

For mass spectrometry, 1 mg protein in 500 µl were loaded onto 0.5 ml Amicon Ultra centrifugal filter devices (3K cutoff) which were centrifuged for 45 min at 4°C and 14 000 × g, to get rid of excess biotin and concentrate the sample volume to approximately 100 µl. Next, the centrifugal filter was washed with 200 µl RIPA buffer which was combined with the concentrated sample and then added to 25 µl pre-washed Pierce Streptavidin Magnetic Beads (Thermo Fisher Scientific). Following a 2 h incubation step at 4°C with overhead shaking, the beads were washed four times with 800 µl RIPA buffer and one time with 800 µl mild wash buffer (20 mM HEPES–KOH (pH 7.9), 137 mM NaCl, 2mM MgCl2, 0.2% Triton X-100, 0.1% NP-40) for 5 min at 4°C with overhead shaking. Biotinylated proteins were eluted in 50 µl 5% SDS supplemented with 20 mM DTT and 3 mM biotin for 15 min at room temperature and 15 min at 96°C. This step was repeated using 25 µl and the two eluates were combined and after

addition of 8.5 µl of 400 mM CAA (to a final concentration of 40 mM) incubated at 55°C for 30 min to alkylate the samples. Lastly, the eluates were incubated in the dark for 30 min.

7.7.3 Label-free mass spectrometry

Both immunoprecipitated and proximity labeled samples were subjected to label-free mass spectrometry (MS). The following procedures were performed by the proteomics core facility at CECAD. A modified version of the single pot solid phase-enhanced sample preparation (SP3) protocol described by (Hughes, Foehr et al. 2014) was used for tryptic digestion of the proteins. The reduced and alkylated protein samples were supplemented with paramagnetic Sera-Mag speed beads (Cytiva), mixed in a 1:1-ratio with 100% acetonitrile (ACN) and incubated at room temperature for 8 min. Afterwards, the protein-beads-complexes were captured using a magnetic rack (in-house built), followed by two washing steps with 70% EtOH and one washing step with 100% ACN. The samples were air-dried and reconstituted in 5 µl 50 mM triethylammonium bicarbonate, which was supplemented with 0.5 µg trypsin and 0.5 µg LysC. Following an overnight incubation at 37°C, the beads were again mixed with 200 µl ACN. After 8 minutes of incubation at room temperature, the samples were placed on the magnetic rack, washed once with 100% ACN, air-dried and dissolved in 4% DMSO. Subsequently, samples were transferred to fresh PCR tubes and acidified with 1 µl of 10% formic acid.

The LC-MS/MS analysis was performed via data-dependent acquisition using an Easy nLC1200 ultra high-performance liquid chromatography (UHPLC) system connected via nano electrospray ionization to a Q Exactive Plus instrument (all Thermo Fisher Scientific) running in DDA Top10 mode. The tryptic peptides were first separated according to their hydrophobicity using a chromatographic gradient of 60 min with a binary system of buffer A (0.1% formic acid) and buffer B (80% ACN, 0.1% formic acid) with a total flow of 250 nl/min. In-house made analytical columns with a length of 50 cm, an inner diameter of 75 µm and filled with 2.7 µm C18 Poroshell EC120 beads (Agilent) were heated to 50°C in a column oven (Sonation) for separation. First, Buffer B was linearly increased from 3% to 30% in 41 min, followed by an increase to 50% in 8 min and a final increase to 95% within 1 min. The columns were washed with 95% buffer B for 10 min. Full MS spectra (300–1,750 m/z) were recorded with a resolution of 70,000, a maximum injection time of 20 ms and an AGC target of 3e6. In each full MS spectrum, the top 10 most abundant ions were selected for HCD fragmentation

(NCE 27) with a quadrupole isolation width of 1.8 m/z and 10 s dynamic exclusion. The MS/MS spectra were then measured with a 35,000 resolution, an injection time of maximum 110 ms and an AGC target of 5e5.

For analysis of MS RAW files, the MaxQuant suite version 1.5.3.8 was used on standard settings. Peptides were identified using the integrated Andromeda scoring algorithm (Cox, Neuhauser et al. 2011, Cox, Hein et al. 2014) that matches them to the human UniProt database (2021). Carbamidomethylation of cysteine was defined as a fixed modification, while methionine oxidation and N-terminal acetylation were variable modifications. The digestion protein was Trypsin/P. With a false discovery rate (FDR) of less than <0.01 peptide-spectrum matches were identified and proteins were quantified. To process the resulting data and for statistical analysis, the Perseus software (version 1.6.15.0) was used (Tyanova, Temu et al. 2016). For classification in R (version 4.1.2) the gene ontology biological process (GOBP) terms were obtained from Uniprot (2021) using the majority Protein ID. and subsequently analyzed for the terms 'splic', 'RNA processing', 'ribonucleoprotein|RNA binding', 'mRNA' to define relevant GOBPs.

The results of the MS analyses can be found in the Supplementary Tables 7 and 8 from (Schlautmann, Lackmann et al. 2022).

7.8 Protein modeling and data visualization

The structures of the ASAP complex (accession number 4A8X on PDB, (Murachelli, Ebert et al. 2012)) and the EJC (accession number 2J0S on PDB, (Bono, Ebert et al. 2006)) were visualized using the Chimera X (version 1.1). Different proteins are depicted in different colors as stated in the figure legend.

The performed experiments were – if not stated otherwise – calculated and quantified using either R (version 4.1.2) or Microsoft Excel (version 1808). Plots were generated using either IGV (version 2.8.12), Graph-Pad Prism 5, or the R packages ggplot2 (version 3.3.5), ComplexHeatmap (version 2.10.0), nVennR (version 0.2.3) and ggsashimi (version 1.0.0, (Garrido-Martín, Palumbo et al. 2018)). Dr. Volker Boehm helped with the illustration of the data using R.

8 RESULTS

Research over the last years has shown that the EJC-associated factor RNPS1 is involved in alternative splicing regulation in multiple species from flies to mammals as part of the ASAP or PSAP complex. Also, a potential role of RNPS1 in NMD activation was documented recently. In this work, further characterization of RNPS1 is carried out to define the molecular functions more precisely. Furthermore, the modes of action of RNPS1-dependent mechanisms will be investigated in more detail.

To address these questions, the results of this thesis can be grouped into three main parts. In the first two parts, RNA sequencing is used to determine the effect of RNPS1 depletion in human cultured cells on (1) the activity of NMD and (2) changes in alternative splicing patterns. To this end, RNA sequencing data were analyzed using various bioinformatic tools. Lastly, in the third part, protein interactors of RNPS1 are identified to establish potential molecular mechanisms of RNPS1's functions.

8.1 RNPS1 only mildly affects NMD

In several previous studies it was indicated that RNPS1 might be involved in the NMD process. For instance, RNPS1 overexpression was shown to decrease the amount of a β-globin NMD reporter (Viegas, Gehring et al. 2007). Additionally, when RNPS1 was artificially attached to an NMD reporter in tethering experiments, reduced expression of the NMD reporter was observed (Lykke-Andersen, Shu et al. 2001, Gehring, Kunz et al. 2005). Multiple endogenous genes that were upregulated in EIF4A3- and UPF1-depleted conditions were also found to be upregulated in the absence of RNPS1 in a more recent study (Mabin, Woodward et al. 2018). Although these studies suggest that RNPS1 can enhance NMD of specific endogenous targets or reporter mRNAs, it is not clear whether RNPS1 is generally essential for NMD. Therefore, this thesis aims to unravel how RNPS1 is involved in NMD and whether it acts on all NMD sensitive transcripts or rather a selected portion.

To investigate the function of a protein, a common strategy is to deplete this protein and observe potential changes in the cells. RNPS1 is listed as "common essential" in DepMap, which means that it is probably not possible to generate viable knockout (KO) cell lines. Therefore, here the method of siRNA-mediated knockdown (KD) was used to rid Flp-In-T-REx-293 (in short: 293) and HeLa Tet-Off (in short: HTO) cells of RNPS1. Since NMD is an mRNA

degradation pathway, RNPS1-depleted cells were subjected to RNA sequencing to estimate changes in gene expression on the transcriptome level (Figure 8A). Additionally, several rescue experiments were sequenced. First, a siRNA-insensitive, FLAG-emGFP-tagged RNPS1 full-length (FL) construct was overexpressed in RNPS1-depleted 293 cells. This rescue should reverse all effects that were specific to RNPS1 depletion, and any remaining effects are potentially due to unspecific side effects of the treatment. Next, RNPS1 KD was rescued by the overexpression of the likewise FLAG-emGFP-tagged RNPS1 RRM domain in 293 and HTO cells. This domain was chosen since it is sufficient and essential for ASAP and PSAP assembly and therefore interaction with the EJC core (Murachelli, Ebert et al. 2012, Wang, Ballut et al. 2018). If the effects of RNPS1 depletion are rescued by the RNPS1 RRM, this could indicate that assembly of the ASAP or PSAP complex is sufficient for functional RNPS1. Moreover, the RNPS1 RRM alone was previously shown to rescue specific alternative splicing events that appear upon RNPS1 KD, indicating that it might also be capable to execute some regulatory functions on its own (Boehm, Britto-Borges et al. 2018). The dataset of this previous study was mostly/only analyzed for alternative splicing changes, motivating the re-analysis with regard to RNPS1's NMD activity. Also, this study used another HeLa cell line, HeLa Flp-In-T-Rex (in short: HFT), and therefore provides information about differences and overlaps in similar cell lines. In the re-analyzed HFT dataset, RNPS1 KD was also rescued using a RNPS1 H176E and D179R (in short: 176) mutant. This mutation lies in the RRM domain and was shown to disrupt ASAP and PSAP assembly (Boehm, Britto-Borges et al. 2018). Thus, it will add to the results by showing whether this assembly is required to enable proper RNPS1 function. Although RNPS1 is hitherto the only ASAP/PSAP component that was indicated to be involved in NMD, ACIN1 and PNN are supposedly required to mediate RNPS1 interaction with the EJC (Wang, Ballut et al. 2018). Therefore, both factors were knocked down in HTO cells, either individually or in combination, to reveal how much the functions of RNPS1 rely on the formation of the ASAP or PSAP complex.

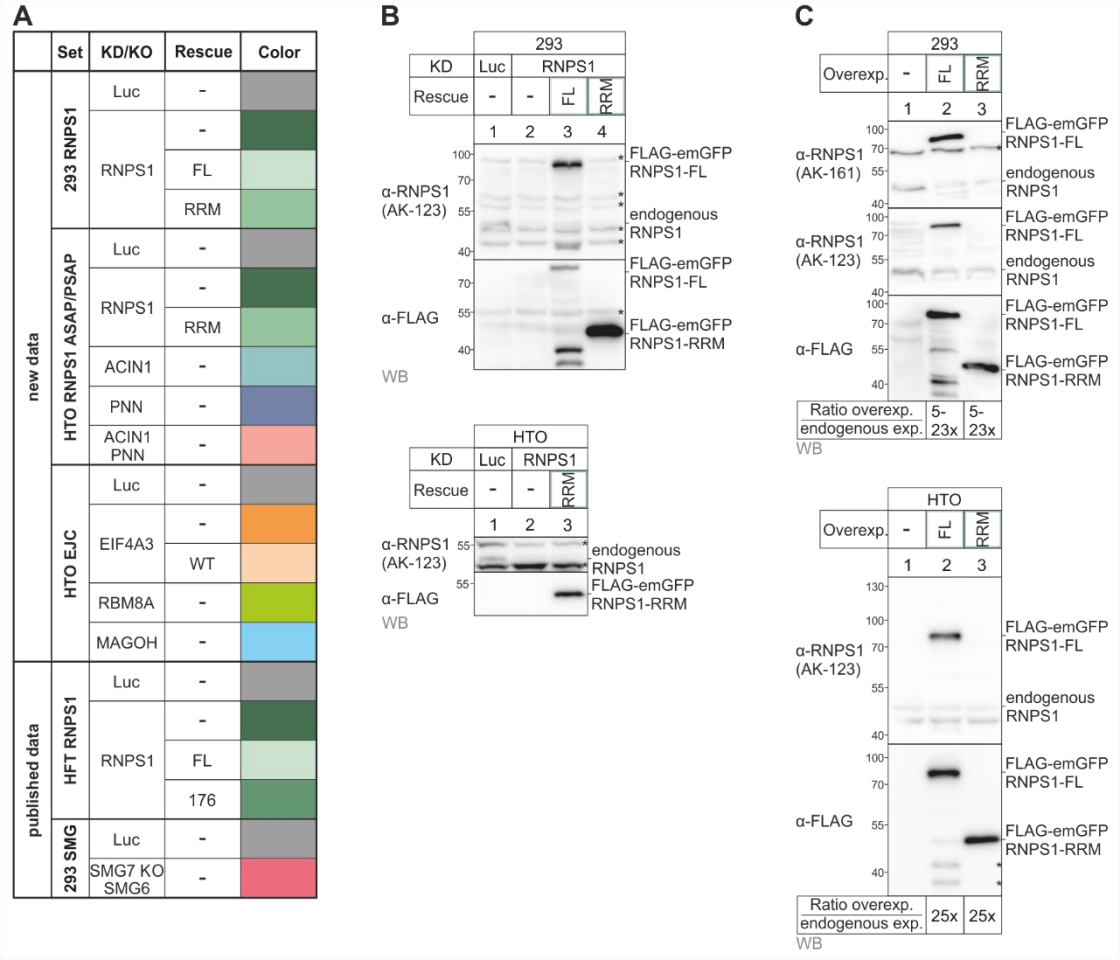


Figure 8: RNPS1 is efficiently depleted and overexpressed in the KD/KD rescue experiments. *(A) Overview of the analyzed new and already published RNA sequencing datasets. The used cell lines, siRNA-mediated knockdowns (KD) and rescues are indicated. The colors will be used throughout the thesis to depict the different conditions. (B) Western blots of RNPS1 in KD rescues in 293 and HTO cells shows that endogenous RNPS1 is robustly depleted and FLAG-emGFP-tagged RNPS1 FL or RNPS1 RRM rescue constructs are strongly expressed. (C) Overexpression of the RNPS1 rescue constructs in 293 and HTO cells was calculated from western blot expression bands.*

As a reference for strongly impaired NMD, the already published dataset of 293 SMG7 KO cells that were depleted of SMG6 by siRNA (SMG6/7 KD/KO) were used (Boehm, Kueckelmann et al. 2021). Lastly, depletions of EJC core factors in HTO cells were sequenced. The EJC interacts with RNPS1 via the ASAP or PSAP complex, which indicates that there might by a functional overlap. Additionally, the EJC core is known to stimulate NMD although it is not considered to be an essential NMD factor (reviewed in (Woodward, Mabin et al. 2017)). Thus, RNA sequencing of SMG6/7 KD/KO cells provides a reference for severely impaired NMD and RNA sequencing of EJC-depleted cells provides a measure for partially impaired NMD. As a negative control, cells were sequenced that were treated with firefly luciferase (Luc) siRNA. Since this

gene is not present in human cells, this siRNA treatment should not affect any endogenous transcripts.

Expression of the two RNPS1 rescue constructs, RNPS1 FL and RNPS1 RRM, was confirmed in western blots using anti-RNPS1 and anti-FLAG antibodies (Figure 8B, C). Both are overexpressed about 5-23x (depending on the antibody) in 293 and about 25x in HTO cells compared to the respective endogenous RNPS1 levels. This strong overexpression suggests that the constructs should be sufficiently expressed to reconstitute RNPS1 function (FL) or draw conclusions of RNPS1 RRM functionality.

8.1.1 Typical NMD targets remain largely unchanged upon RNPS1 depletion

The above-described datasets were analyzed in parallel for differential gene expression (DGE) using the DESeq2 tool which calculates gene up- or down regulation in a specific condition compared to the control condition (Luc KD). Severe NMD impairment would typically lead to less degradation of endogenous NMD targets and therefore result in a globally observable upregulation of these genes. The global expression changes were depicted in heatmaps for the individual RNA sequencing datasets. This displays the amount of up- and downregulated genes and also how much they change in the different conditions. Here, between 288 (HTO) and 656 (293) genes were upregulated upon RNPS1 depletion, which is comparable to the number of upregulated genes upon EJC core KDs (between 364 (EIF4A3) and 691 (RBM8A), Figure 9A-D). However, also a very similar portion of genes were downregulated upon RNPS1 and EJC core KDs, with RNPS1 KD in HTO cells having considerably more down- than upregulated genes (518 down, 288 up, Figure 9B). Interestingly, only 16 genes were commonly upregulated among the three RNPS1-depleted cell types, suggesting that most of the observed changes were cell type specific (Supplementary Figure 1A). Therefore, it was assumed that the expression levels and depletion efficiencies vary between the cell lines. When RNPS1 transcript expression in control cells was visualized as normalized counts, it was highest in 293 cells and lowest in HTO cells (Supplementary Figure 1B). Also, depletion of RNPS1 was the most effective in 293 cells with a log2 foldchange (log2 FC) of -2.79. This difference in basal RNPS1 expression and depletion efficiency might provide an explanation for the observed variations between the different cell lines.

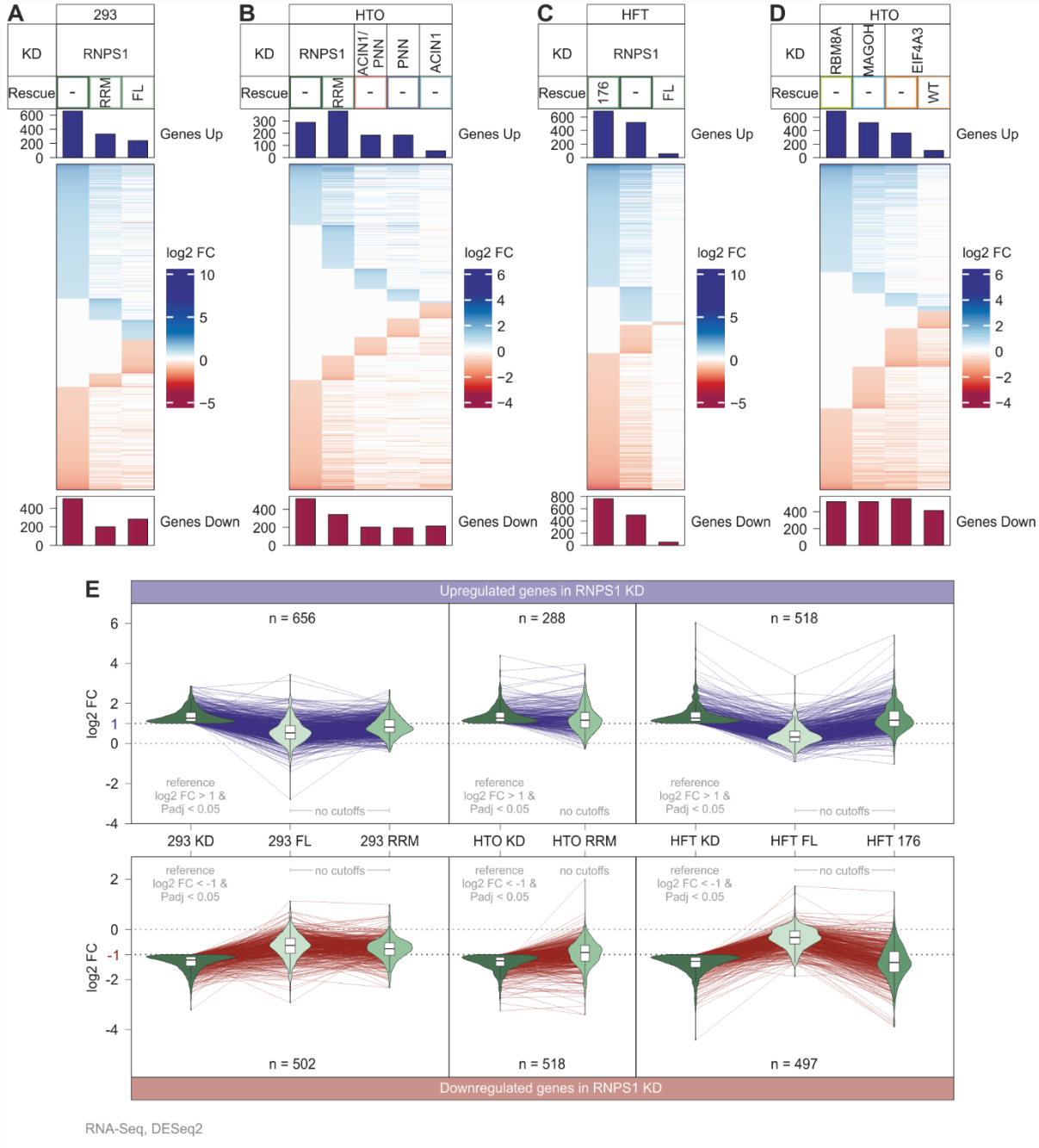


Figure 9: Gene expression changes globally upon depletion of RNPS1. *(A, B, C, D) DGE is depicted as the log2 FC compared to the control KD (Luc) with the cutoffs adjusted P-value (Padj) < 0.05 and |log2 FC| > 1. Total numbers of up- or downregulated genes is indicated as bar graphs on the top or bottom, respectively. (A) 293 RNPS1 set, (B) HTO RNPS1 ASAP/PSAP set, (C) HFT RNPS1 set, (D) HTO EJC set. (E) Genes up- or downregulated with a |log2 FC| > 1 and Padj < 0.05 upon RNPS1 KD are depicted in combined violin and parallel coordinate plots. The log2 FCs of the same genes (that are found in the RNPS1 KD condition) are depicted for the RNPS1 KD rescue conditions (no cutoffs were applied to the rescue conditions).*

Notably, the rescue capability of the RNPS1 RRM also varied strongly among different cell types. In HTO cells, RNPS1 RRM rescue induced upregulation of even more genes than only RNPS1 KD, which could imply that RNPS1 RRM overexpression might have a dominant negative effect on NMD activation (Figure 9B and E). However, overexpression of the RNPS1

RRM in 293 cells effectively reduced the number of upregulated genes, arguing against this hypothesis (Figure 9A and E). This strengthens the hypothesis established above that many gene upregulation events upon RNPS1 depletion were strongly cell type specific.

In 293 and HFT cells, the overexpression of RNPS1 FL restored normal expression of most genes that were upregulated upon RNPS1 depletion. This indicates that the observed upregulations were indeed caused by a lack of RNPS1. While the rescue with RNPS1 FL was quite robust, introducing the 176 mutant to full-length RNPS1 completely disrupted this rescue ability (Figure 9C). In fact, rescue with the RNPS1 176 mutant even increased the number of upregulated genes compared to the RNPS1 KD only. Since the RNPS1 176 mutation abolishes the assembly of the ASAP and PSAP complex, these findings hinted that either this assembly or the ASAP/PSAP mediated EJC interaction might be important for RNPS1-dependent gene upregulation. Contrary, depletion of either ACIN1, or PNN or both, only induced upregulation of fewer than 200 genes (Figure 9B). Also, roughly the same number of genes, if not even more, were downregulated in these conditions. Therefore, there is lacking evidence that ACIN1 or PNN are directly involved in NMD, and deeper investigation would be required to obtain more conclusive results.

A high number of upregulated genes can indicate reduced NMD activity but can also result from other dysregulations in gene expression. Therefore, the overlaps of upregulated genes upon RNPS1 and EJC core KDs with the SMG6/7 KD/KO were investigated (Figure 10A, B). Upon combined depletion of the two NMD factors SMG6 and SMG7, NMD is largely inhibited (Boehm, Kueckelmann et al. 2021). Thus, gene upregulation in this condition is a strong indicator that the afffected genes are targeted by the NMD pathway. Here, a vast upregulation of 2433 genes was observed, which is nearly 4-times as much as measured in the strongest RNPS1-depleted condition (293 cells, 656 upregulated genes). The overlap of RNPS1 KD with SMG6/7 KD/KO showed broad variation depending on the cell type. While 167 genes were consistently upregulated in SMG6/7 KD/KO and RNPS1 KD in 293 cells, the overlap in HeLa cells was considerably lower (67 in HFT and 45 in HTO). The fact that the overlap of RNPS1 KD in 293 cell with the SMG6/7 KD/KO was the strongest appears reasonable, since in these cases the same cell line was used. In addition, the depletion of RNPS1 was most efficient in 293 cells, which might have pronounced the effect.

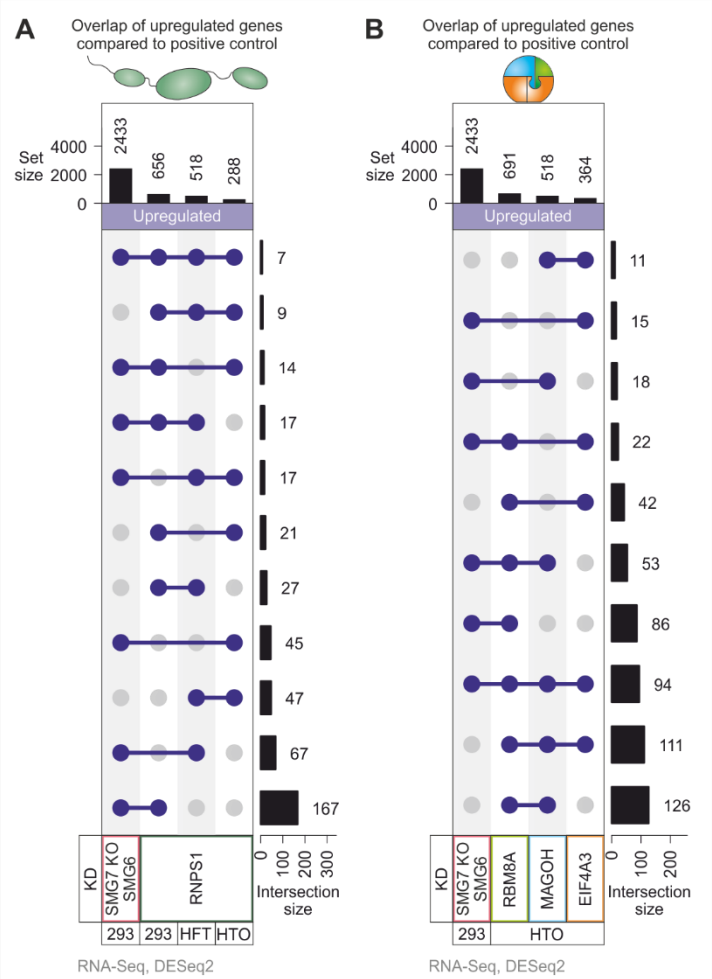


Figure 10: Genes upregulated upon RNPS1 KD overlap with genes upregulated in SMG6/7 KD/KO, but weakly among the different cell types. *(A) Top 15 overlaps of the upregulated genes in the different RNPS1 KD conditions and SMG6/7 KD KO (positive control) are depicted in an UpSet-plot (Cutoffs: Padj < 0.05 and |log2 FC| > 1. (B) Same as in (A), but with EJC KDs in HTO cells.*

In contrast to the RNPS1 KDs, which shared only 16 commonly upregulated genes, the different EJC core depletions shared a high number of upregulated genes (205; 111+94). Furthermore, these commonly upregulated genes overlapped strongly with the SMG6/7 KD/KO. This was especially noteworthy, since the EJC KDs were performed in HTO cells, while the SMG6/7 KD/KO was performed in 293 cells. In case of RNPS1 depletion, using a different cell line strongly reduced the overlap with SMG6/7 KD/KO, but for EJC depletions the overlap was still robust in a different cell type. Potentially, EJC KDs in 293 cells would accordingly overlap even more with the SMG6/7 KD/KO.

The high amount of commonly upregulated genes in EJC core KDs and SMG6/7 KD/KO underlines the central role of these factors in the activation of the NMD pathway. Similarly, the relatively high number of upregulated genes upon RNPS1 depletion in 293 cells that were

shared with SMG6/7 KD/KO might indicate that RNPS1 indeed plays a role in mRNA degradation via NMD. However, the striking differences and low overlap between the RNPS1 KDs in the three used cell types suggests that the observed gene upregulations are widely cell type specific.

These cell type-specific variations complicate the development of a clear statement regarding the involvement of RNPS1 in NMD. Thus, instead of a global approach, the DGE of individual bona fide NMD targets was analyzed in more detail. Among the typical endogenous NMD targets are the small nucleolar RNA (snoRNA) host genes GAS5 and ZFAS1 (Lykke-Andersen, Chen et al. 2014). Plotting of the log2 FC of these targets shows a clear upregulation upon depletion of either EJC core factors or SMG6 and SMG7 (Figure 11A). While RNPS1 depletion did not significantly increase the expression of GAS5, ZFAS1 was mildly upregulated in HTO cells. Still, also in this case the log2 FC was lower than 1 while EJC core depletions or SMG6/7 KD/KO increased expression of ZFAS1 with a log2 FC above 2. The same trend was observed when other known NMD targets of different classes (Regular EJC-dependent, uORF, long 3' UTR) were analyzed (Figure 11B). None of the depicted NMD targets was upregulated with a log2 FC above 1 in any of the RNPS1 KD conditions, but in nearly all EJC KD and SMG6/7 KD/KO conditions. Additionally, a number of putative RNPS1-dependent NMD targets that were identified in 293 cells by a previous study was investigated (Mabin, Woodward et al. 2018). While none of these targets was considerably upregulated upon RNPS1 or SMG6 and SMG7 depletion, two of them were upregulated upon EJC core depletion (ARC, EIF4A2). In addition to the gene upregulation, the -log10 meta_meta scores for the selected genes are depicted. The meta_meta score was calculated by Colombo et al. using various NMD factor KD and rescue conditions and serves as a "confidence" measure of how likely a specific gene is targeted by NMD (Colombo, Karousis et al. 2017). For the potentially RNPS1-dependent NMD targets, the -log10 meta_meta scores were comparably low, while they were much higher for the bona fide NMD targets. This suggests that the potentially RNPS1-dependent NMD targets are probably not usually degraded by NMD. Taken together, these results show that there is little evidence for an essential role of RNPS1 in the NMD pathway.

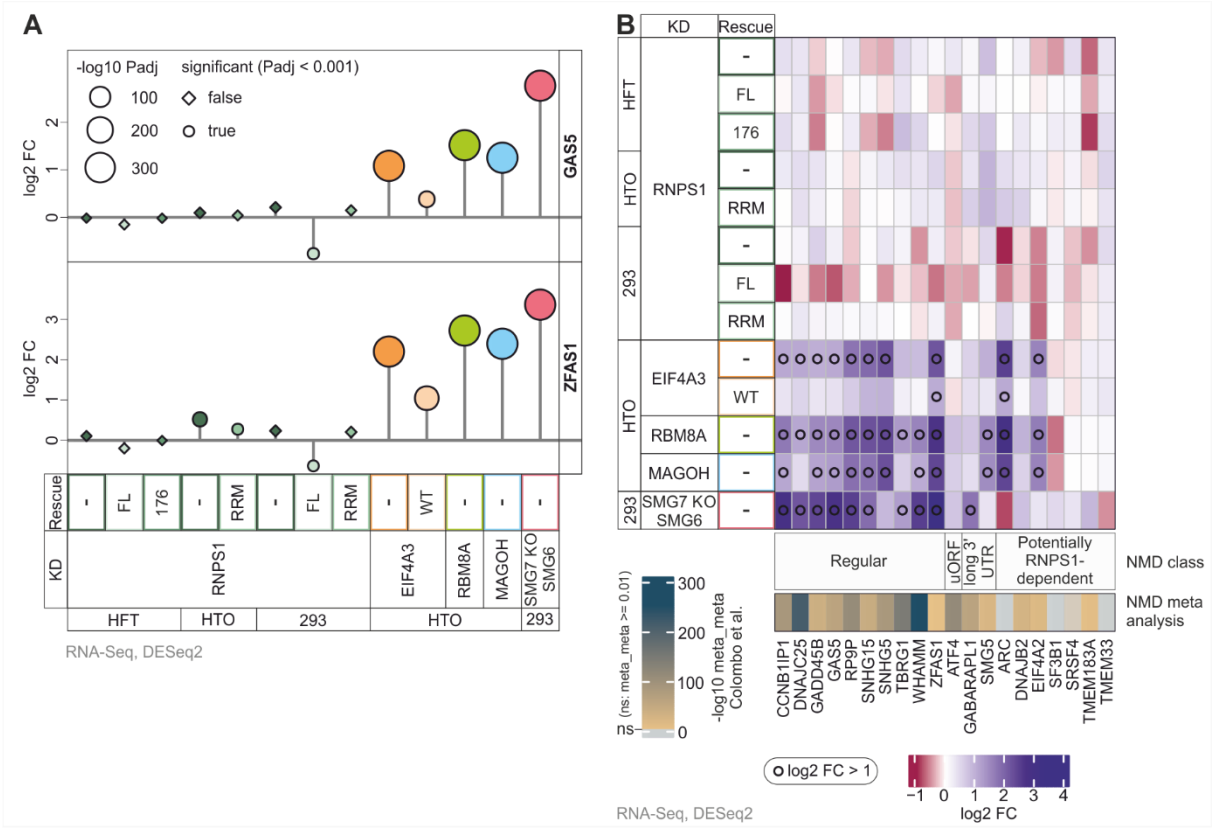


Figure 11: RNPS1 overexpression can decrease the expression of selected NMD targets. *(A) The log2 FCs resulting from the DGE analysis of the snoRNA host genes GAS5 and ZFAS1 is shown in a lollipop-plot. Significant (Padj < 0.001) values are depicted as circles, non-significant values as squares and the size depicts the -log10 Padj. (B) The heatmap shows the log2 FCs of the selected bona fide NMD targets, which are classified as "Regular", "uORF" (upstream open reading frame), or "long 3' UTR". Additionally, a number of "Potentially RNPS1-dependent" NMD targets is shown. Expression changes with a log2 FC > 1 are indicated with a black circle. The values of the NMD meta-analysis indicate the likelihood of the chosen genes to be affected by NMD (Colombo, Karousis et al. 2017).*

Intriguingly, while expression of the RNPS1 RRM had nearly no effect on DGE of the chosen NMD targets, RNPS1 FL overexpression notably downregulated several NMD targets, including GAS5 and ZFAS1, in 293 cells and showed the same trend in HFT cells (Figure 11). Also, in 293 cells the number of downregulated genes was higher in the RNPS1 FL rescue compared to the RNPS1 RRM rescue (Figure 9A). This could indicate that overexpression of RNPS1 can increase NMD activity in 293 cells, which matches older findings, where overexpression of RNPS1 increased NMD efficiency (Viegas, Gehring et al. 2007).

Altogether, the DGE analysis revealed that RNPS1 depletion upregulates the expression of several genes, but that this upregulation is not very consistent among cell types and experiments. Furthermore, RNPS1 seems to not play a general role in NMD, as judged by the analysis of known NMD targets. At the same time, overexpression of RNPS1 was able to downregulate expression of multiple specific genes, indicating that RNPS can act as an NMD

enhancer. To consolidate these hypothetical mechanisms of RNPS1 action, further analyses are needed.

8.1.2 Upon RNPS1 depletion only few NMD-sensitive transcript isoforms are upregulated

The global upregulation of gene expression is a first indicator for general NMD impairment. However, hampered NMD can also have other outcomes, like a dysregulation in transcript usage. Oftentimes, NMD-sensitive (e.g. bearing a PTC) and -insensitive transcript isoforms are produced from the same gene. In such cases, disruption of the NMD pathway would prevent the decay of NMD-sensitive isoforms and therefore increase their expression. When the expression of NMD-insensitive isoforms is decreased at the same time, these effects can cancel each other out on the gene level and would not be observed in standard DGE analyses.

In order to investigate whether upon RNPS1 KD more NMD-sensitive isoforms remain in the cell compared to the control condition, the RNA sequencing datasets were analyzed for differential transcript usage (DTU) using the Isoform Switch AnalyzeR (ISAR, (Vitting-Seerup and Sandelin 2019)). Different from DESeq2, ISAR calculates the delta Isoform Fraction (dIF), which provides a measure of how much a transcript isoform changes in a specific condition compared to the control condition. To identify expression of an isoform, ISAR makes use of the GENCODE (release 33) annotations. Together with the 50 nt rule, these annotations allow ISAR to distinguish between NMD-sensitive and -insensitive isoforms (Nagy and Maquat 1998, Zhang, Sun et al. 1998).

ISAR analysis revealed that upon RNPS1 depletion, 42, 48 and 107 PTC-containing isoforms were differentially expressed in HFT, HTO and 293 cells, respectively (Figure 12A). The RNPS1 FL construct restored normal expression for most of these transcripts, indicating that the observed effects are RNPS1-specific. Expression of the ASAP/PSAP assembly-deficient RNPS1 176 mutant did not reduce the number of differentially expressed transcripts substantially, whereas overexpression of the RNPS1 RRM had a considerable effect. This shows that RNPS1 controls the expression of these transcript isoforms (at least in part) via its interaction with the ASAP or PSAP complex and the EJC.

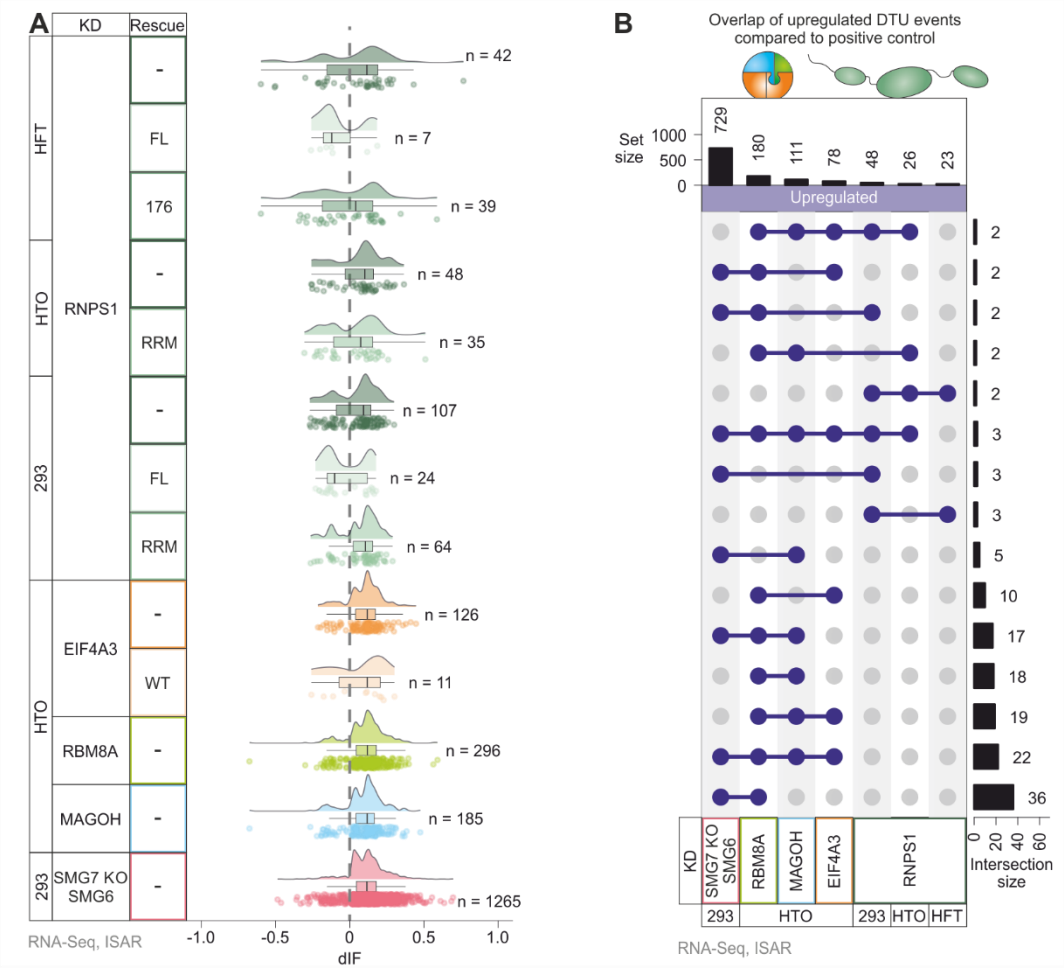


Figure 12: RNPS1 depletion leads to an upregulation of PTC-positive transcript isoforms that are rarely also upregulated in the SMG6/7 KD/KO condition. *(A) The DTU analysis is presented as a raincloud plot that shows the dIF values of the individual transcripts as well as their overall distribution (Cutoff: Padj < 0.001). (B) The UpSet plot depicts the shared upregulated, PTC-positive transcript numbers in the indicated conditions (Cutoffs: Padj < 0.001 and dIF > 0.1)*

EJC core depletions resulted in 126 (EIF4A3) to 296 (RBM8A) differentially expressed NMD sensitive isoforms. While RNPS1 as well as EJC depletion led to more upregulated (dIF > 0) than downregulated (dIF < 0) isoforms, this effect was more pronounced upon EJC depletion (Figure 12A, see density plots and boxplots). As anticipated, depletion of SMG6 and SMG7 again displayed the strongest effect, with more than one thousand differentially expressed transcripts, of which most were upregulated.

As in the DGE analysis, the SMG6/7 KD/KO provides a good measure to identify high-confidence NMD-sensitive isoforms. Therefore, especially the transcripts upregulated upon SMG6/7 KD/KO should be analyzed in the other KD conditions. Investigation of the overlap of upregulated transcripts among the different conditions revealed that the upregulated transcripts found upon RNPS1 KD overlap only mildly with the ones upregulated in the

SMG6/7 KD/KO (Figure 12B). Depletions of the EJC core components yielded noticeably more transcripts that were likewise upregulated in the SMG6/7 KD/KO. Thus, the EJC core factors seem to be more important for the efficient degradation of NMD-sensitive isoforms than RNPS1.

Additionally, the dIF values of shared differentially expressed transcripts were plotted. The RBM8A KD not only shared more differentially expressed transcripts with the SMG6/7 KD/KO but the dIF values of these also correlated much better than with the RNPS1 KDs (Figure 13). Therefore, although the DTU analysis of RNPS1 and SMG6/7 depletion yielded some shared NMD targets, these are not necessarily affected in the same way or to the same extent by the depletion of RNPS1 and the two NMD core factors.

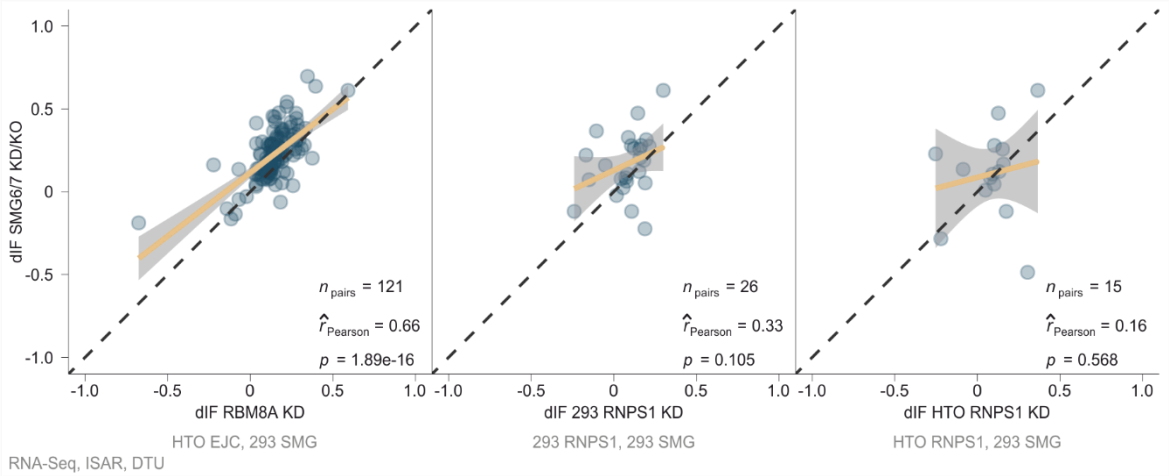


Figure 13: RNPS1 KDs share few differentially expressed transcripts with SMG6/7 KD/KO which do not correlate well. *The dIF of SMG6/7 KD/KO differentially expressed transcripts is plotted against the dIF of the shared transcripts in either RMB8A KD, RNPS1 KD in 293 cells or RNPS1 KD in HTO cells, as indicated (Cutoff: Padj < 0.001). npairs is the number of transcripts, rPearson indicates the Pearson correlation coefficient and p is the corresponding p-value.*

The overlap of upregulated NMD sensitive transcripts that were detected in the SMG6/7 KD/KO with differentially expressed transcripts upon RNPS1 KD in 293 cells was as low as 16 (and even lower in the other cell types). When these 16 genes were manually analyzed, 7 genes indeed showed upregulation of the NMD-sensitive isoform upon RNPS1 KD (Supplementary dataset in (Schlautmann, Lackmann et al. 2022)). Of the remaining 9 genes, some showed downregulation of the NMD-sensitive isoform (PXMP2, PTPMT1, DNAJC2), some showed gene downregulation but unchanged NMD-sensitive isoform levels (H2AX, ASNS) and some showed even incorrectly annotated isoforms (FAM234B, ATG5, TMEM248, HSPA4) (Supplementary Figure 2A, Supplementary Dataset in (Schlautmann, Lackmann et al.

2022)). Interestingly, multiple of these incorrectly annotated isoforms appear to be caused by alternative splicing events that are not yet annotated in GENCODE. This is observed in case of FAM234B: in the NMD-sensitive isoform, an intron is spliced in the 3' UTR, leading to EJC deposition and identification of the normal stop codon as a PTC (Supplementary Figure 2B). Upon RNPS1 depletion however, in addition to the 3' UTR intron, the preceding exon and the first part of the 3' UTR are skipped. Therefore, in this isoform the stop codon is located in the last part of the 3' UTR. This shift of the stop codon places it behind all splicing events, meaning that no NMD-inducing EJCs will be deposited downstream of it. Thus, this transcript isoform cannot be a target of EJC-dependent NMD. As this isoform is not annotated in GENCODE and lacks reads in the 3' UTR intron that is also spliced in the NMD-sensitive isoform, ISAR erroneously assigns reads of this isoform to the NMD-sensitive transcript.

Overall, RNPS1 increased the expression of relatively few genes and NMD-sensitive isoforms (as defined by upregulation upon SMG6/7 KD/KO). Still, some NMD-sensitive isoforms were upregulated to a low extent and RNPS1 FL overexpression decreased the expression of multiple genes. Taken together, this reinforces the hypothesis that RNPS1 is not a general but rather a specific NMD factor that can enhance NMD when overexpressed. Notably, the incorrect identification of several isoforms by ISAR suggests that RNPS1 regulates some thus far unknown or unannotated alternative splicing events.

8.2 RNPS1 depletion affects multiple types of alternative splicing

Though the above-described results allow the conclusion that RNPS1 plays a rather minor role in NMD, it was frequently found to be involved in splicing regulation in multiple species. Hitherto, RNPS1 was shown to be involved in preventing retention of PIWI intron 4 and exon skipping in the MAPK gene in *Drosophila* (Ashton-Beaucage, Udell et al. 2010, Roignant and Treisman 2010, Hayashi, Handler et al. 2014, Malone, Mestdagh et al. 2014). Recently, a mutation in the RNPS1 RRM was demonstrated to increase the number of exon skipping (ES) and intron retention (IR) events in mice (Zhong, Choi et al. 2022). In human cells, RNPS1 can suppress the usage of reconstituted 5' splice sites (Blazquez, Emmett et al. 2018, Boehm, Britto-Borges et al. 2018). However, most of these studies focus on the analysis of one specific target or one specific alternative splicing type. Hence, the effect of RNPS1 on alternative splicing was globally examined in this thesis. To this end, different bioinformatic tools were used to identify alternative splicing in the initially described RNA sequencing datasets. In the

NMD analyses, depletions of the ASAP/PSAP components ACIN1 and PNN were mostly excluded since they are not known for NMD-regulatory functions which was supported by the findings of the initial DGE analysis. In the alternative splicing analyses however, the effects of their depletions were investigated since both ACIN1 and PNN were implicated to have splicing regulatory abilities. The RNPS1 KD and rescue set in HFT cells was already analyzed for alternative splicing in the study it was published in and was therefore excluded from the following analyses (Boehm, Britto-Borges et al. 2018).

8.2.1 The RNPS1 RRM cannot generally rescue all alternative splicing defects resulting from RNPS1 depletion

Initially, alternative splice sites (A3SS/A5SS), ES and exon inclusion (EI) events were detected using the intron-centric LeafCutter tool. LeafCutter first aggregates splicing events with overlapping splice sites to obtain splice clusters. Then LeafCutter calculates the delta Percent Spliced In (dPSI) value by comparing how much a specific splice junction in one splice cluster is used compared to the control condition (Li, Knowles et al. 2018). Therefore, for a given splice cluster, usually at least one splice junction with a positive and one with a negative dPSI value are found, for instance because the canonical splice site is used less, while another splice site is used more.

Upon RNPS1 depletion, LeafCutter detected about 300 alternative splicing events in HTO cells and about 500 in 293 cells (Figure 14A, B, C). This difference might again be explained by the normally higher expression and more effective depletion of RNPS1 in 293 cells. As expected, most of the events that occur upon RNPS1 depletion were rescued by RNPS1 FL completely or at least to a great extent. The RNPS1 RRM on the other hand rescued the alternative splicing resulting from RNPS1 depletion surprisingly poorly, which can be nicely seen in the violin and parallel coordinate plots (Figure 14A). This plot depicts how the alternative splicing events that are up- or downregulated upon RNPS1 depletion behave in the RNPS1 FL and RNPS1 RRM rescues. The comparably weak rescue by the RNPS1 RRM was unexpected, since this domain was previously shown to be sufficient to rescue other alternative splicing events (Boehm, Britto-Borges et al. 2018).

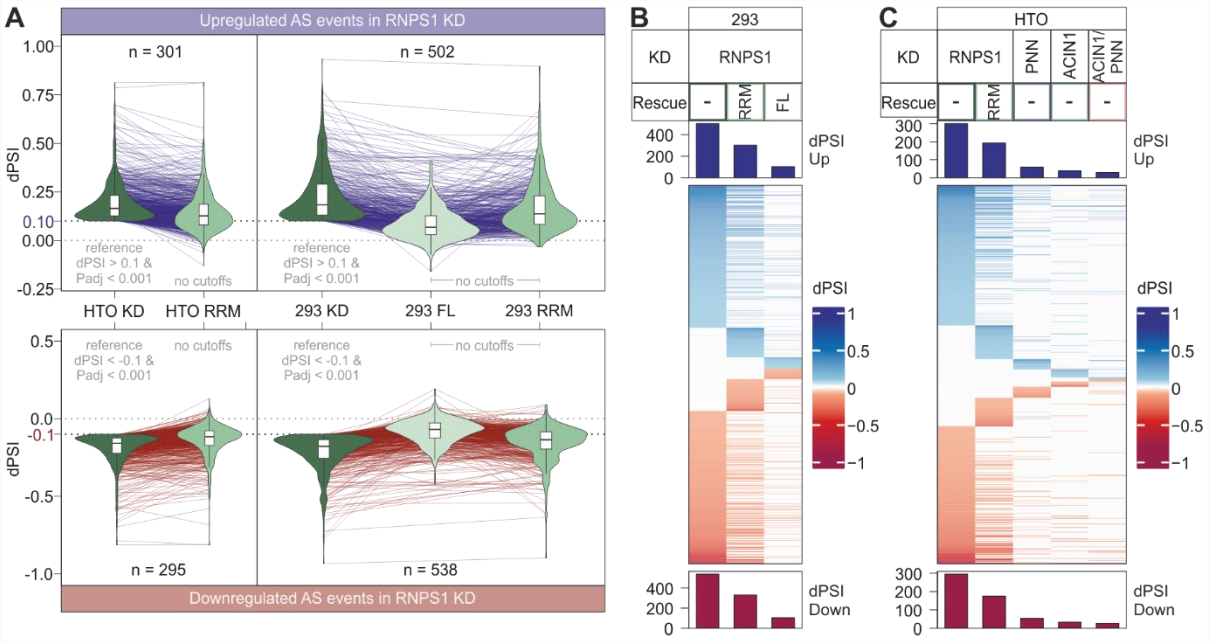


Figure 14: RNPS1 depletion induces many alternative splicing events that are only in part rescued by expression of the RNPS1 RRM. *(A) dPSIs as calculated by LeafCutter are represented in a combination of violin and parallel coordinate plots. The events are filtered for |dPSI| > 0.1 and Padj < 0.001 in the RNPS1 KD condition. No cutoffs were applied to the rescue conditions. (B, C) Heatmaps depicting up- and downregulated splicing events with the cutoffs |dPSI| > 0.1 and Padj < 0.001. The bar graphs above and below indicate the numbers of up- or downregulated events for the conditions. (B) 293 RNPS1 set, (C) HTO RNPS1 ASAP/PSAP set.*

To verify that the RNPS1 RRM indeed does not rescue all alternative splicing events and that this observation was not due to bioinformatic limitations, two RNPS1-dependent alternative splicing events were inspected in more detail. First, RER1 was used, in which an A5SS is used upon RNPS1 KD, that leads to the loss of a complete exon except for one single base. This event was described before to be rescued by expression of the RNPS1 RRM (Boehm, Britto-Borges et al. 2018). RT- and qRT-PCR analyses could confirm that RER1 alternative splicing was rescued by RNPS1 RRM overexpression in both 293 and HTO cells (Figure 15A). Second, A5SS usage of FDPS was examined. This event was identified in the new RNA sequencing datasets and seemed to not be rescued by the RNPS1 RRM. Again, this result could be reproduced in the PCRs (Figure 15B). RT-PCR of two other targets (INTS3 and TAF15) demonstrated that RER1 and FDPS are not just outliers but that the RNPS1 RRM can indeed rescue some events, while others are not rescued at all (Supplementary Figure 3A, B). Both INTS3 and TAF15 exhibited alternative 5’ splicing in the absence of RNPS1 that was only rescued by RNPS1 RRM overexpression in case of INTS3 but not TAF15. Noteworthy, also in the PCR experiments the variability between the cell lines was observed. Overall, the rescue efficiency of the RNPS1 RRM was discernibly better in 293 cells compared to HTO cells.

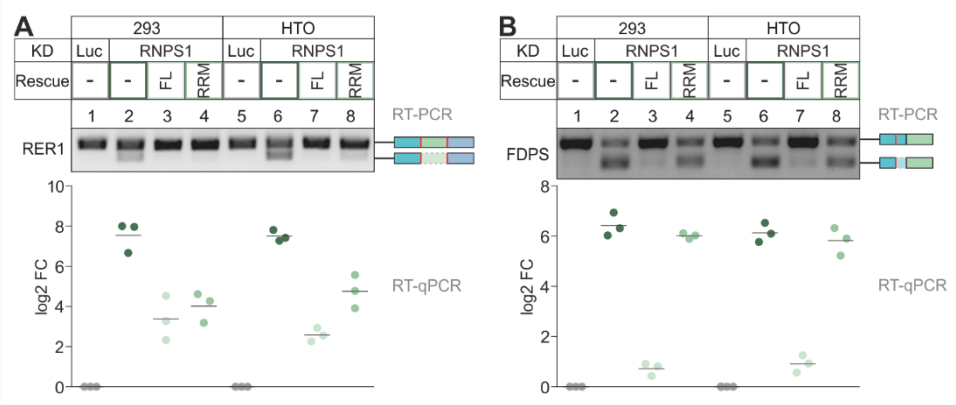


Figure 15: Overexpresion of the RNPS1 RRM rescues many, but not all RNPS1-dependent alternative splicing events. *(A, B) RT-PCR and RT-qPCR of alternatively spliced transcripts of (A) RER1 and (B) FDPS in the indicated conditions (n = 3). On the right, the detected transcripts are depicted.*

Albeit not all RNPS1-dependent alternative splicing events were rescued by the RNPS1 RRM, it still rescued a significant portion of events. RNPS1 is supposedly recruited to the EJC as a component of the ASAP or PSAP complex, which require the RNPS1 RRM for their assembly (Murachelli, Ebert et al. 2012, Wang, Ballut et al. 2018). To examine the importance of ASAP/PSAP assembly and EJC interaction for RNPS1's function as a splicing regulator, the effects of ACIN1 or PNN depletions were investigated. Interestingly, neither depletion of PNN, nor ACIN1 increased the amount of alternatively spliced transcripts considerably (Figure 14C). This could be due to some redundancy of the two complexes, as both can in principle recruit RNPS1 to the EJC. Otherwise, technical limitations, like an inefficient depletion of ACIN1 and PNN could be a reason for the observed low effects. Furthermore, the overlap of the upregulated events in PNN or ACIN1 KDs with the RNPS1 KD was relatively low (31 for PNN, 18 for ACIN1, Supplementary Figure 4A). Although previous studies suggested that the two complexes cannot act redundantly, ACIN1 and PNN might still be able to replace each under certain conditions (Wang, Ballut et al. 2018). In order to test this hypothesis, both factors were depleted simultaneously. Theoretically, this should abolish the interaction of RNPS1 with the EJC, which is supposedly a prerequisite for RNPS1's splicing regulatory function. Unexpectedly, in the double KD, even less alternative splicing events were found compared to the individual KDs (27 up/29 down upon the double KD compared to 33 up/39 down upon ACIN1 KD and 54 up/58 down upon PNN KD). These also showed weak overlap with the events found upon RNPS1 depletion (Figure 14C, Supplementary Figure 4A). Technically, the double KD condition was hard to achieve, since oftentimes the treated cells were dying excessively. Furthermore, examination of the expression levels of both ACIN1 and PNN mRNAs in the different KD

conditions revealed that the depletion of PNN was less efficient in the double KD compared to the PNN-only KD (log2 FC of -1.59 in PNN KD compared to -0.62 in the double KD, Supplementary Figure 4B, C). It appears, that a complete depletion of both ACIN1 and PNN simultaneously might not be achievable by using only siRNA KD. This suggests that the higher residual levels of PNN in the double KD are sufficient to prevent most alternative splicing events and that cells depleted of ACIN1 and PNN would probably not be viable.

8.2.2 RNPS1 depletion widely increases intron retention

The analysis of the RNA sequencing datasets with LeafCutter revealed that RNPS1 depletion induces many alternative splicing events, of which about half are rescued by expression of the RNPS1 RRM. However, LeafCutter detects only A3SS, A5SS, ES and EI events, but not IR events. Thus, to get more comprehensive results, the RNA sequencing datasets were subjected to the IRFinder algorithm which specifically detects IR events (Middleton, Gao et al. 2017).

Interestingly, with IRFinder, over 600 IR events were found in HTO cells, while in 293 cells only 243 events were detected (Figure 16A). Compared to the results of LeafCutter and the different NMD analyses where more events were found in 293 cells, here, HTO cells appear to be more sensitive to RNPS1 depletion. This might be an indicator that NMD regulation as well as the suppression of undesired splicing by RNPS1 has a different underlying mechanism than IR prevention by RNPS1. As in the LeafCutter analysis, depletions of ACIN1, PNN or a combined depletion resulted in very few IR events (37, 25 and 43 events, respectively). Although the double KD induced the most IR events, the differences are still too small to draw insightful conclusions. Again, the minor effects of the ACIN1 and PNN depletions could represent a potential redundancy of the ASAP and PSAP complexes.

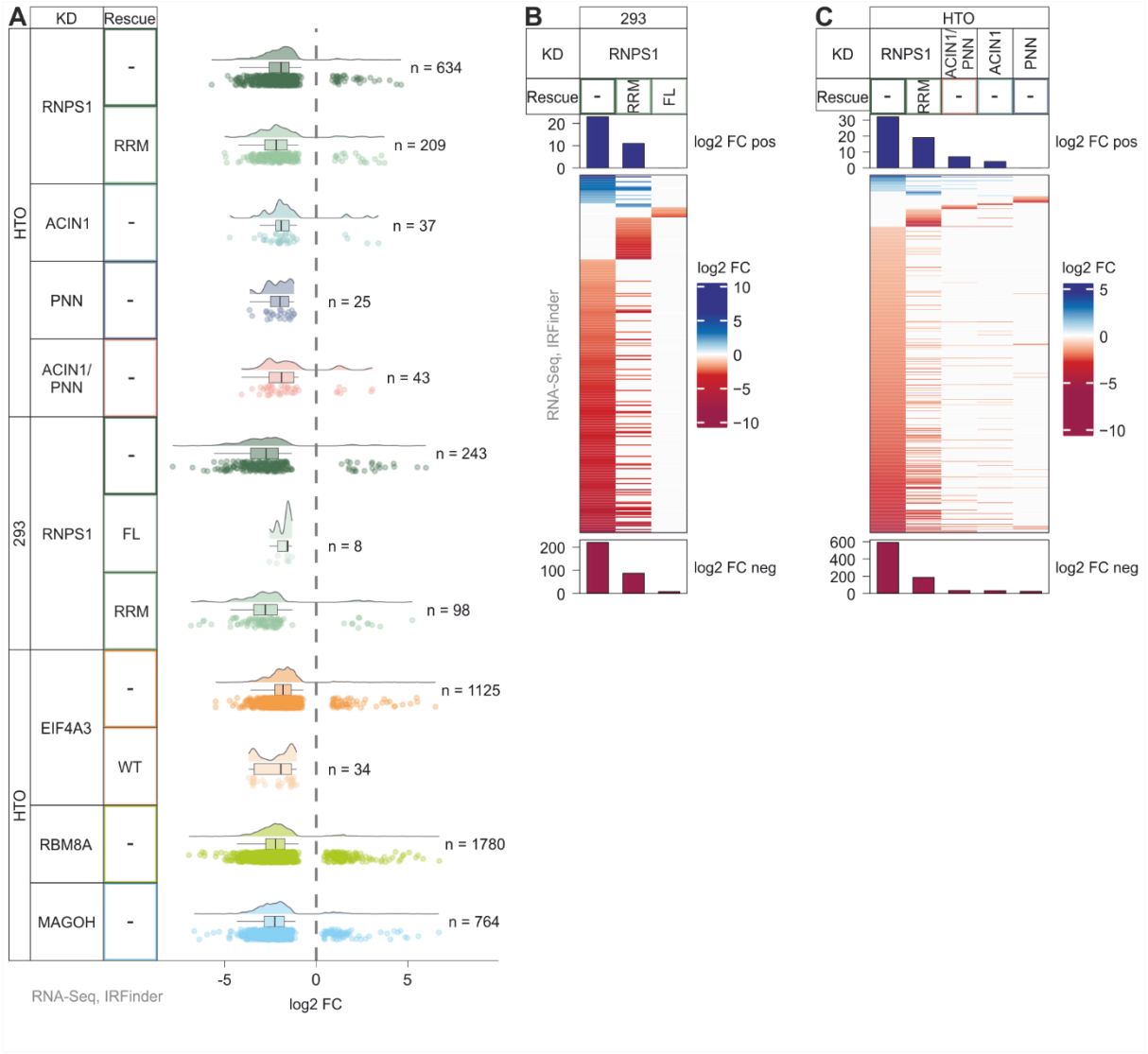


Figure 16: RNPS1 depletion leads to a strong upregulation of IR events. *(A) The raincloud plot depicts the individual log2 FCs of the IR events as well as their distribution in the different KD and KD rescue conditions compared to the control condition (Cutoffs: |log2 FC| > 1 and Padj < 0.001). A negative log2 FC indicates increased retention of an intron, a positive value indicates decreased inclusion of an intron. (B,C) The log2 FCs of IR events in the indicated conditions are shown in a heatmap for the (B) 293 RNPS1 set and (C) HTO RNPS1 ASAP/PSAP set, with the same cutoffs as in (A).*

Another difference between IR and the other analyzed alternative splicing types is that IR events were seemingly better rescued by RNPS1 RRM expression. Compared to the rescue with RNPS1 FL, the RNPS1 RRM rescue was still not complete, but in both 293 and HTO cells more than half of the events were rescued (Figure 16B, C). To confirm that also in case of IR the RNPS1 RRM rescued some events but not others, two IR events were depicted in so-called sashimi plots. These plots can be used to display RNA sequencing reads that span an exon-exon junction and therefore provide insight into the splicing events that took place in the mRNA. Here, retention of RFX5 intron 9 for example was not rescued by overexpression of the RNPS1 RRM, while the IR event in INTS2 was completely abolished in the same condition

(Figure 17). Thus, these events confirmed again that the RNPS1 RRM rescue of alternative splicing events is incomplete.

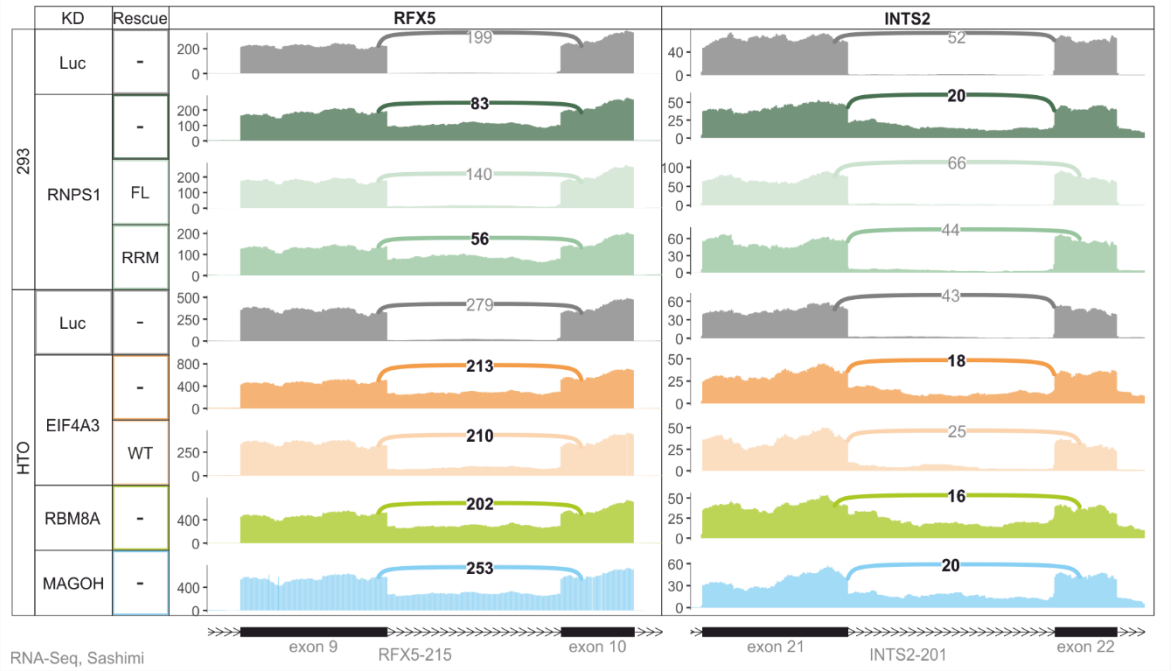


Figure 17: Not all IR events that are induced by RNPS1 depletion were rescued by RNPS1 RRM overexpression. *The sashimi plots display the mean junction coverages for RFX5 intron 9 and INTS2 intron 21 in the indicated conditions. Important values are highlighted.*

Mechanistically, IR is a special case of alternative splicing because it requires the activation of splicing, while in other alternative splicing types like A5SS, unwanted splicing events are repressed. To better understand the mechanism of IR in human cells, RFX5 reporters were designed. Cryptic 5' splice site suppression in human cells was demonstrated to require the presence of the upstream intron (Boehm, Britto-Borges et al. 2018). On the other hand, PIWI intron 4 in *Drosophila* was retained when the downstream intron was removed (Hayashi, Handler et al. 2014) (Malone, Mestdagh et al. 2014). In both cases, RNPS1 is potentially recruited to the EJC as a part of the ASAP or PSAP complex. However, the described events required the positioning of RNPS1 at different exon-exon junctions: upstream or downstream of the to-be-spliced intron, respectively. Since the alternative splicing types and the organism differed for the two events (IR in flies and A5SS in human cells), it was investigated, whether the IR event in RFX5 in human cells rather required splicing of the preceding or the subsequent intron. Therefore, in the reporters either one or both of the neighboring introns of RFX5 intron 9 were removed (Figure 18A). The introns were deleted from the reporters to mimic splicing without inducing EJC deposition and ASAP or PSAP recruitment. After stable transfection into

HFT cells, the reporters were expressed, and RNA was harvested. Splicing of intron 9 was quantified using RT-PCR. In the reporter without surrounding introns, intron 9 was retained in nearly 100 % of the transcripts (Figure 18B). This confirmed that EJC deposition and/or ASAP/PSAP recruitment at neighboring exon-exon junctions are crucial for correct intron 9 splicing in RFX5. Moreover, similar to PIWI intron 4 splicing in *Drosophila*, this effect was mostly due to deletion of the subsequent intron, which leads to IR in roughly 70 % of the transcripts.

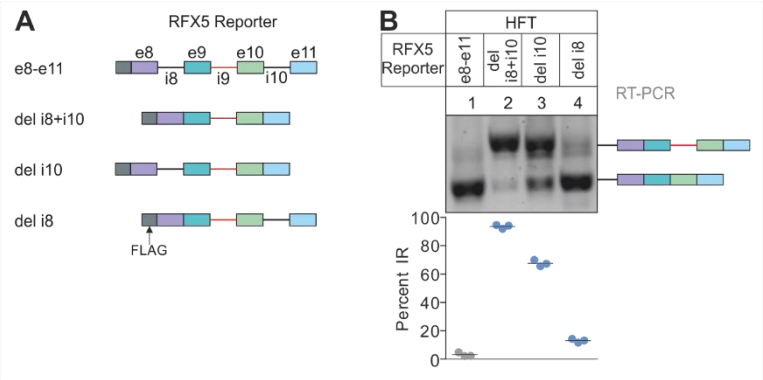


Figure 18: Correct RFX5 intron 9 splicing requires the presence of the subsequent intron. *(A): Scheme of the RFX5 reporter constructs that were stably transfected into HFT cells. (B) RT-PCR analysis and quantification of IR in the RFX5 reporter, the detected transcripts are indicated on the right (n = 3).*

The sashimi plot showed that RFX5 splicing was not only disturbed when RNPS1 was depleted, but also when the cells lacked the core components of the EJC (Figure 17). To determine which of the EJC or ASAP/PSAP components is essential for correct RFX5 intron 9 splicing, a tethering system was used to artificially bring the desired protein to the mRNA reporter. Therefore, the RFX5 reporter with deleted intron 10 was used and the MS2 stem loops were integrated at the end of the exon 10 where the EJC would normally be deposited after splicing (Figure 19). Additionally, the protein of interest was tagged with the MS2 coat protein that binds the MS2 stem loops and a V5 tag for western blot detection (Supplementary Figure 5A). This system can mimic the recruitment of the individual EJC or ASAP/PSAP components without the preceding splicing step. To rule out that any observed effects are due to the presence of an unspecific protein at the exon-exon junction, glutathione-S-transferase (GST), which has no splicing regulatory activity was used as a control. The RFX5 reporter with the MS2 stem loops as well as the MS2V5-tagged genes of interest were cloned into human expression plasmids and transiently transfected into HTO cells. Expression of the reporter/the tagged proteins was

induced and IR was detected using RT-PCR (Figure 19). Western blot was used to confirm the expression of the MS2V5-tagged EJC, ASAP and PSAP components (Supplementary Figure 5B).

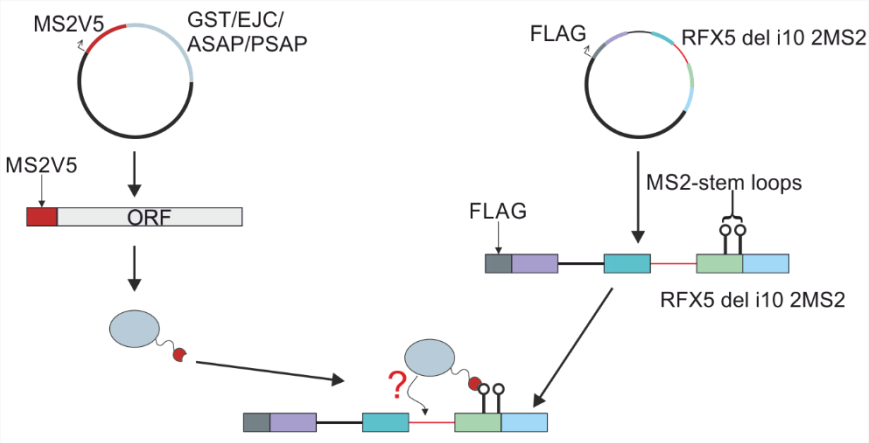


Figure 19: MS2V5-tagged proteins can bind to the MS2 stem loops of the IR reporter. *The plasmids for the reporter as well as the tagged tethering protein were transiently transfected to HTO cells.*

Although depletion of all EJC core factors increased the IR of RFX5 in the RNA sequencing analysis, neither MAGOH nor RBM8A were able to prevent IR of the RFX5 tethering reporter (Figure 20A). This suggests that the individual EJC components cannot rescue RFX5 splicing on their own. Interestingly, tagging of MAGOH or RBM8A with MS2V5 should in principle not interfere with their ability to form the complete EJC core (Gehring, Kunz et al. 2005). Potentially, the function of the EJC in prevention of RFX5 IR is not the direct regulation of splicing but the recruitment of the ASAP or PSAP complex, including RNPS1. So far it was not shown whether the tethered EJC can still recruit the ASAP or PSAP complex. Therefore, defective/abolished interaction with its auxiliary complexes might indeed be the cause of the inefficiency of EJC tethering.

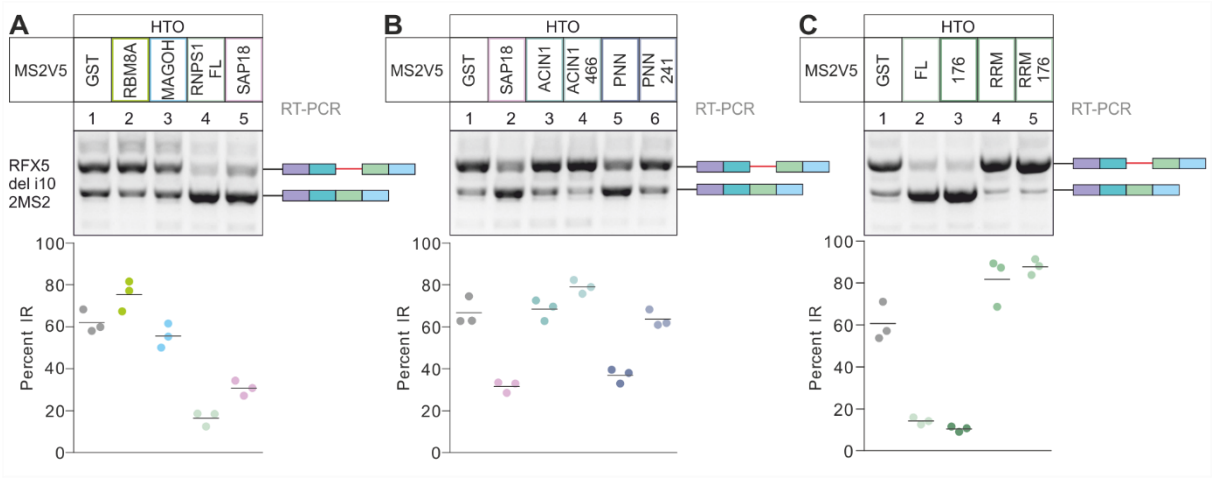


Figure 20: Tethering of RNPS1 but not the EJC components RBM8A or MAGOH rescued IR of the RFX5 tethering reporter. *RT-PCR and the corresponding quantification for the RFX5 tethering reporter is shown (n = 3) for (A) EJC components versus RNPS1 and SAP18, (B) different ACIN1 and PNN tethering constructs, and (C) different RNPS1 and RNPS1 RRM tethering constructs.*

Tethering of the ASAP/PSAP components RNPS1 and SAP18 strongly reduced IR of the RFX5 tethering reporter (Figure 20A). Since RNPS1 and SAP18 are the shared components of both complexes, also ACIN1 and PNN were tethered to the reporter to uncover potential discrepancies between the complexes. To this end, shortened versions of both proteins were used to ensure proper expression (Supplementary Figure 5A). While PNN tethering reduced RFX5 IR, ACIN1 did not alter splicing of the reporter at all (Figure 20B). Inspection of the protein expression levels showed that the MS2V5-tagged ACIN1 was not expressed, so it cannot be ruled out that the ASAP complex is involved in RFX5 intron 9 splicing (Supplementary Figure 5B). A PNN mutant was designed according to an ACIN1 mutation in the RSB (RNPS1-SAP18-binding) motif that was shown to disrupt ASAP assembly (unpublished data). Mutation of the PNN RSB indeed hindered the splicing activation of RFX5 intron 9 by PNN (Figure 20B). This was in accordance with the hypothesis that the PSAP complex (and potentially the ASAP complex) is mostly required to guide RNPS1 to its correct and functional position.

Furthermore, tethering of the RNPS1 176 mutant that cannot assemble the ASAP or PSAP or interact with the EJC, was able to prevent IR (Figure 20C). The RNPS1 RRM did not reduce IR, neither in its wildtype form where it can still assemble the ASAP and PSAP, nor in its mutated form. Combined, these tethering experiments confirm that RNPS1 is the effector molecule that is essential for splicing activation of RFX5 intron 9. Also, it was shown that for correct RFX5 splicing, unlike suppression of cryptic 5’ splice sites, splicing of the subsequent intron was more important than splicing of the preceding intron.

Even though RFX5 IR was not rescued by the RNPS1 RRM, IR seemed to be rescued better by the RRM compared to other splicing types (Figure 16). However, these results were generated using different bioinformatic tools and are thus not directly comparable. This raised the question, whether rescue efficiencies of the RNPS1 RRM actually vary between the different types of alternative splicing or whether this phenomenon was caused by computational limitations. In an attempt to generate comparable results for the alternative splicing types, rMATS was used to detect alternative splicing of all types (A5SS, A3SS, mutually exclusive exons (MXE), ES and IR). As anticipated, also rMATS confirms that the RNPS1 RRM rescue is overall not as efficient as the full-length rescue (Figure 21A). However, when the fractions of the alternative splicing types are calculated, the variation seemed to be negligible, and no clear rescue preference was observed for the RNPS1 RRM (Figure 21B). Yet, the numbers of events found by rMATS ranged up to more than 8000, which was far more than found in LeafCutter and IRFinder combined (Leafcutter 300-500 in total, IRFinder 200-600 in total). Manual inspection of high-ranking rMATS events revealed that these included many false positives. Therefore, although it might seem as if all event types are rescued to the same extent by the RNPS1 RRM, the rMATS data alone are not reliable enough to draw this conclusion. Currently, the available bioinformatic approaches either do not detect all types of alternative splicing or are otherwise unsuited for this type of analysis. Thus, reliable and comparable analyses among different splice types would require the development of new bioinformatic tools.

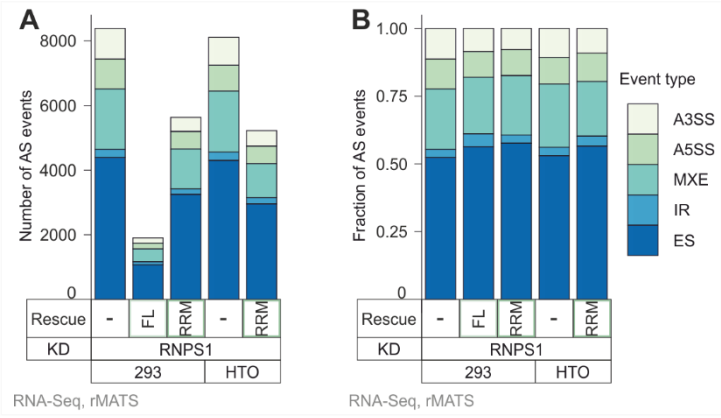


Figure 21: The RNPS1 RRM rescue is incomplete, but apparently has no preference for specific alternative splicing events. *Alternative splicing events as detected by rMATS are color-coded for A3SS, A5SS, MXE, IR andES events (Cutoffs: |dPSI| > 0.2 and Padj < 0.01). In (A) the total number of events is shown while in (B) the fraction of the different event types is depicted.*

8.3 RNPS1 provides a binding hub for splicing factors on the mRNA

8.3.1 The domains of RNPS1 have distinct splicing regulatory abilities

The previous chapters provide an overview of RNPS1's functions in regulating mRNA processing. These functions include the regulation of various alternative splicing types as well as the potential regulation of a subset of NMD targets. It appears that the functions of RNPS1 are carried out by different domains of RNPS1, since overexpression of the RNPS1 RRM rescued many, but by far not all alternative splicing events. As this indicates that at least one other domain of RNPS1 might possess some splicing regulatory activity, different RNPS1 deletion mutants were generated. To assess which of the domains of RNPS1 affects splicing, the rescue abilities of the RNPS1 deletion mutants were examined in RNPS1-dependent alternative splicing events that were not rescued by the RNPS1 RRM. RNPS1 deletion constructs in which either the C-terminus (Del-C; including the RS/P domain), the S domain (Del-S) or the N-terminus (Del-N) were deleted individually or in combination (Del-N+C, Del-S+C) were stably transfected into 293 cells and expression was confirmed using western blot (Figure 22A, Supplementary Figure 6A). Alternative splicing of the transcripts FDPS and TAF15, which were not rescued by RNPS1 RRM expression, was measured by RT-PCR. In both targets, the rescue ability of the different deletion mutants varied noticeably (Figure 22B, C). Deletion of the RNPS1 C-terminus either alone or in combination exhibited decreased rescue ability compared to the rescue with RNPS1 FL. As for FDPS, neither lack of the S domain, nor the N-terminus truly affected the rescue ability of RNPS1. In case of TAF15, RNPS1 Del-S rescued alternative splicing to a lesser extent. Nevertheless, deletions of the S domain or the N-terminus had the strongest effect when combined with lack of the C-terminus.

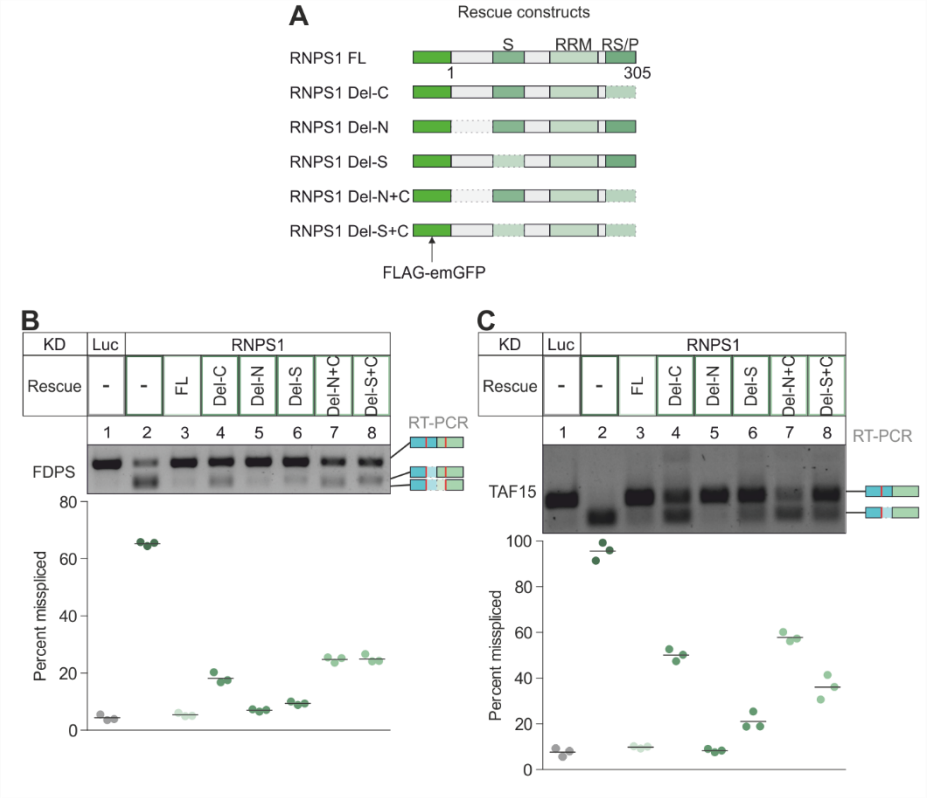


Figure 22: Different deletion mutants of RNPS1 exhibit different rescue abilities. *(A) Schematic depiction of the FLAG-emGFP-tagged RNPS1 deletion mutants that were used in KD rescue experiments. (B,C) RT- PCR and the corresponding quantification of alternative splicing in (B) FDPS and (C) TAF15 (n = 3). The detected PCR fragments are indicated on the right.*

Upon tethering to the RFX5 IR reporter, none of the RNPS1 deletion mutants was able to rescue alternative splicing completely (Figure 23A, Supplementary Figure 6B, C). Deletion of the N- or C-terminus induced a partial rescue of RFX5 IR, while simultaneous deletion of both domains abolished the rescue ability of RNPS1 completely, suggesting that they might have an additive effect. Noteworthy, RNPS1 Del-S did not much increase the production of normally spliced transcript but instead induced an additional exon skipping event of the preceding exon. When the results of the KD rescue experiments were investigated again, it appeared that rescue with the Del-S construct also led to a slight shift of the alternatively spliced band of FDPS in the gel. The sashimi plot of FDPS then resolved this mystery as there was not only one, but two different alternative splicing events upon RNPS1 depletion (Figure 23B). First, an A5SS in exon 4 was used in roughly 30 % of the transcripts in RNPS1 KD conditions. Second, another 50 % of transcripts showed usage of the same A5SS combined with an A3SS in the subsequent exon. This second event would fit to the shift in the Del-S rescue. Interestingly, also rescuing RNPS1 KD with the RNPS1 RRM resulted more often in the production of the shorter isoform with two alternative splice sites (Figure 23B). These findings indicate that in case of FDPS the

S domain and the RNPS1 RRM can cooperate to avoid the usage of this alternative 3’ splice site, while the N- and C-terminus are probably required to ensure correct 5’ splicing.

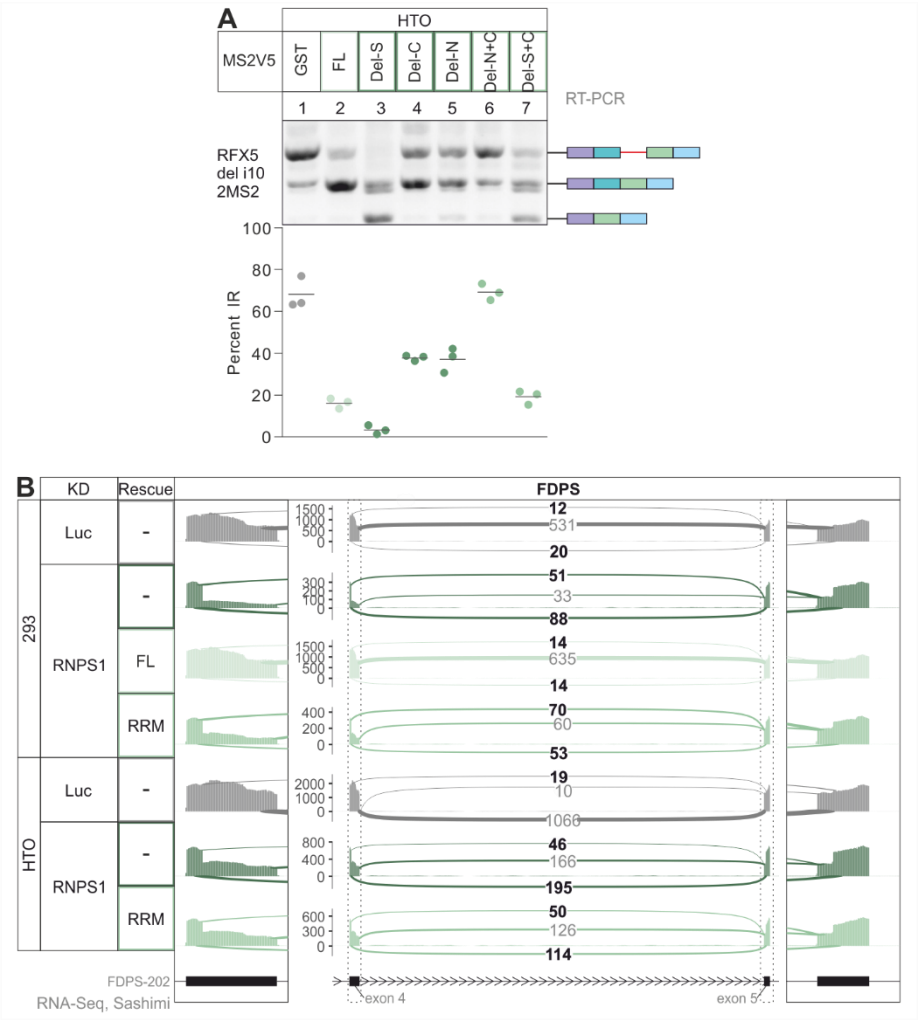


Figure 23: RNPS1 Del-C and RNPS1 Del-S rescue distinct alternative splicing events. *(A) RT- PCR and the corresponding quantification of tethering of RNPS1 deletion mutants to the RFX5 tethering reporter (n = 3). (B) The sashimi plot of FDPS depicts the mean splice junctions detected with RNA sequencing in the indicated conditions. Boxes on the left and right sight show a zoom-in on the exons bearing the alternative splice sites. The alternative splice junction counts are highlighted.*

Taken together, the KD rescues and tethering experiments with the RNPS1 deletion mutants demonstrate that not only the RRM, but also the S domain and the C-terminus, and sometimes even the N-terminus, of RNPS1 can exhibit certain splicing regulatory functions.

8.3.2 RNPS1 domains have individual binding partners

The previous results demonstrate that RNPS1 has multiple functions that are conducted by its different domains. However, it is unlikely that RNPS1 alone is able to provide the required functional variability. Thus, it was assumed that RNPS1 might fulfill its functions by recruiting

other factors to the EJC. To test this hypothesis, the interactome of RNPS1 was identified using mass spectrometry. In addition to the interactome of full-length RNPS1, the interactome of RNPS1 Del-C and the RNPS1 RRM were determined. Both constructs were of interest, as they exhibited distinct rescue potentials in the previous analyses.

Moreover, two different experimental approaches were used that rely on distinct principles. First, the RNPS1 constructs were equipped with a FLAG-emGFP tag and used in a FLAG-Immunoprecipitation (IP), which should co-precipitate direct as well as bridged interactors of RNPS1 (Figure 24). To complement the results of the IP, a proximity labeling approach was applied. The RNPS1 constructs were tagged with either a FLAG-TurboID (TID) or a MYC-UltraID (UID, only RNPS1 FL) tag and exposed to biotin, so that proteins in the vicinity of the tag are biotinylated (Branon, Bosch et al. 2018, Zhao, Bitsch et al. 2021). Co-precipitated or biotin-labeled proteins were purified and identified using label-free mass spectrometry.

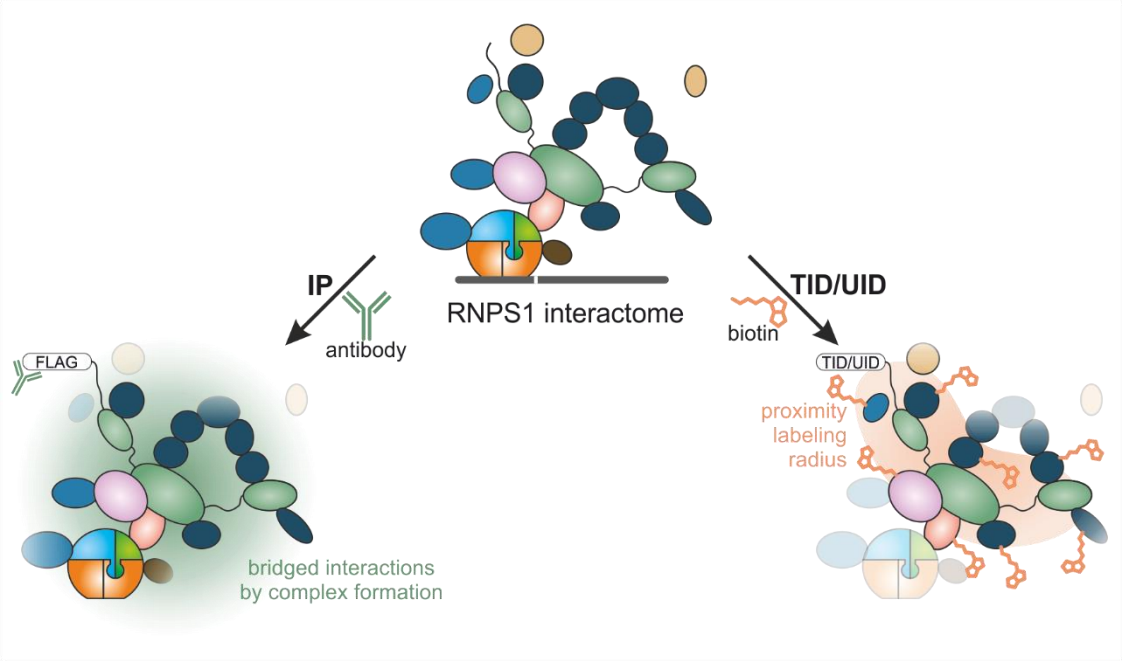
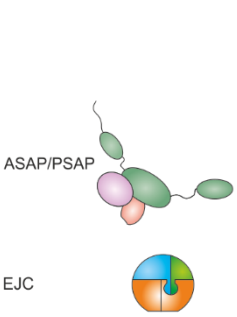


Figure 24: The RNPS1 interactome is identified using IP and proxomity labeling appoaches followed by mass spectrometry. *In the IP, FLAG-emGFP-tagged RNPS1 was pulled down using an anti-FLAG antibody. Direct or bridged interaction partners were co-precipitated using this method. For proximity labeling, RNPS1 was tagged with Turbo/UltraID. Addition of biotin to the cells lead to biotinylation of proteins in the vicinity of RNPS1.*

As expected, all RNPS1 constructs interacted with the other components of the ASAP and PSAP complex (Figure 25). Noticeably, the resulting log2 FCs were often considerably lower in the proximity labeling samples, pointing to a discrepancy of the two methods. Surprisingly, the core EJC factors were only enriched in the IP data but not in the proximity labeling data. This

might be a result of the suggested indirect binding of RNPS1 to the EJC via ACIN1 or PNN (Wang, Ballut et al. 2018). Therefore, the distance between the EJC core and the TurboID- or UltraID-tag might be too long for proximity labeling or the other ASAP/PSAP components might even sterically hinder the tag from biotinylating the EJC core. The interaction of SAP18 and EIF4A3 with the RNPS1 constructs was confirmed using IP and western blot (Supplementary Figure 7A, B).



	log2 FC FL/Ctrl			log2 FC Del-C/Ctrl		log2 FC RRM/Ctrl	
Name	IP	TID	UID	IP	TID	IP	TID
SAP18	5.18	2.91	3.52	6.43	2.77	10.25	1.95
ACIN1	6.12	3.58	7.82	7.27	3.47	8.08	3.09
PNN	6.54	6.72	6.80	7.44	6.56	6.31	5.39
MAGOHB	6.52	n.d.	n.d.	7.34	n.d.	6.25	n.d.
RBM8A	3.73	n.d.	n.d.	4.36	n.d.	5.51	n.d.
EIF4A3	3.25	-0.10	0.39	3.66	-0.03	4.47	-0.02

Figure 25: The interaction of RNPS1 with the EJC could only be detected in the IP. *The log2 FCs of the individual ASAP/PSAP or EJC core proteins are listed for the indicated RNPS1 constructs as compared to the control.*

Notably, except for the EJC core factors, the only other NMD factor identified in the MS experiments was UPF1. Moreover, UPF1 was only found in proximity labeling samples of RNPS1 Del-C and RNPS1 RRM but neither in the IP, nor in RNPS1 FL samples. This suggests that UPF1 does not bind to RNPS1 under normal conditions, which strengthens the view that RNPS1 is not an essential but rather a supporting NMD factor.

Next, the gene ontology terms for biological processes (GOBP) of the RNPS1 interaction partners were determined using the Perseus software (version 1.6.15.0) (Tyanova, Temu et al. 2016). These terms were used to classify the proteins as “splicing”, “RNA processing”, “RBP”, “mRNA” or “other”. In all datasets and all RNPS1 constructs, remarkably many interactors had splicing related functions. Moreover, 63 of these splicing related interactors were found in all three RNPS1 FL conditions and 58 more were found in at least two conditions (Figure 26A, Supplementary Figure 7C). Besides these splicing related proteins, also 12 proteins were enriched that were classified as relevant for RNA processing or mRNA in general.

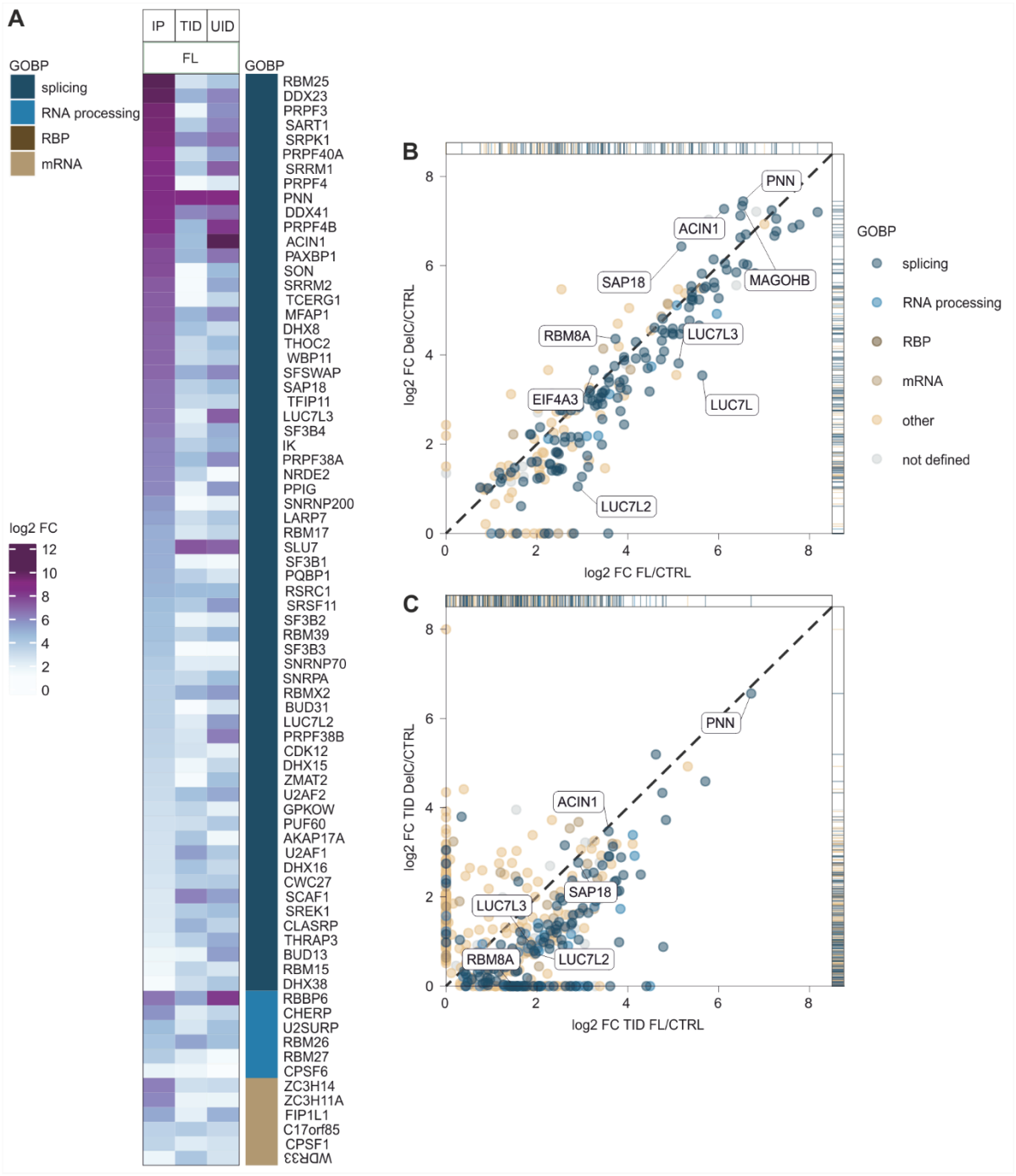


Figure 26: Many splicing related proteins that were enriched in RNPS1 FL conditions were slightly reduced in RNPS1 Del-C conditions. *(A) The heatmap depicts proteins that were enriched in all three RNPS1 FL datasets and that are classified as either "splicing", "RNA-processing", "RBP" or "mRNA" according to their gene ontology biological process (GOBP) terms with the cutoffs q-value < 0.05 and log2 FC > 1. (B, C) The log2 FCs of proteins enriched in RNPS1 FL is plotted against the log2 FCs of the proteins enriched in RNPS1 Del-C for (B) IP and (C) TurboID (TID) (Cutoff: q-value < 0.05 in at least one condition).*

The log2 FCs of the interaction partners found in RNPS1 Del-C were plotted against those found in RNPS1 FL for the IP as well as for the TID data. Overall, the fold changes spread far more in the IP conditions compared to the TID conditions, where they are mostly lower.

The interactors overlap relatively well between the RNPS1 FL and RNPS1 Del-C in both experiments. However, especially splicing related interactors were often less enriched in the RNPS1 Del-C experiments (Figure 26B, C). This observation was confirmed for the splicing factor Luc7-like protein 3 (LUC7L3) and the U1 component small nuclear ribonucleoprotein A (SNRPA) using an IP experiment, followed by western blotting (Supplementary Figure 7D).

The interactome of the small RNPS1 RRM was also still enriched for splicing associated proteins, including the 3' splice site binding U2AF1 (Figure 27A, B). Interestingly, among all identified proteins in general only 28 were found in both the IP and the TID data but 16 of these overlapping interactors were classified as splicing related (Figure 27C, D). Of the remaining proteins, 5 were classified as mRNA or RNA processing related and only 7 were not related to splicing or did not have a GOBP term. This shows that the interactome of RNPS1 was strongly enriched in factors involved in splicing regulation, even when only certain domains of RNPS1 were present.

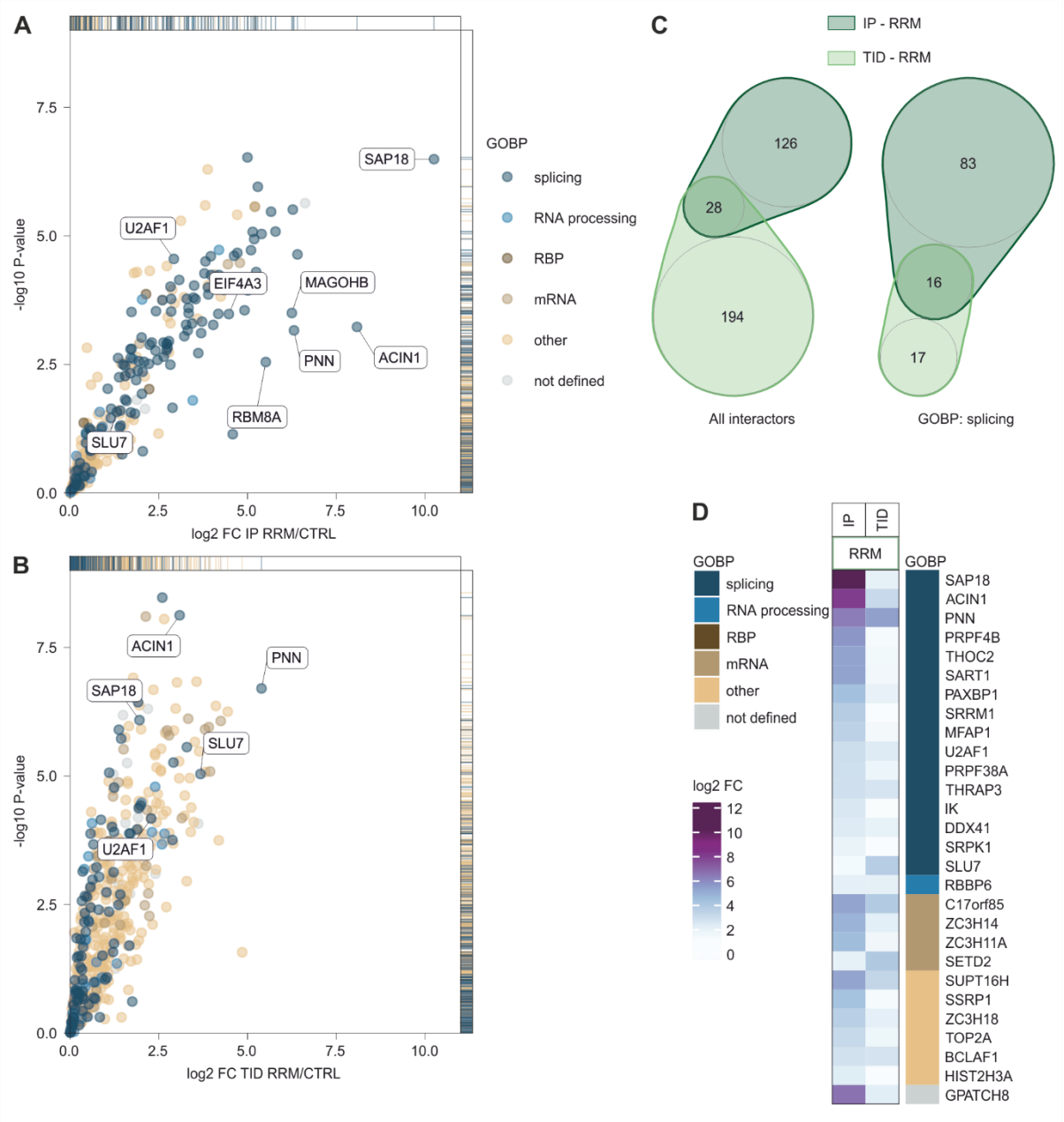


Figure 27: The RNPS1 RRM interacts with many splicing related proteins. *(A, B) The log2 FCs of proteins enriched in RNPS1 RRM conditions as compared to control conditions is plotted against the -log10 P-value (Cutoff: log2 FC ≥ 0) for (A) IP and (B) TID. (C) The venn diagrams depict the overlap of the RNPS1 RRM interactomes as identified by IP or TID for either all interactors or only those classified as splicing related. (D) The heatmap depicts the log2 FCs of the 28 interactors that were enriched in both RNPS1 RRM conditions and indicates their classification according to the GOBP terms.*

In both RNPS1 Del-C and RNPS1 RRM interactomes the EJC core proteins and ASAP/PSAP proteins were equally well or even higher enriched compared to full-length RNPS1 (Figure 25). This demonstrates that the reduced interactions with splicing associated proteins cannot be caused by reduced interaction with the EJC or the ASAP/PSAP complex. For direct comparison of the different RNPS1 interactomes, the absolute numbers of interactors found in the mass spectrometry datasets were plotted in a bar graph (Figure 28A). As expected, the overall

number of interaction partners decreased with the size of the RNPS1 fragment in both IP data and TID data. Furthermore, especially in the TID data, the decrease in splicing or RNA processing related interaction partners was stronger compared to the total decrease of interactions.

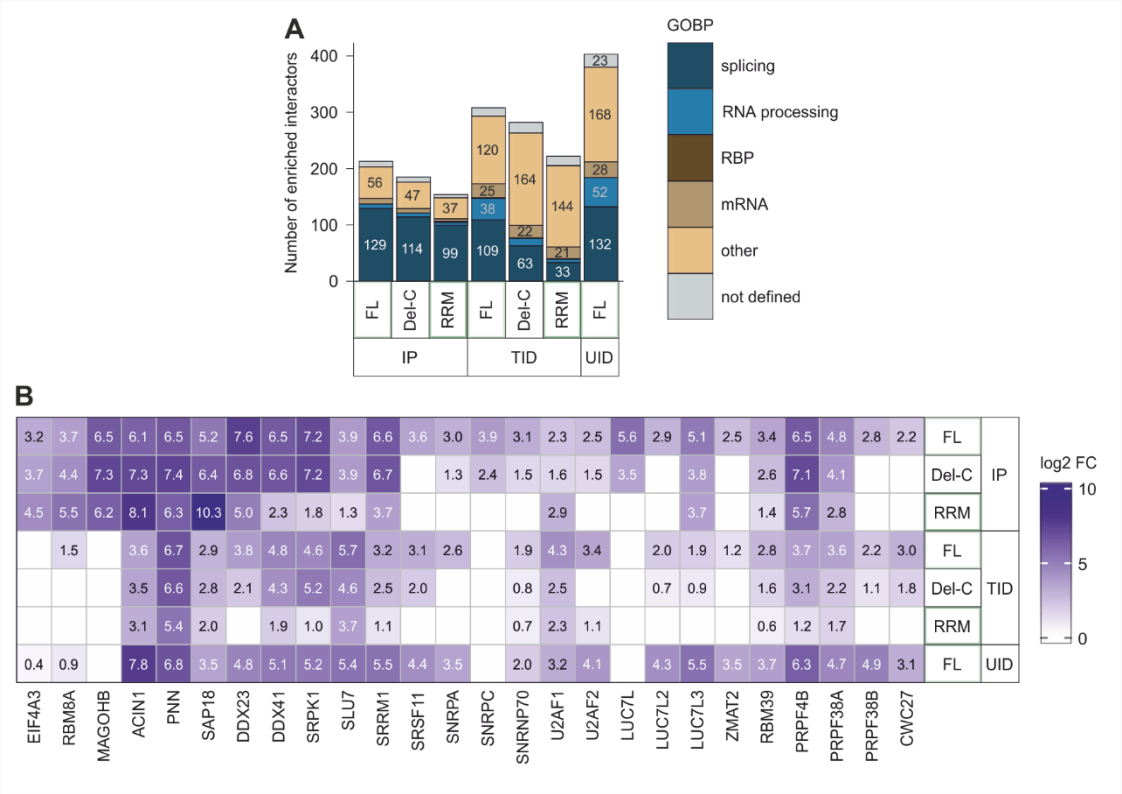


Figure 28: The number of splicing related interactors decreases with the size of the RNPS1 fragment. *(A) The absolute numbers of RNPS1 interaction partners for the indicated conditions is plotted in a bar graph. The number of interactors of the different classes is indicated by color-coding (Cutoffs: q-value < 0.05 and log2 FC > 1). (B) The heatmap depicts the log2 FCs of selected RNPS1 interactors in the indicated conditions (Cutoff: q-value < 0.05).*

When log2 FCs of selected RNPS1 interactors are depicted in a heatmap, many were less enriched or not found at all, when RNPS1 was truncated (Figure 28B). This again displays a proportionality of splicing related interactors and the size of the RNPS1 fragment. In summary, RNPS1 was demonstrated to interact with a large number of splicing factors and even some spliceosomal components. The different domains of RNPS1 furthermore seem to be responsible to enable distinct interaction patterns.

9 DISCUSSION

The complex processes of gene expression are regulated on multiple levels by various mechanisms, involving splicing regulation and the degradation of PTC-containing mRNAs via the NMD pathway. Many proteins that bind to the mRNA either in a sequence-dependent or -independent manner ensure proper splicing and NMD execution. Among the sequence-independent RBPs, the EJC is deposited by the spliceosome close to exon-exon junctions. The EJC participates in the regulation as well as activation of splicing and NMD by interacting with a multitude of auxiliary factors. In this work, new insights into the role of the EJC-associated factor RNPS1 during splicing regulation and NMD activation are provided.

9.1 RNPS1 acts as a minor enhancer of NMD for specific targets

To study the involvement of RNPS1 in NMD, it was examined how NMD activity changed when RNPS1 was depleted in three different cell lines. While depletion of RNPS1 did not completely disrupt NMD, as the simultaneous depletion of the NMD factors SMG6 and SMG7 did, it still led to the upregulation of several genes (Figure 10 (Boehm, Kueckelmann et al. 2021)). However, these gene upregulations were mostly cell type specific as they were not consistently found in the three cell types subjected to RNPS1 KD. The examination of differential gene expression (DGE) of bona fide NMD targets showed very little or no effect of RNPS1 depletion in any cell type (Figure 11). Despite these weak effects upon RNPS1 depletion, several of the canonical NMD targets were downregulated when RNPS1 was overexpressed in 293 cells. This was in good agreement with an older study that found NMD to be enhanced when RNPS1 was highly expressed (Viegas, Gehring et al. 2007). Furthermore, the differential usage of PTC-positive transcripts upon RNPS1 correlated only weakly with SMG6/7 KD/KO upregulated, high-confidence NMD targets (Figure 12, Figure 13). Depletion of EJC core factors yielded noticeably more DTU events and correlated much better with the SMG6 and SMG7 depletion, which probably displays the known NMD-inducing function of the EJC. Moreover, in agreement with the DGE findings, it suggests that the effect of RNPS1 on NMD is rather mild.

A model was developed, where the availability of RNPS1 determines how efficiently a rather minor subset of NMD targets is degraded (Figure 29). At the same time, the decay of the

majority of NMD targets remains unaffected and therefore RNPS1-independent. Conclusively, RNPS1 rather is a specific enhancer of NMD than a core NMD factor.

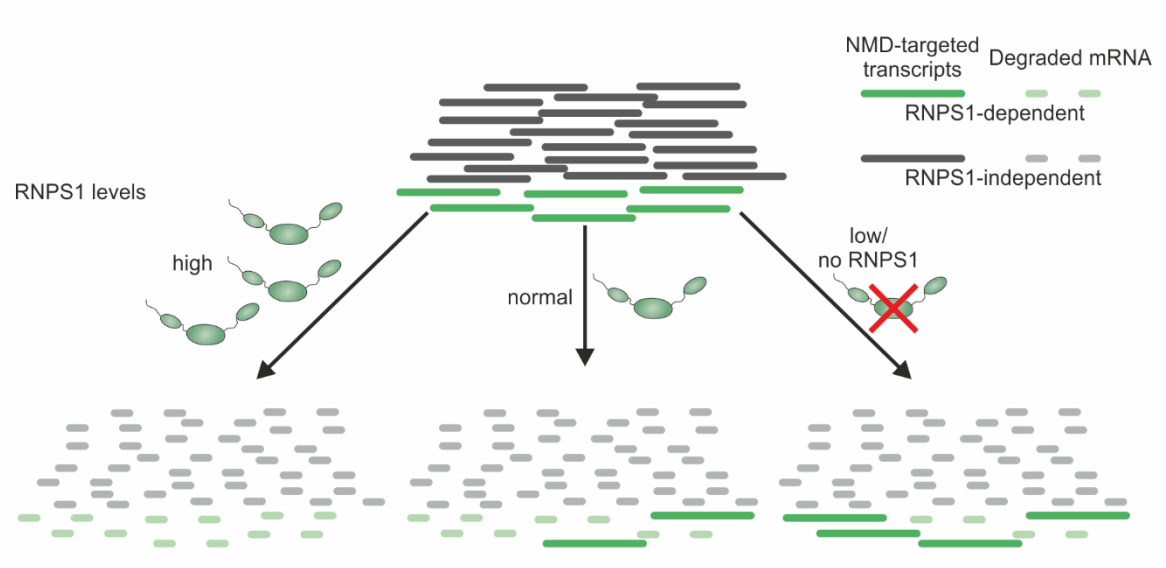


Figure 29: RNPS1-dependent NMD targets are degraded less efficiently in the absence of RNPS1. *A pool of NMD targeted transcripts is depicted that either are RNPS1-dependent (green) or -independent (gray). When RNPS1 is overexpressed (high), RNPS1-dependent NMD targets are degraded more efficiently than under normal RNPS1 expression levels. Depletion of RNPS1 reduces the degradation efficiency of RNPS1-dependent NMD targets. RNPS1-independent NMD targets are not affected by a reduction or an increase of RNPS1 levels.*

Although the overall transcriptomic effects of RNPS1 on NMD were comparably weak, it still appeared to regulate the degradation of a subset of specific targets. Therefore, the question was raised whether the activation of NMD by RNPS1 might be mediated by interaction with core NMD factors. Previous studies found an interaction between RNPS1 and the NMD factors UPF2 and UPF3B using IP (Lykke-Andersen, Shu et al. 2001, Gehring, Kunz et al. 2005). However, none of the typical NMD factors (SMG or UPF proteins) were detected in the IP or proximity labeling interactome of RNPS1 in this work or in a more recent study (Mabin, Woodward et al. 2018).

It can therefore be hypothesized, that RNPS1 potentially interacts with the NMD machinery in an indirect manner. The previous NMD-activating functions were detected by tethering RNPS1 to the 3' UTR of an NMD reporter (Lykke-Andersen, Shu et al. 2001, Gehring, Kunz et al. 2005). Thus, it is possible that this artificial binding of RNPS1 to the target mRNA activated NMD because it led to EJC recruitment to the 3' UTR which is known to activate NMD. Together these findings indicate that the supplemental NMD-activating function of RNPS1 could be mediated through its interaction with the EJC. To test this hypothesis, the RNPS1 176 mutant

could be used in a tethering approach similar to the ones used in the older studies. Since this RNPS1 mutant is deficient of ASAP/PSAP assembly, it also cannot recruit the EJC core. If the RNPS1-dependent enhancement of NMD requires the EJC, tethering of the RNPS1 176 mutant to the 3' UTR of an NMD reporter should not activate its degradation.

9.2 RNPS1 ensures proper transcript maturation by activation or suppression of certain splicing events

When RNPS1 was depleted from 293 and HTO cells, this induced severe changes in the splicing pattern. All canonical types of alternative splicing were affected, including the previously investigated alternative 5' splicing (Figure 30A, (Blazquez, Emmett et al. 2018, Boehm, Britto-Borges et al. 2018)). In this thesis, the prevention of IR was studied in more detail. This type of alternative splicing was particularly interesting, because it differs from the other alternative splicing types and supposedly requires another underlying mechanism. For the other types, the spliceosome identifies the wrong splice sites and therefore mediates the excision of an alternative intronic sequence. In case of IR, the spliceosome fails to detect the correct splice sites and as a consequence does not splice the intron at all. Thus, RNPS1-dependent IR regulation must be attained via a different mechanism than the repression of undesired splice sites.

Alternative splicing regulation by the EJC often relies on the preceding splicing of surrounding introns. Intriguingly, repression of A5SS in human seemed to require splicing of the upstream intron, whereas IR in *Drosophila* was prevented by downstream intron splicing (Hayashi, Handler et al. 2014, Malone, Mestdagh et al. 2014, Blazquez, Emmett et al. 2018, Boehm, Britto-Borges et al. 2018). Here, a reporter system was used to investigate RNPS1-dependent retention of RFX5 intron 9. To prevent RFX5 IR, splicing of the downstream intron was crucial (Figure 18). This demonstrated that the mechanism underlying RNPS1-dependent IR is conserved in human and flies. Thus, two molecularly more defined mechanisms for alternative splicing regulation by RNPS1 in human can be proposed (Figure 30B).

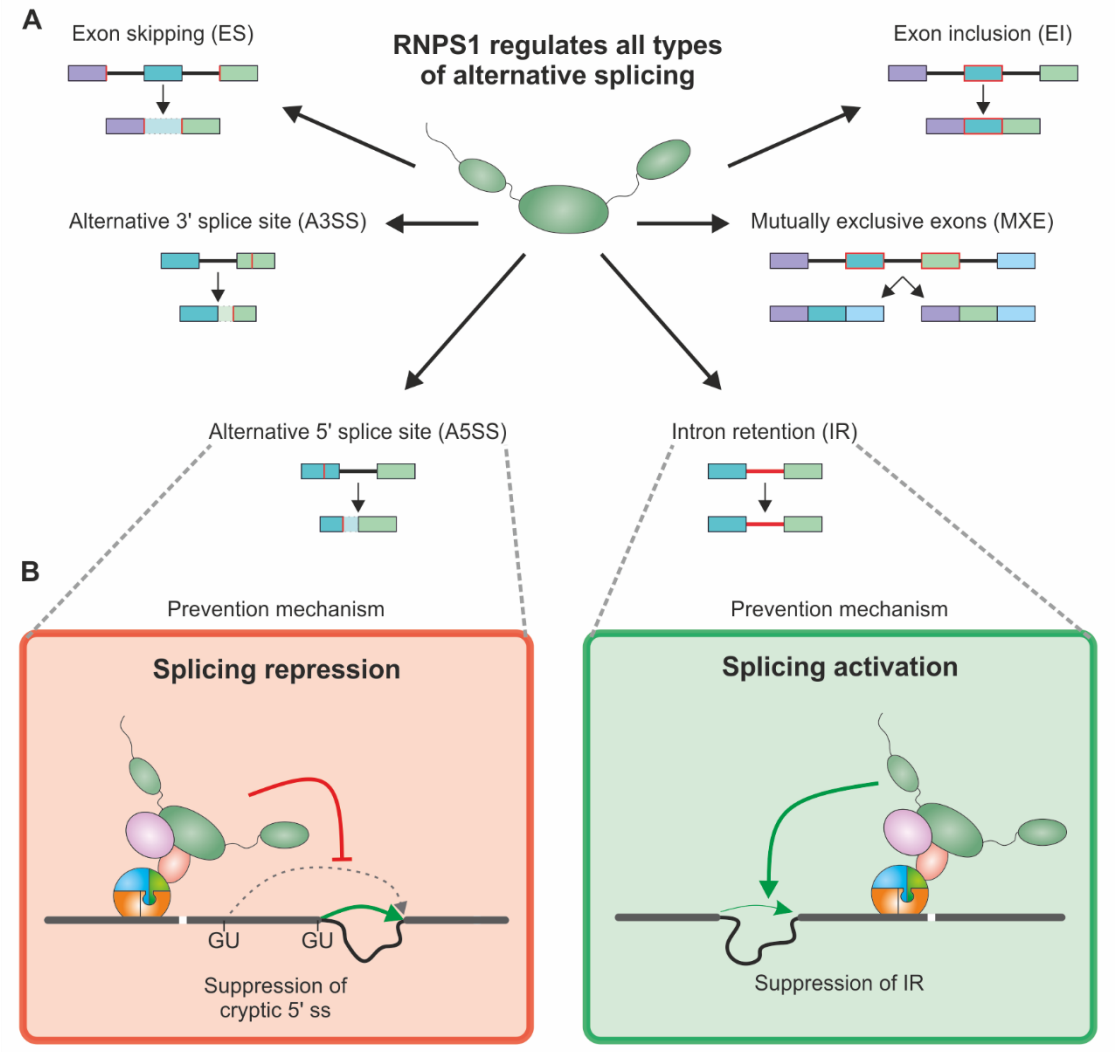


Figure 30: RNPS1 regulates all types of alternative splicing and can function either as a repressor of undesired splice sites or an activator of desired splicing. *(A) The outcomes of the alternative splicing types regulated by RNPS1 are depicted schematically with the important splice sites highlighted in red. (B) Two distinct RNPS1-dependent mechanisms for alternative splicing regulation are illustrated. RNPS1 represses the usage of A5SS but activates splicing of otherwise retained introns.*

Furthermore, the involvement of the RNPS1 RRM in alternative splicing regulation was investigated in more detail. Previously, it was assumed that the RNPS1 RRM, which is essential and sufficient for ASAP/PSAP assembly, might also be sufficient for alternative splicing regulation (Boehm, Britto-Borges et al. 2018). Rescuing the depletion of RNPS1 with the RNPS1 RRM however revealed that the RRM only rescued about half of the detected events (Figure 31). Although slight differences were observed between the RRM's ability to avert specific types of alternative splicing, this does not necessarily display a true preference but might be caused by discrepancies between the bioinformatic methods. Unfortunately, other currently available bioinformatic approaches were unsuitable or could not identify all types of alternative splicing with high confidence and high comparability. Even the usage of the

popular tool rMATS lead to many false-positive results that could not be verified by manual inspection. Of note, rMATS was used with the currently experimental setting "—novelSS" to detect also unannotated splicing events, which might explain the suboptimal reliability of the detected events. In order to answer the question of preferentially rescued alternative splicing types and similar future questions, the development of new splicing detection tools would be required. Nonetheless, it was consistent in all the performed alternative splicing analyses (LeafCutter, rMATS, IRFinder) that the RNPS1 RRM did not rescue the effects of RNPS1 depletion completely, neither in sheer numbers nor concerning the effect strength (Figure 31).

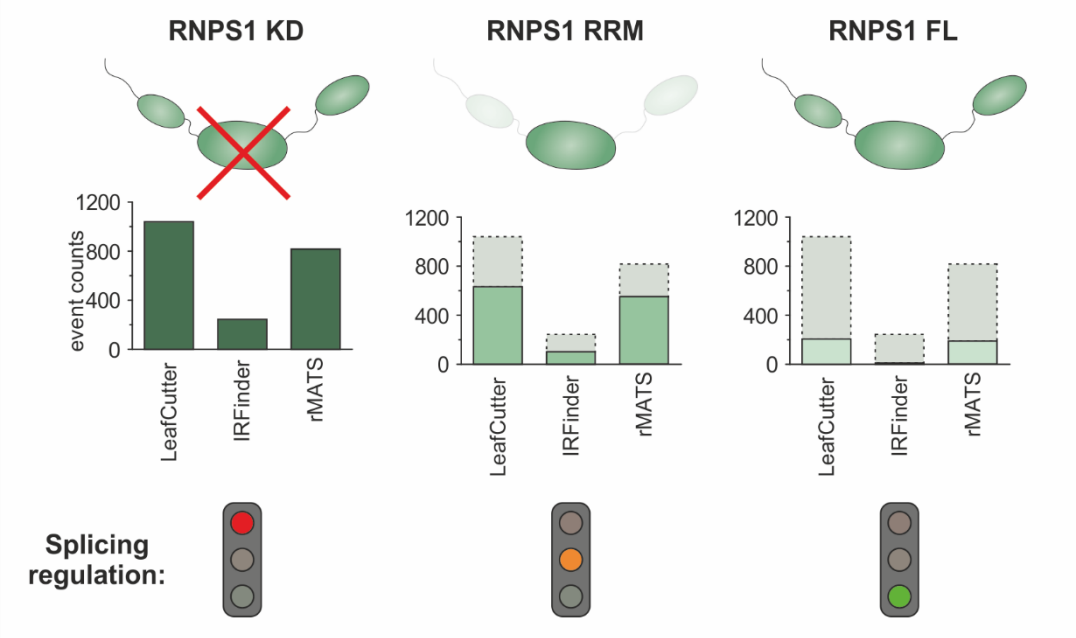


Figure 31: The RNPS1 RRM cannot confer the complete splicing-regulatory potential of full-length RNPS1. *RNPS1 depletion is compared with rescue expression of the RNPS1 RRM or RNPS1 FL in 293 cells. The bar graphs depict the absolute numbers of alternative splicing events that were detected with the indicated tools (Cutoffs: LeafCutter: |dPSI| > 0.1 and Padj < 0.001; IRFinder: |log2 FC| > 1 and Padj < 0.001; rMATS: |dPSI| > 0.2 and Padj < 0.01). The traffic lights indicate the state of alternative splicing regulation where green indicates (nearly) full activity, orange indicates reduced activity and red indicates no regulatory activity.*

Although the RNPS1 RRM rescued not all alternative splicing events, it still had a considerable rescue effect on many events that were upregulated upon RNPS1 depletion. Since the RNPS1 RRM is sufficient to assemble the ASAP or PSAP complex, it was hypothesized that the other complex components are, at least in part, responsible for this activity. Also, the ASAP/PSAP components ACIN1 and PNN are required for RNPS1 recruitment to the EJC (Wang, Ballut et al. 2018). Thus, investigation of ASAP/PSAP involvement was expected to give some insight into the mechanism of RNPS1 RRM rescue and the importance of RNPS1 recruitment to the

EJC. If RNPS1 would rely on the ASAP/PSAP-mediated recruitment, the depletion of ACIN1 or PNN should induce alternative splicing patterns similar to RNPS1 depletion. Contrary, neither the depletion of ACIN1 nor PNN resulted in a dysregulation of alternative splicing that was comparable to RNPS1 depletion (Figure 14, Figure 16). Albeit previous findings argue against it, this result could be explained by a redundancy of the two complexes (Figure 32, (Wang, Ballut et al. 2018)). In order to test this hypothesis, the cells were simultaneously treated with ACIN1 and PNN siRNA mediated KD. However, this could neither confirm nor deny a redundancy of the complexes, due to high residual amounts of PNN in the double KD (Supplementary Figure 4).

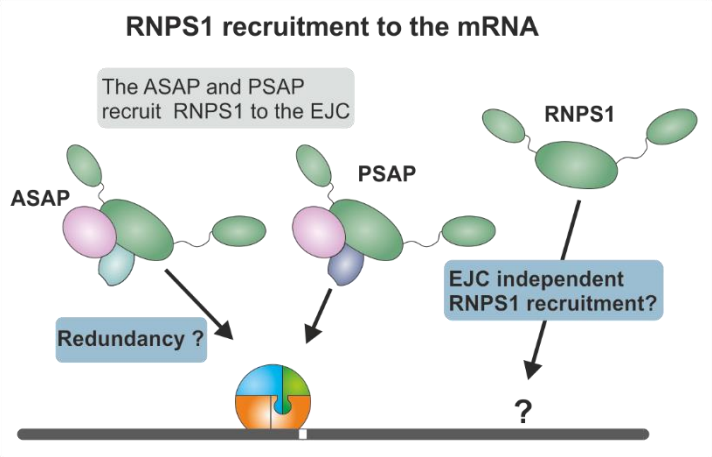


Figure 32: RNPS1 can be recruited to the mRNA via the ASAP/PSAP complex. *The potential mechanisms of RNPS1 recruitment to the mRNA are depicted schematically. Gray squares indicate already known mechanisms and blue squares indicate still open questions.*

Nevertheless, nearly all of the RNPS1-dependent alternative splicing events that were analyzed in more detail required the additional presence of either ACIN1 or PNN, suggesting they are usually necessary for RNPS1 recruitment to the EJC (Hayashi, Handler et al. 2014, Malone, Mestdagh et al. 2014, Blazquez, Emmett et al. 2018, Boehm, Britto-Borges et al. 2018). The use of a tethering reporter showed that although RNPS1 can regulate splicing on its own when it is directly brought to the required position, its recruitment by PNN was functional as well (Figure 20). This further reinforces the model that RNPS1 is the downstream effector molecule that is essential for splicing regulation but requires ACIN1 and PNN to mediate the interaction with the EJC.

If their main task is the recruitment of RNPS1, the functional difference between the ASAP and PSAP complex might lie in a distinct mRNA and EJC binding ability. Intriguingly, PNN

supposedly has a higher binding affinity for RNPS1 and SAP18 than ACIN1 (Murachelli, Ebert et al. 2012). Whether this means that more PSAP complexes are found in the cell compared to ASAP complexes and whether they have better binding affinity to the EJC remains unclear. Furthermore, it would add greatly to the clarification of RNPS1 recruitment to examine whether ACIN1 or PNN exhibit a binding preference to exon-exon junctions with distinct characteristics. This type of analysis could for instance provide information whether the different complexes recruit RNPS1 to regulate different types of alternative splicing.

Theoretically, RNPS1 could also be recruited to the mRNA by other mechanisms, either independent of the EJC as part of the ASAP/PSAP complex, or via a completely different mechanism that involves neither of these complexes. An indicator for the first mechanism is the finding that ACIN1 can bind to the pyrimidine tract of introns (Rodor, Pan et al. 2016). However, much more ACIN1 binding was observed that matched the exonic position of the EJC, suggesting that recruitment to the EJC is the more prominent pathway compared to direct recruitment of the ASAP to the mRNA. In principle, RNPS1 could also bind to the mRNA alone but as most investigated alternative splicing events require the additional presence of ACIN1 or PNN, this is potentially not one of the main RNPS1 recruitment pathways. Still, more research would be required to investigate the details of RNPS1 recruitment and confirm or deny the existence of EJC- and ASAP/PSAP-independent RNPS1-regulated splicing.

However, assembly of the ASAP/PSAP complex via the RNPS1 RRM was, as mentioned above, not always sufficient to rescue RNPS1-dependent splicing defects. This implied that also other domains of RNPS1 might assist in RNPS1-dependent alternative splicing regulation. Further studies using KD rescue experiments and a tethering reporter revealed that the deletion of the RNPS1 N-terminus barely affected alternative splicing, while both deletion of the S domain and the C-terminus did. However, the outcomes of splicing upon deletion of the S domain or the C-terminus differed, as observed in FDPS alternative splicing. Conclusively, the intact domain structure of RNPS1 is crucial for alternative splicing regulation and different disruptions of this composition leads to distinct splicing defects.

9.3 RNPS1 enables the formation of splicing competent complexes

As the rescue efficiency of RNPS1 deletion mutants is diminished in comparison to RNPS1 FL rescue, it was assumed that the different domains harbor some splicing regulatory activity themselves, potentially by interacting with further splicing factors. The interactome of RNPS1

FL and two deletion mutants was identified and enabled the establishment of a mechanistic model of RNPS1 functions. Among the interaction partners of RNPS1 were not only many splicing factors but also spliceosomal proteins (Figure 26). Deletions of RNPS1 domains resulted in decreased interaction, especially with splicing related factors. The U1 components SNRPA and SNRP70 for instance interacted well with RNPS1 FL but considerably less with RNPS1 Del-C. Mechanistically, this could indicate that the C-terminus might be involved in the detection of 5' splice sites by the spliceosome. This model would be supported by the finding that upon RNPS1 Del-C rescue, the A5SS of FDPS was still used (Figure 22).

Interestingly, the relatively small RNPS1 RRM still interacted with several splicing regulatory proteins. However, from the available data, it cannot be distinguished whether these proteins directly interact with the RNPS1 RRM or indirectly via the ASAP/PSAP complex. This problem could be overcome by introducing the 176 mutation into the RRM, which abolishes the ability to assemble the ASAP or PSAP complex. To clearly define the tasks of the different RNPS1 domains, it would furthermore be valuable to identify the interactome of an RNPS1 Del-S construct, as also the RNPS1 S domain was found to be involved in splicing regulation. Further point mutations in the different domains might give insight which residues are especially important for the interaction of the individual proteins with RNPS1.

Overall, the interactors of RNPS1 included markedly many SR or SR-like proteins. Proteins of these families were demonstrated to often contain so-called low complexity regions (Haynes and Iakoucheva 2006, Kwon, Xiang et al. 2014). These regions enable the interaction of the SR and SR-like proteins among each other and might explain the formation of high molecular weight complexes around the EJC (Singh, Kucukural et al. 2012, Mabin, Woodward et al. 2018). The detection of such high molecular weight complexes also matches the large number of interactors found with RNPS1. Potentially, RNPS1, which itself is an SR-like protein, provides the initial anchoring point on the mRNA for the establishment of these complexes around the EJC. The interaction of RNPS1 with many SR proteins thereby might serve as the basis for the recruitment of further splicing factors and the spliceosome itself (Figure 33). This was also supported by the large number and variety of binding partners of RNPS1. Many splicing factors also have their own binding motifs in the mRNA. This suggests that they can in principle bind to the mRNA by themselves, but RNPS1 probably facilitates this binding or recruits them specifically to the region surrounding the exon-exon junction. It can be hypothesized, that RNPS1 provides the basal point of attachment for complicated, multilayered splicing

regulatory networks. The dual approach that was used to identify the RNPS1 interactome thereby ensures that the detected interactions are reliable and are not just a byproduct of the formation of such large splicing networks. If only the IP approach would have been used, the interactions could have often been indirect, but the proximity labeling approach ensures that the proteins must at least be in the close vicinity of RNPS1. A good example here is the EJC, which was detected in the IP dataset but not in the proximity labeling dataset. The EJC is known to interact with RNPS1 in an indirect manner and therefore is probably not reached by the TurboID or UltraID tags. Splicing regulatory proteins that were detected in the IP as well as the proximity labeling interactomes can therefore be seen as reliable RNPS1 interactors.

For the cell, these complex splicing regulatory networks would provide an important advantage, as they would probably be more resistant to the alteration of individual factors. Potentially, depletion of only one factor would only affect the most sensitive alternative splicing targets. An example for this is the interaction between RNPS1 and SRSF11, which was detected in both IP and TID data. In a previous study, it was shown that depletion of SRSF11 for instance had no effect on the alternative splicing event in RER1 that was strongly affected upon RNPS1 depletion (Boehm, Britto-Borges et al. 2018). This would furthermore imply, that linear mechanisms that were proposed in earlier studies, might be too simple. For example in 2004, the inclusion of exons was proposed to be achieved by a cooperative effect of RNPS1 and P54 that is bound to the S domain (Sakashita, Tatsumi et al. 2004). Another example but from the opposite point of view is the regulation of neuronal microexons that was dependent on SRRM4 and its interaction with SRSF11 and RNPS1 (Gonatopoulos-Pournatzis, Wu et al. 2018, Gonatopoulos-Pournatzis, Niibori et al. 2020). Overexpression of SRRM4 induced the inclusion of microexons, however, depletion of neither RNPS1 nor SRSF11 impeded this function (unpublished data). This suggests that in this case, the overrepresentation of SRRM4 can compensate for the loss of other factors in the network and that for the regulation of these events the connection to the EJC via RNPS1 might be negligible.

The findings generated in this thesis indicate that functional linear interactions, where RNPS1 interacts with only one downstream effector, are probably rare. This type of interaction can be required for the regulation of individual, sensitive transcripts but is supposedly less important for the RNPS1 regulatory function in general.

A model was proposed, in which RNPS1 enhances the definition of otherwise insufficiently defined introns by enabling the formation of large networks composed of many splicing factors (Figure 33). The interaction with the splicing factors is mediated by the different domains of RNPS1, which can be seen as a multiadapter to connect splicing proteins to the EJC. Thereby, the correct splicing of these introns is promoted. Depletion of RNPS1 in such a model would leave the other splicing factors without their connection to the EJC. Potentially, some would bind by themselves to their binding motifs in the mRNA and others would not bind at all. This could induce either faulty splicing by the usage of non-canonical splice sites or abolish splicing completely, leading to intron retention.

Collectively, this thesis provides detailed knowledge about the functions of RNPS1 in NMD and splicing regulation. During NMD, RNPS1 acts as an activator or enhancer for the degradation of specific NMD targets. During splicing, RNPS1 provides the basis for vast splicing-regulatory networks. This model might furthermore serve as a basic frame for splicing regulation by other splicing factors, as they supposedly have frequent interconnections. Therefore, this thesis is an important contribution to unravel the complex mechanisms of gene expression regulation on the mRNA level.

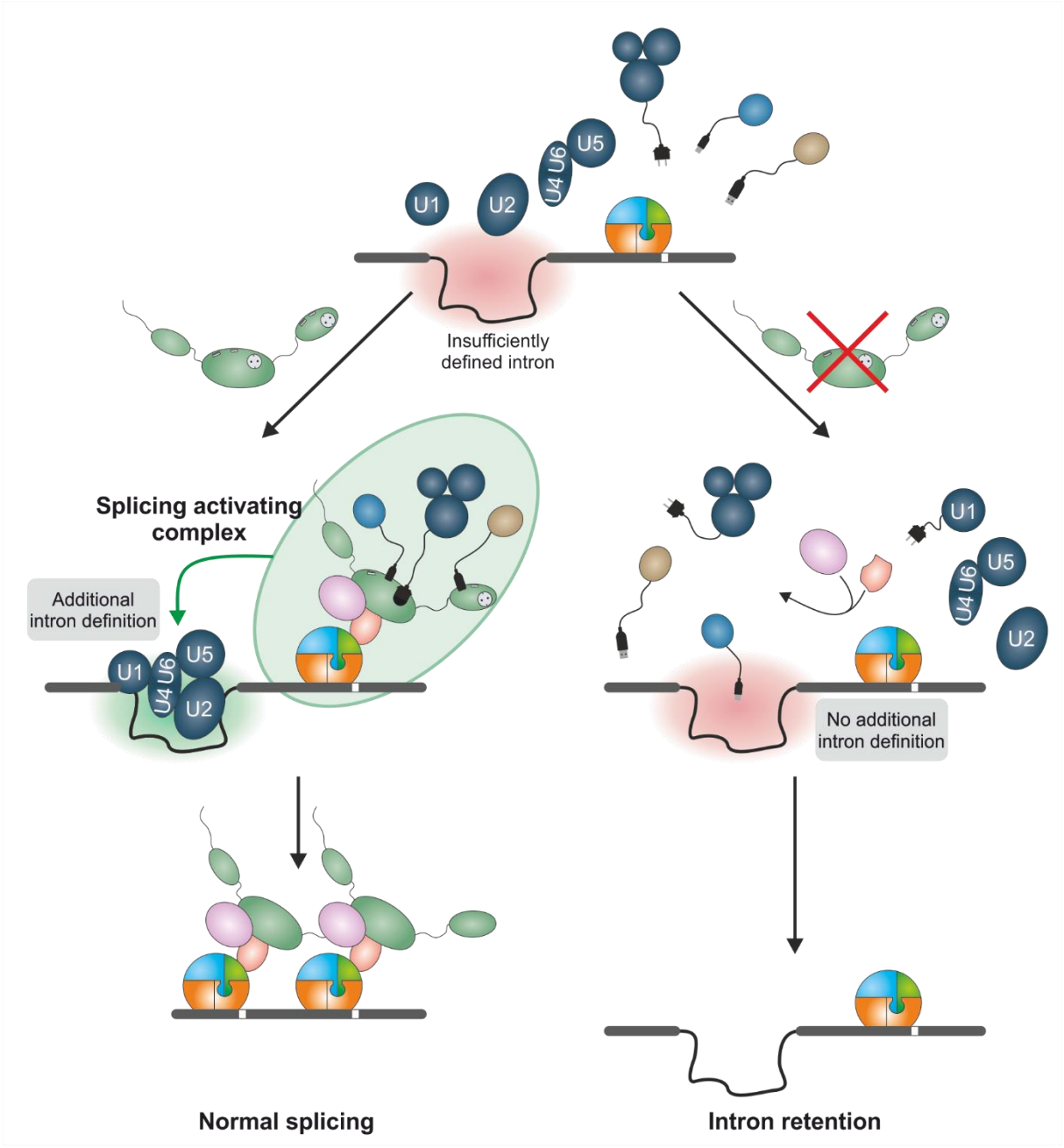


Figure 33: RNPS1 initiates the formation of splicing activating complexes. *During splicing, the EJC is recruited to an exon-exon junction and is bound by RNPS1 in the ASAP/PSAP complex. By functioning as a multiadapter, RNPS1 enables the formation of splicing enhancing/activating complexes and thereby facilitates the splicing of nearby, insufficiently defined introns. In the absence of RNPS1, no splicing activating complex is formed and the intron remains unspliced.*

10 LIST OF FIGURES

The following lists describes which parts of the respective figure was generated for this thesis or taken from previous publications or a previous master's thesis.

Figure 1: Parts of this figure were taken from (Schlautmann and Gehring 2020). Otherwise, this figure was generated for this thesis.

Figure 2: This figure was modified from (Schlautmann 2018).

Figure 3: This figure was generated for this thesis.

Figure 4: This figure was produced for this thesis. The EJC core structure can be accessed on PDB with the accession number 2J0S (Bono, Ebert et al. 2006).

Figure 5: Parts of this figure were taken from (Schlautmann and Gehring 2020). Otherwise, this figure was generated for this thesis.

Figure 6: This figure was produced for this thesis. The structure of the minimal ASAP complex can be accessed on PDB with the accession number 4A8X (Murachelli, Ebert et al. 2012).

Figure 7: Parts of this figure were taken from (Schlautmann, Lackmann et al. 2022). Otherwise, this figure was generated for this thesis.

Figure 8: The panels (B) and (C) of this figure were taken from (Schlautmann, Lackmann et al. 2022), the panel (A) was taken from the same study and complemented.

Figure 9: The panels (A) to (D) were produced for this thesis, panel (E) was modified from (Schlautmann, Lackmann et al. 2022).

Figure 10: This figure was taken from (Schlautmann, Lackmann et al. 2022).

Figure 11: This figure was taken from (Schlautmann, Lackmann et al. 2022).

Figure 12: Panel (A) was taken from (Schlautmann, Lackmann et al. 2022), panel (B) was recalculated for this thesis.

Figure 13: This figure was taken and modified from (Schlautmann, Lackmann et al. 2022).

Figure 14: Panel (A) was taken from (Schlautmann, Lackmann et al. 2022), panels (B) and (C) were produced for this thesis.

Figure 15: This figure was taken from (Schlautmann, Lackmann et al. 2022).

Figure 16: This figure was taken from (Schlautmann, Lackmann et al. 2022) and complemented for this thesis.

Figure 17: This figure was taken from (Schlautmann, Lackmann et al. 2022).

Figure 18: This figure was taken from (Schlautmann 2018, Schlautmann, Lackmann et al. 2022).

Figure 19: This figure was generated for this thesis.

Figure 20: This figure was taken from (Schlautmann 2018, Schlautmann, Lackmann et al. 2022).

Figure 21: This figure was taken from (Schlautmann, Lackmann et al. 2022).

Figure 22: This figure was taken from (Schlautmann, Lackmann et al. 2022).

11 PUBLICATION

Title:

"Exon junction complex-associated multi-adapter RNPS1 nucleates splicing regulatory complexes to maintain transcriptome surveillance"

Published in:

Nucleic Acids Research, Vol. 50, No. 10

Publishing date:

30 May 2022

Accession:

(Schlautmann, Lackmann et al. 2022)

DOI: 10.1093/nar/gkac428

Author contributions:

Conceptualization: Lena P. Schlautmann, Volker Boehm, Niels H. Gehring

Methodology: Lena P. Schlautmann, Volker Boehm, Niels H. Gehring

Software: Lena P. Schlautmann, Volker Boehm

Investigation: Lena P. Schlautmann, Volker Boehm and Jan-Wilm Lackmann

Resources and Data Curation: Volker Boehm, Janine Altmüller and Jan-Wilm Lackmann

Writing – Original Draft, Review & Editing: Lena P. Schlautmann, Volker Boehm, Niels H. Gehring

Visualization: Lena P. Schlautmann and Volker Boehm

Supervision: Volker Boehm and Niels H. Gehring

Funding Acquisition: Christoph Dieterich and Niels H. Gehring

12 ERKLÄRUNG

„Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht."

Teilpublikation: “Exon junction complex-associated multi-adapter RNPS1 nucleates splicing regulatory complexes to maintain transcriptome surveillance”

Datum, Name und Unterschrift

Köln, den Lena Schlautmann

Digital unterschrieben von Lena Schlautmann
DN: cn=Lena Schlautmann, o, ou, email=lschlau2@uni-koeln.de, c=DE
Datum: 2023.02.09 11:24:59 +01'00'

Lena Pia Schlautmann

13 REFERENCES

(2021). "UniProt: the universal protein knowledgebase in 2021." Nucleic Acids Res **49**(D1): D480-d489.

Albers, C. A., D. S. Paul, H. Schulze, K. Freson, J. C. Stephens, P. A. Smethurst, J. D. Jolley, A. Cvejic, M. Kostadima, P. Bertone, M. H. Breuning, N. Debili, P. Deloukas, R. Favier, J. Fiedler, C. M. Hobbs, N. Huang, M. E. Hurles, G. Kiddle, I. Krapels, P. Nurden, C. A. Ruivenkamp, J. G. Sambrook, K. Smith, D. L. Stemple, G. Strauss, C. Thys, C. van Geet, R. Newbury-Ecob, W. H. Ouwehand and C. Ghevaert (2012). "Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome." Nat Genet **44**(4): 435-439, s431-432.

Alexandrov, A., D. Colognori, M. D. Shu and J. A. Steitz (2012). "Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay." Proc Natl Acad Sci U S A **109**(52): 21313-21318.

Amrani, N., R. Ganesan, S. Kervestin, D. A. Mangus, S. Ghosh and A. Jacobson (2004). "A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay." Nature **432**(7013): 112-118.

Anders, S., A. Reyes and W. Huber (2012). "Detecting differential usage of exons from RNA-seq data." Genome Res **22**(10): 2008-2017.

Andersen, C. B., L. Ballut, J. S. Johansen, H. Chamieh, K. H. Nielsen, C. L. Oliveira, J. S. Pedersen, B. Séraphin, H. Le Hir and G. R. Andersen (2006). "Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA." Science **313**(5795): 1968-1972.

Aravind, L. and E. V. Koonin (2000). "SAP – a putative DNA-binding motif involved in chromosomal organization." Trends in Biochemical Sciences **25**(3): 112-114.

Ashton-Beaucage, D., C. M. Udell, H. Lavoie, C. Baril, M. Lefrancois, P. Chagnon, P. Gendron, O. Caron-Lizotte, E. Bonneil, P. Thibault and M. Therrien (2010). "The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in Drosophila." Cell **143**(2): 251-262.

Ballut, L., B. Marchadier, A. Baguet, C. Tomasetto, B. Seraphin and H. Le Hir (2005). "The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity." Nat Struct Mol Biol **12**(10): 861-869.

Barbosa, I., N. Haque, F. Fiorini, C. Barrandon, C. Tomasetto, M. Blanchette and H. Le Hir (2012). "Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly." Nat Struct Mol Biol **19**(10): 983-990.

Blazquez, L., W. Emmett, R. Faraway, J. M. B. Pineda, S. Bajew, A. Gohr, N. Haberman, C. R. Sibley, R. K. Bradley, M. Irimia and J. Ule (2018). "Exon Junction Complex Shapes the Transcriptome by Repressing Recursive Splicing." Mol Cell **72**(3): 496-509 e499.

Boehm, V., T. Britto-Borges, A. L. Steckelberg, K. K. Singh, J. V. Gerbracht, E. Gueney, L. Blazquez, J. Altmuller, C. Dieterich and N. H. Gehring (2018). "Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity." Mol Cell **72**(3): 482-495 e487.

Boehm, V., S. Kueckelmann, J. V. Gerbracht, S. Kallabis, T. Britto-Borges, J. Altmuller, M. Kruger, C. Dieterich and N. H. Gehring (2021). "SMG5-SMG7 authorize nonsense-mediated mRNA decay by enabling SMG6 endonucleolytic activity." Nat Commun **12**(1): 3965.

Bono, F., J. Ebert, E. Lorentzen and E. Conti (2006). "The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA." Cell **126**(4): 713-725.

Boucher, L., C. A. Ouzounis, A. J. Enright and B. J. Blencowe (2001). "A genome-wide survey of RS domain proteins." Rna **7**(12): 1693-1701.

Brandner, J. M., S. Reidenbach and W. W. Franke (1997). "Evidence that "pinin", reportedly a differentiation-specific desmosomal protein, is actually a widespread nuclear protein." Differentiation **62**(3): 119-127.

Branon, T. C., J. A. Bosch, A. D. Sanchez, N. D. Udeshi, T. Svinkina, S. A. Carr, J. L. Feldman, N. Perrimon and A. Y. Ting (2018). "Efficient proximity labeling in living cells and organisms with TurboID." Nature Biotechnology **36**(9): 880-887.

Buchwald, G., J. Ebert, C. Basquin, J. Sauliere, U. Jayachandran, F. Bono, H. Le Hir and E. Conti (2010). "Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex." Proc Natl Acad Sci U S A **107**(22): 10050-10055.

Busetto, V., I. Barbosa, J. Basquin, E. Marquenet, R. Hocq, M. Hennion, J. A. Paternina, A. Namane, E. Conti, O. Bensaude and H. Le Hir (2020). "Structural and functional insights into CWC27/CWC22 heterodimer linking the exon junction complex to spliceosomes." Nucleic Acids Res **48**(10): 5670-5683.

Colombo, M., E. D. Karousis, J. Bourquin, R. Bruggmann and O. Mühlemann (2017). "Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways." Rna **23**(2): 189-201.

Costa, E., S. Canudas, I. Garcia-Bassets, S. Pérez, I. Fernández, E. Giralt, F. Azorín and M. L. Espinás (2006). "Drosophila dSAP18 is a nuclear protein that associates with chromosomes and the nuclear matrix, and interacts with pinin, a protein factor involved in RNA splicing." Chromosome Research **14**(5): 515-526.

Cox, J., M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj and M. Mann (2014). "Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ." Mol Cell Proteomics **13**(9): 2513-2526.

Cox, J., N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen and M. Mann (2011). "Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment." Journal of Proteome Research **10**(4): 1794-1805.

Czaplinski, K., M. J. Ruiz-Echevarria, S. V. Paushkin, X. Han, Y. Weng, H. A. Perlick, H. C. Dietz, M. D. Ter-Avanesyan and S. W. Peltz (1998). "The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs." Genes Dev **12**(11): 1665-1677.

De Conti, L., M. Baralle and E. Buratti (2013). "Exon and intron definition in pre-mRNA splicing." Wiley Interdiscip Rev RNA **4**(1): 49-60.

Dever, T. E. and R. Green (2012). "The elongation, termination, and recycling phases of translation in eukaryotes." Cold Spring Harb Perspect Biol **4**(7): a013706.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras (2013). "STAR: ultrafast universal RNA-seq aligner." Bioinformatics **29**(1): 15-21.

Eberle, A. B., S. Lykke-Andersen, O. Mühlemann and T. H. Jensen (2009). "SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells." Nat Struct Mol Biol **16**(1): 49-55.

Favaro, F. P., L. Alvizi, R. M. Zechi-Ceide, D. Bertola, T. M. Felix, J. de Souza, S. Raskin, S. R. Twigg, A. M. Weiner, P. Armas, E. Margarit, N. B. Calcaterra, G. R. Andersen, S. J. McGowan, A. O. Wilkie, A. Richieri-Costa, M. L. de Almeida and M. R. Passos-Bueno (2014). "A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects." Am J Hum Genet **94**(1): 120-128.

Frankish, A., M. Diekhans, A. M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J. M. Mudge, C. Sisu, J. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, S. Carbonell Sala, J. Chrast, F. Cunningham, T. Di Domenico, S. Donaldson, I. T. Fiddes, C. García Girón, J. M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, T. Hunt, O. G. Izuogu, J. Lagarde, F. J. Martin, L. Martínez, S. Mohanan,

P. Muir, F. C. P. Navarro, A. Parker, B. Pei, F. Pozo, M. Ruffier, B. M. Schmitt, E. Stapleton, M. M. Suner, I. Sycheva, B. Uszczynska-Ratajczak, J. Xu, A. Yates, D. Zerbino, Y. Zhang, B. Aken, J. S. Choudhary, M. Gerstein, R. Guigó, T. J. P. Hubbard, M. Kellis, B. Paten, A. Reymond, M. L. Tress and P. Flicek (2019). "GENCODE reference annotation for the human and mouse genomes." Nucleic Acids Res **47**(D1): D766-d773.

Fukumura, K., S. Wakabayashi, N. Kataoka, H. Sakamoto, Y. Suzuki, K. Nakai, A. Mayeda and K. Inoue (2016). "The Exon Junction Complex Controls the Efficient and Faithful Splicing of a Subset of Transcripts Involved in Mitotic Cell-Cycle Progression." Int J Mol Sci **17**(8).

Garrido-Martín, D., E. Palumbo, R. Guigó and A. Breschi (2018). "ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization." PLoS Comput Biol **14**(8): e1006360.

Gatfield, D. and E. Izaurralde (2004). "Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila." Nature **429**(6991): 575-578.

Gehring, N. H., J. B. Kunz, G. Neu-Yilik, S. Breit, M. H. Viegas, M. W. Hentze and A. E. Kulozik (2005). "Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements." Mol Cell **20**(1): 65-75.

Gehring, N. H., G. Neu-Yilik, T. Schell, M. W. Hentze and A. E. Kulozik (2003). "Y14 and hUpf3b form an NMD-activating complex." Mol Cell **11**(4): 939-949.

Gerbracht, J. V., V. Boehm, T. Britto-Borges, S. Kallabis, J. L. Wiederstein, S. Ciriello, D. U. Aschemeier, M. Kruger, C. K. Frese, J. Altmuller, C. Dieterich and N. H. Gehring (2020). "CASC3 promotes transcriptome-wide activation of nonsense-mediated decay by the exon junction complex." Nucleic Acids Res **48**(15): 8626-8644.

Gonatopoulos-Pournatzis, T., R. Niibori, E. W. Salter, R. J. Weatheritt, B. Tsang, S. Farhangmehr, X. Liang, U. Braunschweig, J. Roth, S. Zhang, T. Henderson, E. Sharma, M. Quesnel-Vallieres, J. Permanyer, S. Maier, J. Georgiou, M. Irimia, N. Sonenberg, J. D. Forman-Kay, A. C. Gingras, G. L. Collingridge, M. A. Woodin, S. P. Cordes and B. J. Blencowe (2020). "Autism-Misregulated eIF4G Microexons Control Synaptic Translation and Higher Order Cognitive Functions." Mol Cell **77**(6): 1176-1192 e1116.

Gonatopoulos-Pournatzis, T., M. Wu, U. Braunschweig, J. Roth, H. Han, A. J. Best, B. Raj, M. Aregger, D. O'Hanlon, J. D. Ellis, J. A. Calarco, J. Moffat, A. C. Gingras and B. J. Blencowe (2018). "Genome-wide CRISPR-Cas9 Interrogation of Splicing Networks Reveals a Mechanism for Recognition of Autism-Misregulated Neuronal Microexons." Mol Cell **72**(3): 510-524 e512.

Gromadzka, A. M., A. L. Steckelberg, K. K. Singh, K. Hofmann and N. H. Gehring (2016). "A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs." Nucleic Acids Res **44**(5): 2348-2361.

Haremaki, T. and D. C. Weinstein (2012). "Eif4a3 is required for accurate splicing of the Xenopus laevis ryanodine receptor pre-mRNA." Dev Biol **372**(1): 103-110.

Hayashi, R., D. Handler, D. Ish-Horowicz and J. Brennecke (2014). "The exon junction complex is required for definition and excision of neighboring introns in Drosophila." Genes Dev **28**(16): 1772-1785.

Haynes, C. and L. M. Iakoucheva (2006). "Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins." Nucleic Acids Res **34**(1): 305-312.

Hughes, C. S., S. Foehr, D. A. Garfield, E. E. Furlong, L. M. Steinmetz and J. Krijgsveld (2014). "Ultrasensitive proteome analysis using paramagnetic bead technology." Mol Syst Biol **10**(10): 757.

Huntzinger, E., I. Kashima, M. Fauser, J. Saulière and E. Izaurralde (2008). "SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan." Rna **14**(12): 2609-2617.

Jo, B. S. and S. S. Choi (2015). "Introns: The Functional Benefits of Introns in Genomes." Genomics Inform **13**(4): 112-118.

Kashima, I., S. Jonas, U. Jayachandran, G. Buchwald, E. Conti, A. N. Lupas and E. Izaurralde (2010). "SMG6 interacts with the exon junction complex via two conserved EJC-binding motifs (EBMs) required for nonsense-mediated mRNA decay." Genes Dev **24**(21): 2440-2450.

Kashima, I., A. Yamashita, N. Izumi, N. Kataoka, R. Morishita, S. Hoshino, M. Ohno, G. Dreyfuss and S. Ohno (2006). "Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay." Genes Dev **20**(3): 355-367.

Kim, V. N., N. Kataoka and G. Dreyfuss (2001). "Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex." Science **293**(5536): 1832-1836.

Kishor, A., S. E. Fritz and J. R. Hogg (2019). "Nonsense-mediated mRNA decay: The challenge of telling right from wrong in a complex transcriptome." Wiley Interdiscip Rev RNA **10**(6): e1548.

Kornblihtt, A. R., I. E. Schor, M. Allo, G. Dujardin, E. Petrillo and M. J. Munoz (2013). "Alternative splicing: a pivotal step between eukaryotic transcription and translation." Nat Rev Mol Cell Biol **14**(3): 153-165.

Kwon, I., S. Xiang, M. Kato, L. Wu, P. Theodoropoulos, T. Wang, J. Kim, J. Yun, Y. Xie and S. L. McKnight (2014). "Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells." Science **345**(6201): 1139-1145.

Le Hir, H., E. Izaurralde, L. E. Maquat and M. J. Moore (2000). "The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions." Embo j **19**(24): 6860-6869.

Li, C., R. I. Lin, M. C. Lai, P. Ouyang and W. Y. Tarn (2003). "Nuclear Pnn/DRS protein binds to spliced mRNPs and participates in mRNA processing and export via interaction with RNPS1." Mol Cell Biol **23**(20): 7363-7376.

Li, Y. I., D. A. Knowles, J. Humphrey, A. N. Barbeira, S. P. Dickinson, H. K. Im and J. K. Pritchard (2018). "Annotation-free quantification of RNA splicing using LeafCutter." Nat Genet **50**(1): 151-158.

Loh, B., S. Jonas and E. Izaurralde (2013). "The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2." Genes Dev **27**(19): 2125-2138.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biol **15**(12): 550.

Lykke-Andersen, J., M.-D. Shu and J. A. Steitz (2001). "Communication of the Position of Exon-Exon Junctions to the mRNA Surveillance Machinery by the Protein RNPS1." Science **293**(5536): 1836-1839.

Lykke-Andersen, J., M. D. Shu and J. A. Steitz (2000). "Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon." Cell **103**(7): 1121-1131.

Lykke-Andersen, S., Y. Chen, B. R. Ardal, B. Lilje, J. Waage, A. Sandelin and T. H. Jensen (2014). "Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes." Genes Dev **28**(22): 2498-2517.

Mabin, J. W., L. A. Woodward, R. D. Patton, Z. Yi, M. Jia, V. H. Wysocki, R. Bundschuh and G. Singh (2018). "The Exon Junction Complex Undergoes a Compositional Switch that Alters mRNP Structure and Nonsense-Mediated mRNA Decay Activity." Cell Rep **25**(9): 2431-2446 e2437.

Mäkitie, A. A., P. P. d. Reis, S. Arora, C. MacMillan, G. C. Warner, M. Sukhai, I. Dardick, B. Perez-Ordonez, R. Wells, D. Brown, R. Gilbert, J. Freeman, P. Gullane, J. Irish and S. Kamel-Reid (2005). "Molecular characterization of salivary gland malignancy using the Smgb-Tag

transgenic mouse model." Laboratory Investigation 85(8): 947-961.

Malone, C. D., C. Mestdagh, J. Akhtar, N. Kreim, P. Deinhard, R. Sachidanandam, J. Treisman and J. Y. Roignant (2014). "The exon junction complex controls transposable element activity by ensuring faithful splicing of the piwi transcript." Genes Dev 28(16): 1786-1799.

Maniatis, T. and B. Tasic (2002). "Alternative pre-mRNA splicing and proteome expansion in metazoans." Nature 418(6894): 236-243.

Mao, H., L. J. Pilaz, J. J. McMahon, C. Golzio, D. Wu, L. Shi, N. Katsanis and D. L. Silver (2015). "Rbm8a haploinsufficiency disrupts embryonic cortical development resulting in microcephaly." J Neurosci 35(18): 7003-7018.

Mayeda, A., J. Badolato, R. Kobayashi, M. Q. Zhang, E. M. Gardiner and A. R. Krainer (1999). "Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing." Embo j 18(16): 4560-4570.

McCallum, S. A., J. F. Bazan, M. Merchant, J. Yin, B. Pan, F. J. de Sauvage and W. J. Fairbrother (2006). "Structure of SAP18: a ubiquitin fold in histone deacetylase complex assembly." Biochemistry 45(39): 11974-11982.

McMahon, J. J., E. E. Miller and D. L. Silver (2016). "The exon junction complex in neural development and neurodevelopmental disease." Int J Dev Neurosci 55: 117-123.

Middleton, R., D. Gao, A. Thomas, B. Singh, A. Au, J. J. Wong, A. Bomane, B. Cosson, E. Eyras, J. E. Rasko and W. Ritchie (2017). "IRFinder: assessing the impact of intron retention on mammalian gene expression." Genome Biol 18(1): 51.

Miller, J. N. and D. A. Pearce (2014). "Nonsense-mediated decay in genetic disease: friend or foe?" Mutat Res Rev Mutat Res 762: 52-64.

Moore, M. J. and N. J. Proudfoot (2009). "Pre-mRNA processing reaches back to transcription and ahead to translation." Cell 136(4): 688-700.

Murachelli, A. G., J. Ebert, C. Basquin, H. Le Hir and E. Conti (2012). "The structure of the ASAP core complex reveals the existence of a Pinin-containing PSAP complex." Nat Struct Mol Biol 19(4): 378-386.

Nagy, E. and L. E. Maquat (1998). "A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance." Trends in Biochemical Sciences 23(6): 198-199.

Nguyen, L. S., H.-G. Kim, J. A. Rosenfeld, Y. Shen, J. F. Gusella, Y. Lacassie, L. C. Layman, L. G. Shaffer and J. Gécz (2013). "Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neuro-developmental disorders." Human Molecular Genetics 22(9): 1816-1825.

Okada-Katsuhata, Y., A. Yamashita, K. Kutsuzawa, N. Izumi, F. Hirahara and S. Ohno (2012). "N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD." Nucleic Acids Res 40(3): 1251-1266.

Ouyang, P. and S. P. Sugrue (1992). "Identification of an epithelial protein related to the desmosome and intermediate filament network." J Cell Biol 118(6): 1477-1488.

Ouyang, P. and S. P. Sugrue (1996). "Characterization of pinin, a novel protein associated with the desmosome-intermediate filament complex." J Cell Biol 135(4): 1027-1042.

Patro, R., G. Duggal, M. I. Love, R. A. Irizarry and C. Kingsford (2017). "Salmon provides fast and bias-aware quantification of transcript expression." Nat Methods 14(4): 417-419.

Pedrotti, S. and T. A. Cooper (2014). "In Brief: (mis)splicing in disease." J Pathol 233(1): 1-3.

Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi and G. K. Smyth (2015). "limma powers differential expression analyses for RNA-sequencing and microarray studies." Nucleic Acids Res 43(7): e47.

Robinson, M. D. and A. Oshlack (2010). "A scaling normalization method for differential expression analysis of RNA-seq data." Genome Biol 11(3): R25.

Rodor, J., Q. Pan, B. J. Blencowe, E. Eyras and J. F. Caceres (2016). "The RNA-binding profile of Acinus, a peripheral component of the exon junction complex, reveals its role in splicing regulation." RNA **22**(9): 1411-1426.

Roignant, J. Y. and J. E. Treisman (2010). "Exon junction complex subunits are required to splice Drosophila MAP kinase, a large heterochromatic gene." Cell **143**(2): 238-250.

Roth, M. B., A. M. Zahler and J. A. Stolk (1991). "A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription." J Cell Biol **115**(3): 587-596.

Sahara, S., M. Aoto, Y. Eguchi, N. Imamoto, Y. Yoneda and Y. Tsujimoto (1999). "Acinus is a caspase-3-activated protein required for apoptotic chromatin condensation." Nature **401**(6749): 168-173.

Sakashita, E., S. Tatsumi, D. Werner, H. Endo and A. Mayeda (2004). "Human RNPS1 and its associated factors: a versatile alternative pre-mRNA splicing regulator in vivo." Mol Cell Biol **24**(3): 1174-1187.

Sauliere, J., V. Murigneux, Z. Wang, E. Marquenet, I. Barbosa, O. Le Tonqueze, Y. Audic, L. Paillard, H. Roest Crollius and H. Le Hir (2012). "CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex." Nat Struct Mol Biol **19**(11): 1124-1131.

Schlautmann, L. P. (2018). Mechanistic details of alternative splicing regulation by the exon junction complex. Master of Science, University of Cologne.

Schlautmann, L. P. and N. H. Gehring (2020). "A Day in the Life of the Exon Junction Complex." Biomolecules **10**(6).

Schlautmann, L. P., J. W. Lackmann, J. Altmuller, C. Dieterich, V. Boehm and N. H. Gehring (2022). "Exon junction complex-associated multi-adapter RNPS1 nucleates splicing regulatory complexes to maintain transcriptome surveillance." Nucleic Acids Res **50**(10): 5899-5918.

Schwerk, C., J. Prasad, K. Degenhardt, H. Erdjument-Bromage, E. White, P. Tempst, V. J. Kidd, J. L. Manley, J. M. Lahti and D. Reinberg (2003). "ASAP, a novel protein complex involved in RNA processing and apoptosis." Mol Cell Biol **23**(8): 2981-2990.

Shen, S., J. W. Park, Z.-x. Lu, L. Lin, M. D. Henry, Y. N. Wu, Q. Zhou and Y. Xing (2014). "rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data." Proceedings of the National Academy of Sciences **111**(51): E5593-E5601.

Shepard, P. J. and K. J. Hertel (2009). "The SR protein family." Genome Biology **10**(10): 242.

Silver, D. L., D. E. Watkins-Chow, K. C. Schreck, T. J. Pierfelice, D. M. Larson, A. J. Burnetti, H. J. Liaw, K. Myung, C. A. Walsh, N. Gaiano and W. J. Pavan (2010). "The exon junction complex component Magoh controls brain size by regulating neural stem cell division." Nat Neurosci **13**(5): 551-558.

Singh, G., A. Kucukural, C. Cenik, J. D. Leszyk, S. A. Shaffer, Z. Weng and M. J. Moore (2012). "The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus." Cell **151**(4): 750-764.

Somers, J., T. Pöyry and A. E. Willis (2013). "A perspective on mammalian upstream open reading frame function." Int J Biochem Cell Biol **45**(8): 1690-1700.

Soneson, C., M. I. Love and M. D. Robinson (2015). "Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences." F1000Res **4**: 1521.

Steckelberg, A. L., J. Altmueller, C. Dieterich and N. H. Gehring (2015). "CWC22-dependent pre-mRNA splicing and eIF4A3 binding enables global deposition of exon junction complexes." Nucleic Acids Res **43**(9): 4687-4700.

Steckelberg, A. L., V. Boehm, A. M. Gromadzka and N. H. Gehring (2012). "CWC22 connects pre-mRNA splicing and exon junction complex assembly." Cell Rep **2**(3): 454-461.

Tange, T., T. Shibuya, M. S. Jurica and M. J. Moore (2005). "Biochemical analysis of the EJC

reveals two new factors and a stable tetrameric protein core." Rna **11**(12): 1869-1883.

Tyanova, S., T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann and J. Cox (2016). "The Perseus computational platform for comprehensive analysis of (prote)omics data." Nat Methods **13**(9): 731-740.

Viegas, M. H., N. H. Gehring, S. Breit, M. W. Hentze and A. E. Kulozik (2007). "The abundance of RNPS1, a protein component of the exon junction complex, can determine the variability in efficiency of the Nonsense Mediated Decay pathway." Nucleic Acids Res **35**(13): 4542-4551.

Viphakone, N., I. Sudbery, L. Griffith, C. G. Heath, D. Sims and S. A. Wilson (2019). "Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome." Mol Cell **75**(2): 310-323 e318.

Vitting-Seerup, K. and A. Sandelin (2017). "The Landscape of Isoform Switches in Human Cancers." Mol Cancer Res **15**(9): 1206-1220.

Vitting-Seerup, K. and A. Sandelin (2019). "IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences." Bioinformatics **35**(21): 4469-4471.

Wahl, M. C., C. L. Will and R. Luhrmann (2009). "The spliceosome: design principles of a dynamic RNP machine." Cell **136**(4): 701-718.

Wallmeroth, D., J. W. Lackmann, S. Kueckelmann, J. Altmüller, C. Dieterich, V. Boehm and N. H. Gehring (2022). "Human UPF3A and UPF3B enable fault-tolerant activation of nonsense-mediated mRNA decay." Embo j **41**(10): e109191.

Wang, F., K. J. Soprano and D. R. Soprano (2015). "Role of Acinus in regulating retinoic acid-responsive gene pre-mRNA splicing." J Cell Physiol **230**(4): 791-801.

Wang, G.-S. and T. A. Cooper (2007). "Splicing in disease: disruption of the splicing code and the decoding machinery." Nature Reviews Genetics **8**(10): 749-761.

Wang, P., P. J. Lou, S. Leu and P. Ouyang (2002). "Modulation of alternative pre-mRNA splicing in vivo by pinin." Biochem Biophys Res Commun **294**(2): 448-455.

Wang, Y., M. Ma, X. Xiao and Z. Wang (2012). "Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules." Nature Structural & Molecular Biology **19**(10): 1044-1052.

Wang, Z., L. Ballut, I. Barbosa and H. Le Hir (2018). "Exon Junction Complexes can have distinct functional flavours to regulate specific splicing events." Sci Rep **8**(1): 9509.

Wang, Z., V. Murigneux and H. Le Hir (2014). "Transcriptome-wide modulation of splicing by the exon junction complex." Genome Biol **15**(12): 551.

Wilkinson, M. E., C. Charenton and K. Nagai (2020). "RNA Splicing by the Spliceosome." Annu Rev Biochem **89**: 359-388.

Woodward, L. A., J. W. Mabin, P. Gangras and G. Singh (2017). "The exon junction complex: a lifelong guardian of mRNA fate." Wiley Interdiscip Rev RNA **8**(3).

Zahler, A. M., W. S. Lane, J. A. Stolk and M. B. Roth (1992). "SR proteins: a conserved family of pre-mRNA splicing factors." Genes Dev **6**(5): 837-847.

Zhang, J., X. Sun, Y. Qian and L. E. Maquat (1998). "Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm." Rna **4**(7): 801-815.

Zhang, Y., R. Iratni, H. Erdjument-Bromage, P. Tempst and D. Reinberg (1997). "Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex." Cell **89**(3): 357-364.

Zhang, Z., M. Guo, Y. Liu, P. Liu, X. Cao, Y. Xu and X. Zhu (2020). "RNPS1 inhibition aggravates ischemic brain injury and promotes neuronal death." Biochem Biophys Res Commun **523**(1): 39-45.

Zhao, X., S. Bitsch, L. Kubitz, K. Schmitt, L. Deweid, A. Roehrig, E. C. Barazzone, O. Valerius, H. Kolmar and J. Béthune (2021). "ultraID: a compact and efficient enzyme for proximity-dependent biotinylation in living cells." bioRxiv: 2021.2006.2016.448656.

Zhong, X., J. H. Choi, S. Hildebrand, S. Ludwig, J. Wang, E. Nair-Gill, T.-C. Liao, J. J. Moresco, A. Liu, J. Quan, Q. Sun, D. Zhang, X. Zhan, M. Choi, X. Li, J. Wang, T. Gallagher, E. M. Y. Moresco and B. Beutler (2022). "RNPS1 inhibits excessive tumor necrosis factor/tumor necrosis factor receptor signaling to support hematopoiesis in mice." Proceedings of the National Academy of Sciences **119**(18): e2200128119.

14 ACKNOWLEDGEMENTS

Zuallererst geht ein herzliches Dankeschön an Prof. Niels Gehring. Er hat mir als mein Doktorvater jederzeit, auch während der Corona-Pandemie, mit seinem Rat zur Seite gestanden. Nicht nur Diskussionen über wissenschaftliche Themen, aber auch über das aktuelle Weltgeschehen haben es mir ermöglicht, mich auf vielen Ebenen weiterzuentwickeln.

Ich möchte auch Prof. Kay Hofmann und Prof. Ulrich Baumann danken, dass sie mir als Zweitgutachter und Prüfungsvorsitzender zur Verfügung stehen.

Ebenso bedanke ich mich bei der Graduate School for Biological Sciences, insbesondere bei Isabell Witt, für die Unterstützung bei organisatorischen Dingen, sowie bei Prof. Andreas Beyer und Dr. Martin Denzel, die mir als Teil meines Thesis Advisory Committee wertvolle Ratschläge mit auf den Weg gegeben haben.

Mein besonderer Dank gilt den aktuellen und ehemaligen Kollegen des Gehring-Labs. Die entspannte und freundschaftliche Arbeitsatmosphäre, Hilfsbereitschaft und der Zusammenhalt im Labor sind wirklich außergewöhnlich. Ich habe die Zeit meiner Thesis im Gehring-Lab sehr genossen. Außerdem danke ich Damaris Wallmeroth, Sabrina Kückelmann und Volker Böhm für das Korrekturlesen dieser Arbeit.

Zuletzt geht noch ein großer Dank an meine Familie und Freunde, die mich während meines gesamten universitären Werdegangs unterstützt haben.

15 SUPPLEMENT

15.1 Tables

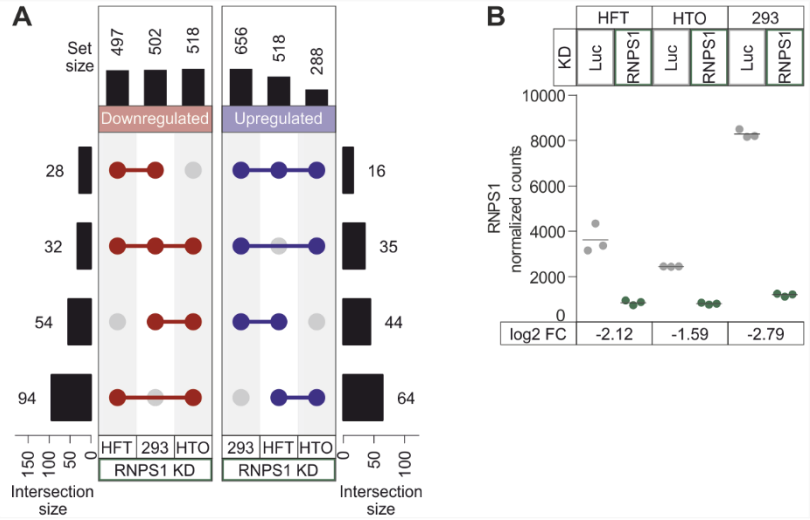
Table 1: Additional Plasmids used in this thesis.

Name	Vector	Internal ID
pCI MS2V5 MAGOH	pCI-neo	1985
pCI MS2V5 RNPS1 Del C (1-255)	pCI-neo	2830
pCI MS2V5 RNPS1 Del S (69-121)	pCI-neo	2829
pCI-MS2V5 ACIN1 (101-Ter)	pCI-neo	879
pCI-MS2V5 ACIN1 (101-Ter, 465-GYA)	pCI-neo	880
pCI-MS2V5 PNN 1-381 241Mut	pCI-neo	2674
pCI-MS2V5 PNN 1-381 WT	pCI-neo	2673
pCI-MS2V5 RNPS1 Del N (siRes)	pCI-neo	2462
pCI-MS2V5 RNPS1 Del N+C (siRes)	pCI-neo	2463
pCI-MS2V5 RNPS1 Del S+C (siRes)	pCI-neo	2464
pCI-MS2V5 SAP18 siRes	pCI-neo	876
pCI-MS2V5 Y14	pCI-neo	1396

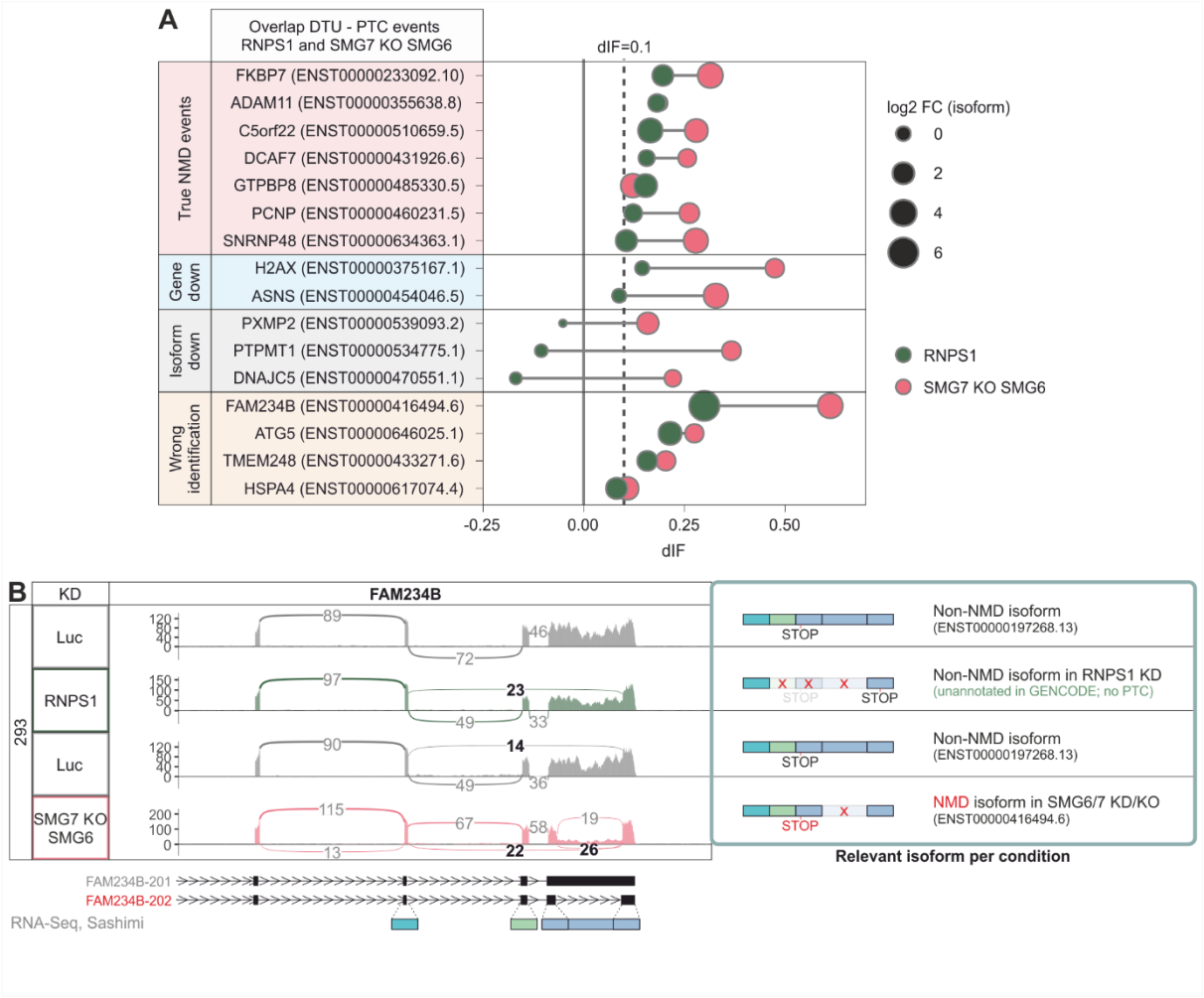
Table 2: Additional siRNAs used in this thesis.

Name	Target Sequence in Gene/ sense oligo	Source	Internal ID
ACIN1_1	CUGCAGAGCAUGAAGUAAAUU	Integrated DNA Technologies	S021
ACIN1_13.1	UCAGUAUCACCACUGAAUCACUAAA	Integrated DNA Technologies	S060
ACIN1_13.2	rCrArArGrUrGrArArArArCrArGrArCrCrUrGrArArArArUGA	Integrated DNA Technologies	S061
PNN_1	GAUUUCUUGAUAAAAAAGGAUUACC	Integrated DNA Technologies	S033
PNN_2	AUACUUCAGGACUAGAAAGAAGUCA	Integrated DNA Technologies	S034

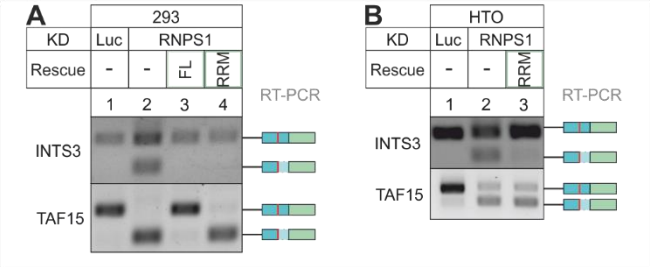
15.2 Supplementary figures



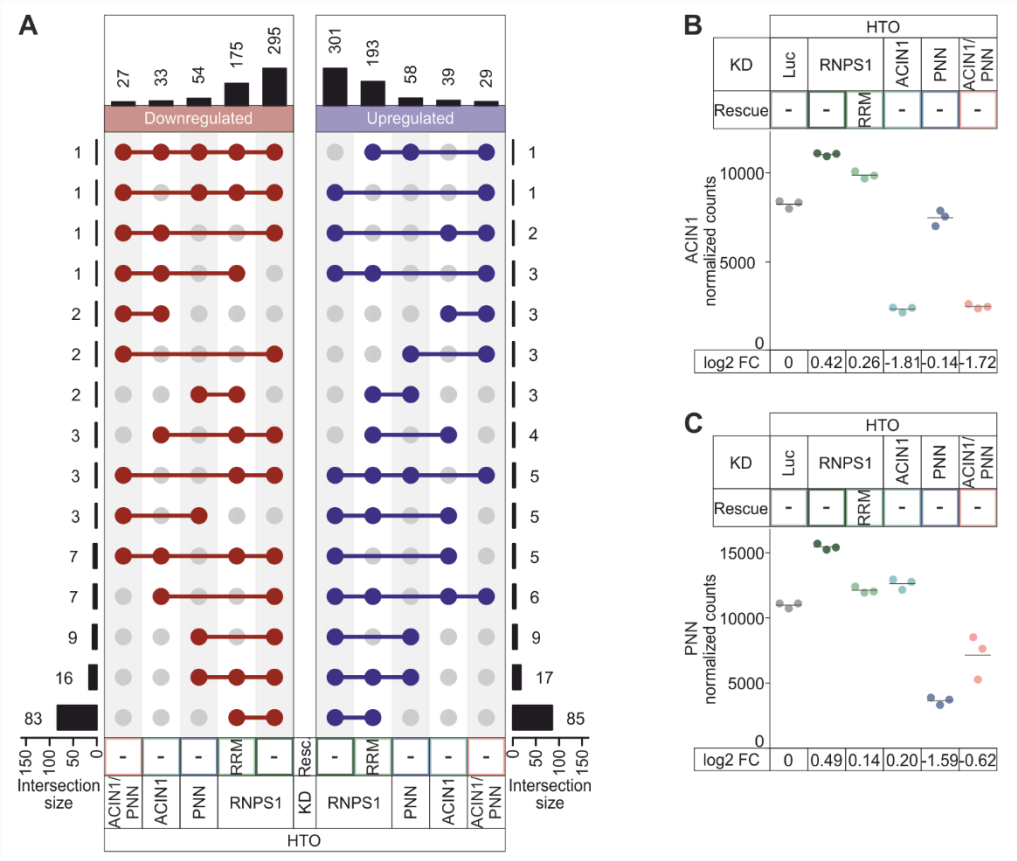
Supplementary Figure 1: Only few genes were consistently upregulated upon RNPS1 depletion in the different cell types. *(A) The UpSet plot depicts the commonly found differentially expressed genes upon RNPS1 KD in the indicated cell lines (Cutoffs: Padj < 0.05 and |log2 FC| > 1). (B) RNPS1 expression differences are displayed as normalized counts in the indicated conditions. On the bottom, the log2 FCs of RNPS1 expression in the KD condition compared to control condition is shown.*



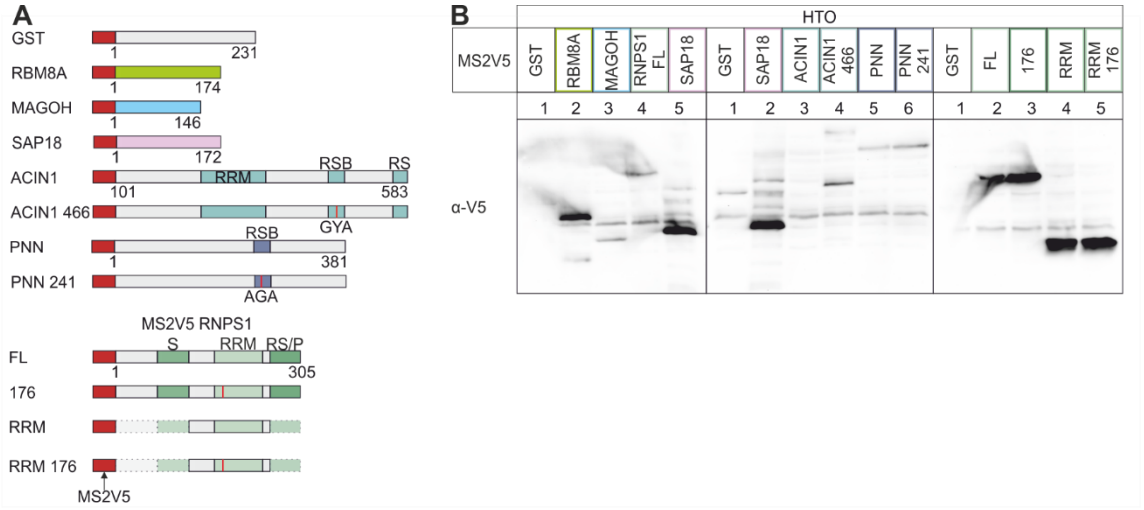
Supplementary Figure 2: Only few high-confidence NMD targets were upregulated in 293 cells depleted of RNPS1. *(A) Of the PTC-positive transcripts that were upregulated upon SMG6/7 KD/KO and found in RNPS1 KD, the dIF for both conditions is displayed. The transcripts are classified as: "True NMD events", "Gene down", "Isoform down" and "Wrong identification" according to their differential expression upon RNPS1 depletion. (B) On the left, the Sashimi plot of FAM234B depicts the read and junction coverage in RNPS1 or SMG6/7-depleted cells. The relevant junction reads are highlighted, and the NMD-sensitive isoform is labeled in red. On the right, the resulting transcripts are drawn schematically, and the corresponding transcript ID is noted.*



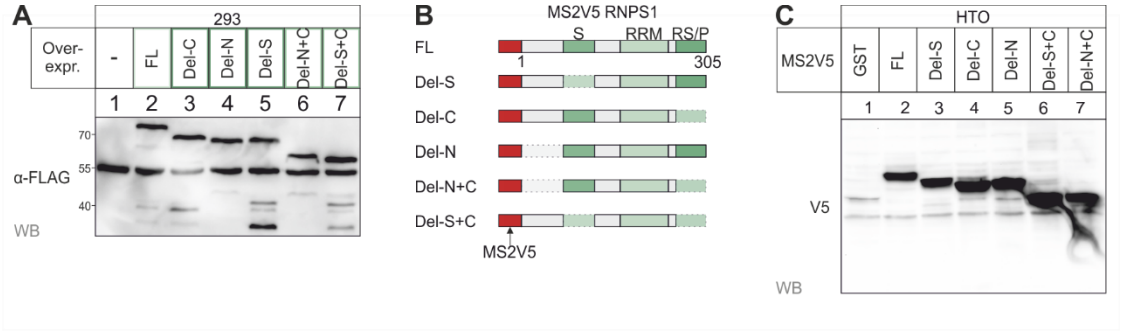
Supplementary Figure 3: The RNPS1 RRM rescues not all RNPS1-dependent alternatives splicing events. *Alternative splicing of the two RNPS1-dependent events in INTS3 and TAF15 is evaluated using RT-PCR in the indicated KD rescue conditions in (A) 293 and (B) HTO cells. A representative image of three replicates is shown. The corresponding PCR fragments are indicated on the right.*



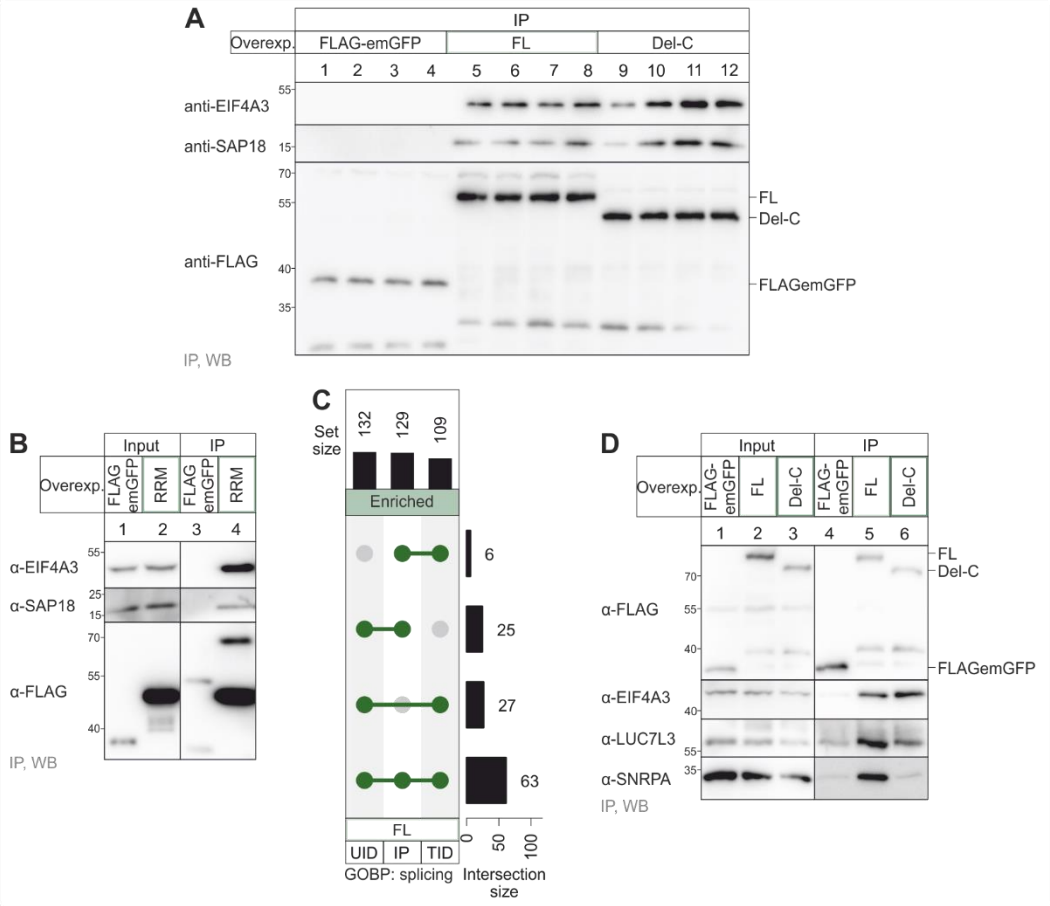
Supplementary Figure 4: ACIN1 and PNN depletions induce alternative splicing patterns that are distinct from RNPS1 KD induced patterns. *(A) The UpSet plot depicts the shared alternative splicing events among the ASAP/PSAP KD conditions (Cutoffs: Padj < 0.001 and |dPSI| > 0.1). (B, C) Expression is shown as normalized counts for (B) ACIN1, and (C) PNN. The log2 FCs for the conditions as compared to the control conditions are indicated below.*



Supplementary Figure 5:Expression of most MS2V5-tagged tethering constructs was confirmed using western blot. *(A) Schematic of the used control, EJC and ASAP/PSAP tethering constructs. (B) Western blot with antibody against V5 confirmed the expression of the indicated MS2V5-tagged constructs.*



Supplementary Figure 6: Rescue and tethering constructs of RNPS1 deletion mutants were expressed. *(A) The expression of the indicated RNPS1 deletion mutants was detected in a western blot. (B) Schematic and (C) western blot detection of the MS2V5-tagged RNPS1 deletion mutants used for the tethering assay.*



Supplementary Figure 7: RNPS1 interacts with EJC, ASAP/PSAP proteins and further splicing related factors. *(A) FLAG-IP of replicates of the indicated FLAG-emGFP control, FLAG-emGFP-tagged RNPS1 FL and FLAG-emGFP-tagged RNPS1 Del-C. Antibodies against FLAG, EIF4A3 and SAP18 were used to validate the expression of the constructs and verify EJC and ASAP/PSAP pulldown. (B) Same as in (A) but for the RNPS1 RRM. (C) Shared interaction partners in the three RNPS1 FL datasets that are classified as GOBP "splicing" are depicted in an UpSet plot. (D) Same as in (A) but pulldown of LUC7L3 and SNRPA was tested.*