JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Institute of Molecular Biology

INVESTIGATING THE MECHANISM OF THE EXON JUNCTION COMPLEX IN PRE-mRNA SPLICING AND ITS IMPLICATION IN THE piRNA PATHWAY

Dissertation zur Erlangung des Grades Doktor

der Naturwissenschaften am Fachbereich Biologie

der Johannes Gutenberg-Universität Mainz

Claire Mestdagh

Geboren am 27.08.1988 in Paris

Mainz, den 21. September 2015

ABSTRACT

The exon junction complex (EJC) is a highly conserved ribonucleoprotein complex which binds RNAs at a late stage of the splicing reaction and remains associated with them following export to the cytoplasm. This complex is involved in several cellular post-transcriptional processes including mRNA localization, translation and degradation. The EJC plays an additional role in the splicing of a subset of genes in *Drosophila* and in human cells but the underlying mechanism remains to be elucidated.

The goal of my PhD project was to decipher the mechanism of the EJC in the splicing process. To this end, we have searched for new developmentally regulated EJC targets. We have found a novel function for the EJC and its splicing subunit RnpS1 in preventing transposon accumulation in both *Drosophila* germline and surrounding follicular cells. We have shown that this function appears to be mediated through the control of the splicing of the *piwi* transcript. Piwi is involved in the piRNA pathway, a mechanism leading to transposon repression. In absence of RnpS1, one of the *piwi* intron is retained. This intron contains a weak 5' splice site as well as degenerate transposon fragments, reminiscent of heterochromatic introns. Furthermore, we identified a small A/T rich region, which alters its polypyrimidine tract and confers dependency to RnpS1. We demonstrated that the removal of this intron by RnpS1 requires the initial splicing of the flanking introns, suggesting a model in which the EJC facilitates the splicing of weak introns following its initial deposition to adjacent exon junctions.

This work provides an interesting mechanism underlying the cooperation between introns in facilitating their splicing. In addition, it opens new avenues in the investigation of the regulation of the piRNA pathway as well as in the understanding of the polarity and kinetics of the splicing reaction.

ZUSAMMENFASSUNG

Der Exon Junction Complex (EJC) ist ein hochkonservierter Ribonukleoproteinkomplex, der während der letzten Phase einer Splicingreaktion RNA bindet. Er bleibt mit der RNA verbunden und folgt ihr ins Zytoplasma, wo der Komplex eine Rolle in mehreren post-transkriptionellen Abläufen wie z.B. in der mRNA-Lokalisierung, der Translation und dem mRNA-Abbau spielt. Der EJC hat zudem eine Funktion beim Splicing verschiedener Gene in *Drosophila* und in menschlichen Zellen; die Aufklärung des genauen Mechanismus ist Gegenstand der aktuellen Forschung.

Das Projekt meiner Doktorarbeit zielte darauf ab, die Rolle des EJC beim Splicing weiter aufzuklären. Zu diesem Zweck haben wir neue Zielgene untersucht, die unter EJC-Regulierung stehen. Wir haben eine neue Funktion für den EJC und seine Splicinguntereinheit RnpS1 in der Vorbeugung der Transposonakkumulation in der Keimbahn und den umliegenden follikularen Zellen von *Drosophila* gefunden. Wir konnten zeigen, dass dies durch die Kontrolle des *piwi*-Transkript-Splicings passiert. Piwi ist in den piRNA-Prozess involviert, eine Reaktionskette, die zur Transpositionsunterdrückung führt. In Abwesenheit von RnpS1 bleibt eines der *piwi*-Introns erhalten, welches eine schwache 5‘-Splice-Site und degenerierte Transposonfragmente enthält. Dies sind charakteristische Eigenschaften heterochromatischer Introns. Wir haben eine kleine A/T-reiche Region identifizieren können, die den Polypyrimidine-Trakt ändert und die RnpS1-Abhängigkeit überträgt. Wir haben desweiteren zeigen können, dass das Entfernen dieses Introns durch RnpS1 das vorhergehende Splicing der umliegenden Introns voraussetzt. Das lässt ein Modell vermuten, in welchem der EJC nach vorheriger Deposition an flankierenden Exon-Exon Kontakten das Splicing von schwachen Introns erleichtert.

Diese Arbeit hat wichtige Fragen über piRNA-Prozess Regulation und Polarität und Kinetik in Splicingsreaktionskette eröffnet. Sie stellt auch eine interessantes Mechanismus zur Verfügung, über die Frage, wie Intron kooperieren, um ihnen Splicing zu verbessern.

TABLE OF CONTENT

ACKNOWLEDGMENTS

My first thanks go to my PhD supervisor, Jean-Yves Roignant, for giving me the opportunity to work in his lab and for guiding me during all these 3 years. Jean-Yves, je te remercie vivement pour ce que tu as pu m'apporter au cours de ma thèse: tes conseils, ta rigueur scientifique et ton orientation constante vers des pistes qui ont permis à des résultats captivants de voir le jour. Que ce soit d'un point de vue scientifique ou plus large, je ne regrette pas d'avoir signé pour ce drôle de métier et d'avoir mené mon projet de thèse à terme.

I would like to thank the members of my thesis advisory committee Gerhard Technau and Jean-René Huynh for taking the time to follow this project and for the helpful discussions during our annual meetings. I am also much honored that they took part of the examination panel to judge my work. I would also like to acknowledge the other committee members René Ketting and Joachim Urban for the interest they gave while revising my work.

This work would not have been possible without Colin Malone and Ravi Sachidanandam, two collaborators from New York. I would like to thank them for the technique of piRNA analysis they have brought to the publication we have co-authored.

I am also grateful to the members of the IMB core-facilities. In particular Jasmin Cartano and Nastasja Kreim did a terrific job in RNA-sequencing and computational analysis, which turned out to be very important for the success of this thesis.

I would like to acknowledge the Roignant group with whom I shared working space, buffers and good mood during working time and coffee breaks:

Junaid Akhtar (Junaid, I will probably not miss your horizontal and intrusive organization, but who will put me in a good mood and play "Daddy cool" overloud in my future working place?)

Maria Carla Antonelli (Maria Carla, hai saputo prendere la fiaccola del "progetto AGO3" con successo e hai svolto un ottimo lavoro. Ti auguro davvero il meglio per la tua carriera futura.)

Christine Bauer (Tine, wie schön es war, das Labor mit dir einzurichten! Ich muss sagen, dass seit du nicht mehr mit uns arbeitest, keiner auf „Kanne" gedrückt hat… aber warrrum?)

Laura Götzinger (Laura, danke für deine fleißige Arbeit bei den Fliegen. Ohne dich könnten sie mehrmals sterben… überhaupt meine armen 18°C Stöcke, die ich viel zu häufig vergessen habe.)

Giriram Kumar (Giri, I learnt so much from your constant calm! I was not conscious of this at that time, but it becomes now very useful for me, to overcome baby crying and tears.)

Anna Lena Leifke (Anna Lena, ich war immer beeindruckt von der Beharrlichkeit, mit der du die Vertreter behandelt hast. Als du da warst alles ist wieder ganz schön Flott gelaufen. Das war eine Freude, mit dir zu arbeiten, auch wenn „der ganze Käse im Labor in die Hose ging".)

Tina Lence (Tina, I really enjoyed our scientific discussions that were certainly the best I had during my PhD. I want to thank you also for your generosity, for giving so much and expecting so little.)

Nadja Schmid (Nadja, danke für deine Kritikfähigkeit, die mir geholfen hat, um meine letzten Experimenten für das „AGO3 Projekt" zu machen. Dank den Gelen, die du während meiner Schwangerschaft gegossen hast, ist der Raphaël ganz fit und wach!)

Many people helped in one way or another to make this thesis successful: explanation of a technique, lending materials, deal with the German bureaucracy, corrections, emotional support or "after-work blackout"... It is not realistic to mention all of you although a few names come to my mind immediately: Achim, Alexey, Andreas, Anke, die rote Anke, Bianca, Carina, Carine, Christina, Christoph, Doris, Elmar, Federico, Florie, Hanna, Hannah, Hans, Heike, Irina, Johann, Julian, Katharina, Katrin, Kerstin, Laura, Laure, Liliana, Linda, Michaela, Pierre, Raymond, Ronald, Simon, Sondes, Steffi, Ulli, Valentina... The list seems quite long but not enough ... Please don't take it too personally. The important thing: I thank all of you.

Finally I address the simplest but the strongest words to my family, accompanied by a particular sweet thought to the man who is sharing my life. The distance during all these (too many) years away from their intelligence, their affection and their trust made me realize that life is not only done of problems that one could resolve by simple lab experiments. Et à toi, le petit homme en herbe qui a éclairé ma dernière année de thèse sans le savoir, je te donne le plus précieux conseil que j'ai tiré de cette expérience : Même si tu te perds, ton temps ne sera jamais perdu. Chaque expérience a plusieurs facettes qui gagnent à être découvertes. Comprendre tes échecs et apprendre à les surmonter en regardant ce qu'il y a autour de toi te mènera toujours, certes parfois par des chemins détournés, là où tu veux vraiment aller. Et tu y prendras d'autant plus de plaisir.

ABBREVIATION LIST

ATP: adenosine tri-phosphate

BP: branch-point

BSA: bovine Serum Albumin

Btz: Barentsz, or MLN51: metastatic lymph node 51

CLIP-Seq: Crosslink immunoprecipitation -sequencing

C-ter/N-ter: C-terminus/ N-terminus

DNA: Deoxyribonucleic acid

eIF4A3: eukaryotic Initiation Factor A3

EJC: Exon Junction Complex

ESE: exonic splicing enhancer

ESS: exonic splicing silencer

hnRNP: heterogeneous nuclear ribonucleoprotein

ISE: intronic splicing enhancer

ISS: intronic splicing silencer

KD: knock down

kDa: kilo Dalton

LSm proteins (2 to 8)

Mago: Mago nashi

NLS: nuclear localization site

NMD: Nonsense Mediating Decay

piRNA: PIWI interacting RNA

PRP19: pre-mRNA Processing Factor 19

RNA pol II: RNA polymerase II, contains the CTD: C-terminal Domain

RNA: ribonucleic acid, mRNA: messenger RNA, pre-mRNA: pre-messenger RNA

RnpS1: RNA binding Protein S1

snRNP: small nuclear ribonucleoprotein particle

SR proteins: Serin Rich proteins composed from RRM (RNA recognition motif) and RS domain (arginine serine rich domain)

SS: Splice site

Tsu: Tsunagi, Y14

U2AF: U2 Auxiliary Factor

INTRODUCTION

PREAMBLE: PRE-mRNA SPLICING IS AN ESSENTIAL STEP OF GENE REGULATION

Protein synthesis is essential for cell division and survival. This process consumes an important amount of energy and its implementation is regulated by the available energy resources.

The piece of information that is necessary for protein production is carried by DNA (deoxyribonucleic acid) fragments, the so-called genes. The different steps of gene expression, giving rise to a functional protein, are conserved across all domains of life. During the initial step of transcription, the sequence of genes is copied from DNA into pre-mRNA (pre-messenger ribonucleic acid). Subsequently, the pre-mRNA undergoes a maturation stage giving rise to a mature mRNA. As a final step, the mRNA is used as a template for the translation process. This step is termed "translation" because the molecular template that carries the information has changed. The nucleic acid template switches to an amino acid sequence, which is the unit-component of the protein.

Most of the pre-mRNA maturation steps take place in the nucleus. The addition of specific sequences or modifications at the extremities of the pre-mRNA enhances its stability. Another important transformation, called splicing, consists in the removal of non-coding fragments present in the pre-mRNA (the introns), and the ligation of the remaining coding sequences (the exons). Altogether, these modifications give rise to the mature mRNA.

During my PhD studies, I focused my interest on the mechanism of the exon junction complex (EJC) in pre-mRNA splicing using *Drosophila melanogaster* as a model organism. The EJC is a macromolecular complex involved in posttranscriptional regulation of gene expression. In the introduction I will first describe the main actors involved in pre-mRNA splicing, their molecular functions and their regulations. Then, I will present the current knowledge regarding the molecular and physiological functions of the EJC. Finally, as I found a novel role for the EJC in the so-called piRNA pathway, I will also introduce the roles and mechanisms of this small RNA pathway.

1. MECHANISMS AND REGULATIONS OF PRE-mRNA SPLICING

In eukaryotes, transcription gives rise to pre-mRNA that undergoes at least three major maturation steps before being exported into the cytoplasm and translated into proteins. These transformations consist in -1- the addition of a methyl group at the 5' end (capping) -2- the excision of the introns (splicing) and -3- the cleavage and addition of multiple adenosine residues at the 3' end (polyadenylation).

A macromolecular complex called spliceosome, composed of both proteins and small non-coding RNA, catalyzes the splicing reaction. Conserved sequence elements at the exon-intron boundaries, the 5' and 3' splice sites of each exon, and the branchpoint facilitate spliceosome recruitment. Despite the apparent simplicity of the reaction, which consists in the recognition of exon-intron boundaries and the removal of introns, this process is highly regulated. Indeed, the spliceosome is one of the largest, if not the largest, molecular complex in the cell. Aberrant disruption of the splicing reaction is responsible for a large number of genetically transmitted diseases as well as a variety of pathologies, including inflammation and cancer.

Historical aside

At the end of the 70's of the 20st century, Richard Roberts and Phillip Sharp, two scientists working on the adenovirus, discovered that regions present in the mRNA were excised, and that the remaining sequences were joined together before being exported to the cytoplasm [1,2]. The view of the gene as a continuous segment was radically changed after this finding. In 1993, both researchers received the Nobel Prize for this major discovery.

1.1. BIOCHEMICAL REACTION OF THE SPLICING

The splicing reaction leading to intron removal involves two sequential transesterification steps (Fig. 1). First, a nucleophilic attack by the 2'-hydroxyl of the adenosine of the branch-point, a sequence situated within the intron, occurs on the phosphodiester bond of the nucleotide present at the end of the upstream exon. This generates two intermediates: a free-ended upstream exon and a lasso-shaped intron bound to the downstream exon (structure called lariat). During the second step, the free hydroxyl end of the upstream exon attacks the phosphodiester bond of the nucleotide at the end of the intron, leading to exon ligation and lariat excision.

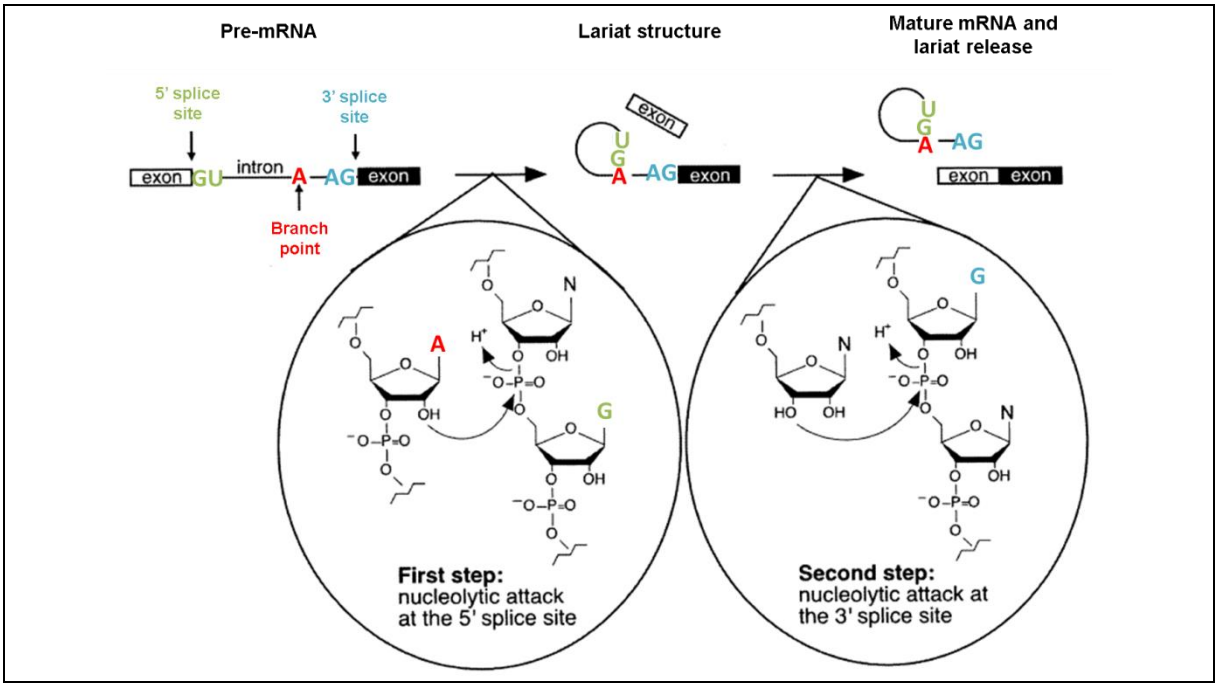

Figure 1: The two transesterification splicing steps
Scheme of the splicing reaction of a pre-mRNA containing two exons. Splice sites and branchpoint are colored. Transesterification steps are detailed in the circles. Adapted from Jurica et al., 2002.

1.2. THE MAIN ACTORS OF PRE-mRNA SPLICING

The splicing reaction is carried out by the coordinated action of several constitutive actors, which include:

a) **The cis-acting elements of the pre-mRNA,** which define the positions of the exons and introns.
b) **The spliceosome,** the basal machinery that catalyzes the two step-reaction of splicing.

A. THE CIS-ACTING ELEMENTS OF THE PRE-MRNA

Despite the fact that intron size can vary from several to thousands of nucleotides, the splicing machinery must recognize precisely intron borders. Several conserved elements of the pre-mRNA are involved in this recognition (Fig. 2). Most introns have common consensus sequences at their 5' and 3' ends that are recognized by the spliceosome. These sequences, called 5' and 3' splice sites, were first observed in viruses [3]. The presence of the nucleotides GU at the 5' end of the intron and the nucleotides AG at its 3' end constitute the so-called GU-AG rule.

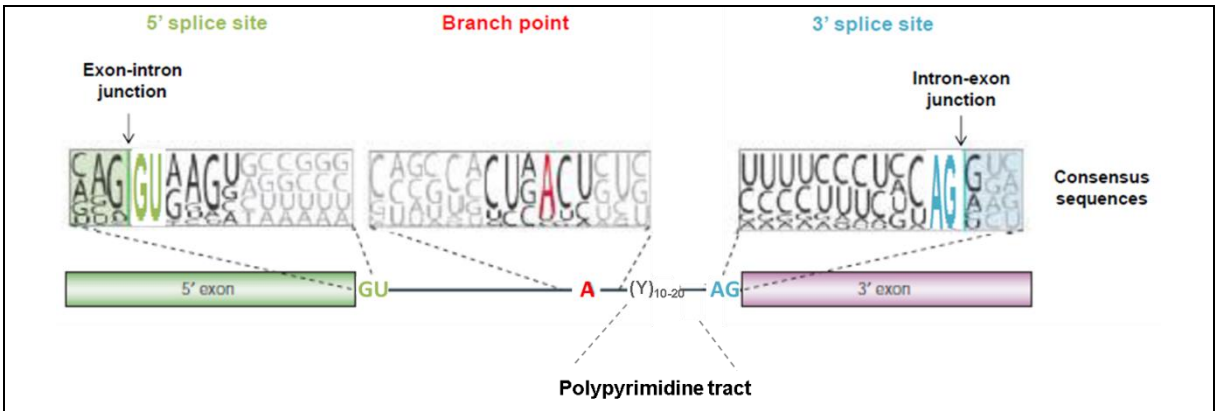

Figure 2: Splicing consensus sequences
5' splice site, branch point and 3' splice site consensus sequences of the majority of introns are shown from the left to the right. Letter height represents the occurrence frequency of corresponding nucleotide at that position. Nucleotides thought to be implicated in intron recognition are shown in black, the other in gray. Adapted from Patel et al., 2003.

5' splice site: it corresponds to the junction between the upstream exon and the intron. It consists in an 8-nucleotides consensus sequence: AG-**GU**RAGU (where R is a purine and – the exon-intron junction). The two GU nucleotides in bold are highly conserved. The recognition of this site is the first key-step of the splicing reaction. This sequence can be relatively degenerated, and several binding factors can either help or inhibit its recognition. Thus, the use of the 5' splice site is not only dependent of its proper sequence but also of the surrounding sequences that can recruit additional regulators.

3' splice-site: this sequence corresponds to the junction between the intron and the downstream exon. It is defined by a relatively well-conserved consensus motif X**AG**-G (where X is U or G). As mentioned for the 5' splice site, additional splicing regulators can also influence the use of the 3' splice site.

Branch-point: The branch-point is a sequence located 18 to 40 nucleotides upstream of the 3' splice site. Its consensus sequence is highly degenerated

YNYUR**A**C (where Y is a pyrimidine), although the adenosine residue is conserved and corresponds to the branch-point of the intron-lariat structure.

Polypyrimidine tract: It corresponds to 10 to 20 pyrimidine residues, with a majority of U, immediately upstream of the 3' splice site. The length is variable and can influence the recognition of the branch-point and the 3' splice-site.

B. THE SPLICEOSOME

The spliceosome is a large complex that involves about a hundred of proteins and several small RNAs. The canonical assembly of the spliceosome on pre-mRNA requires 5 snRNPs (small nuclear ribonucleoprotein particles) U1, U2, U4, U5 and U6. These snRNPs are the basic components of the spliceosome. They act in association with an array of additional splicing factors that modulate the splicing outcome [4].

a. Description and biosynthesis of SnRNPs

Each snRNP is composed of an UsnRNA particle associated with seven proteins of the Sm-family (Fig. 3A) [5]. They are organized as a heptameric ring around the UsnRNA. In addition, each snRNP is decorated with a specific set of proteins unique for each UsnRNA [6].

UsnRNAs undergo several modifications after transcription. Sm proteins are first pre-assembled on UsnRNAs in the cytoplasm before being re-localized into the nucleus, where they follow a maturation step inside small nuclear domains called the Cajal bodies. Mature snRNPs are then either directly loaded into the spliceosomal machinery or stored into interchromatin regions, called nuclear speckles, generally localized next to transcribed genes (Fig. 3B) [6].

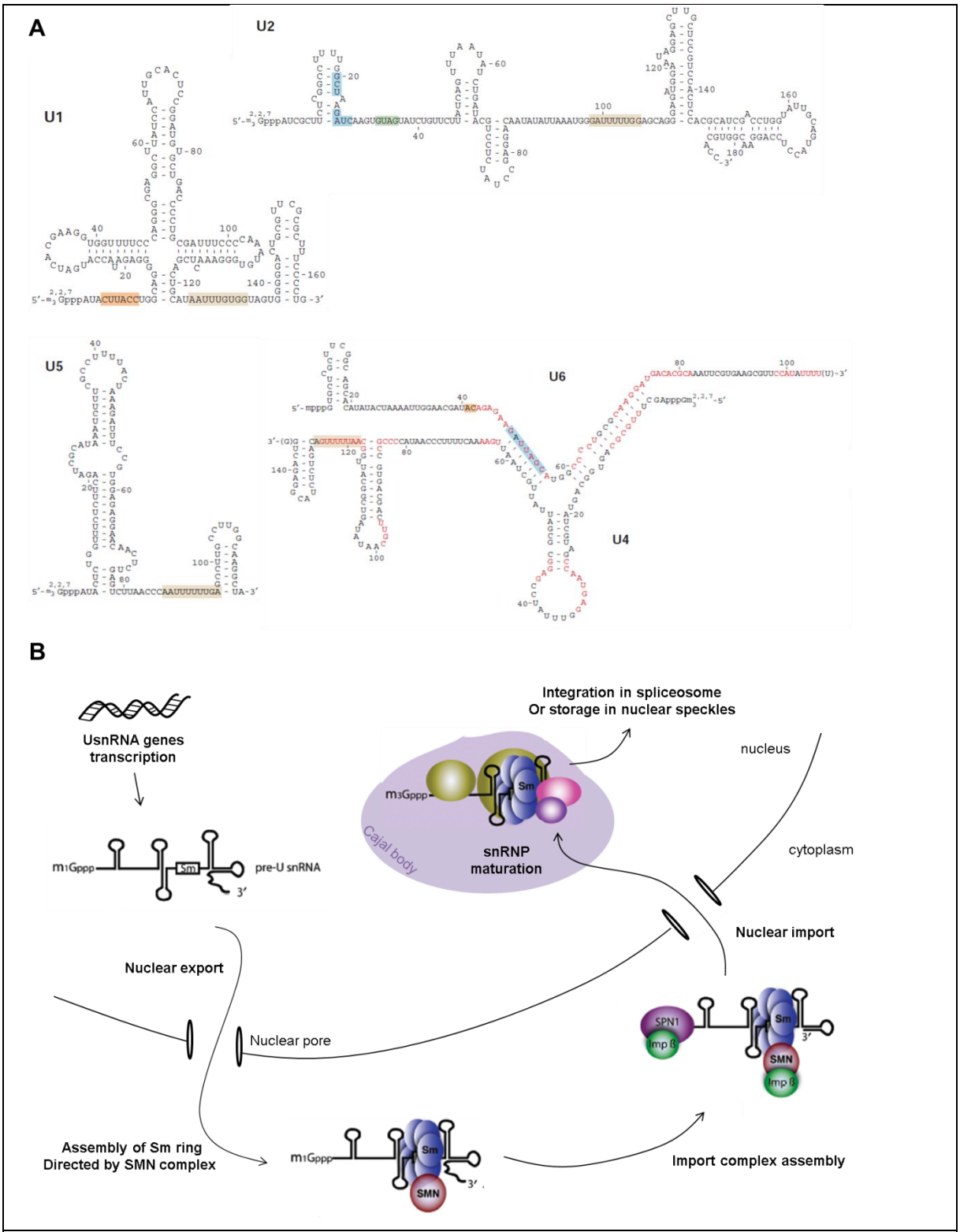

Figure 3: synthesis of snRNP

A. Sequences of UsnRNA and predicted secondary structures. Sm binding site is indicated in light brown. Orange boxes indicate sequences involved in RNA-RNA interactions at the 5' splice site, green boxes at the branch-point and blue boxes for U2-U6 interaction. Red sequences represent stretches of four or more identical nucleotides between U4 and U6. Adapted from Patel et al., 2003. B. Assembly of snRNP and maturation pathway. U2 snRNP used here as a representative snRNP. Adapted from Patel et al., 2008.

U1 snRNP: The U1 snRNP is composed of the U1 snRNA and the seven Sm proteins (B/B', D1, D2, D3, E, F and G), forming a ring surrounding the U1 snRNA. Three other U1 specific proteins make up the U1 snRNP: U1-70kDa, U1-A and U1-C (Fig. 4A). U1-70kDa and U1-A associate with the U1snRNA via a stem-loop structure. U1-C is linked to the U1 snRNP by interactions with U1-70kDa and Sm proteins (Fig. 4B) [7]. In addition, U1-70kDa contains Arginine/Serine-rich domains (RS) that can interact with several regulating splicing factors including the SR proteins.

U1 snRNP is involved in the recognition of 5' splice sites. A 9-nucleotide sequence at the 5' end of U1 snRNA is conserved from *Drosophila* to vertebrates (ACUUACCUG) (Fig. 4B,C)[8]. This extremity is complementary to the 5' splice site, between the nucleotides -3 to +6, with respect to the exon-intron junction [9,10].

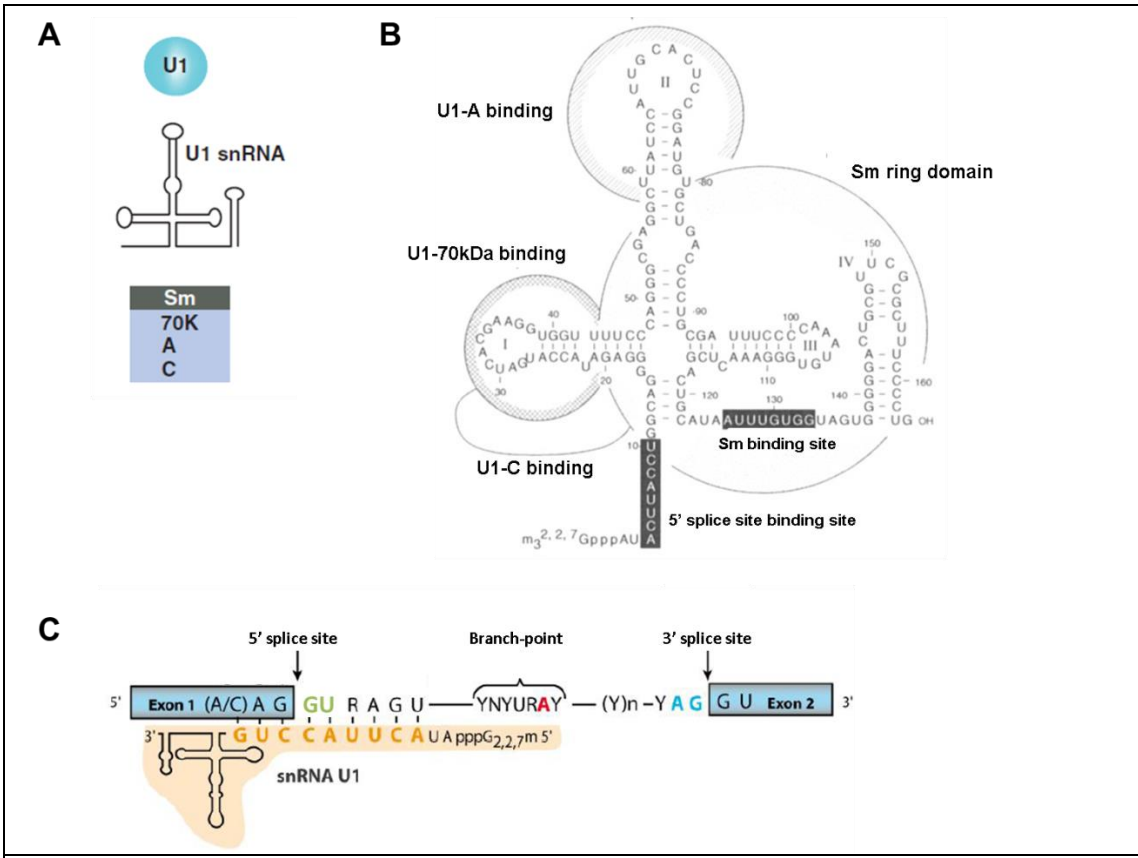

Figure 4: U1 snRNP
A. Protein composition of U1 snRNP. All the seven Sm proteins (B/B', D1, D2, D3, E, F, and G) are indicated by Sm at the top of the box showing U1 snRNP associated proteins. Adapted from Will et al., 2011. B. U1 snRNA secondary structure and specific protein associations. U1 snRNA folds into a secondary structure containing four stem-loops (I, II, III and IV). Sm ring assembles in a globular manner around the Sm binding site. Binding of specific proteins U1-A and U1-70kDa around loops II and I respectively are represented. U1-C protein is incorporated into U1 snRNP through U1-70kDa interaction. The conserved sequence at the 5' end of U1 snRNA pairs with the 5' splice site. Adapted from Nagai et al., 2001. C. Base-pairing between U1 snRNP (in orange) and the 5' splice site of a pre-mRNA. Nucleotides R and Y correspond respectively to a purine and a pyrimidine. For review, see Will et al., 2001.

U2 snRNP: U2 snRNP is composed of the Sm heptameric ring around the U2 snRNA. In addition, U2 snRNP associates with two specific proteins, U2A' and U2B", and the complex SF3. SF3 is formed by two subunits: SF3a, which contains three proteins (SF3a60, SF3a66 and SF3a120) and SF3b constituted by seven proteins (SF3b10, SF3b14a/p14, SF3b14b, SF3b49, SF3b130, SF3b145 and SF3b155) (Fig. 5A,B). The SF3 complex targets U2snRNP to the branch-point [11]. This interaction is stabilized mostly by the SF3b subunits 14 and 155 (Fig. 5B,C).

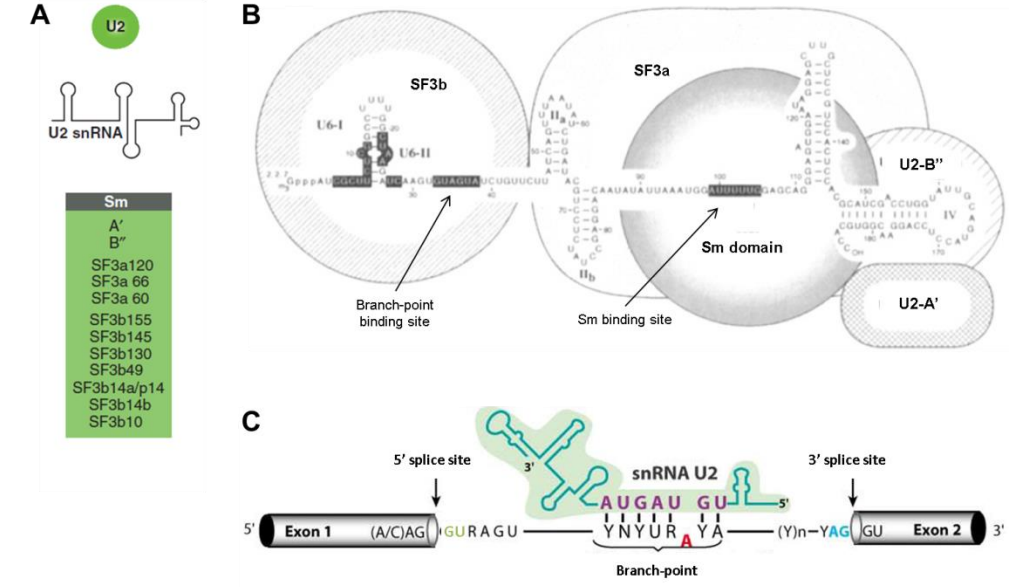

Figure 5: U2 snRNP
A. Protein composition of U2 snRNP. All the seven Sm proteins (B/B', D1, D2, D3, E, F, and G) are indicated by Sm at the top of the box showing U2 snRNP associated proteins. Adapted from Will et al., 2011. B. U2 snRNA secondary structure and specific protein associations. U2 snRNA folds into a secondary structure containing five stem-loops (I, IIa, IIb, III and IV). Sm ring assembles in a globular manner around the Sm binding site. U2-A' and B" bind the loop IV. SF3a and b bind the 5' end of U2 snRNA. The conserved GUAGUA sequence of U2 snRNA pairs with the branch point. U6-I and U6-II pair with sequences within U6 snRNA during splicing process. Adapted from Nagai et al., 2001. C. Base-pairing between U2 snRNP (in blue) and the branch point of a pre-mRNA. Nucleotides R and Y correspond respectively to a purine and a pyrimidine. For review, see Will et al., 2001.

U4/U6.U5: The snRNPs U4, U6 and U5 are organized into a higher order particle that consists in their functional form U4/U6.U5 tri-snRNP, where the di-particle U4/U6 associates with the U5 snRNP [12].

- snRNP U4/U6 contains both snRNA U4 and U6, as well as the seven Sm proteins associated with U4 snRNA, and seven LSm proteins associated with U6 snRNA (Fig. 6A). These LSm proteins (LSm 2 to 8) are of the same family than the Sm proteins. They also form a heptameric ring around the U6 snRNA [13]. U6 snRNA contains a conserved binding sequence for the recycling factor SART3/p110, which is necessary for the pairing of U4/U6 to form its Y-shaped

structure [14,15]. There are at least five additional proteins specific of U4/U6, which are the subunits of molecular sizes 15.5, 20, 60, 61 and 90 kDa (Fig. 6A,B).

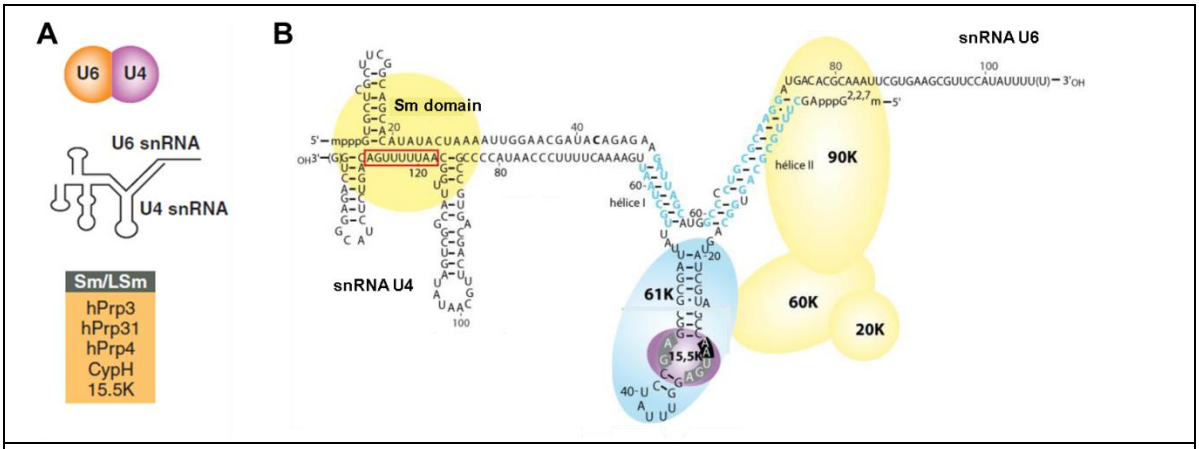

Figure 6: U4/U6 snRNP
A. Protein composition of U4/U6 di-snRNP. All the seven sm proteins (B/B', D1, D2, D3, E, F, and G) and LSm proteins (2 to 8) are indicated by Sm/LSm at the top of the box showing U4/U6 di-snRNP associated proteins. Adapted from Will et al., 2011. B. U4/U6 di-snRNA secondary structure and specific protein associations. U4 and U6 snRNA associate into a Y-shaped secondary structure stabilized by two helices (I and II). Sm ring assembles in a globular manner around the Sm binding site (red box). Interactions of the 61kDa protein (in blue), 15.5kDa protein (in purple) and the heterodimeric complex 20kDa/60kDa/90kDa (in orange) with the RNAs are represented. Adapted from Patel et al., 2003.

- U5 snRNP is the biggest spliceosomal snRNP. In addition to the U5 snRNA and the Sm proteins, it contains at least nine specific factors of 15, 40, 52, 100, 102, 110, 116, 200 and 220 kDa (Fig. 7A,B) [16]. U5 snRNP facilitates exon ligation by binding and aligning the 3' and 5' extremities of the upstream and the downstream exons, respectively, just before the second transesterification reaction [17,18].

U4/U6 and U5 are linked via the U4/U6-61 and U5-102 kDa subunits. The newly formed U4/U6.U5 tri-snRNP then associates with the spliceosomal machinery and with the mRNA. Three proteins (27K, hSad1 and hSnu66) specifically interact with U4/U6.U5 tri-snRNP. They contain RS domains and are called SR-like proteins (Fig. 7C) [19].

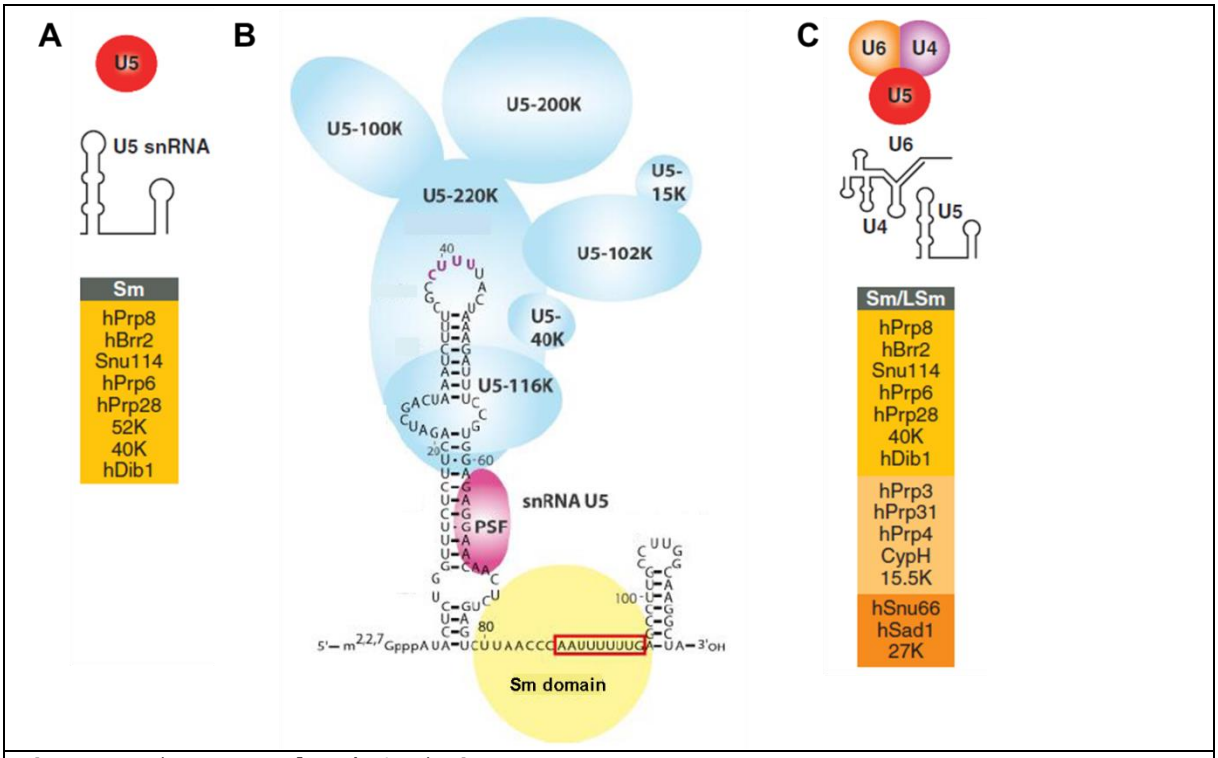

Figure 7: U5 snRNP and U4/U6.U5 tri-snRNP
A, C. Protein composition of U5 snRNP (A) and U4/U6.U5 tri-snRNP (C). All the seven Sm proteins (B/B', D1, D2, D3, E, F, and G) are indicated by Sm at the top of the box showing snRNP associated proteins. Adapted from Will et al., 2011. B. U5 snRNA secondary structure and specific protein associations. U5 snRNA folds into a secondary structure containing two stem-loops. Sm ring assembles in a globular manner around the Sm binding site (red box), between the two loops. U5 specific proteins associate around the 5'end loop via RNA-protein and protein-protein interactions. Nucleotides in purple are implicated in pairing with both exons. Adapted from Patel et al., 2003.

b. Splicing factors (other than UsnRNP)

The spliceosome contains roughly 170 proteins, of which 45 are snRNP-associated proteins [4]. Given that the spliceosome is a highly dynamic machinery and is largely remodeled during the splicing process, a remarkable exchange of proteins from one splicing step to another is observed. The exact function of many of these co-factors remains unrevealed yet. A portion of them is thought to be involved in the coupling of the splicing machinery to other molecular processes such as transcription, or other downstream RNA processing events [20]. Despite the current lack of knowledge on many associated non-snRNP proteins, the composition and functions of two major complexes, U2AF and PRP19, have been well described:

U2AF is a heterodimer composed of two subunits: U2AF65 and U2AF35. The complex is involved in the recognition of the 3' end of the intron by U2AF35 binding at the 3' splice site and U2AF65 binding at the poly-pyrimidine tract (Fig. 8, top) [21–23]. During U2 snRNP assembly, U2AF65 interacts with SF3b155 guiding U2 snRNP to the branch-point (Fig. 8 bottom). This non-snRNP splicing factor is thus essential for the correct recognition of the 3' splice site, the polypyripidine tract and the branchpoint.

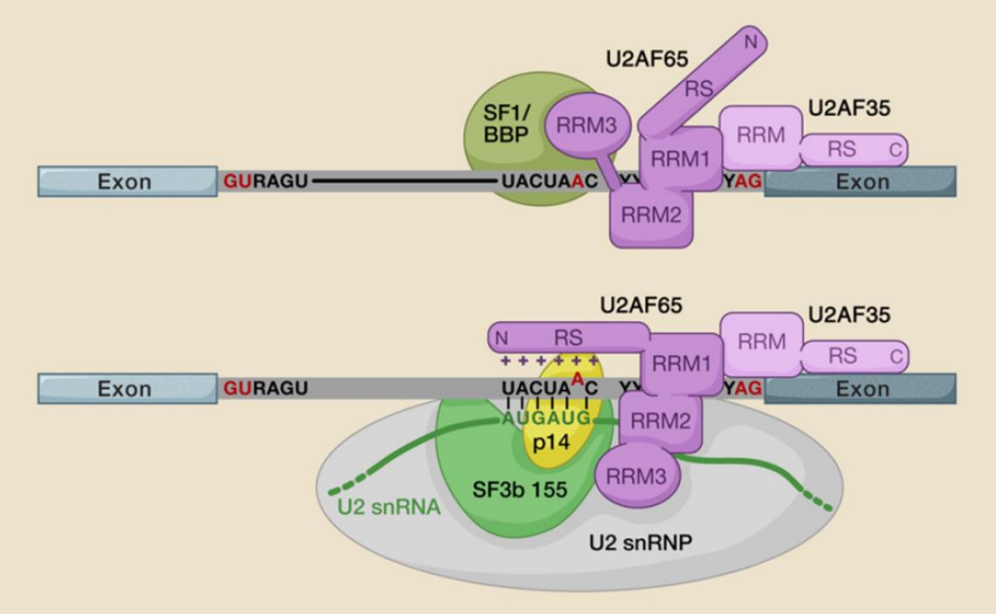

Figure 8: Interactions of U2AF with the branch point and 3' splice site
Top: before recognition of the splicing acceptor site by U2 snRNP, pre-mRNA (exons in blue, intron in gray) branch point is bound by SF1/BBP, whereas the polypyrimidine tract and the 3' splice site are bound by U2AF subunits U2AF65 and U2AF35 respectively. U2AF65 interacts with both SF1/BBP and U2AF35. Bottom: Upon U2 snRNP recognition, SF1/BBP is displaced, allowing U2-p14 specific subunit to contact the branch point and U2AF65 to interact with SF3b155. U2-branch point interaction is stabilized by components of the U2 snRNP and U2AF65. From Wahl et al., 2009.

PRP19 complex is composed of 8 core proteins (Prp19, Cef1, Snt309, Syf1, Syf2, Clf1, Isy1, Ntc20) and 13 associated factors (Cdc5, Spf27, Prl1, AD-002, Skip, Syf3, ECM2, KIAA0560, GCIPp29, MGC23918, G10, Cyp-E, PPIase-like1) [24]. This complex joins the spliceosome together with the U4/U6.U5 tri-snRNP, before the two transesterification steps. During the splicing reaction, the spliceosome is largely remodeled and results in the release of U4 snRNP. Although the precise mechanisms remain unknown, PRP19 complex likely facilitates spliceosomal rearrangements by its ubiquitination activity [25]. It stabilizes U5 and U6 snRNP in the catalytically active form and remains bound to them until the end of the second transesterification reaction (Fig. 9) [26].

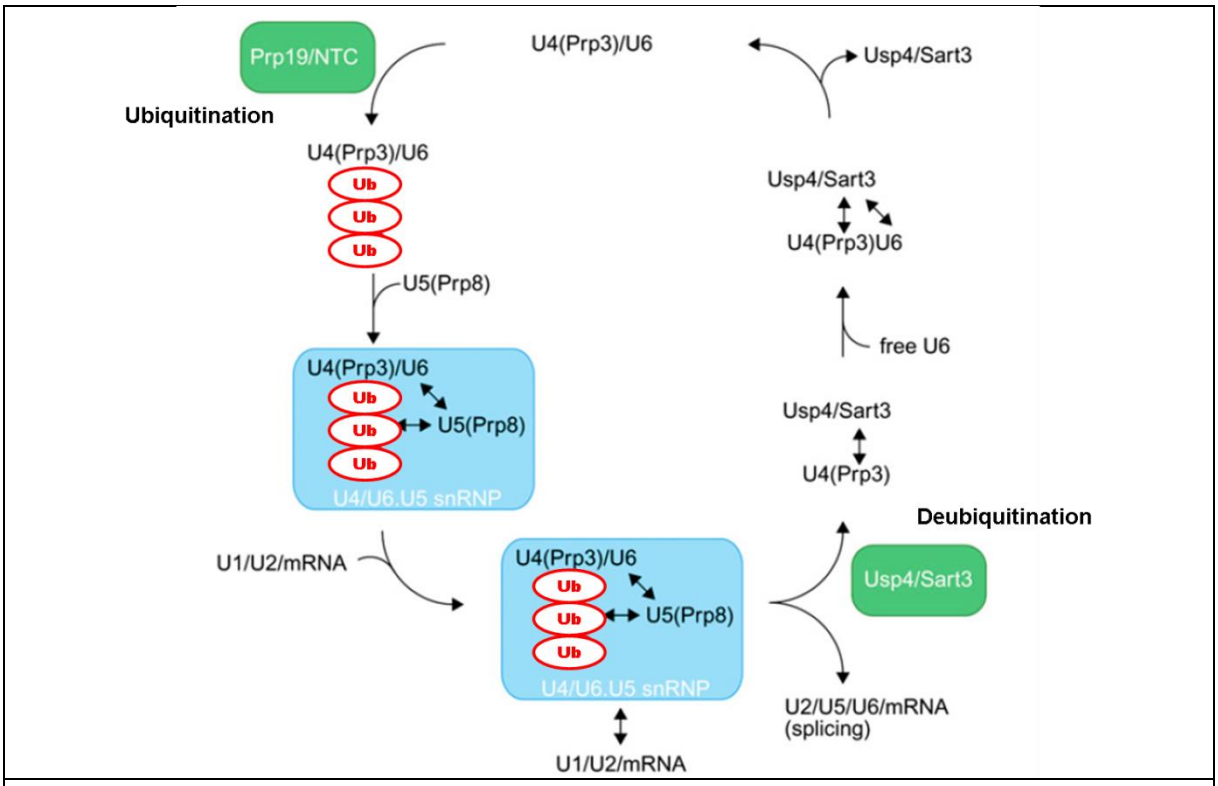

Figure 9: Model of ubiquitin-dependent regulation of U4/U6.U5.
PRP19 (NTC) promotes the ubiquitination (red circles) of the U4 snRNP specific component Prp3. this modification can be recognized by the U5 snRNP subunit Prp8, allowing U4/U6.U5 stabilization. After loading of U4/U6.U5 on the spliceosome, U4 snRNP dissociates. De-ubiquitination of Prp3 by Usp4/Sart3 decreases the affinity to Prp8 and facilitates dissociation. Sart3 also promotes the re-annealing of U4 and U6 snRNP, for another round of splicing. Adapted from Song et al., 2010

1.3. STEPS OF THE SPLICING REACTION

The spliceosome is a highly dynamic machinery that assembles in a stepwise manner. During the assembly, a multitude of rearrangements are observed. UsnRNP are heavily remodeled, changing numerous times of RNA-RNA, RNA-protein and protein-protein interactions via the activity of ATP-dependent RNA helicases. Although the precise molecular rearrangements of transitory interactions remain to be determined, the major steps have been identified [4,27,28]. The splicing process can be viewed as a cycle, where the spliceosome first assembles on the pre-mRNA, excises an intron, releases the mature mRNA and is then recycled and *de novo* assembled on the pre-mRNA to excise a new intron (Fig. 10).

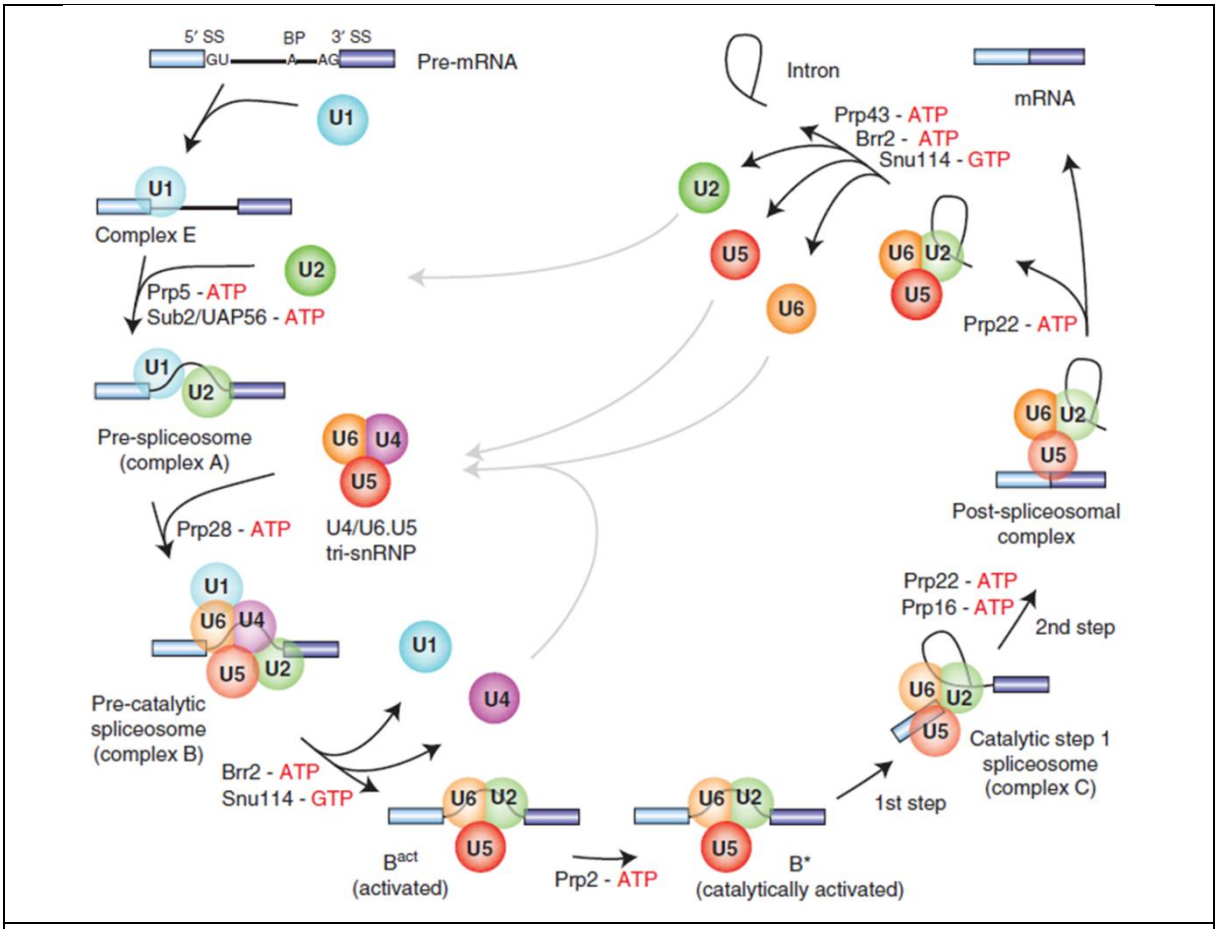

Figure 10: Assembly and disassembly of the spliceosome machinery.
The stepwise interaction of snRNP (colored circles) in the splicing of an mRNA containing two exons (in blue) is depicted. Eight conserved ATPases/helicases (Sub2/UAP56, Prp5, Prp28, Brr2, Prp2, Prp16, Prp22 and Prp 43) act in different steps of the splicing cycle to catalyze RNP remodeling events. The GTPase Snu114 also functions in several steps of the cycle. From Will and Lührmann, 2011.

A. THE H COMPLEX

The H complex is not specific to splicing substrates as it assembles on RNA lacking functional splice sites. It is composed mainly of hnRNP (heterogeneous nuclear ribonucleoproteins) [29]. The role of this early recruitment is unclear but these proteins may antagonize or promote the subsequent interaction of spliceosomal components with alternative splice sites. The recruitment of hnRNP at that stage does not require ATP hydrolysis [30].

B. THE E COMPLEX

The E complex, also called early or commitment complex, is the first specific complex of the splicing process. It is formed by U1 base pairing on the 5' splice-site, in an ATP-independent manner. This association is stabilized with proteins specific of the U1 snRNP and members of SR proteins (Serine-Rich proteins). In parallel, loose associations of U2AF with the poly-pyrimidine tract and the 3' splice site are observed. U2AF65 subunit binds the poly-pyrimidine tract and interacts with SF1/BBP that is recruited to the branch-point. U2AF35 subunit associates with the conserved AG nucleotides of the 3' splice site. Altogether, these interactions form the spliceosomal complex E that plays a crucial role in the recognition of the 5' and 3' splice sites.

C. THE A COMPLEX

The A complex, or pre-spliceosome, is the first ATP-dependent intermediary. U2AF recruits U2snRNP in an ATP-dependent manner, displacing SF1/BBP from the branch point. A tight base pairing of U2 snRNA with the branch-point keeps the conserved adenosine such in a position that the nucleophilic attack by the 5' splice site is possible. This base pairing between U2 snRNA and the pre-mRNA is stabilized by U2 snRNP specific proteins SF3a and SF3b [31] and by the U2AF subunit U2AF65 [32], still associated with U2AF35.

D. THE B COMPLEX AND CATALYTICALLY ACTIVE B COMPLEX*

The formation of the B complex is also ATP-dependent and requires the presence of 5' and 3' splice sites for its assembly. It is formed by the addition of the pre-assembled tri-snRNP U5:U4/U6 on the A complex. Although all snRNPs are present in the B complex, it remains still catalytically inactive. It requires major rearrangements in composition and conformations for the catalytic activation to facilitate the first transesterification step. During the remodeling of the spliceosome, partly helped by the action of the PRP19 complex, U1 and U4 snRNPs are destabilized [33]. U1snRNP is exchanged by U6 snRNP at the 5' splice site, leading to a new base pairing between U2 and U6 snRNAs. The release of U1 and U4 snRNP and subsequent activation of the DEAH-box helicase Prp2 give rise to the activated spliceosome, the so-called B* complex. This complex undergoes the first catalytic step of splicing leading to the lariat formation.

E. THE C COMPLEX

The conversion of B* complex into C requires ATP hydrolysis. After the first transesterification reaction, the resulting C complex is composed by U2, U5 and U6 snRNP in association with a free-ended upstream exon and a downstream exon ligated with the lariat. Before the second step of splicing, many structural and compositional rearrangements will occur [34]. Nevertheless, the C complex has a short lifetime. Splicing intermediaries are rapidly converted into spliced products. After the second transesterification step, U2, U5 and U6 snRNPs are released with the intron lariat and form the post-spliceosomal complex. The splicing components are then rapidly dissociated and recycled.

1.4. INTRON DEFINITION VS EXON DEFINITION

During evolution, an increase of the intron size is observed [35]. This phenomenon is due to insertion of transposable elements within introns, the presence of regulatory elements, or the presence of other types of non-coding sequences, such as snoRNA [36,37]. Two modes of recognition by the spliceosome can be distinguished, based on intron size. In genomes having small introns, like yeast or *Drosophila*, the intron is recognized as a unit and the spliceosome machinery is assembled across its borders. This recognition mechanism is called intron definition. However, beyond 250 nucleotides, the intron is considered too large for spliceosome assembly. Instead of spliceosome assembly across intron, the spliceosome assembled first across the exon (Fig. 11). In this case, the 5' end and 3' end of the exon are recognized by U2 snRNP and U1 snRNP, respectively. A cross talk between both snRNPs across the exon is facilitated by SR proteins and other splicing regulators [29,38,39]. This mode of recognition is called exon definition and is the main recognition mechanism in vertebrates. The precise mechanism of exon definition is poorly understood due to the difficulty to reconstitute an in vitro system using large introns, and also, because each recognition is likely context-dependent.

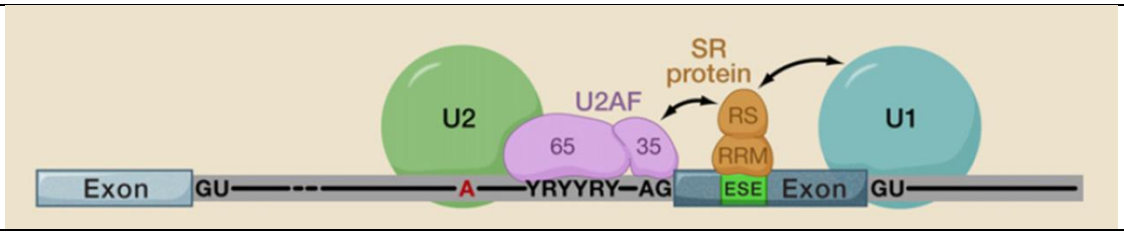

Figure 11: Model of cross-exon interactions for exon definition
Binding of SR protein containing RS and RRM domains to an exonic splicing enhancer (ESE). The U1 snRNP (in blue) and the U2 snRNP (in green) as well as the two U2AF subunits (U2AF65 and U2AF35) are also shown interacting with the splice sites flanking the exon. SR proteins facilitate stable association of U1 and U2AF with their respective cis-elements. From Wahl et al., 2009.

1.5. THE MINOR CLASS SPLICEOSOME

In the mid-1990s, a class of rare introns containing non-canonical consensus sequences has been discovered [40,41]. Their excision requires a distinct splicing machinery called the minor class spliceosome. This spliceosome contains four low abundant snRNPs (U11, U12, U4atac and U6atac) and the U5 snRNP. This machinery presents a highly similar two-step mechanistic, with the formation of the lariat intermediate. U11, U12, U4atac and U6atac can base pair with non-canonical consensus splice sites, but their secondary structure mimics those of U1, U2, U4 and U6, respectively. U12 snRNP functionally replaces U2 snRNP. It can associate with introns lacking poly-pyrimidine tract. The disruption of U12 in fruit fly leads to lethality, indicating that the minor-class spliceosome is essential for this organism [42].

1.6. ALTERNATIVE SPLICING

The short size of splice sites and their sequence variability imply that other factors than components of the spliceosome might be implicated in the recognition of exon-intron junctions. In addition, depending on the cellular context, an exon can be included or excluded from the final transcript. This phenomenon is called alternative splicing. The process of alternative splicing gives rise to several proteins isoforms that result from a single gene. These isoforms can have differential functions, locations, or post-translational modifications [43]. Besides, alternative splicing also allows high degree of flexibility, which permits a rapid response to regulatory signals. Several types of alternative splicing events can be distinguished and are described in Figure 12.

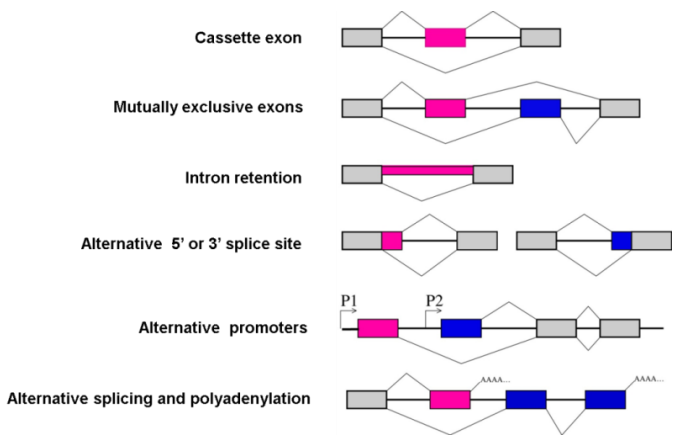

Figure 12: Several types of alternative splicing events.
An exon can be either present or absent of the mature mRNA, depending of the efficiency of its splice site recognition (cassette exon). In the same way, two exons can be mutually excluded, giving rise to two mature mRNA isoforms, containing either one or the other exon (mutually exclusive exons). Splice sites can be sub-optimal. Thus, 5' or 3' splice sites can enter in competition or not be recognized at all, giving rise to isoform presenting variable exon sizes (alternative 5' or 3' splice site) or intron retention. In some cases, Alternative exons can be found at the beginning or at the end of the transcript, because of a differential choice in the transcription start site or the polyadenylation site.)

Alternative splicing is regulated at different levels. A part of the information is brought by *cis*-auxiliary sequences involved in both constitutive and alternative splicing [44–46]. These elements are classified according to their intronic or exonic localization and their enhancing or inhibiting functions. These *cis*-regulatory sequences recruit specific *trans*-acting regulators that enhance or repress splicing [47]. Other elements can influence splicing such as secondary structures of the RNA, the elongation rate of RNA Polymerase II and the chromatin structure.

A. CIS- AND TRANS-ACTING REGULATORY ELEMENTS

a. Description

Exonic splicing enhancers (ESE) and their associated factors, the SR proteins

ESE are the most studied splicing cis-regulatory sequences. They are usually purine-rich and are often recognized by SR proteins (serine/arginine-rich proteins) [48]. All SR proteins present structural similarities: they contain one or two RNA recognition motifs (RRM) at their N-terminal domain, which can bind a large set of RNA consensus sequences [49]. They also have an RS domain (arginine and serine rich domain) at their C-terminus that ensures protein-protein interaction, and thereby facilitates spliceosome recruitment. They are involved in both constitutive and alternative splicing. One SR protein can recognize a large set of RNA sequences. Conversely, different SR proteins can recognize a similar sequence, permitting the fine-tuning of the splicing reaction.

Exonic splicing silencers (ESS) and their associated factors, the hnRNP

ESS are generally recognized by hnRNP. hnRNP are the most abundant mRNA associated proteins. They can bind nascent mRNA in a rather aspecific manner (complex H). They are characterized by their size and classified by letters ranging from the smallest to the largest molecular weight (hnRNP A1 to hnRNP U). They are implicated in splicing, but also in numerous other biological processes such as telomere synthesis, RNA stability or translation. Among the hnRNP known to be involved in the splicing reaction are hnRNP A1, A2, F, H, I (also known as PTB), K, G, and L. However, the specificity and precise functions of these proteins remain poorly described.

Intronic splicing enhancers (ISE) and silencers (ISS)

The features of intronic splicing sequences and their regulation are less studied. They can also be bound by hnRNP but not necessarily [50]. The best characterized intronic motif is the G-tract, found generally near 5' or 3' splice sites [51]. The G-tract acts as an ISE and participates in the recruitment of U1 snRNP via hnRNP F and H interactions [52]. However, these motifs have an inhibitory activity when they are present within exons [53].

b. Mechanisms

As explained above, SR proteins preferentially bind ESE. They stimulate exon inclusion by at least two mechanisms. The first one is exon-dependent. In this case, SR proteins bind ESE via the RRM. This facilitates the recruitment of splicing factors like U2AF and U1 snRNP and interferes with the inhibitory activity of hnRNP (Fig. 13A). The second model is exon-independent. In this model, SR proteins are involved in protein-protein interactions via their RS domains, leading to a bridge formation between splice sites, which facilitate their recognition by the spliceosome. According to this second model, SR proteins play also an important role in the recruitment of the tri-snRNP U4/U6.U5 via protein-protein interactions (Fig. 13B) [54–56].

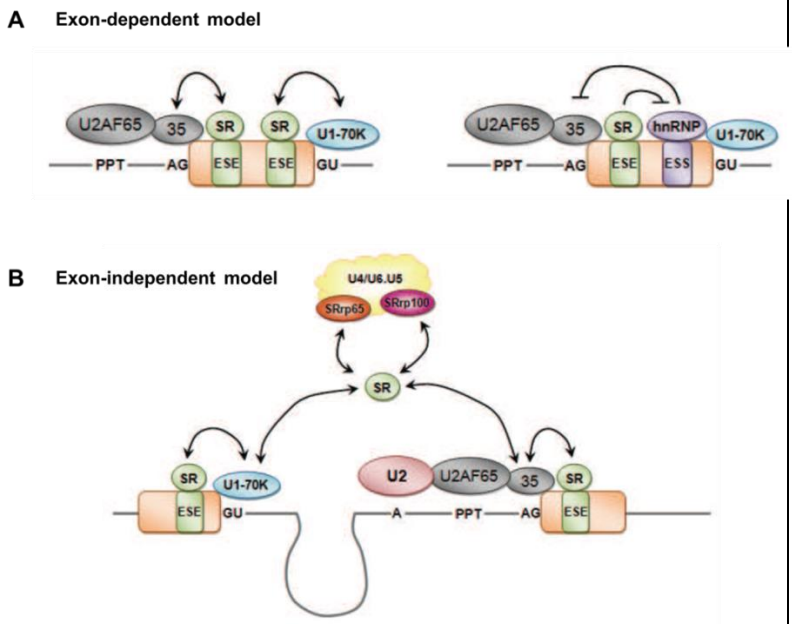

Figure 13: Model of splicing regulation by ESE
A. Exon-dependent model. Left: SR proteins (green circles) are depicted to bind ESE (in green on an orange exon). This facilitates the recruitment of splicing factors like U2AF (gray) and U1-70kDa subunit (blue). Right: SR proteins can also counteract the inhibition effect of hnRNP (purple circles) that bind to ESS (purple). B. Exon-independent model: SR proteins build up protein-protein interaction network (arrows). This can help to bring splice sites closer. Besides, SR proteins play an important role in the recruitment of the tri-snRNP U4/U6.U5 (yellow) via SRrp65 (dark orange) and SRrp100 (magenta). Adapted from Graveley et al., 2000 and Long et al., 2009

Three repression models have been proposed. The first model involves the direct competition between splicing activators (SR proteins) and repressors (hnRNP) for their binding due to ESE and ESS overlapping sequences (Fig. 14A). In a second related model, hnRNP bind ESS motifs and dimerize, hiding SR protein binding-sites. Conversely, tight binding of SR proteins can prevent hnRNP polymerization and stimulate exon inclusion (Fig. 14B). In both models, exon inclusion or exclusion is dependent of the binding affinity of splicing regulators, and also, to their relative concentration. Changes in expression level in splicing regulators can indeed affect splicing pattern in certain conditions. In the third model, hnRNP prevent exon inclusion by binding to flanking ISS. These interactions trap the exon in a loop, excluding it from the mature transcript (Fig. 14C) [57].

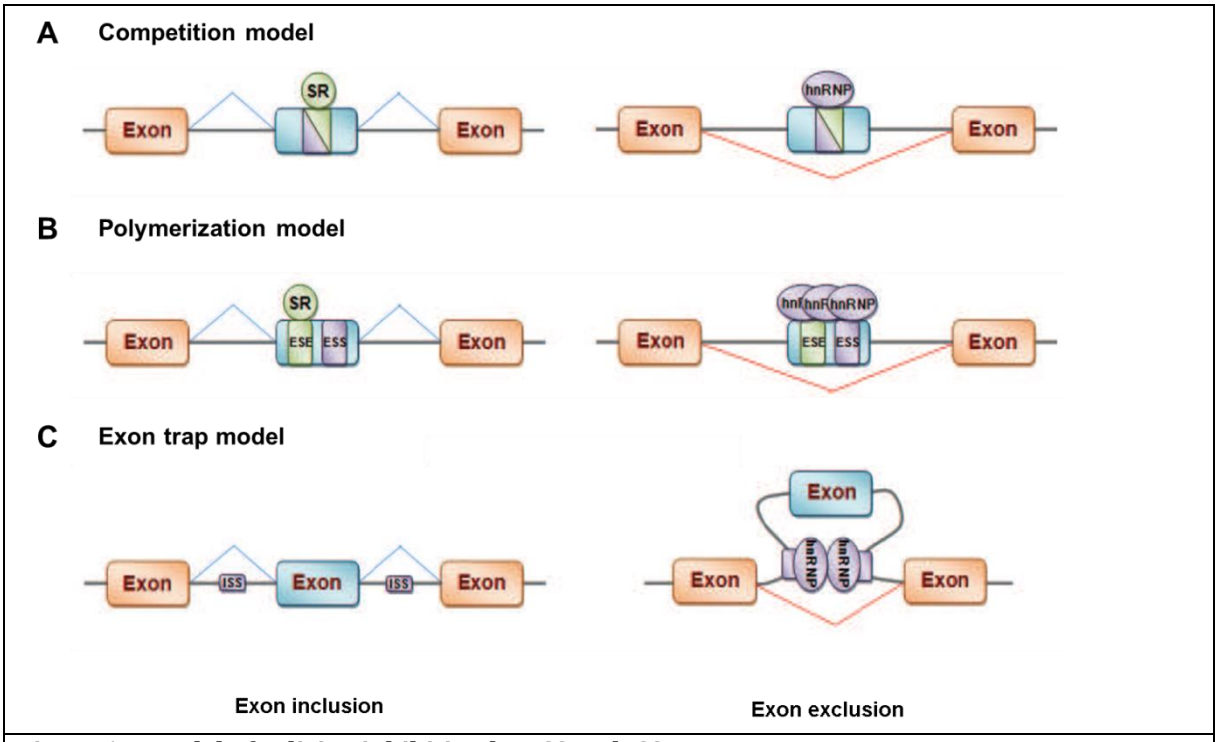

Figure 14: Model of splicing inhibition by ESS and ISS
A. Competition model. Establishment of a direct competition between splicing activators (SR proteins in green) and repressors (hnRNP in purple) by ESE and ESS overlapping sequences (depicted on the blue alternative exon with green and purple triangles) enforcing an exclusive binding. B. Polymerization model. Left: tight binding of SR proteins avoids hnRNP binding and polymerization and stimulates exon inclusion. Right: hnRNP bind ESS motifs and dimerize, hiding SR protein binding-sites. This inhibits recognition of the exon, which is excluded from the mature mRNA. C. Exon trap model: ISS (in purple boxes within introns) are bound by hnRNP that dimerize and exclude the alternative exon (in blue) by a loop formation. Adapted from Cartegni et al. 2002).

B. SECONDARY STRUCTURES IN THE mRNA

mRNAs are usually represented as linear, but in reality they are folded in secondary structures, which can form tertiary interactions. These conformations can influence splice-site recognition and modulate spliceosome assembly. RNA structures can influence splicing in at least two ways [58]. Secondary structures can impair the accessibility of splicing factors to their target sequences, by hiding a splice-site or a *cis*-regulatory sequence (Fig. 15A). Conversely, RNA structures can generate novel splicing sites that are not recognized by splicing factors in their linear form. Alternatively, RNA structures can vary the distance between two splice sites and indirectly impact the binding of splicing regulators (Fig. 15B). A recent study seeking for correlation between alternative splicing events and RNA secondary structures proposes that RNA structures can be involved in 4% of alternative splicing control [59].

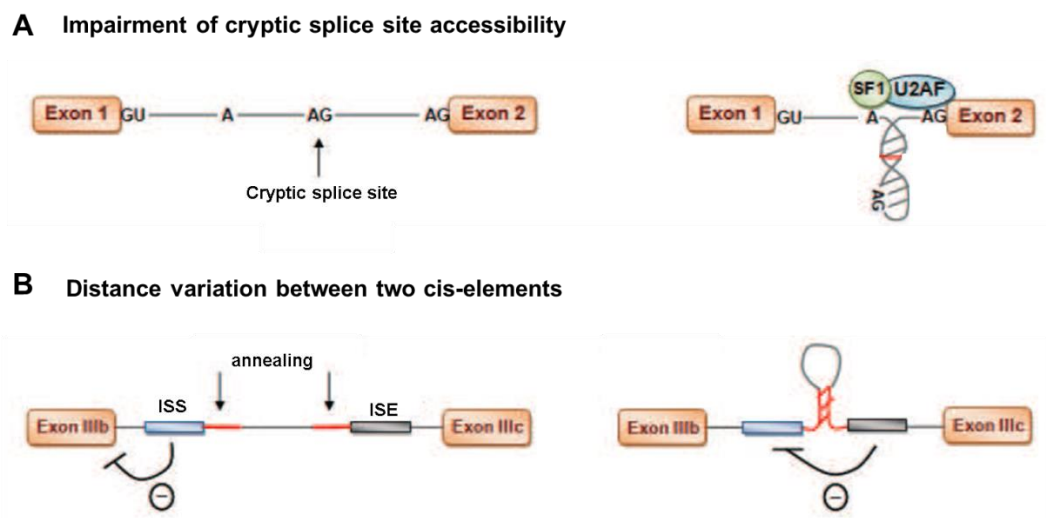

Figure 15: mRNA secondary structures in splicing regulation.
A. A secondary structure in the mRNA can hide a cryptic 3' splice site and help to bring the branch point and a downstream 3' splice site closer, thereby allowing SF1 (green) and U2AF (blue) binding. B. A secondary structure formed through annealing of the two indicated nucleic sequences (red) modulates the distance between an ISS (in blue) and an ISE (in gray) leading to the inhibition by the ISE of the ISS. Adapted from Buratti et al., 2004.

C. COUPLING BETWEEN SPLICING AND TRANSCRIPTION

The first clue of the existence of a coupling between transcription and splicing arisen about 25 years ago after the visualization of mRNA-DNA hybrids from *Drosophila* embryo preparations using electron microscopy [60]. Loops in the RNA representing the formation of lariats could be seen. More current approaches using single molecule imaging [61] and next generation sequencing [62–64] confirmed that most splicing events

occurs co-transcriptionally in metazoan. However, despite that splicing complexes are recruited to most exon-intron junctions during transcription, this does not necessarily imply that every intron is excised during this process. Additional regulators can facilitate or delay the splicing reaction. This is the difference between splicing commitment and intron removal. Interestingly, introns flanking alternative exons are usually removed later. Examples of functional coupling between transcription and splicing are described in Bentley et al., 2014 [65] and briefly illustrated below:

1) **The identity of promoters can affect alternative splicing events**.
   It has been shown that the promoter itself can recruit splicing factors at the transcription start site via transcription factors and remain associated with the nascent transcript and the RNA Pol II (RNA polymerase II) complex during transcription elongation [66–70].

2) The rate of **transcription elongation can modulate alternative splicing**.
   This mechanism is called the kinetic coupling model. The current view is that a slow elongation rate gives enough time to splicing factors to recognize the 3' splice site of an intron before the downstream intron is even synthetized and available to the splicing machinery. Then, even a "weak" 3' splice site will be recognized since it does not have to compete with a stronger downstream splice site. This follows the rule "first comes first serve" [71]. In general, the fastest the RNA polymerase II (RNA pol II) is, the less alternative exons is included in the final mRNA. This model is supported by decreasing RNA pol II elongation rate with drugs or mutants, and examining the effect on alternative splicing [72–74]. However, this is not a general rule as slowing down the speed of Pol II can in some contexts has an opposite effect [75,76].

3) The carboxy-terminal domain (CTD) of **RNA polymerase II influences co-transcriptional splicing**.
   The CTD of RNA pol II plays a major role in RNA processing. By its interaction with different factors, the CTD can influence capping, splicing and polyadenylation. The CTD is a repetition of seven amino-acids ($Y_1S_2P_3T_4S_5P_6S_7$). Post-translational modifications of these amino acids control the interaction of the CTD with its cofactors and play a determinant role in the modulation of transcription regulation and RNA maturation [77]. In the case of splicing, it has been shown that the CTD affects splice-site selection by directly interacting with splicing factors such as SR proteins and U2AF[78,79]. Hence, The CTD can serve as a platform to recruit splicing factors involved in splice site recognition and spliceosome assembly. This mechanism has been named "the recruitment coupling model".

In conclusion, two non-exclusive mechanisms can explain the coupling of transcription and splicing by RNA Pol II. On one hand, the rate of RNA Pol II can influence the kinetic of recognition of splice site by the splicing machinery, and

therefore modulates splicing. On the other hand, the direct binding of the CTD to splicing factors can facilitate the recruitment of these factors to the nascent transcript and modulate the splicing outcome.

4) **Chromatin structure and histone modifications can regulate splicing pattern.**

The coupling between transcription and splicing allows novel layers of regulation. For instance, chromatin structure can also influence alternative splicing. This has been indirectly suggested few years ago by using drugs that inhibit histone acetyl deacetylases that showed a strong effect on alternative splicing. More recent studies have shown that nucleosomes are not distributed evenly within the gene body but exhibit enrichment at the exonic level. It is thought that this enrichment slows down RNA Pol II and facilitates the recognition of splice sites by the spliceosome. Supporting this hypothesis, exons containing weak splice sites or surrounded by large introns have higher nucleosome enrichment compared to others.

Histone modifications can also modulate alternative splicing by influencing the binding of splicing regulators to nascent transcripts. For instance, Luco and Mistelli demonstrated an enrichment of H3K36me3 at an alternative exon of the FGFR2 gene. They showed that this mark interacts with the chromatin adaptor MGR15, which in turn recruit the splicing inhibitor PTB to the nascent FGFR2 transcripts [80]. As PTB is a splicing inhibitor, this recruitment ultimately prevents inclusion of the alternative exon. In addition, genetically modifying the level of this mark alters the outcome of exon inclusion. These experiments nicely demonstrate a direct role for a specific histone mark in regulating splicing outcome. However, due to the lack of genome wide studies, it is currently unclear how general is the influence of histone modifications on the regulation of alternative splicing.

2. THE EXON JUNCTION COMPLEX

In order to be translated into functional proteins, mRNAs undergo a succession of processing and quality control events. After transcription and maturation steps in the nucleus, mRNA transit to the cytoplasm, where they have to be addressed to their final location. Only correct mRNAs will be translated while aberrant mRNAs, for instance those containing premature stop codons, will be degraded by the cellular machinery. Intimate coupling between all these events ensures fidelity of gene expression. One crucial molecular complex ensuring these coupling is called the Exon Junction Complex (EJC). The EJC has received a lot of attention these past years for its ability to bind RNAs at an early stage and to influence their fate via its interaction with diverse mRNA-related cellular machineries. In this second section of the introduction, I will detail the mechanisms and functions of this crucial complex.

2.1 COMPOSITION OF THE EXON JUNCTION COMPLEX

The EJC is a ribonucleoprotein complex, which is deposited to mRNAs during splicing. It remains stably bound to mRNAs 20 to 24 nucleotides upstream of a subset of exon-exon junctions in a sequence independent manner [81–83]. After the first round of translation, the EJC is removed from mRNA by the ribosomal machinery and is recycled into the nucleus. The EJC plays important roles in many cellular processes, which include splicing, nuclear export, nonsense-mediated decay, RNA localization and translation [84–89]. This functional diversity is accomplished by the dynamic nature of the EJC and its transient interactions with numerous effector proteins [90–92]. A stable complex composed of four proteins (eIF4A3, Mago-nashi, Y14 and Barentsz) forms the core EJC, which acts as a recruitment platform to the other transitory interacting proteins.

A. CORE EJC COMPONENTS

EIF4A3:

Eukaryotic initiation factor 4A3 (eIF4A3) anchors the EJC to the mRNA [89,90,93,94]. This member of the eIF4A family belongs to the DEAD-box helicase group, whose components are involved in many aspects of RNA metabolism via their RNA binding domains and ATPase activity. As every eIF4A family members, eIF4A3 has two Rec-A-like domains separated by a flexible central region [90,95], which can be divided into twelve conserved motifs (Q, I, Ia, Ib, Ic, II, III, IV, IVa, V, Va, VI) (Fig. 1A) [96]. The motifs Q, I, II, and

VI are required for ATPase activity, whereas Ia, Ib, Ic, IV, IVa and V are involved in RNA binding. Motifs III and Va are required for both RNA binding and ATP-hydrolysis [97–99]. The protein can adopt two conformations depending on the folding of the flexible part. The extended central region confers to eIF4A3 an open conformation in absence of its ligands [92]. After binding to the heterodimer Mago/Y14, eIF4A3 adopts a closed conformation, which increases its affinity to the RNA (Fig. 1B) [90]. This conformation change requires ATP hydrolysis [92]. Binding of Mago and Y14 to the eIF4A3 C-terminal domain also inhibits its ATPase hydrolysis activity. eIF4A3 is thus unable to return to an open conformation and remains tightly associated with the mRNA [100]. In contrast, Barentsz (Btz) is able to activate ATP-hydrolysis activity by binding eIF4A3 at its both Rec-A-like domains. When Mago and Y14 are released from the complex, the stimulation by Btz allows eIF4A3 to adopt an open conformation in order to be recycled [101].

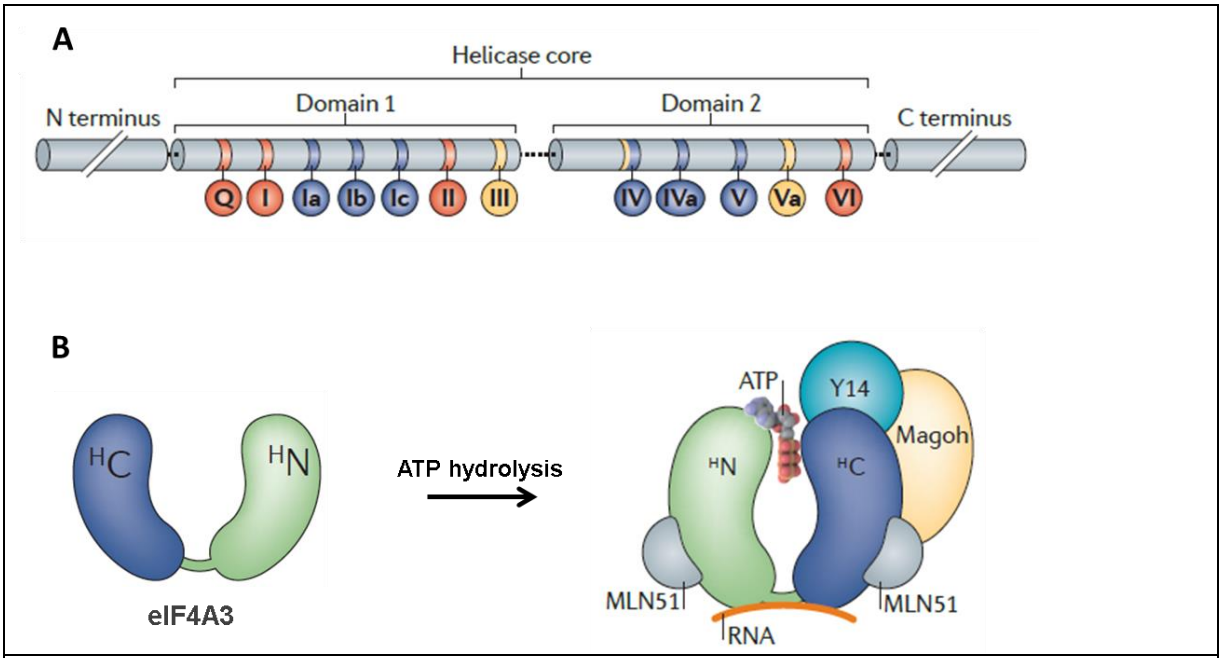

Figure 1: Structure of the DEAD box helicase eIF4A3 and substrate interactions.
A. The helicase core region of eIF4A3 consists of two recombinase A (RecA)-like helicase domains (domains 1 and 2). The conserved sequence motifs within these domains are shown, with color coding that corresponds to the primary function of the domain: ATP binding and hydrolysis (red); RNA binding (blue); communication between ATP-binding and RNA-binding sites (yellow). B. Schematic structural representation of eIF4A3 conformation changes upon ATP hydrolysis and binding of EJC core factors (MLN51 (also called Btz), MAGOH and Y14) to the two helicase core domains of eIF4A3. HN represents the amino-terminal helicase core domain and HC represents the carboxy-terminal helicase core domain (adapted from Linder et al., 2011).

<u>Mago-nashi/Y14 dimer</u>:

The two EJC core proteins Mago and Y14 (also known as Tsunagi and RBM8), form a highly stable heterodimer. This dimer formation is essential for Mago and Y14 functions (Fig. 2A,B,C) [102–104]. A slow co-evolution of the two proteins seems to be required for the

maintenance of their obligate heterodimerization mode [105]. Both Y14 and Mago are nuclear-cytoplasmic shuttling proteins [106–108]. Mago/Y14 bind to eIF4A3 during EJC assembly and are removed by active translation in the cytoplasm [109]. Mago/Y14 are then re-imported into the nucleus as a dimer by Importin 13 (Fig. 2D) [110,111].

Y14 contains a central RNA binding domain that is not involved in RNA binding but in protein-protein interaction with Mago [106,112,113]. Mago is an 18kDa protein highly conserved among eukaryotes with no recognizable sequence motif. It has an α/β structure composed of 6 β-sheets and 2 α-helices [114]. The 2 α-helices are involved in the binding of Y14 via its RNA binding domain while the β-sheet structure is likely involved in the binding with other EJC subunits. The interaction of Mago with Y14 blocks the RNA-binding ability of Y14 [114]. This observation suggests that Mago/Y14 does not contact mRNA directly but via its interaction with the eIF4A3 C-terminal domain (Fig. 2A,B,C) [90].

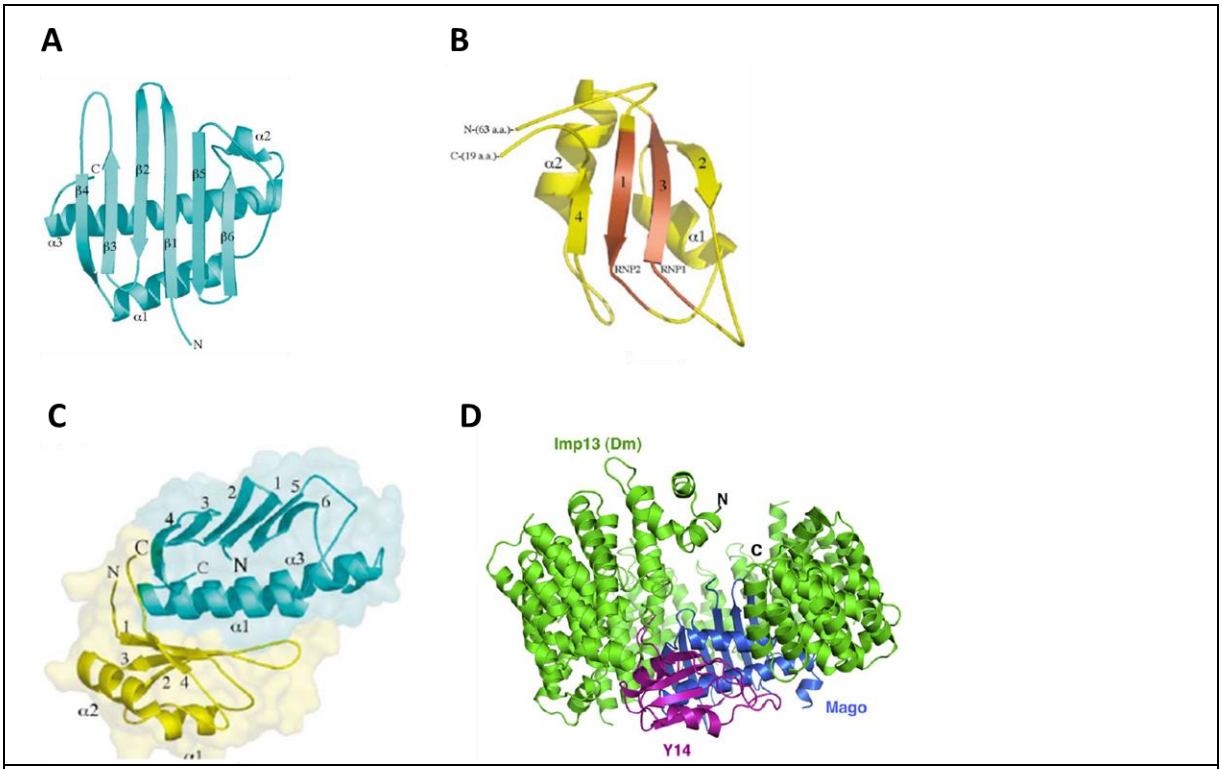

Figure 2: Structure of the Mago/Y14 heterodimer.
A. Structure of the Mago protein, β-sheets and α-helices are indicated (Lau et al., 2003). B. Structure of Y14, β-sheets and α-helices are indicated. RNA binding motif is colored in red. (Lau et al., 2003). C. Structure of Mago (blue)/Y14 (yellow) heterodimer, β-sheets and α-helices are indicated. Mago contacts Y14 at its RNA-binding motif. (Lau et al., 2003). D. Structure of Mago (blue)/Y14 (purple) surrounded by importin13 (green) in *D.melanogaster.* The interaction of Mago/Y14 with Importin 13 blocks association with other proteins (Bono et al., 2010).

Barentsz:

Btz, also known as MLN51 (metastatic lymph node 51) in human is a nucleoplasmic protein. Btz and MLN51 are structurally very similar but their sequences are poorly conserved. Btz N-terminal domain contains a coil-coiled motif followed by two nuclear localization signals. Its C-terminal domain mediates the cytoplasmic retention of the protein (Fig. 3A)[115]. Both the C-terminus and N-terminus domains are conserved across species. They are connected by a long and flexible linker. This structure, called the SELOR domain, enables Btz to wrap the surface of eIF4A3 to its both domains (Fig. 3B)[101]. The binding of Btz to eIF4A3 triggers major rearrangements in the complex, which leads to the activation of eIF4A3 ATPase activity. The association of Btz with the mRNA stabilizes further the clamping of the EJC on its substrate [90].

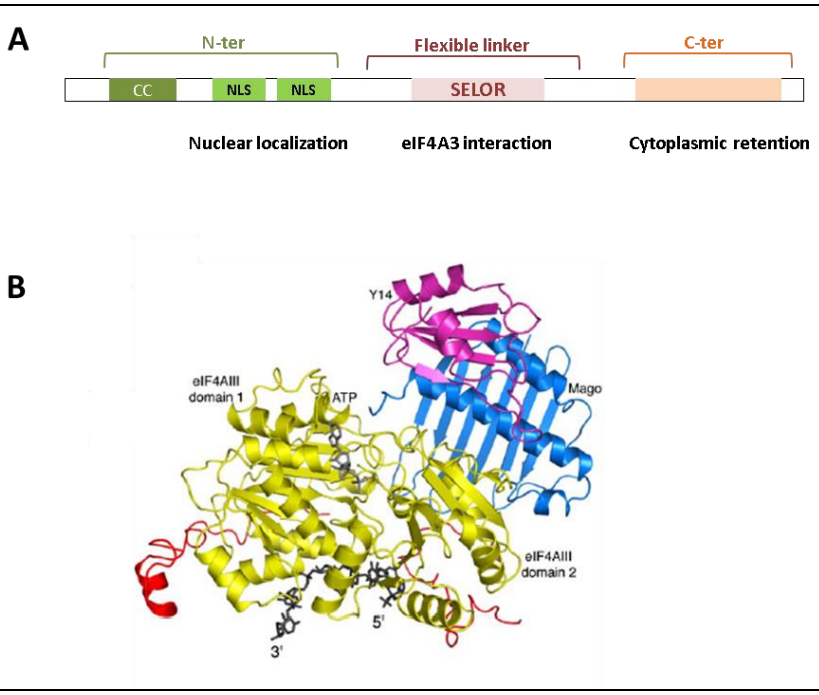

Figure 3: Structure of Barentsz in association with the EJC core.
A. Btz N-ter domain (green) contains a coil-coiled motif (CC) and two nuclear localization sites (NLS). The flexible linker (purple) contains the SELOR domain interacting with eIF4A3. The C-ter (orange) mediates the cytoplasmic retention. B. View of the human EJC. In the complex, Btz/MLN51 (shown in red) stretches around eIF4AIII (in yellow). Both proteins interact with RNA (in black), which is bound at a cleft formed between the two RecA-like domains of eIF4AIII. ATP (in gray) binds at an interface between the two domains of eIF4AIII, distinct from the RNA binding cleft. The other two protein components of the EJC, Mago (blue), and Y14 (purple), bind eIF4AIII (Bono et al. 2006).

B. PERIPHERALLY ASSOCIATED PROTEINS

The core EJC serves as a platform for other protein interactions, which enable the EJC to perform different functions. Depending on their functions, peripherally associated proteins can be classified into four groups (Fig. 4)[81,106–108,112,116,117]:

1.1. Proteins involved in **splicing function**: RnpS1, Acinus, SRm160 and Pinin
2.1. Proteins involved in **export**: UAP56, REF/Aly
3.1. Proteins involved in **translation efficiency**: PYM, Skar
4.1. Proteins involved in **mRNA quality control**: Upf1, Upf2 and Upf3

These interactions are often transitory, as it is the case for SRm160 or REF/Aly which are released from the complex just after mRNA export [107,108,112]. However, this is not a strict rule. For instance, the factor UPF3 involved in RNA quality control is assembled during splicing and is only released when the first round of translation strips off the entire complex [118,119].

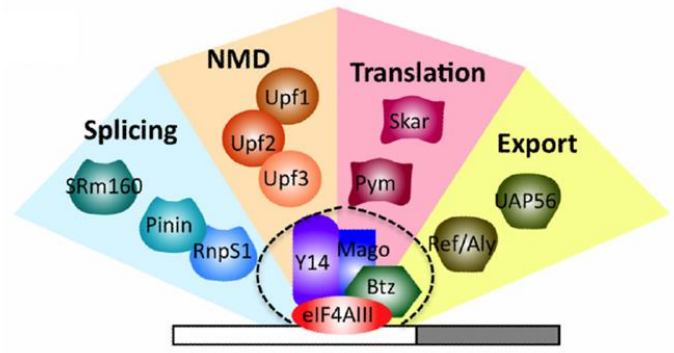

Figure 4: Proteins associated with the core EJC and their known functions.
Four proteins are tightly associated with the mRNA and constitute the core complex. Other proteins associate transiently with the complex and modulate its post-transcriptional functions. Abbreviations: RnpS1 (RNA-binding protein S1), SRm160 (the SR nuclear matrix protein of 160 kDa), Upf (Up-frameshift protein) (Roignant et al., 2010).

2.2. ASSEMBLY AND RECYCLING OF THE EJC

The assembly of the EJC on mRNA begins in the nucleus, concomitantly with the splicing reaction [120–122]. The complex travels to the cytoplasm, where it influences mRNA fate. After the first round of translation, the EJC is disassembled and is recycled into the nucleus in order to be reassembled on another transcript [88].

A. DEPOSITION OF EJC

The recruitment of the EJC occurs in a sequential manner. It takes place in the periphery of the nuclear speckle domains, which is the location where splicing factors are stored, and at proximity of pol II active transcription sites [123].

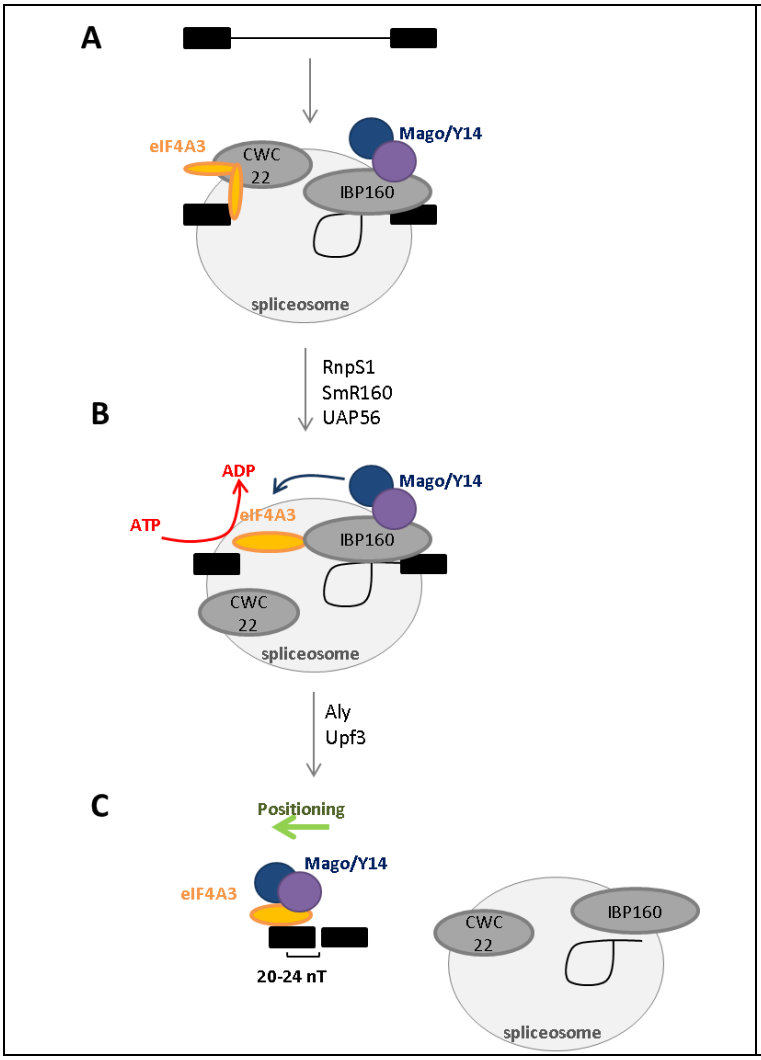

Figure 5: Assembly of the EJC on the mRNA
A. During splicing eIF4A3 (yellow) is targeted to the splicing location in an opened conformation by CWC22. Mago/Y14 (blue) is recruited by IBP160. B. under ATP hydrolysis (red), eIF4A3 adopts a locked conformation. Binding of Mago/Y14 on eIF4A3 clumps the complex on the mRNA. C. After removal of the lariat, EJC is repositioned 20 to 24 nucleotides upstream of the exon-exon junction. (For clarity issue EJC peripheral proteins are indicated but not represented on the cartoons).. (adapted from Steckelberg et al., 2012, Ideue et al., 2007, Gehring et al., 2009).

Recruitment of eIF4A3 followed by Mago/Y14 binding

The first protein to be recruited is eIF4A3. It is recruited at an early step of the splicing reaction via its binding to the spliceosome component CWC22. This interaction brings the helicase held into an opened conformation at proximity of the RNA [124–126]. The intron binding adaptor IBP160 acts then as a platform to bring EJC components together through the recruitment of eIF4A3 and RnpS1 on the one hand and the dimer Mago/Y14 on the second hand at the splicing location [127]. This scaffold organization takes place before the second step of splicing, when the mRNA structure present a free upstream exon and a lariat (Fig. 5A, B)[127].

Clumping of the complex on the mRNA and repositioning

An ATP-dependent step locks eIF4A3 in a closed conformation around the mRNA [122]. This clump results in the binding of Mago/Y14 to eIF4A3, which inhibits the ATP-binding ability of the helicase and its switch to an open conformation. This binding operates just before or during the second step of cleavage. After this second cleavage, the EJC undergoes a repositioning to its final location, 20 to 24 nucleotides upstream of exon-exon junctions (Fig. 5B,C) [121].

Recruitment of peripheral EJC factors

Peripheral EJC components join the complex at different steps of the EJC assembly. The EJC-associated splicing factors SRm160, RnpS1 and UAP56 assemble on the complex before the second step of the splicing reaction while the export factor Aly/REF and the mRNA quality control factor Upf3 join the complex at a late step of the splicing reaction, during its repositioning at the exon-exon junctions (Fig. 5) [33,112,120,128–131].

Binding of Btz

The timing of Btz association with the EJC remains controversial [88]. It has been reported that Btz localizes to nuclear speckles where the EJC is assembled via its SELOR domain [123,132]. However, cellular localization studies showed predominant cytoplasmic staining supporting Btz recruitment after nuclear export [132–134]. Therefore, more experiments are clearly required to decipher the timing of Btz assembly. It should be noticed that while Btz is required for EJC stability in the cytoplasm its function appears dispensable in the nucleus [90,122].

B. DISASSEMBLY AND RECYCLING

In the cytoplasm, the first round of translation removes most of mRNA binding proteins to their template, including the EJC [109]. The passage of ribosomes is associated with conformational changes that lead to the release of both eIF4A3 and Btz from the mRNA [92,101]. Another release mechanism involves the dissociation of Mago/Y14 from the complex by the factor PYM. PYM directly associates with Mago/Y14 by displacing Upf3 [135,136]. Subsequently, eIF4A3 ATPase activity becomes activated leading to its destabilization from the RNA (Fig. 6) [101,104,137].

In order to be re-loaded on mRNA, EJC components have to be re-imported into the nucleus. The dimer Mago/Y14 is specifically loaded to Importin 13, which binds the dimer in the cytoplasm and releases it into the nucleus in a RanGTP-dependent manner [110,111]. Importin 13 masks Mago/Y14 binding surfaces, preventing its interaction with other proteins like PYM or Upf3 (Fig. 6). It still remains to be understood how eIF4A3 is re-imported into the nucleus.

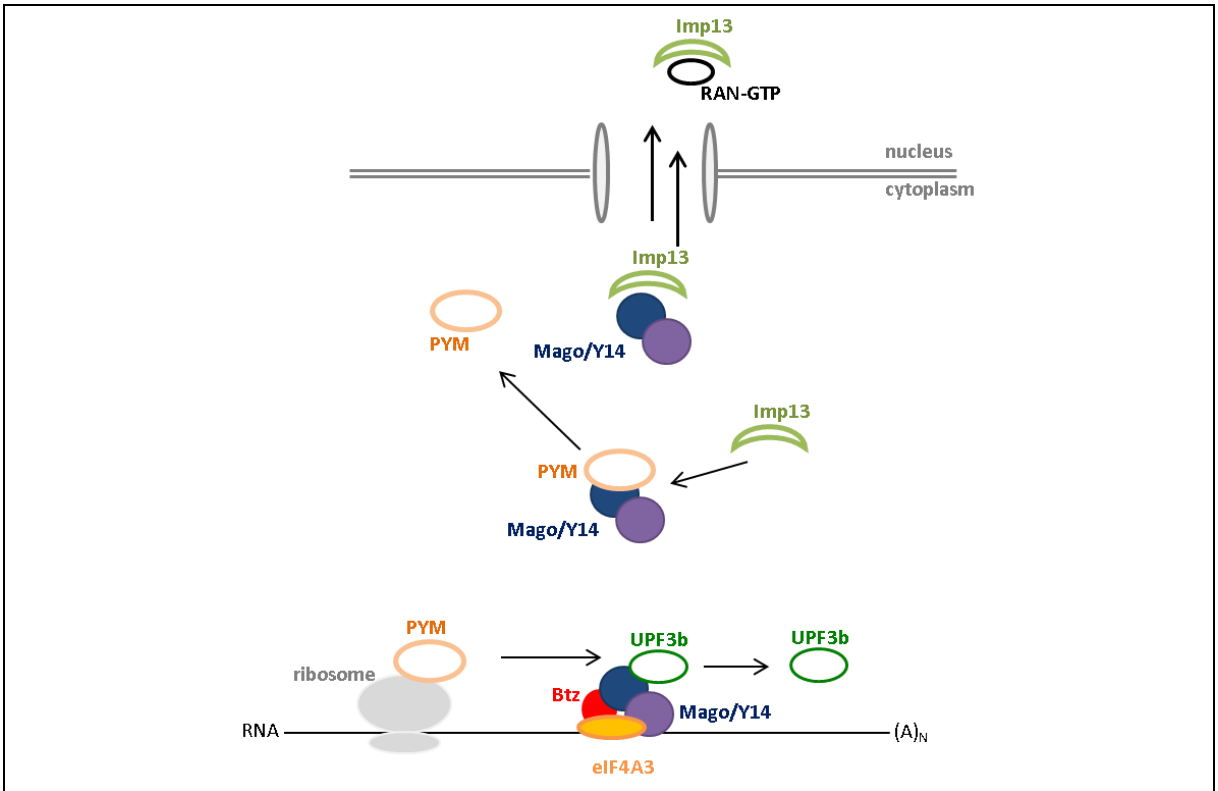

Figure 6: Disassembly of the EJC.
During the first round of translation, the ribosomal factor PYM (orange) destabilizes Upf3b (green) and binds to Mago/Y14 (blue) leading to its dissociation of the mRNA. Importin 13 (Imp13, light green) then shuttles the dimer in the nucleus in a RAN-GTP dependent manner.

2.3. FONCTIONS OF THE EJC

A. mRNA SPLICING

Initially, the EJC has been described as a mark of spliced RNA that can influence their fate in the cytoplasm. However, several recent studies challenged this view by providing evidence for a direct role of EJC in pre-mRNA splicing. The first evidence for an implication of the EJC in splicing has been provided in *Drosophila* by two independent studies [138,139], which were later confirmed in human cells [84,140].

In *Drosophila*, the nuclear core EJC regulates the splicing of a subset of genes containing long introns and preferentially expressed from heterochromatic location. It is unclear why these features confer the EJC dependency. Several hypotheses have been raised. For instance, the presence of repeated elements and satellite sequences within heterochromatic introns might render their excision particularly challenging and therefore an additional mechanism would be required to facilitate their splicing [141]. Alternatively, the specific nature of chromatin state and histone modifications of heterochromatin domains might also interfere with splicing efficiency. The fact that large introns are particularly sensitive to EJC depletion indicates that EJC plays an important role in exon definition. This has been confirmed recently in human cells where many exon skipping were detected [140].

There are currently two non-exclusive models by which the EJC might regulate exon definition (Fig. 7) [87]. On the one hand, the EJC could associate with the spliceosome machinery during exon definition and facilitates the bridging between U1 and U2 snRNPs, allowing splice site recognition and faithful splicing (Fig. 7 model 1). In absence of the EJC, exons would be poorly defined, triggering random splice site recognition and leading to exon skipping. In the second model, the action of the EJC would occur indirectly, after its initial deposition to some exon-exon junctions (Fig. 7 model 2). In this later model, the complex would stabilize larger segments formed by a combination of spliced exons. In the absence of the EJC, this stretch of exons would be poorly recognized, leading to exon skipping. Interestingly, Wang et al. recently suggested that the core EJC could modulate exon definition by influencing the rate of transcription elongation providing an alternative model for the role of the EJC in exon definition.

To conclude, more experiments are clearly required to establish the precise mechanisms of EJC during pre-mRNA splicing. It would be also interesting to determine whether the splicing function of the EJC is subject to regulation and modulation. Can the EJC regulate alternative splicing in a tissue dependent manner or is this function purely constitutive?

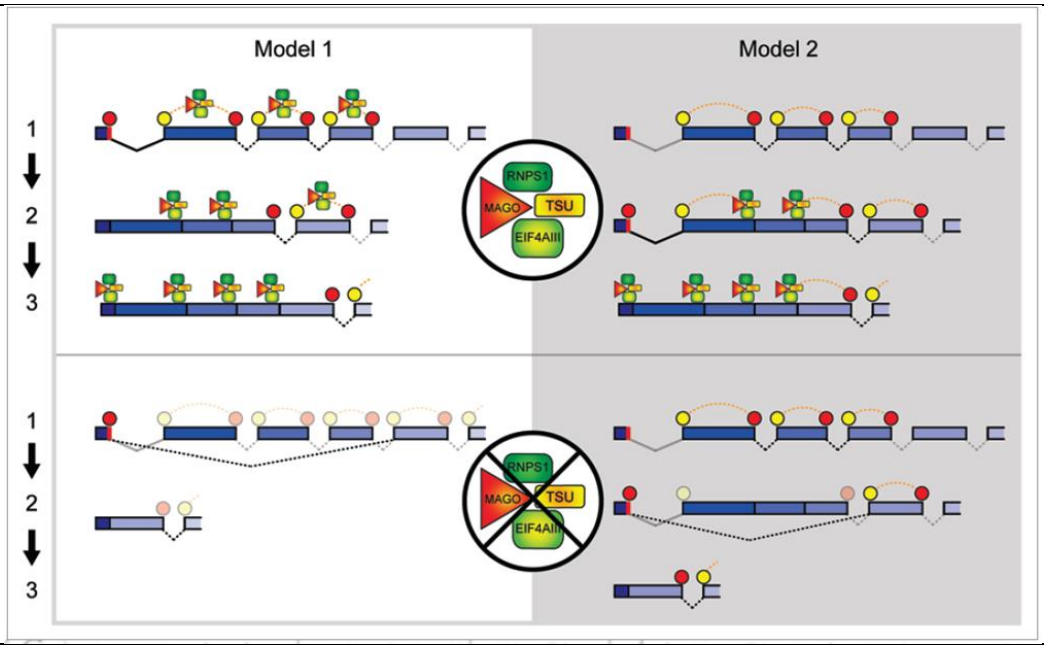

Figure 7: Two models of the EJC's function in the splicing of long introns.
The spliceosomal U1 and U2 complexes are represented by shaded circles binding to the 5'SS and 3'SS of the *mapk* transcript as splicing progresses. Exon definition interactions between U1 and U2 complexes are represented by curved dotted lines linking the U1 and U2 complexes. In model 1, the EJC physically associates with the spliceosomal complex that forms during exon definition. The EJC serves to strengthen exon definition interactions and thus allows large introns to be efficiently recognized and spliced out. Without the EJC, exon definition does not proceed as efficiently resulting in random exon skipping. In model 2, the EJC, deposited on splice junctions after splicing, serves as an extension to allow exon definition to reach across the larger exon formed by multiple joined exons. Exon definition requires that exon length be short, presumably so that U1 and U2 complexes can contact each other. In the larger exons formed in splice intermediates where multiple consecutive exons are joined together, the EJC thus would serve as a platform to maintain exon definition. In the absence of the EJC, there is a loss of exon definition of the larger exons formed in splice intermediates resulting in skipping of two or more exons (Ashton-Beaucage et al., 2011).

B. mRNA EXPORT TO THE CYTOPLASM

The first insight into a potential role of the EJC in mRNA export was brought out by the observation that spliced products were more efficiently exported to the cytoplasm compared to transcripts that did not undergo the splicing process [142]. This has been later explained by the rapid recruitment of export factors after the transcription and splicing steps have occurred [143]. The EJC-peripherally-associated protein REF/Aly, which contacts the core component Y14 [144], recruits the heterodimer TAP/NXF1:p15 (Fig. 8, top) [145–148]. This interaction makes mRNA competent for their export to the cytoplasm. However, if the EJC enhances mRNA export, it is clearly non-essential since mRNA that did not undergo splicing can also interact with REF/Aly, UAP56 and TAP/NXF1:p15 (Fig. 8, bottom) [149]. Furthermore, depletion of REF/Aly and other EJC components by RNAi do not generally affect mRNA export neither in *Drosophila* nor in *Caenorhabditis elegans*, suggesting that the EJC might regulate export of only a subset of mRNAs [150,151].

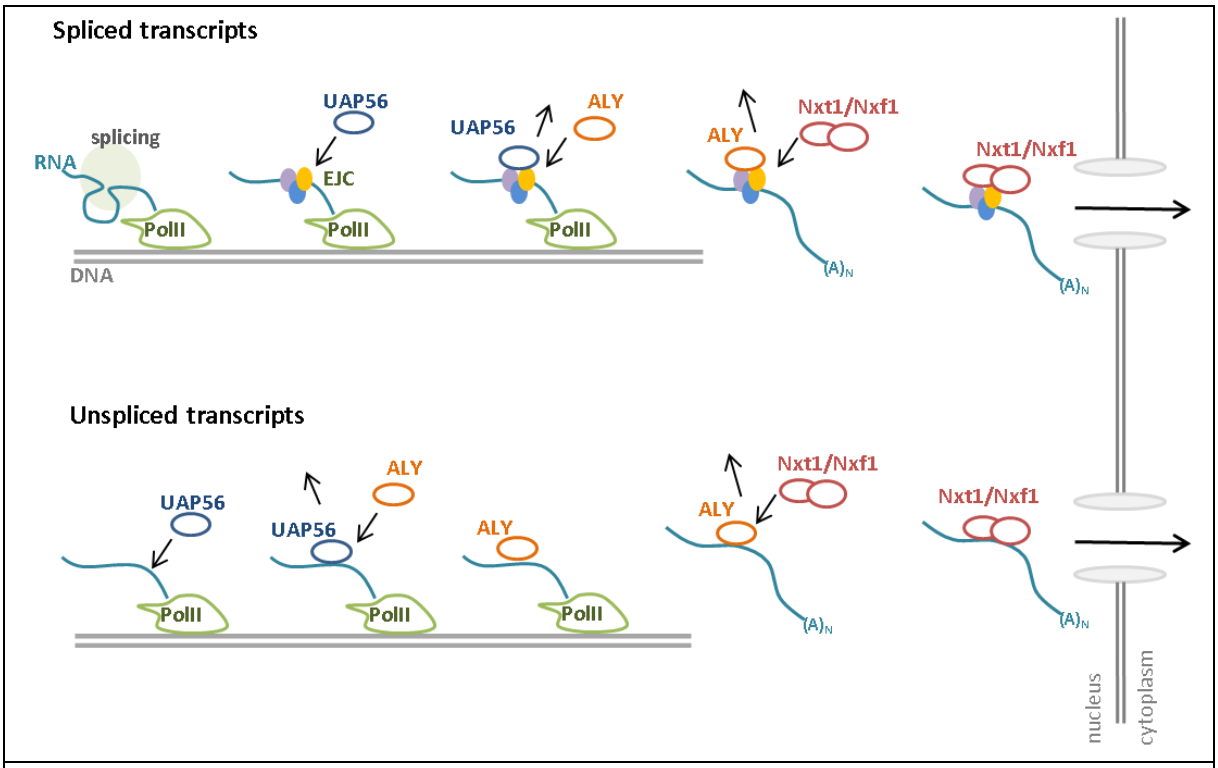

Figure 8: mRNA export to the cytoplasm
Upper panel: model of export for spliced mRNA. After co-transcriptional splicing EJC binds RNA and binds UAP56 (blue). This interaction is exchanged with Aly (orange) and then with Nxt1/Nxf1 (red) that lead the transcript to nuclear pore. Lower panel: model of export of unspliced mRNA. During transcription UAP56 binds directly mRNA. Interaction exchanges with Aly and then Nxt1/Nxf1 lead the transcript to nuclear pore. (adapted from Carmody et al., 2009).

C. mRNA TRANSLATION

The evidence of an existing link between splicing and translation has been first observed in *Xenopus* [152,153], and later confirmed in Mammals. More proteins were produced from a spliced transcript compared to its otherwise non-spliced form [154–156]. Several studies showed that the EJC has a direct implication in this phenomenon [157,158]. Two models have been proposed. On the one hand, the EJC enhances translation via its association with the SKAR protein (Fig. 9). SKAR can interact and targets the kinase S6K1 to spliced mRNAs [159], leading to the phosphorylation of factors required for translation and an increase in protein production. As S6K1 is activated in response to mTOR signaling, SKAR was proposed to link the EJC to the mTOR stimulating translational activity [160,161]. Interestingly, this mechanism concerns only non-already translated RNA, since the EJC is released from the mRNA after the first round of translation. Therefore, translation will be enhanced only for genes recently turned on or for untranslated mRNAs. This would trigger a rapid response to stimuli. The second mechanism involved the EJC component Btz. Btz was shown to enhance translation by binding directly to the translation initiation factor eIF3 via its SELOR domain [162].

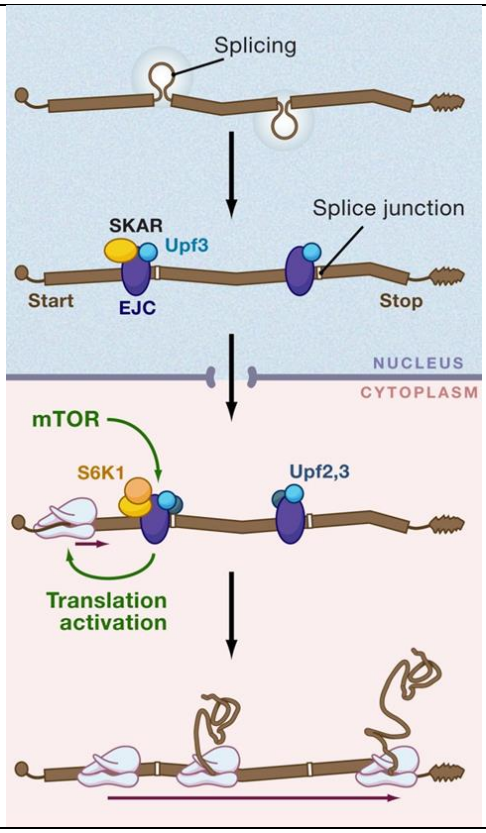

Figure 9: role of the EJC in translation activation
Splicing of a pre-mRNA in the nucleus generates an mRNA bound to two exon junction complexes (EJCs, dark blue). SKAR (yellow) joins one of these two EJCs before export of the mRNA to the cytoplasm. It is currently unclear whether SKAR joins all or only a subset of all EJCs (as depicted here). In the cytoplasm, S6K1 (orange) joins SKAR and stimulates translation upon activation by mTOR, at least during the pioneer round of translation (Le Hir et al., 2008).

D. NONSENSE MEDIATED DECAY

The process of nonsense-mediated mRNA decay (NMD) is an mRNA quality control mechanism that protects eukaryotic cells from potentially deleterious effects of premature stop codons, leading to the synthesis of truncated proteins [163]. A stop codon is recognized as premature if it is placed more than 50 to 55 nucleotides upstream of an EJC [164–168].

The factors Upf1, Upf2 and Upf3 are considered as core components of NMD [168–170]. Upf3 joins EJC in the nucleus during the splicing reaction [112,130,131,171]. Upf2 is recruited on Upf3 immediately after RNA export [107,118]. A ribosome that encounters a stop codon has the status of translation terminating ribosome. It interacts with the eukaryotic translation release factors eRF1 and eRF3, with the kinase SMG1 and the helicase Upf1. Together, they form the SURF complex [172,173]. If during the first round of translation, a premature stop codon is detected, Upf1 bridges the terminating ribosome to EJC by interacting with Upf2 and Upf3 [174,175]. Following these interactions, Upf1 is phosphorylated by SMG1, which triggers the subsequent degradation of aberrant mRNAs by decapping and 3'-5' degradation as well as translation inhibition by transient interaction with the translation initiation factor eIF3 (Fig. 10) [173,176].

NMD can also be triggered in *Drosophila, C. elegans* and plants; however it does not rely on EJC in *C. elegans* and in plants, but instead depends on the size of the 3' UTR. The exact mechanism by which the NMD occurs in these organisms remains to be determined [83,177].

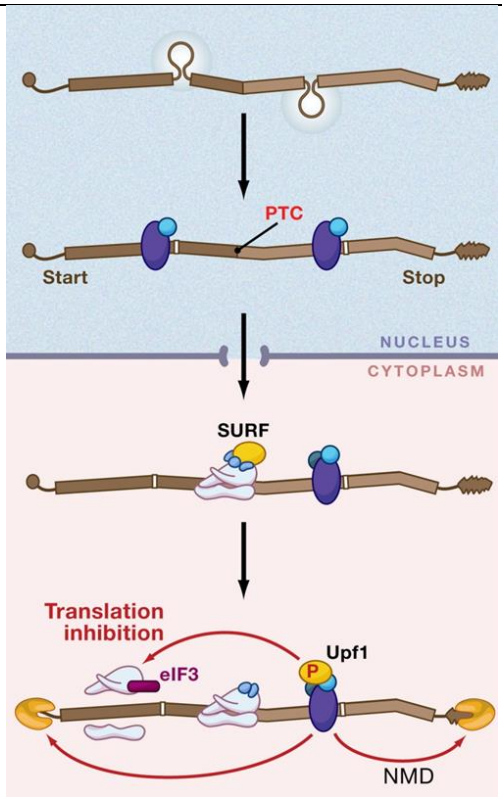

Figure 10: Role of the EJC in the nonsense-mediated decay
Transcription of a pre-mRNA containing a premature termination codon (PTC) leads to the generation of a nonfunctional mRNA bound by two EJCs (dark blue) in association with Upf3 (cyan). Following its transport to the cytoplasm, the first translational event leads to the ribosome terminating at the PTC. The SURF complex (yellow, containing unphosphorylated Upf1), which is assembled on this ribosome, then interacts with Upf2 and Upf3 (cyan and light blue) that have been recruited on a downstream EJC. This induces Upf1 phosphorylation, which inhibits further rounds of translation initiation through interaction with eIF3 before or concomitant with the induction of mRNA decay by the NMD pathway (Le Hir et al., 2008).

E. mRNA LOCALIZATION

In *Drosophila*, the four core EJC components are essential for the localization of *oskar* mRNA at the posterior pole of the oocyte [134,178,179]. This localization permits the formation of the germ plasm in the oocyte, which is a structure required to produce the germ cells in the future progeny [180]. The splicing of *oskar* is critical for EJC deposition and *oskar* localization [181]. However, only the splicing of the first intron is necessary and sufficient to ensure correct localization, while the two more downstream introns are dispensable (Fig. 11A). The reason for this specific requirement is unclear. This is not due to specific sequences within the first intron of *oskar* since it is interchangeable with intron two or three. Rather, the location of EJC deposition appears important. Recently, it was shown that the splicing of the first intron results in the formation of a stem-loop structure in the mRNA, due to the joining of exons one and two. This secondary structure was called SOLE (spliced oskar localization element) (Fig. 11B). Both EJC components and the SOLE are necessary for *oskar* subcellular localization [182]. How these

two entities functionally interact is still uncertain. The proximity of the EJC and the stem-loop both in time and space seems to be important, suggesting that the two units stabilize each other. Alternatively, a SOLE binding factor might cooperate with the EJC for *oskar* mRNA localization.

To date, *oskar* is the only mRNA known to have an EJC-dependent localization. It would be of interest to determine if this mechanism is an exception, or if it is also used for instance to localize mRNA in highly asymmetric cells such as neurons.

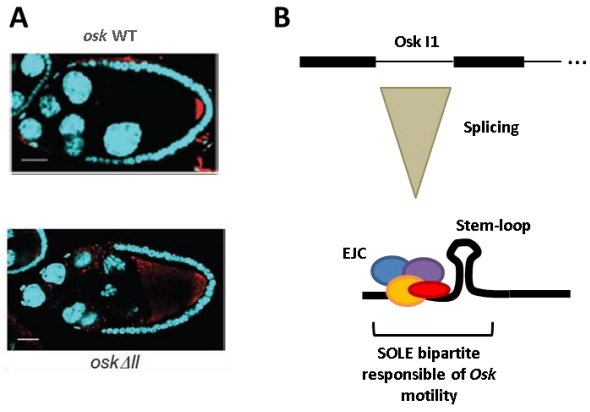

Figure 11: EJC's role in *oskar* mRNA export
A. Egg chamber stained for DNA (cyan) and *oskar* mRNA (Red) in wild type context (upper panel) and in *oskar* deleted for the first intron (Gosh et al.; 2012). B. scheme of the formation of the bipartite SOLE responsible for *oskar* localization. Under splicing of *oskar* intron 1 (*Osk* I1), the mRNA forms a stem-loop and binds the EJC at its proximity allowing correct *oskar* localization.

2.4. SPECIFICITY OF EJC BINDING AND LINK WITH THE PhD PROJECT

in vitro experiments studies have led to the suggestion that the EJC might bind every exon-exon junctions [108,127]. Recent CLIP-Seq experiments in human cells partially confirmed this assumption, demonstrating that about 80% of exon boundaries are covered by the EJC. However, this broad distribution contrasts with the specific functions observed in *Drosophila*. For instance, the EJC plays a specific role during photoreceptor differentiation, principally by regulating *MAPK* splicing [138,139]. Similarly, only *oskar* mRNA has been shown to depend on the EJC for its localization [179,181–184]. The EJC plays also a specific role in controlling axis polarity in embryo [185]. These observations suggest that the EJC binds only a fraction of exon-exon junctions in *Drosophila*. An alternative hypothesis is that despite prominent binding only a fraction of bound EJC would have a physiological function.

In order to get insights into these questions I studied the mechanistic role of the EJC in axis polarity. I found that the EJC controls this process by regulating the piRNA pathway via *piwi* splicing [85,86]. I provided evidence that splicing of only weak introns is stimulated by the EJC. We therefore propose that the EJC's specificity mainly relies on the properties of the mRNA cis-regulatory elements rather than a specific recruitment at particular exon-exon junctions.

3. piRNA PATHWAY

Small RNAs regulate gene expression at different levels. They promote post-transcriptional gene silencing by translational repression or mRNA decay, but also trigger heterochromatin nucleation leading to stable transcriptional silencing. Three major classes of small RNAs exist: the microRNAs (miRNAs), small-interfering RNAs (siRNAs) and PIWI-interacting RNAs (piRNAs). Although the mechanisms that involve small RNAs are distinct, they all implicate interactions with members of the Argonaute (AGO) family.

piRNAs are a class of small RNAs involved in the so-called piRNA pathway. They interact with the PIWI clade of the AGO family, which include Piwi, Aubergine (Aub) and Argonaute 3 (Ago3) in flies [186,187]. These non-coding RNAs of 24 to 30 nucleotides are slightly longer than miRNA and siRNA and are predominantly expressed in gonads. piRNAs were discovered in *Drosophila* [188], where they were shown to derive from transposons repeated sequences. For this reason, they were initially termed rasiRNA (repeat-associated RNA). Later, piRNA were also found in other species including worms, fishes and mice [189–191]. If their binding to PIWI proteins was conserved, it was later shown that they did not necessarily derive from repeated sequences but also from mRNA encoding genes. They were then renamed piRNAs (PIWI-interacting RNAs) [192,193]. It exists thousands of unique piRNA sequences that are not conserved among species.

The major function of the piRNA pathway is the repression of transposon expression in the germline. This can be achieved at both transcriptional and post-transcriptional level (Fig. 1) [194–200]. This pathway contributes to the protection of reproductive success. Indeed, active and uncontrolled transposition leads to DNA damage [201], which can be highly deleterious when it happens in the germ cells since those are the only cells that contribute to the future generation. In the next sections, I will mainly discuss about the mechanisms and functions of the piRNA pathway in *Drosophila*, although the general mechanisms are relatively conserved in other organisms.

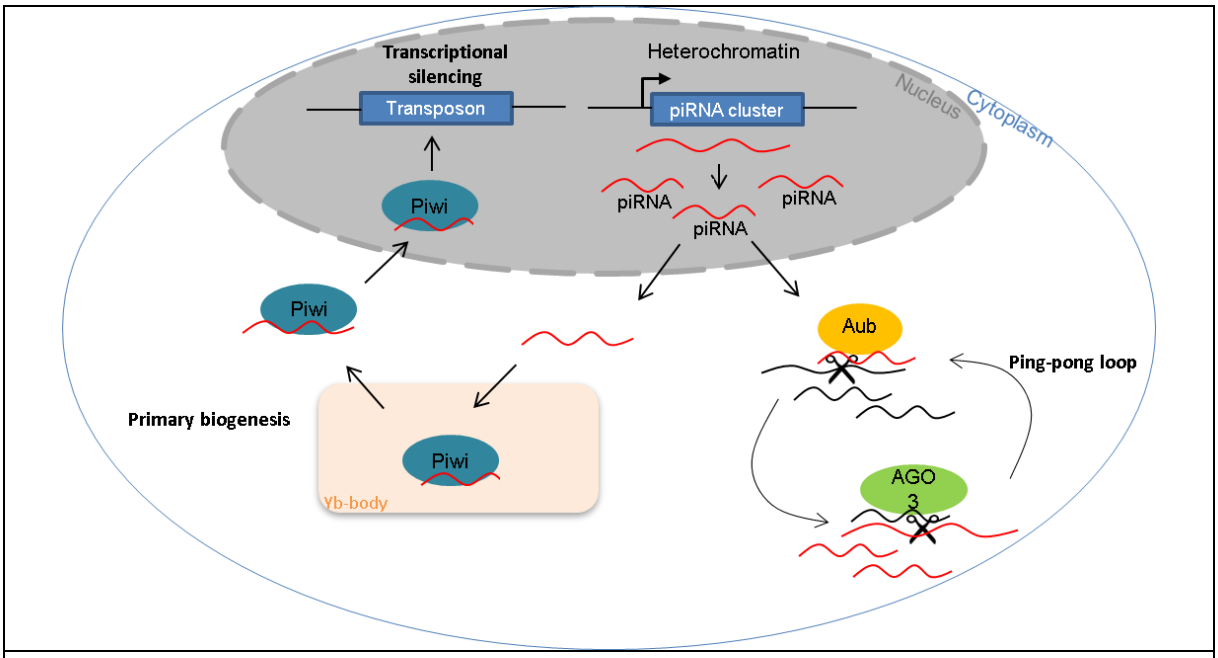

Figure 1: An overview of the piRNA pathway in *Drosophila* ovaries.
piRNAs are derived from heterochromatic piRNA clusters and processed in the nucleus. They are subsequently loaded onto Argonaute proteins in the cytoplasm, where they can target and degrade transposons through the slicing activity of their bound protein. Ago3 and Aub are expressed exclusively in the germ cells, whereas Piwi is also expressed in somatic gonadal cells. The Piwi-piRNA complex has an additional nuclear function in transcriptional silencing through heterochromatin nucleation. Other components assist the function of Argonaute proteins in this pathway.

3.1. JUMPING DNA OR THE SO-CALLED TRANSPOSONS:

Transposons are genomic sequences able to change their position within the genome. This phenomenon is called transposition. There are two classes of transposons:

(1) **Class 1 transposons:** these transposons are also called retro-transposons. Their transposition implicates a transcription of the initial locus, followed by a reverse transcription of the resulting transcript and a re-integration of the corresponding cDNA to another place in the genome. This mode of transposition is also referred to a "copy-past"-manner mechanism. These transposons are of two types. Those that are able to self-propagate by encoding their own reverse-transcriptase and are therefore autonomous. The others are non-autonomous and rely on the cellular machinery for their transposition.

(2) **Class 2 transposons:** These transposons are also called DNA transposons. They transpose via a mechanism involving a transposase, which is a DNA nuclease. The transposase excises the transposon locus, that will be subsequently re-integrated somewhere else in the genome. In this case, the mechanism of transposition acts as a "cut-past"-manner. These transposons can either be autonomous (encoding their own transposase) or not.

The piRNA pathway represses mainly the retrotransposons and to less extent DNA transposons. For instance, in the follicular cells, the piRNA pathway is important for repression of retrotransposons such as Gypsy [202–204]. Gypsy repression is important for the organism, since it is capable to form viral-like particles to invade the neighboring oocyte via a non-autonomous cell effect [205–208].

It is important to keep in mind that the piRNA pathway is not the sole mechanism leading to transposon repression. In particular, somatic tissues do not produce, or only marginally, piRNAs, suggesting additional mechanisms. Furthermore, piRNAs in zebrafish, rat and mouse are not particularly enriched in transposon sequences. This suggests that this pathway is involved in other functions that do not rely on transposon defense [209–214]. After a description of piRNA biosynthesis and involvement in transposon silencing, I will present evidences for a role of the piRNA pathway in controlling gene expression.

3.2. MECHANISMS OF piRNA BIOSYNTHESIS IN *DROSOPHILA*

A. TRANSCRIPTION OF piRNA CLUSTERS INTO LONG RNA PRECURSORS

Drosophila ovaries are structured in 16 ovarioles. An ovariole consists in the succession of egg chambers, maturating from the germarium, which is the location of germinal stem cells, to the mature oocyte. Each egg chamber is composed of an oocyte at the posterior pole and 15 nurse cells that produce RNAs and proteins for the developing oocyte. The nurse cells are polyploid and are integral components of the germline. They are surrounded by a single layer of somatic cells, the so-called follicular cells (Fig. 2A).

piRNA are transcribed and processed from discrete loci in the genome termed piRNA clusters. While these regions are mostly located in pericentromeric and telomeric heterochromatin, few piRNAs also derived from 3'UTR regions of protein coding genes [194,215] (Fig. 2B). In *Drosophila*, piRNA clusters are largely enriched in non-functional transposon sequences and degenerate repeated elements [194]. Expression of these clusters is cell-type specific. For instance, the piRNA cluster that accounts for a large proportion of piRNAs in follicular cells is called *Flamenco* [202,216]. This locus contains transposon remnants distributed only in the antisense orientation. Other types of piRNA clusters expressed in the germ cells have repeated sequences randomly oriented in sense and antisense direction [194,196].

piRNA clusters can be divided into two groups: uni-strand clusters that give rise to piRNAs and map predominantly to only one strand (e.g Flamenco) and dual-strand clusters where piRNAs originate from both DNA strands (e.g. 42B). Dual strand clusters are exclusively expressed in the germline [194,196].

The mechanisms of piRNA cluster expression remain unclear [194,217]. Promoters and regulatory sequences have not yet been clearly defined. While single strand long precursors are generated by RNA pol II transcription [200,218,219], it is not yet known what distinguished these transcripts from canonical pol II transcripts in order to be correctly processed into mature piRNAs.

Recent studies gave insights into the mechanisms of dual-strand piRNA cluster expression. These clusters do not have their own promoters. Instead, their expression relies on read-through transcription from flanking protein-coding genes. The HP1 homolog Rhino binds to the heterochromatin mark H3K9me3 of the piRNA clusters and somehow facilitates this non-canonical transcription. In addition, Rhino recruits Cutoff and Deadlock to the cleaved nascent piRNA transcripts in order to prevent their degradation due to uncapping. Finally, Rhino also recruit UAP56, which inhibits the splicing of piRNA clusters transcripts (Fig. 2C) [217,220–223]. This inhibition seems important to distinguish piRNA clusters from mRNA protein coding genes. Supporting this hypothesis, tethering Rhino to a GFP reporter containing intron located in euchromatin is sufficient to block splicing and generate piRNAs, as long as this construct is expressed from both directions. This indicates that Rhino is the main determinant of piRNA production from dual-strand clusters. However, given that Rhino binding depends on its association with H3K9me3 it is still unclear why Rhino binds only to dual strand clusters, as this mark is also present on single strand clusters and other genomic regions. More experiments are required to elucidate this paradox.

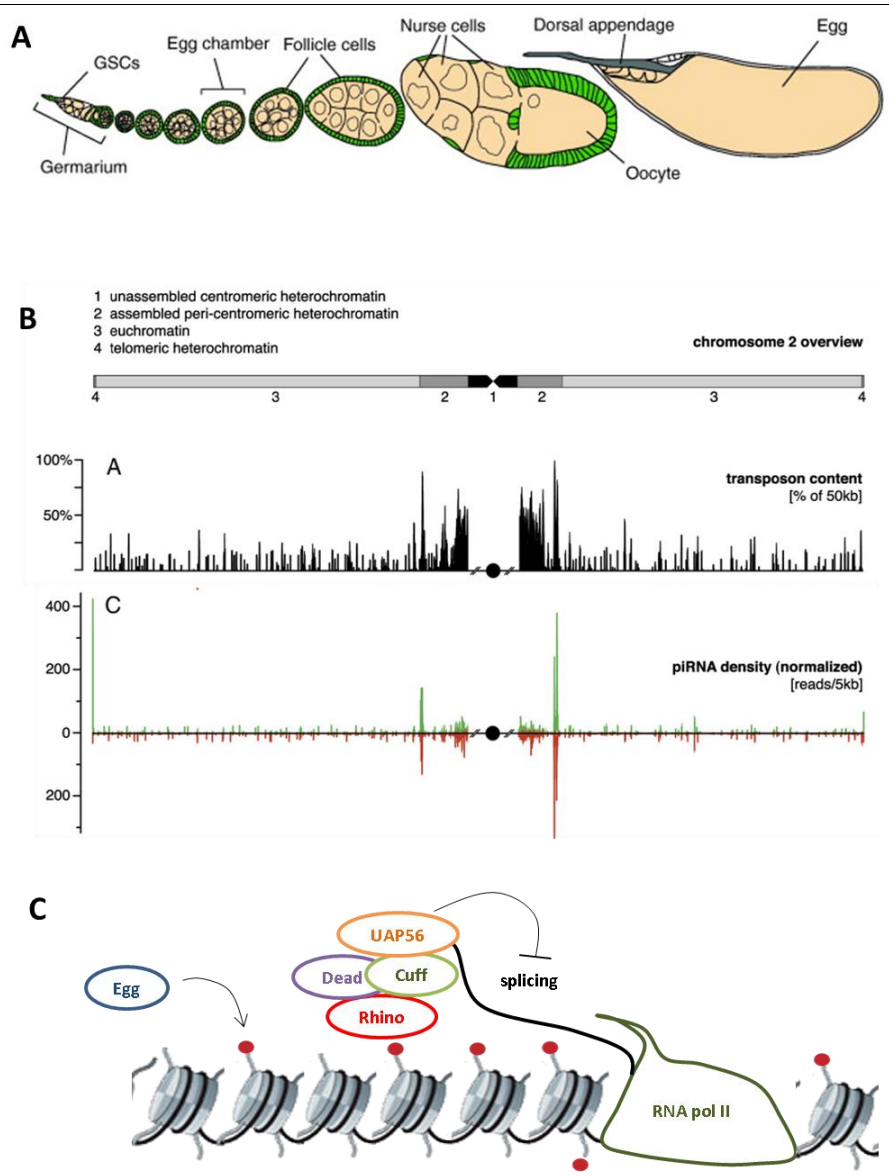

Figure 2: Transcription of heterochromatic piRNA clusters in long RNA precursors
A. Scheme of *Drosophila* ovariole that contain maturating egg chambers from the germarium on the left to the mature egg on the right. These egg chambers are composed of somatic follicular cells (green), and the germ cells (orange) (adapted from Senti et al. 2010). B. *Drosophila* piRNAs map to discrete genomic loci. Top: sketch of chromosome 2 with the major chromatin domains. In black, density of annotated transposons. Bottom: density of piRNAs (green: sense, red: antisense) along the chromosome 2 (normalized for the number of times it maps to the genome). Axis indicate relative units, the centromere is shown as a circle (adapted from Brennecke et al. 2007). C. Model of piRNA cluster transcription in the germline. Eggless (egg, blue) deposits H3K9me3 histone mark that is read by Rhino (red). In complex with Deadlock (Dead, purple) and Cutoff (Cuff, green), it binds the uncapped nascent piRNA cluster transcript and protects it from degradation. Binding of UAP56 (orange) leads to splicing inhibition.

B. PROCESSING OF PRECURSOR TRANSCRIPTS INTO piRNAS

After transcription, long precursor transcripts are exported into the cytoplasm in order to be processed into smaller piRNAs. Mechanistic insights of their export in the germ cells have recently been reported. The splicing and export factor UAP56 is recruited co-transcriptionaly and accompanies piRNA transcripts to the nuclear pore [223,224]. UAP56 delivers the transcripts to the helicase Vasa, which is present at the cytoplasmic side of the pore. Then, Vasa transports the transcript to a structure called nuage, which is the location of piRNA processing in the germline (Fig. 3) [224].

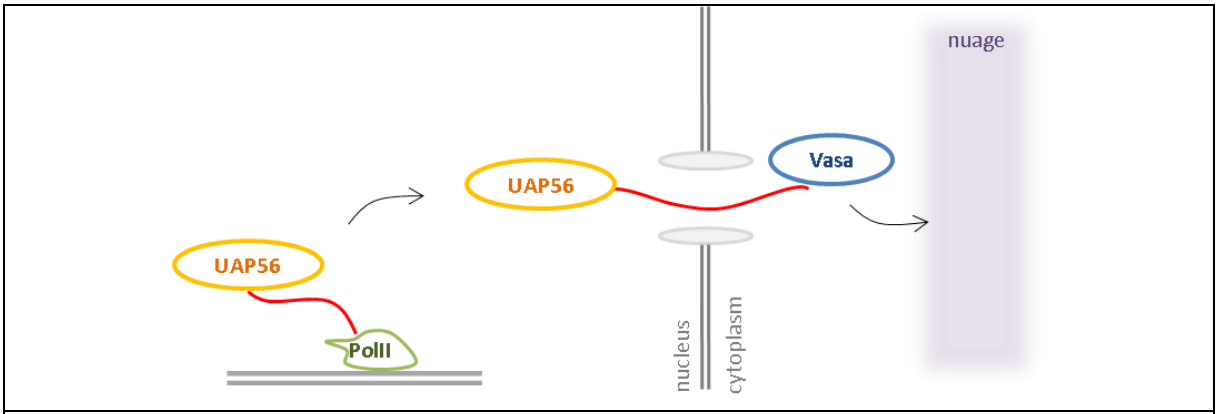

Figure 3: Transport of long piRNA precursors into the cytoplasm in the germline.
UAP56 (orange) binds to the nascent transcript (red) and localizes to nuclear pore. The transcript is taken off by the helicase Vasa (blue) that shuttles the transcript to the nuage (purple shadow), where it will be processed.

C. PRIMARY BIOSYNTHESIS

The processing of long piRNA precursors consists of their cleavage into smaller parts followed by their subsequent loading onto PIWI proteins. Their cleavage is mediated by an endonuclease located at the mitochondrial surface called Zucchini (Zuc) [198,225–227]. Zuc cuts long piRNA intermediates in a non-selectively manner and in size-variable pieces. The cleavage results in the presence of a monophosphate group on the 5' end of the processed RNA product, which is a characteristic mark of mature piRNAs [228–232]. The monophosphate-5'-ended fragments are then loaded onto PIWI proteins. A bias for the presence of a uridine in the 5' end of mature piRNAs has been observed [196]. Since Zuc cuts non-specifically, this suggests that the binding of processed piRNAs to Piwi is selected in 5'-Uridine-ended piRNAs. Other unloaded fragments are likely unstable and degraded. Piwi-bound piRNAs are then cut to their appropriate size (24 to 30 nucleotides) by an unknown endonuclease. This endonuclease is probably not Piwi itself since its slicer activity does not seem to be required for piRNA biogenesis [233]. piRNAs are

subsequently 2'-O-methylated by the methylase Hen1 [234–238]. This methylation prevents piRNA degradation (Fig. 4) [239].

The primary piRNA biogenesis in follicular cells occurs in cytoplasmic granules, called Yb-bodies [198,229,240]. In these granules, Armitage and Vreteno, are required for piRNA loading onto PIWI (Fig. 4) [198,227,229,241]. In principle, the Piwi-RISC complex could bind by base-paring mRNAs that contain sequences antiparallel to bound piRNA and degrade them via its endonuclease activity. However, the post-transcriptional silencing triggered by Piwi after primary biosynthesis is controversial. Indeed, the slicer activity of Piwi appears dispensable for its function [233]. In contrast, it is now clearly established that Piwi is involved in nuclear transcriptional silencing via a mechanism that relies on heterochromatin formation. The details of this mechanism are discussed below (Fig. 1)[242].

In the germ cells, the primary biogenesis is also active but presents some differences. Armitage and Vreteno are also required for Piwi loading [227,229,241]. In contrast, Yb is not expressed and the loading occurs in a perinuclear structure called nuage. The Yb function might be carried out by the two Yb homologs, Sister and Brother of Yb (So-Yb and Bo-Yb) (Fig. 4) [227,228,243].

After Piwi-dependent primary piRNA biogenesis, mature piRNAs are either used in the ping-pong amplification loop, or remain loaded onto Piwi for transcriptional silencing.

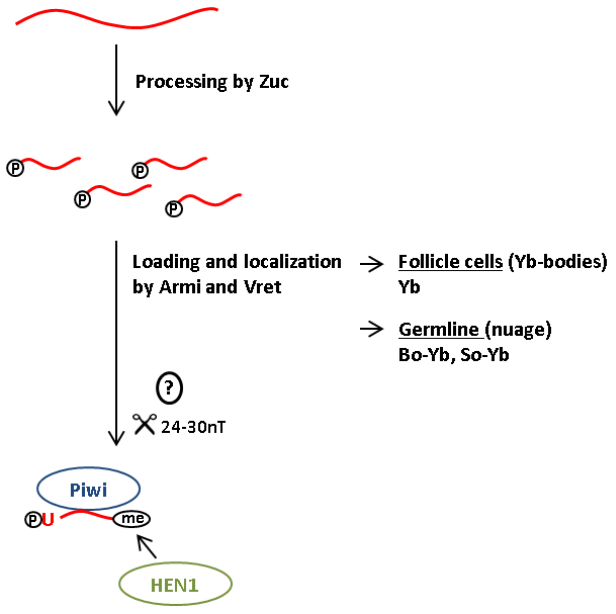

Figure 4: Processing of long precursors into mature piRNAs.
In the cytoplasm, long precursors are cut by Zucchini (Zuc) that results in a monophosphate-5'-ended small RNAs (encircled P). 5'U-ended small RNAs are loaded onto Piwi (blue), with the help of Armitage (Armi) and Vreteno (Vret). An unknown endonuclease cuts the small RNA to 24 to 30 nucleotides. Methylation is performed by the methyltransferase Hen1 (green) giving rise to mature piRNAs. Loading of piRNAs occurs in Yb-bodies (follicular cells) where Yb is expressed, or in the nuage (germ cells) where Sister of Yb (So-Yb) and Brother of Yb (Bo-Yb) are expressed).

D. PING PONG AMPLIFICATION LOOP

While piRNA primary biogenesis occurs in both cell types, piRNA amplification is only triggered in the germline by a mechanism called the ping-pong amplification loop. This loop takes place in the perinuclear nuage and involves the two PIWI-proteins Aubergine (Aub) and Argonaute 3 (AGO3). It leads to the production of a large amount of piRNAs involved in post-transcriptional silencing of transposon.

In this mechanism, Aub-bound piRNAs anneal to transposon transcripts by base pairing, which results in the cleavage of these transcripts by the slicer activity of Aub. This processing gives rise to a second set of piRNAs called secondary piRNAs [194]. These secondary piRNAs, which are in sense orientation relative to transposons transcripts, are loaded onto AGO3 and base pair with piRNA cluster transcripts, leading to their cleavage by AGO3. This produces another set of piRNAs that are loaded onto Aub to feed the amplification loop (Fig. 5).

The set of piRNAs produced from AGO3 slicing presents two distinct features. These piRNAs present an adenine residue at the 10th position. Furthermore, a 10-bases overlap is observed between both piRNA types derived from the ping-pong amplification loop. This results from the assembly of Aub into mature 5'-U-ended primary piRNAs and the cleavage of transposon transcripts after the 10th nucleotide. AGO3 slices according to a similar scheme, giving rise to overlapping piRNA sequences [194,195].

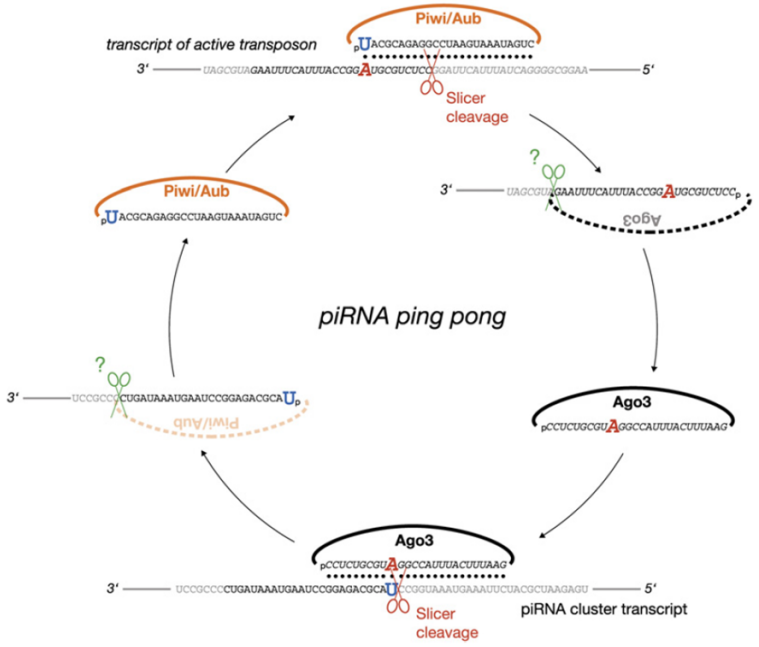

Figure 5: piRNA ping-pong model.
Ping-pong amplification loop consisting of piRNPs, piRNA cluster transcripts, and transcripts of active transposons. Nucleotide cleavage events are shown as scissors. Potential sources of primary piRNAs are piRNA cluster transcripts and maternally inherited piRNA complexes (adapted from Brennecke et al. 2007).

E. CLASSIFICATION OF piRNA PATHWAY COMPONENTS BASED ON THEIR FUNCTION IN piRNA BIOGENESIS

In addition to PIWI-proteins, numerous other factors are implicated in the piRNA pathway. These factors can be classified based on their roles in the follicular and germ cells. Factors that are required in both cell types are distinguished from those required only in the germ cells [244]. For instance, components required in both cell types are Zucchini (Zuc), which is involved in piRNA processing [198,225–227], as well as Armitage (Armi), Vreteno (Vret), Brother and Sister of Yb (Bo-Yb and So-Yb) and Shutdown, involved in Piwi loading and primary biosynthesis [198,227,241,244,245]. In contrast, components only required in the germline include Vasa, which is implicated in the shuttling of piRNA to the Nuage [224] as well as Spindle-E and Krimper, which act together with Aub and AGO3 in the amplification loop. Many other proteins have been identified but their precise mechanisms still remain to be elucidated.

3.3. MECHANISMS OF TRANSCRIPTIONAL SILENCING

In addition to the primary piRNA biogenesis, Piwi has also a role in the transcriptional silencing of transposons. In the current model, the action of Piwi implicates epigenetic mechanisms triggering heterochromatin formation at the transposon loci. This is followed by the inhibition of transcription initiation (Fig. 6).

After primary biogenesis, Piwi can remain loaded with piRNA. This Piwi-RISC can translocate to the nucleus by an undetermined mechanism. In the nucleus, the Piwi-RISC is recruited to transposon loci via an RNA-recognition mode. piRNAs serve as specific guides to target nascent transposon transcripts by sequence complementarity. With the help of the cofactor Maelstrom, Piwi recruits the epigenetic machinery to the transposon loci. This requirement promotes heterochromatin formation by deposition of repressive histone marks [242,246–248]. H3K9me3 is deposited, but other histone modifications are probably also involved [242]. Heterochromatin nucleation results in the inhibition of transcription initiation accompanied by a decrease of Pol II occupancy at transposon loci [242,249]. How precisely Piwi influences the chromatin state remains elusive. While Piwi can bind Hp1a, the domain of Piwi requires for this interaction seems dispensable, suggesting additional mechanisms. Nonetheless, the binding of HP1a to H3K9me3 appears to stabilize the newly formed heterochromatin (Fig. 6) [246,250,251].

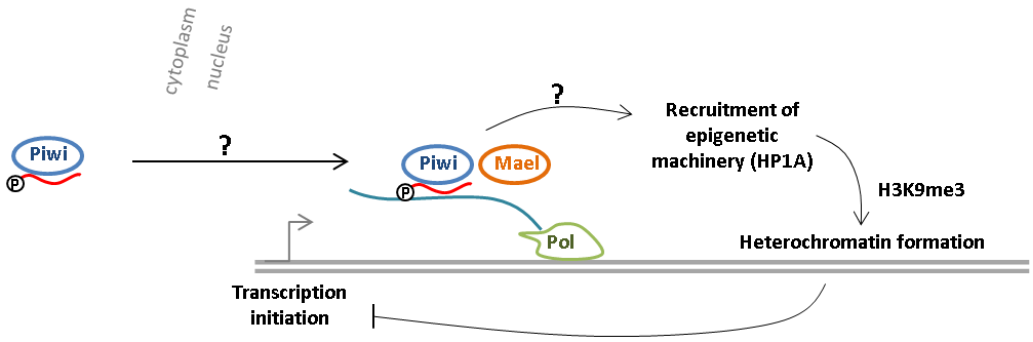

Figure 6: Transcriptional silencing.
Piwi-RISC (blue) is translocated to the nucleus by an unknown mechanism. Nascent transposon transcripts (blue) are recognized by the Piwi-RISC by base-pairing. Together with Maelstrom (Mael, orange), the complex recruits the epigenetic machinery by an unknown mechanism, which deposits H3K9me3 marks on histones, thereby triggering heterochromatin formation and transcription inhibition.

3.4. ROLE OF THE piRNA PATHWAY IN GENE REGULATION

In *Drosophila,* not all piRNAs target transposons. Some of them are indeed mapping to protein coding genes [193,194,252–255] suggesting an additional role in the regulation of gene expression. In several species, a somatic role for piRNAs has been proposed. For instance, piRNAs have been shown to regulate synaptic plasticity in Aplysia [256]. Moreover, some studies reported an up-regulation of human Piwi in cancer [257]. However whether this up-regulation is linked to changes in gene expression remains to be investigated. Here, I will describe different roles of the *Drosophila* piRNA pathway in controlling gene regulation, independently of its function in transposon silencing.

A. TRANSCRIPTIONAL REGULATION OF PROTEIN CODING GENES

Two studies described the role of the piRNA pathway in inhibiting gene expression at the transcriptional level. The first used a genome-wide approach to describe how insertion of transposons in a protein coding gene lead to Piwi-dependent silencing of that gene by spreading of H3K9me3 repressive mark [242]. However, although this mechanism leads to inhibition of gene expression it is still mediated via the presence of transposons. A more direct mechanism described by the group of Siomi involved the repression of the *Fas II* genes via piRNAs derived from the 3'UTR of the gene *traffic-jam.* The mechanism involved also appears to rely on H3K9 methylation [215]. Finally, our lab has shown that the piRNA pathway can also be required for gene activation. Many genes expressed from heterochromatin location appear to depend on the piRNA pathway for their expression. This is the case for *Ago3* for instance. The mechanism also seems to rely of H3K9me3 deposition, which was previously shown to be important for heterochromatin gene expression. The mechanism underlying the positive activity of this repressive mark at heterochromatic loci remained to be elucidated.

Maternal mRNA decay by the piRNA pathway

Maternal mRNAs deposited in the oocyte during oogenesis are later degraded and replaced by zygotic mRNAs during a transition termed maternal-to-zygotic transition. A study in *Drosophila* showed a role for the piRNA pathway in the decay of the maternal transcript *nanos* during this transition [258]. *nanos* mRNA contains two transposon sequences in its 3' UTR that are targeted by piRNAs in complex with Aub. Base-pairing between pi-RISC and *nanos* transcript leads to *nanos* degradation. At present, it is still unclear whether other mRNAs are regulated by the same mechanism.

piRNAs derived from pseudogenes

A pseudogene on the Y chromosome, called Suppressor of Stellate Su(Ste), produces piRNAs in the testis that target *Stellate* mRNA for degradation (Fig. 7A). If this pathway is impaired, Stellate accumulates leading to infertility [259]. In the testis, 70% of Aub-associated piRNAs map to the Su(Ste) sequence, suggesting that *Stellate* mRNA is the main target of the piRNA pathway in this tissue [261]. Pseudogene-derived piRNAs have also been described in other species [262].

piRNAs formation by *cis*-natural antisense transcript derivation

Cis-natural antisense transcripts consist in the formation of piRNA clusters in antisense of an endogenous gene. They can produce piRNA targeting the mRNA of that gene (Fig. 7B). This piRNA-cluster formations have been observed subsequently to recent insertions of transposable elements in antisense within the gene [263]. These piRNA may affect the mRNA expression transcriptionally or post-transcriptionally.

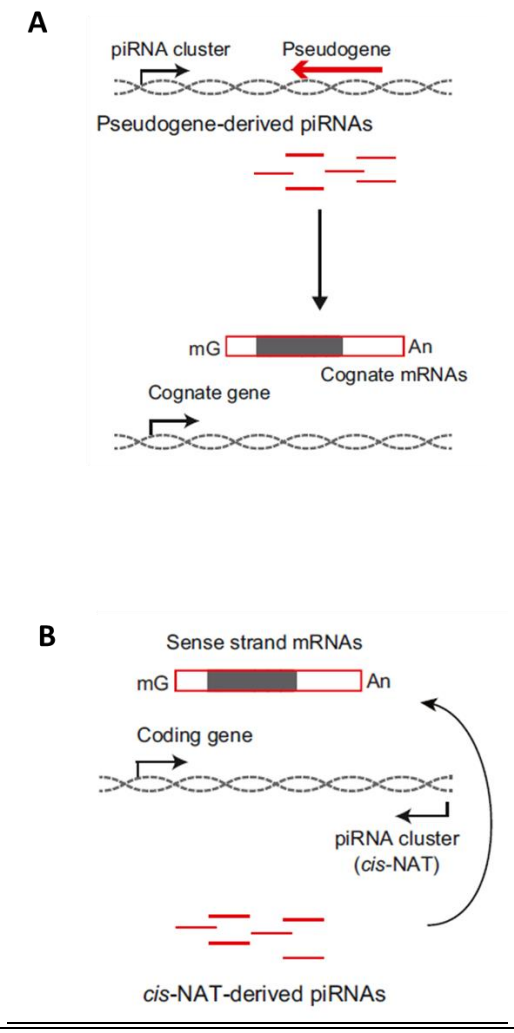

Figure 7: Regulation of mRNAs by piRNAs derived from pseudo-genes and *cis*-natural antisense transcripts.
A. Pseudogene-derived piRNAs regulate mRNAs. Pseudogenes are located in piRNA clusters in an antisense orientation to piRNA cluster transcription. Therefore, piRNAs derived from pseudogenes are in antisense to the target genes. These piRNAs can then target the cognate mRNAs. B. cis-natural antisense-transcript-derived piRNAs regulate sense mRNA. Newly defined piRNA-clusters are transcribed in antisense to the target gene such that piRNA derivated from them can regulate the sense mRNA (adapted from Watanabe et al., 2014).

AIM OF THE PhD PROJECT

When I started my PhD, our group and others had recently discovered a novel role for the EJC in pre-mRNA splicing, in particular for transcripts containing large introns and expressed from heterochromatin location [138,139]. However, the exact mechanism underlying this function remained unknown. Therefore, the objective of my PhD was to decipher the mechanism(s) of the EJC in pre-mRNA splicing and to understand the basis of its specificity. To this end, our approach has been to search for novel developmentally regulated EJC targets.

It had been previously demonstrated a role for the EJC in controlling egg axis polarity [264]. Interestingly, this function was observed only for Mago and Y14 but not for the cytoplasmic subunit Btz, suggesting that the EJC controls this developmental process via a nuclear mechanism, potentially via pre-mRNA splicing. Furthermore, it was clearly established that an impairment of the piRNA pathway leads to axis specification defects [185,189]. Altogether, these observations suggested a potential link between the EJC, pre-mRNA splicing and the piRNA pathway. Therefore, the immediate objective of my PhD has been to investigate a potential implication of the EJC in this pathway.

We found that the EJC, together with its splicing subunit RnpS1, were required to prevent transposon accumulation in *Drosophila* ovaries. In addition, we demonstrated that this function is mediated via the control of the splicing of *piwi* transcripts.

I then characterized in more details the mechanism by which the EJC controls *piwi* splicing. Based on our experiments, we proposed a model in which the EJC facilitates the splicing of *piwi* intron and other non-canonical introns after its binding to adjacent exon-exon junctions. This model implies that the EJC plays an essential role in the kinetics of intron removal.

RESULTS

1. THE EXON JUNCTION COMPLEX MODULATES THE piRNA PATHWAY

1.1. THE NUCLEAR CORE OF THE EJC PREVENTS TRANSPOSON DEREPRESSION

In order to investigate a possible role of the EJC in the piRNA pathway we examined whether the specific knockdown of individual EJC components in the ovaries alter the level of transposons. To perform these knockdowns, I took advantage of Drosophila stocks bearing double stranded RNA (dsRNA) constructs from the VDRC (Vienna Drosophila Resource Center, Vienna, Austria) and the TRiP (Transgenic RNAi project, Boston, US) collections. By driving specifically their expression in the tissue of interest with an appropriate driver, these strains produce dsRNA directed against the targeted gene, leading to its silencing. We first confirmed the functionality of these constructs by driving their expression in eye imaginal discs, where our group previously shown a requirement for the EJC in specification of photoreceptor cells [138]. As expected, flies expressing dsRNA-*Mago* or dsRNA-*Tsu* (Tsunagi/Y14) failed to properly differentiate photoreceptor, whereas flies expressing dsRNA *Btz* (Barentsz) displayed a wild type phenotype (Fig. 1A).

I then expressed these dsRNA in either the germline or in the surrounding follicle cells using appropriate drivers. To perform knockdown in follicular cells, I used the specific follicular driver *traffic jam* (*tj*)-GAL4, and for germline knockdown I used the specific driver *nanos* (*nos*)-GAL4 (Fig. 1B). I compared the effects of EJC depletion with the depletion or the loss of function of *piwi* by examining both the ovarian morphology and the level of representative transposons.

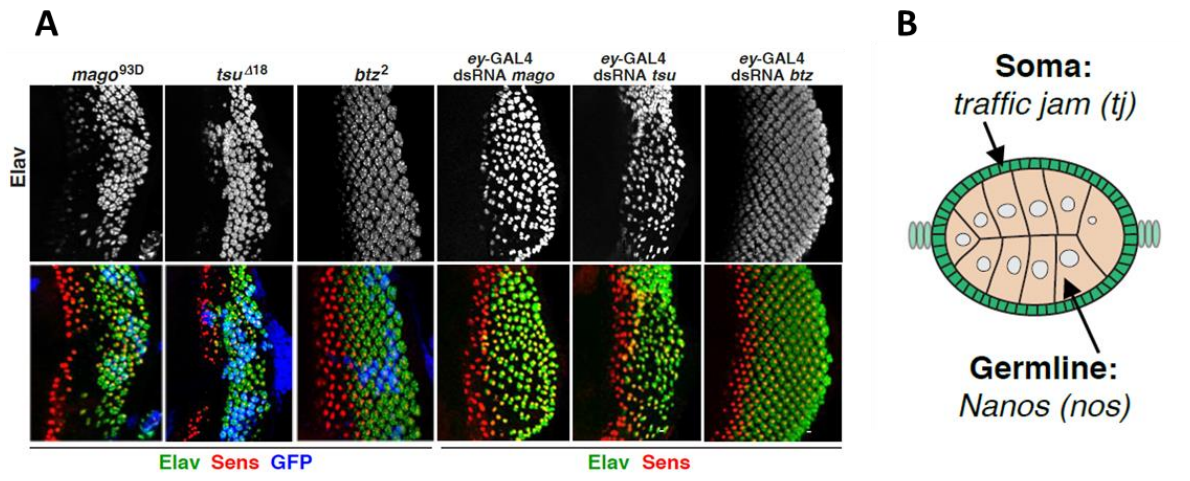

Figure 1: strains used for EJC knock down
A. Knockdown of pre-EJC components recapitulate their mutant phenotypes. Eye discs containing mutant clones of the indicated genotypes marked by the absence of GFP (blue) (left panels) or expressing the indicated double stranded RNAs (right panels). Eye discs are stained with anti-Elav (green) to label all photoreceptors and anti-Senseless (red) to label R8 photoreceptors. Mutations or KD of *mago* and *tsunagi* lead to similar defects in photoreceptor differentiation while loss of *btz* has no effect. B. Schematic representation of an egg chamber showing somatic cells (green) expression traffic jam, and germline cells (orange) expressing nanos.

A. KNOCKDOWNS OF EJC NUCLEAR CORE LEAD TO RUDIMENTARY OVARIES IN FOLLICULAR CELLS

piwi mutant animals were previously shown to be sterile due to failure in oogenesis [265]. We confirmed this result, showing that *piwi* mutant (*piwi1/piwi2*) animals have atrophic ovaries. Similarly, I found that the somatic knockdown of *mago* and to a lesser extent *tsu* (tsunagi) in follicle cells gave rise to rudimentary ovaries, which were devoid of mature egg chambers (Fig. 2A). In these knockdown ovaries, follicle cells formed multicellular layers and failed to encapsulate developing egg chambers, similarly to the phenotypes described with the *piwi* knockdown and *piwi* loss of function alleles (Fig. 2A and [265,266]).

To gain insights into these developmental defects I stained the ovaries with the Hts (Hu-li tai shao) antibody, which marks the spectrosome and the cell cortex. The spectrosome is a spherical, germinal stem cell-specific organelle located in the germarium. Germinal stem cells divide in an asymmetrical manner to produce cystoblasts that undergo four rounds of incomplete divisions leading to a group of interconnected cells. Under differentiation of the cystoblast, the spectrosome becomes increasingly branched and forms a fusome. In wild type ovaries, spectrosomes are

present in the two or three stem cells at the tip of the germarium (Fig. 2A). In contrast, ovaries in which follicle cells were depleted for *mago* or *tsu* showed an increase in the number of spectrosome-like structures, indicating an accumulation of germline stem cells that had failed to differentiate (Fig. 2A). This is reminiscent to the germline stem cell tumor-like phenotype found in *piwi* knockdown and *piwi* loss of function animals (Fig. 2A, [265]).

In contrast to the depletion of *mago* and *tsu*, the depletion of *btz* had no effect on either the morphology of ovaries or the differentiation of germline stem cells (Fig. 2A). This strongly suggests that Btz is unlikely to play a major role in the piRNA pathway. We could not test the implication of eiF4AIII since its depletion is cell-lethal.

B. KNOCKDOWNS OF EJC NUCLEAR CORE COMPONENTS INCREASE TRANSPOSON LEVELS IN FOLLICULAR CELLS AND THE GERMLINE

Given the similarity of *mago* and *tsu* knockdown phenotypes to the *piwi* loss of function, we next wanted to address whether the level of transposons was also impaired upon EJC knockdown.

The activity of transposons can be specific to follicular cells or germ cells or both cell types. For instance, the activity of the retrotransposons gypsy and ZAM is restricted to follicle cells and the piRNA pathway inhibits their proliferation. In contrast, Het-A and GATE are only expressed in the germ cells.

I tested the expression level of transposons by RT-qPCR after knocking down EJC subunits specifically in follicle cells. I found that *gypsy* and *ZAM* RNA were dramatically increased. These results were similar to the large increase of transposon activity observed upon *piwi* depletion (Fig. 2B). In contrast, transposon levels in *btz* knockdown were similar to the control condition.

In order to test if the core EJC plays also a role in transposon repression in the germline, I used the specific germline driver *nos*-GAL4 to deplete their products (Fig. 2C). I found that germline knockdowns of *mago* and *piwi* displayed significant de-repression of the predominant germline transposons *HeT-A* and *GATE*. As shown with the somatic transposons, their levels remained unchanged after *btz* depletion (Fig. 2C).

Taken together, these results strongly suggested a novel role for the EJC in the piRNA pathway. In addition, the fact that Btz appeared to be dispensable for this process was suggestive of a nuclear function.

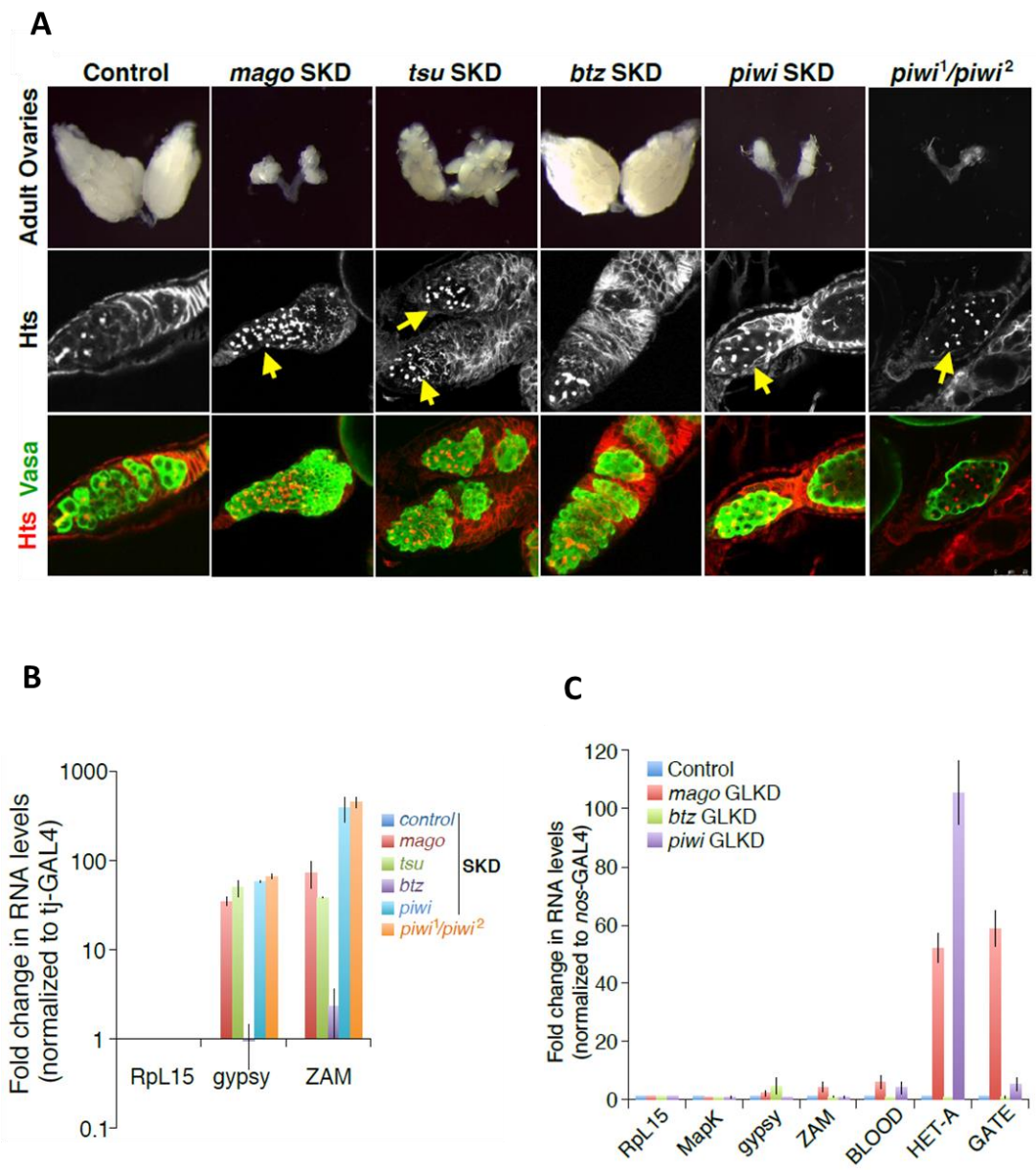

Figure 2: The pre-EJC prevents transposon mobilization in *Drosophila* ovaries
A. Adult ovaries from the indicated genotypes were dissected (upper panel) and stained (lower panels) with anti-Hts (white, red), anti-Vasa (green). Yellow arrows show extra stem cells. B. qRT-PCR of *gypsy* and *ZAM* in ovarian samples depleted for EJC subunits in follicle cells (*mago* (red), *tsu* (green) and *btz* (purple)). qRT-PCR from *piwi* mutant (orange), *piwi* somatic knock down (SKD, dark blue) and wild type (light blue) ovaries are shown as controls. C. qRT-PCR from several germline knock down ovarian extracts.

1.2. ROLE OF THE EJC ON THE LOCALIZATION AND EXPRESSION LEVEL OF ARGONAUTE PROTEINS.

Defects in the piRNA pathway often lead to a mislocalization and/or reduction in the level of PIWI-clade Argonaute proteins [196,229,244]. Indeed, PIWI proteins that are not loaded with piRNAs were shown to be less stable than PIWI-RISC.

piRNA pathway factors have been categorized based on their effects on the localization pattern of Argonaute proteins and whether they preferentially alter primary (implemented by Piwi), secondary (involving Aubergine (Aub) and Argonaute 3 (AGO3)) piRNA biogenesis or both (Olivieri et al. 2012). To confirm that the EJC regulates transposon expression by promoting the activity of the piRNA pathway, we assessed PIWI-clade protein localization in several EJC mutant contexts.

In order to directly compare the distribution of Argonaute proteins in wild type versus EJC-depleted cells I used the Flp-FRP technique (Fig. 3). This technique involves a somatic recombination after the expression of a flipase enzyme. Under mitosis, this enzyme binds to the so-called FRT sites giving rise to trans-chromosomal recombination. After tissue growth, mutant *versus* wild type clones are obtained. To distinguish the clones from each other, the wild type allele generally contains the GFP sequence, which allows the wild type clones to express GFP, whereas mutant clones are devoid of GFP expression.

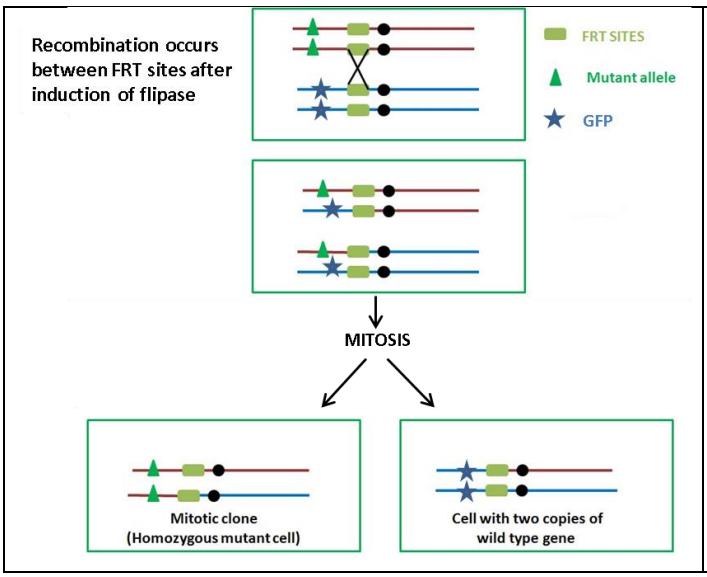

Figure 3: FLP-FRP technique for homologous recombination
Depiction of a somatic recombination occurring between two FRT sites (green squares) under flipase expression, in a cell containing two chromosomes (a red-one mutated for the gene of interest (green triangle), a blue-one containing the wild type allele and a gene coding for GFP (blue star)). After mitosis, daughter cells can be homozygous, the first contains the mutant allele, the second contains the wild type version of the gene expressing GFP.

A. AUBERGINE AND AGO3 ARE NOT LOCALIZED PROPERLY UNDER KNOCKDOWN OF EJC CORE SUBUNITS

We assessed the level of Aub and AGO3 in mutant context for EJC nuclear core components by immunostaining using specific antibodies. Aub and AGO3 are expressed exclusively in the germline in a ring structure called the nuage where the secondary piRNA biogenesis takes place. In *mago* and *tsu* mutant clones, the level of Aub did not appear significantly changed compared to wild type cells, indicating that EJC nuclear core is not required for its stability (Fig. 4A). The same result was found for AGO3 in *tsu* null mutant clones (Fig. 4B). However, in *mago* and *tsu* mutant clones, these proteins did not appeared to be restricted to the nuage, but were partially re-localized in a diffuse manner in the cytoplasm (Fig. 4A, B). This suggests either an impairment of the integrity of the nuage or a specific defect to their localization to this structure.

B. PIWI PROTEIN LEVEL IS STRONGLY REGULATED BY EJC CORE SUBUNITS

I next tested the expression level of Piwi, the third Argonaute protein, which is expressed in both the germline and follicular cells. In contrast to what I observed for AGO3 and Aub, I detected a significant decrease of Piwi levels in both *mago* and *tsu* mutant clones (Fig. 4C). Interestingly, the control of Piwi protein level by Mago and Tsu is not cell-type dependent, since the defect was observed in both germ and follicle cells (Fig. 4C).

Collectively, these results indicate that the nuclear EJC core components Mago and Tsunagi are required for normal Piwi protein levels, but have only a minor effect on AGO3 and Aub localization.

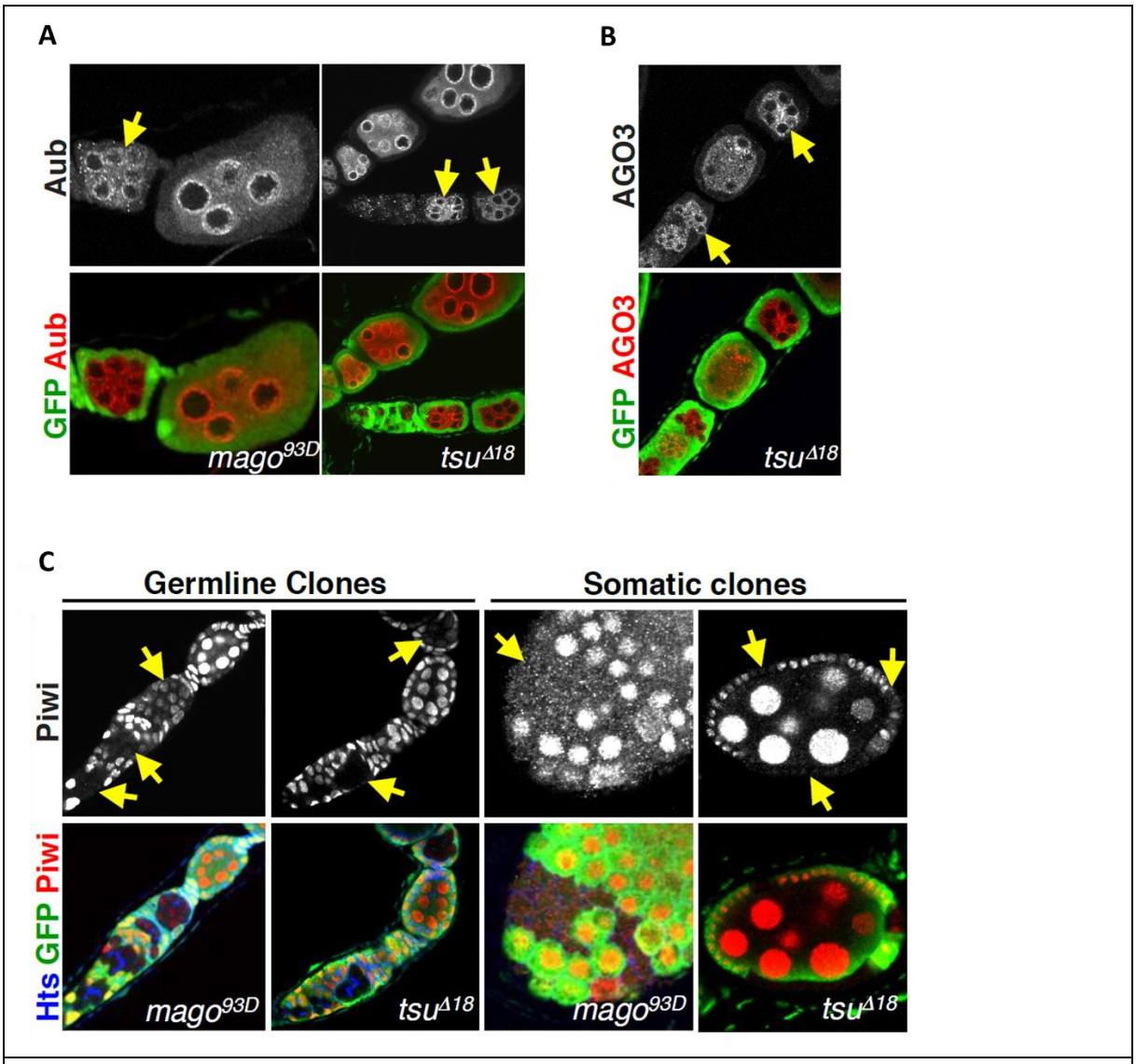

Figure 4: The pre-EJC specifically controls Piwi protein level
A-C. Ovaries containing clones homozygous for the *tsu* null allele *tsuΔ18* or for the *mago* loss of function allele (*mago93D*) are marked by the absence of GFP (green) and stained with antibodies against Aub (A), AGO3 (B), Piwi (red) and Hts (blue) (C).

1.3. THE KNOCK DOWN OF THE SPLICING SUBUNIT RNPS1 RECAPITULATES PHENOTYPES OF THE KNOCK DOWN OF EJC CORE SUBUNITS.

A. RNPS1 CONTROLS PIWI PROTEIN LEVELS

Our results above indicate that the EJC likely regulates the piRNA pathway through a nuclear mechanism. Two nuclear functions have previously been assigned to the EJC:

one in splicing [84,138,139] and one in the control of mRNA export [107,108,150,267]. In order to investigate which of these nuclear functions was involved in the piRNA pathway, I knocked down a representative set of EJC accessory factors and examined their effects on Piwi levels.

Interestingly, I found that depletion of the splicing subunit *RnpS1* reduced Piwi levels in both follicle and germ cells (Fig. 5A,B). This effect was further confirmed by western blot analysis. I observed a significant decrease of Piwi under *RnpS1* depletion indicating that the protein was not simply delocalized but reduced (Fig. 5C).

RnpS1 is a member of the apoptosis- and splicing- associated protein (ASAP) complex [268]. ASAP complex promotes a link between apoptosis and splicing by association of the apoptotic and splicing factor Acinus (Acn) with the splicing factors SAP18 and RnpS1. The ASAP complex differs from the PSAP complex [269] by the replacement of Acn with Pinin. In order to investigate which RnpS1 complex was involved in the control of piwi I knocked down each representative factor. I found that only Acn, in addition to RnpS1, was required for correct Piwi levels (Fig. 5A). This indicates that only a subset of ASAP is required for this function.

In contrast to the depletion of the splicing factors RnpS1 and Acn, knocking down the export factor Aly/REF had no impact on Piwi levels in follicle cells (Fig. 5A).

Altogether, These results strongly suggest that the role of the nuclear core EJC in the regulation of the piRNA pathway involved its splicing function and is mediated via the factors RnpS1 and Acn.

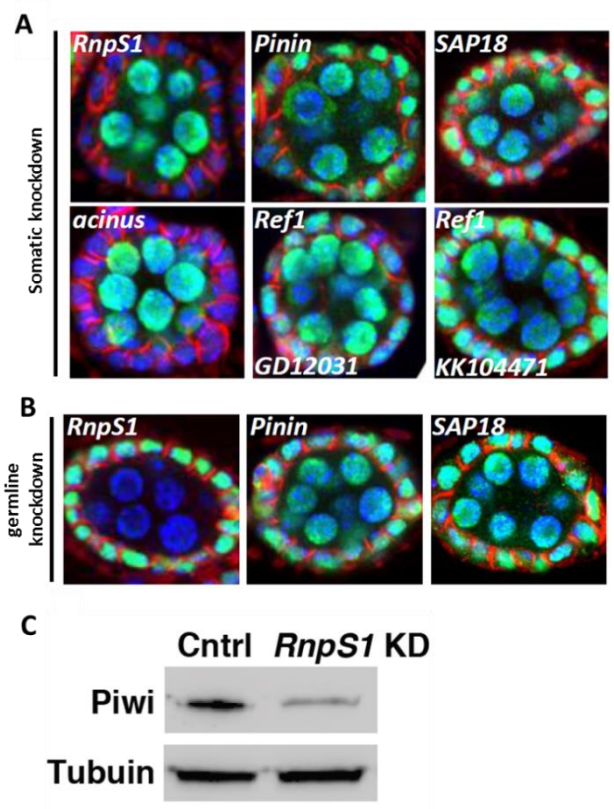

Figure 5: The EJC subunit RnpS1 controls Piwi protein levels
A-B. Egg chambers stained for Piwi (green), Hts (red) and DNA (blue) under somatic knockdown (A) or germline knockdown (B) of indicated factors. B. Western blot analysis of control and *RnpS1* knockdown ovarian extracts using anti-Piwi and anti-Tubulin antibodies.

B. RNPS1 KD LEADS TO TRANSPOSON UP-REGULATION IN BOTH GERMLINE AND FOLLICLE CELLS

To further characterize the role of RnpS1 in the piRNA pathway, I investigated the effect of its knockdown on transposon levels. To this end, I used several complementary approaches. First, as I did previously for EJC core subunits, I drove the expression of dsRNA targeting *RnpS1* transcripts in both germline or follicle cells, and tested transposon activity by RT-qPCR. I observed that upon depletion of *RnpS1* in the follicle cells, the somatic transposons Gypsy and Blood were strongly de-repressed (Fig. 6A). Up-regulation was also observed upon RnpS1 knockdown in the germline for the transposons Het-A and GATE (Fig. 6B).

In parallel, I took a genome-wide approach to obtain a comprehensive view of the impact of *RnpS1* knockdown on transposon level. This was done in collaboration with the Lehmann Laboratory at the Skirball Institute for Biomolecular Medicine in New

York. Total RNAs derived from either control or *RnpS1* knockdown ovaries (in both germline and follicle cells) were depleted of ribosomal RNAs, cloned and sequenced. Colin Malone from the Lehmann lab and Nastasja Kreim from IMB genomic core facilities have performed the computational analysis. They confirmed that many transposons were significantly up-regulated upon *RnpS1* depletion. All classes of elements, which include those predominantly expressed in the soma (like Idefix and Gypsy) or germ cells (such as HeTA or GATE), showed increased expression (Fig. 6C). Altogether, these results indicate that RnpS1 is required to silence transposons and acts similarly to the nuclear core EJC and piRNA pathway components [270].

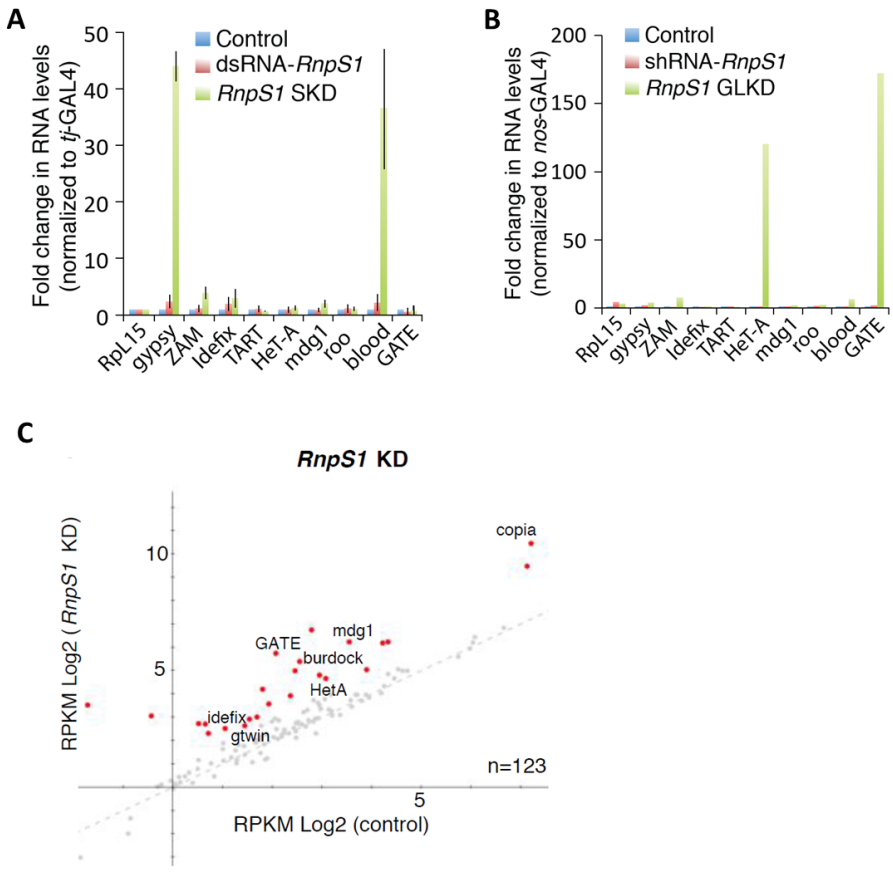

Figure 6: RnpS1 prevents transposon mobilization in both germline and follicular cells
A. Quantitative RT-PCR measuring transposon expression in control (blue), non-expressed dsRnpS1 strain (red) and RnpS1 somatic knockdown (green) ovaries. B. Quantitative RT-PCR measuring transposon expression in control (blue), non-expressed dsRnpS1 strain (red) and RnpS1 germline knockdown (green) ovaries. C. plot showing the log2 average expression values per transposon for *RnpS1* knock down versus control. Highlighted transposons (in red) have an average fold change of at least 2 towards control.

1.4. THE EJC NUCLEAR CORE AND RNPS1 CONTROL PIWI RNA LEVELS

A failure in the loading of piRNAs into Piwi triggers a defect in Piwi nuclear accumulation, which subsequently leads to its degradation [229,244]. Therefore, the decrease in Piwi protein level observed upon depletions of RnpS1 and of the core EJC could be an indirect consequence of reduced piRNA levels.

Alternatively, since the EJC and RnpS1 are involved in mRNA metabolism, a direct effect of the complex on the processing of *piwi* RNAs could also be plausible.

To discriminate between these two possibilities I examined *piwi* mRNA levels in *RnpS1* knockdown *versus* wild type conditions. To this end, I performed RT-qPCR experiments using ovarian RNA extracts. I measured RNA level using several amplicons spanning the *piwi* transcript. *RnpS1* depleted ovaries led to a 40 to 80% decrease in *piwi* transcript levels compared to the control condition (Fig. 7A). The strongest effect was observed with the amplicon spanning the exons four and five. Furthermore, I found that the *mago* knockdown led to a similar effect on *piwi* level (Fig. 7B).

All together, these results indicate that the nuclear EJC together with its RnpS1 splicing subunit have an influence on *Piwi* RNA level. This suggests a direct effect of these components on *piwi* RNA processing rather than an effect on Piwi protein stability.

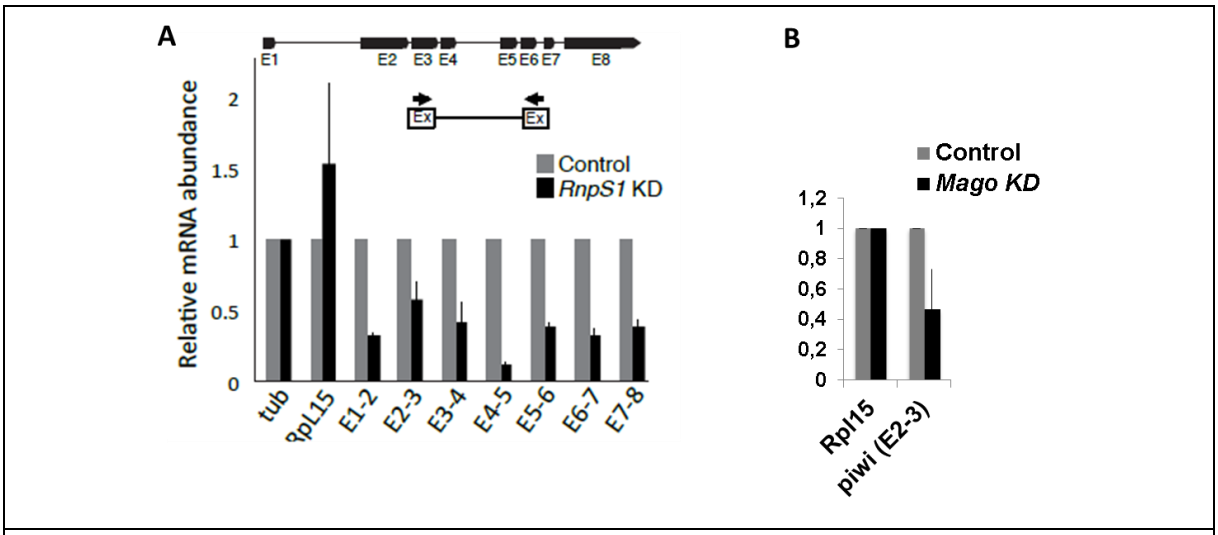

Figure 7: the pre-EJC and its RnpS1 co-factor control piwi RNA level
A. Quantitative RT-PCR measuring *piwi* RNA level at each exon (E) junction of the transcript, normalized with *tubulin* (tub) and *RpL15* transcript levels in control (gray) and *RnpS1* knock down (black) ovaries. B. Quantitative RT-PCR measuring *piwi* RNA level at the exon 2 to exon 3 junction (E2-3) of the transcript, normalized with *RpL15* transcript in control (gray) and *mago* knock down (black) ovaries.

1.5. PIWI IS THE MAIN TARGET OF RNPS1 IN THE piRNA PATHWAY

In order to determine whether Piwi was the only piRNA pathway component affected upon EJC depletion I tested the impact of RnpS1 KD on the RNA level of twenty validated components of the piRNA pathway. I depleted RnpS1 both in follicle cells and in the germline, by combining *tj*-GAL4 and *nos*-GAL4 drivers. Among the twenty tested transcripts, only *piwi* RNA level and to a lesser extent, *Ago3,* displayed a significant decrease (Fig. 8A). The RNA levels of other piRNA components remained unchanged. As *Ago3* is a large gene expressed from heterochromatic location the slight decrease of its levels may relate to the previously described function of the EJC in the splicing of heterochromatin transcripts [138,139].

Piwi has been shown to repress only a subset of transposons, the others being repressed by the two other Argonaute proteins. Therefore, specific de-repression of Piwi-regulated transposons provides a clear signature of its effect. We wondered whether a similar class of transposons would be de-repressed upon RnpS1 knockdown. If the same signature were observed for RnpS1 depletion, this would suggest that Piwi is the main target of RnpS1 in the piRNA pathway. We normalized the transposon levels from *RnpS1* knockdown ovaries (Fig. 8B) with those of *piwi* mutant ovaries (Fig. 8C,D). Interestingly, we found that all up-regulated transposons under RnpS1 knockdown were also found in *piwi*-depleted ovaries. We did not find any transposons that were not also regulated by Piwi (Fig. 8D). These results indicate that *piwi* reduction is sufficient to explain the effects of RnpS1 on the piRNA pathway, strongly suggesting that the function of RnpS1 is predominantly mediated via its effect on Piwi.

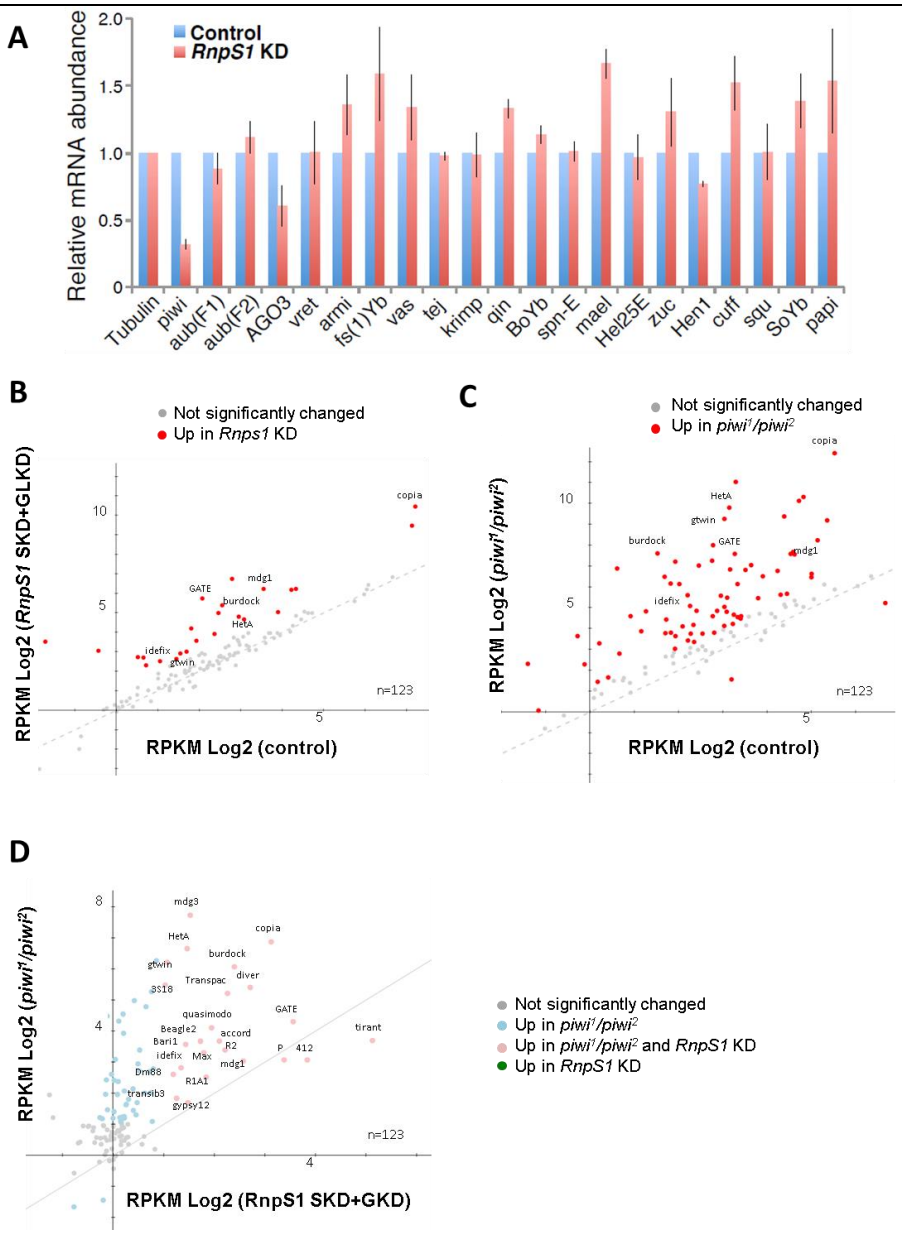

Figure 8: *piwi* RNA is the main target of RnpS1 in the piRNA pathway

A. Quantitative RT-PCR from control (blue) and *RnpS1* knock down (red) ovaries assessing the expression of 20 piRNA pathway transcripts, normalized to *tubulin* expression. B-C. 2 fold or greater, statistically significant transposon de-repression in *RnpS1* knock down (B) and *piwi* mutant (C) ovaries compared to control are highlighted in red. D. Transposon de-repression in *RnpS1* knock down and in *piwi* mutant ovaries are compared to each other (non-significant (gray), up-regulation in *piwi* mutant only (blue), up-regulation in both *piwi* mutant and *RnpS1* knock down (pink) and up-regulation in *RnpS1* knock down only (green)). No transposon showed more than 2 fold enrichment in *RnpS1* knock down only (green).

2. THE EJC IS REQUIRED FOR THE SPLICING OF NON CANONICAL INTRONS

2.1. THE EJC AND ITS SPLICING SUBUNIT RNPS1 FACILITATE THE SPLICING OF THE PIWI TRANSCRIPT

A. THE EJC CONTROLS PIWI AT A POST-TRANSCRIPTIONAL LEVEL

The previous results demonstrated a role for the EJC in the repression of transposition in the *Drosophila* germline. Furthermore, we found that the level of the *piwi* transcript, which encodes a critical component of the piRNA pathway, was strongly reduced upon EJC depletion. We next wanted to address how the EJC controls *piwi* RNA level. Since RnpS1 was also involved, the most logical explanation would be a defect in pre-mRNA splicing. However, we wanted to rule out a potential indirect effect at a transcriptional level. To this end, we took advantage of an enhancer trap P-element inserted in the first intron of *piwi* that expresses LacZ under *piwi* regulatory sequences [266]. If *piwi* transcription was impaired, the *lacZ* reporter should not be expressed. Using the Flp-FRT technique, we compared LacZ expression in *tsu* mutant cells with their wild type sister clones. As previously demonstrated, Piwi protein was strongly decreased in *tsu* mutant clones. However, *LacZ* levels were not reduced, strongly suggesting that EJC loss does not interfere with *piwi* transcription (Fig. 9A). Intriguingly, the LacZ level seemed to be even increased, suggesting the existence of a negative feedback loop that controls the overall level of Piwi.

B. RNPS1 KNOCKDOWN LEADS TO DEFECTS IN THE SPLICING OF PIWI INTRON FOUR

Since *piwi* was not affected transcriptionally, we then examined whether its splicing was defective. To this end, we took advantage of our RNA-Seq datasets from *RnpS1* knockdown ovaries that we compared to transcriptomic data from control ovaries.

When we examined the *piwi* locus, we observed a general decrease in the number of reads mapping to the exons, consistent with the reduced RNA level previously detected by qPCR. In contrast, the number of reads mapping to the intron four was significantly increased, indicating an accumulation and retention of this intron in *piwi* transcripts in the absence of RnpS1 (Fig. 9B).

To confirm these observations, I performed an RT-PCR experiment using wild type and RnpS1-depleted ovarian RNAs with primers spanning the entire transcript (Fig. 9C). Using this approach, I detected an additional larger band on agarose gel compared to control. To test whether this increase in size was due to the retention of intron four I used primers flanking this intron. I found a dramatic accumulation of this intron. I quantified this defect by qPCR and found a 10-fold increase compared to control condition (Fig. 9C). Interestingly, I also observed some retention of this intron in the wild type condition, indicating that this intron is normally inefficiently spliced. Besides, I also detected an intermediate band that corresponds to a partial splicing of the intron four that accumulates upon *RnpS1* knockdown (Fig. 9D). This was due to an increase usage of a cryptic splice site within this intron.

Finally, I also observed a weak defect in the splicing of the first intron. Activation of a cryptic splice site occurs in wild type ovaries at a low level (Fig. 9B) and occurs more frequently in RnpS1 knockdown conditions (Fig. 9B).

All together, these data showed that the intron four of *piwi* (and to a lesser extent the intron one) are sub-optimally spliced under normal conditions, and that RnpS1 and the EJC facilitate intron removal in order to maintain faithful Piwi protein levels.

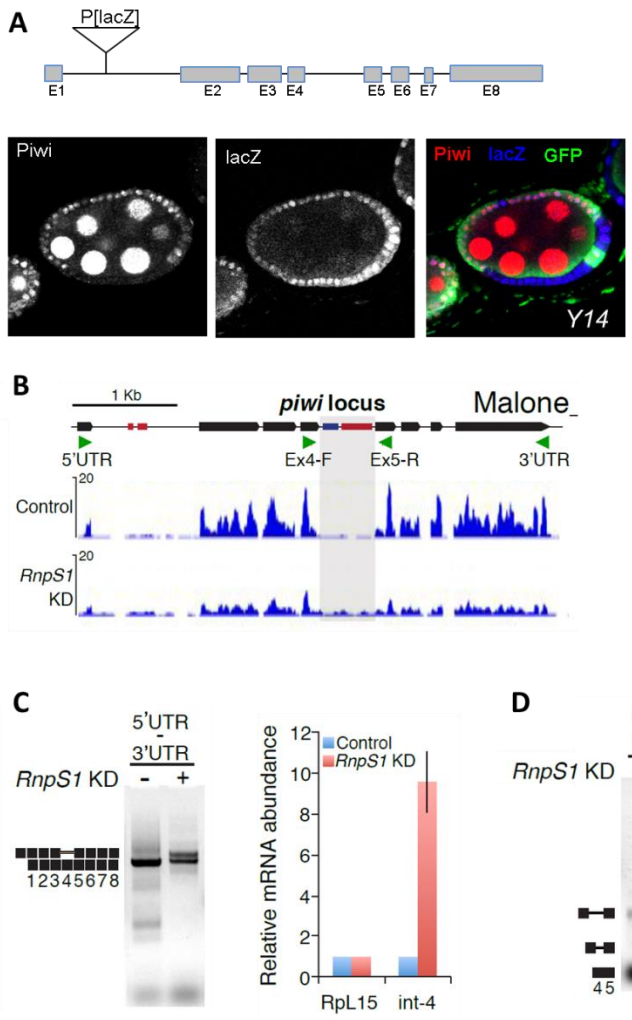

Figure 9: *piwi* is the main target of RnpS1 in the piRNA pathway
A. Assessment of *piwi* transcription using a P element containing *LacZ* reporter. Top: *piwi* locus with the location of *P(LaZ)* reporter within the first intron. Bottom: immunostaining of egg chambers with *tsunagi* mutant somatic clones marked by a lack of GFP (green), Piwi is stained in red. B. RNA transcript mapping over the *piwi* locus from control and RnpS1 knock down ovaries. Intronic transposons Cr1A (blue box) and DINE (red box) are shown, as well as read accumulation in intron 4 (gray box). C. Left: RT-PCR of full-length *piwi* (primers shown by green arrowhead in B) with and without RnpS1 knock down. Right: Quantification by quantitative RT-PCR of *piwi* fourth intron retention. D. RT-PCR of *piwi* intron 4 (primers shown by green arrowhead in B) with and without RnpS1 knock down.

C. A PIWI cDNA PARTIALLY RESCUES THE TRANSPOSON MOBILIZATION DEFECT OF RNPS1 KNOCK DOWN OVARIES

Our work so far showed that RnpS1 plays an important role in controlling *piwi* splicing. Therefore, we asked whether we could restore a low level of transposons by expressing a form of *piwi* that lacks introns, and whose expression should be independent of the EJC. To this end, I used an integrated, inducible Piwi cDNA tagged with GFP, (UAS-GFP-Piwi, Fig. 10A) as well as a Piwi-GFP that lacks its nuclear localization signal as negative control (UAS-GFP-Piwi[ΔN], Fig. 10A) [242]. These stocks were generously provided by J. Brennecke (IMBA, Vienna).

I first validated these constructs by expressing them in follicle cells using the *tj*-GAL4 driver, and examined their sub-cellular localization by immunostaining. As expected,

UAS-GFP-Piwi produced a nuclear protein, whereas Piwi[ΔN] remained cytoplasmic (Fig. 10B). I repeated the experiment in *armitage* (*armi*) knock down context. Armi is a piRNA component that is required for Piwi nuclear localization [229,230,271]. In this condition, I found that GFP-Piwi now failed to enter the nucleus (Fig. 10B). In contrast, expressing these constructs in an *RnpS1* knockdown condition does not impair the nuclear localization of Piwi derived from UAS-GFP-Piwi (Fig. 10B). This result clearly demonstrates that the EJC is not required for Piwi localization.

To examine whether expression of GFP-Piwi was sufficient to restore transposon level in the absence of RnpS1, I examined the steady-state levels of several transposable elements. After expression of the cDNA in follicle cells, I detected a somatic restoration of *piwi* transcript levels, indicating that its expression was independent of RnpS1 (Fig. 10C). In addition, I found a partial suppression of ZAM level, and to a lesser extent, Blood, whereas rescue did not occur with the non-functional GFP-Piwi[ΔN] (Fig. 10C). Taken together, this experiment demonstrates that Piwi is a critical effector of the EJC in preventing transposon mobilization. I noticed that the cDNA constructs were expressed in a variegated pattern in follicle cells (Fig. 10B), potentially preventing a more robust rescue of transposon silencing.

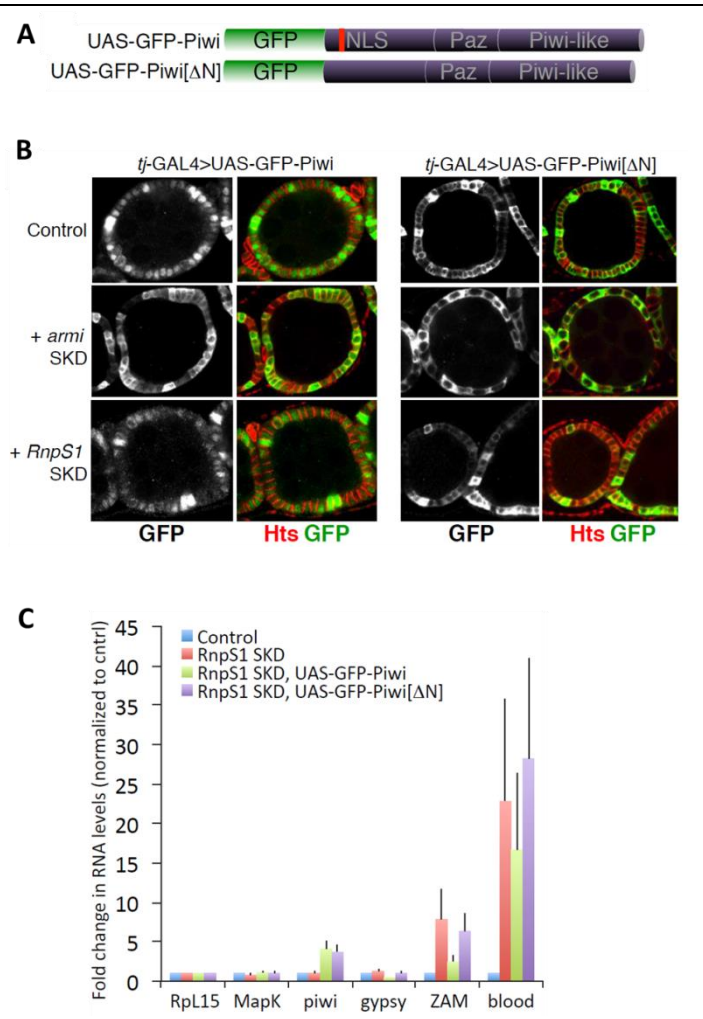

Figure 10: Expression of Piwi cDNA partially restores transposon up-regulation caused by *RnpS1* knock down.
A. Schematic of UAS promoter driven GFP-Piwi (left) and GFP-Piwi with the nuclear localization signal removed to serve as a negative control (right). B. Somatic expression of Piwi constructs in control and *armi* or *RnpS1* knockdown egg chambers stained for anti-Hts (red) and anti-GFP (green). C. Quantitative RT-PCR of somatic transposons in *RnpS1* SKD ovaries with and without co-expression of Piwi constructs.

D. REMOVAL OF INTRON FOUR REQUIRED THE FLANKING INTRONS

So far, our results demonstrated a role for the EJC and its co-factor RnpS1 in *piwi* splicing. We next wanted to gain more insight into the molecular mechanism underlying this regulation. To this end, we first attempted to recapitulate our observations in a cell culture based system. I designed minigene constructs containing regions of the *piwi* locus encompassing the intron four and expressed them in *Drosophila* embryonic Schneider cells (S2R+ cells). Importantly, this cell line has no detectable endogenous *piwi* expression, allowing us to assign Piwi level directly to the expression of the transfected constructs. For each minigene construct, we tested the retention level of intron four in both wild type and *RnpS1* knock down conditions by RT-PCR using specific primers flanking the intron.

First, I transfected a fragment of *piwi* containing exon 3 to exon 6 (Fig. 11A,B, Construct 1). Interestingly, its expression in S2R+ cells recapitulated the RnpS1 dependence of intron 4 removal observed in the gonads. While intron 4 of the construct 1 was almost entirely spliced in control condition, its retention was increased by 20% upon *RnpS1* depletion (Fig. 11B', Construct 1). This result indicates that the regulation of *piwi* splicing by RnpS1 is not specific to the gonads and can occur even in cells where *piwi* is not normally expressed.

I then transfected a construct expressing a fragment encompassing exon 4 to exon 5 (Fig. 11A,B, Construct 2). Interestingly, intron four excision was strongly impaired not only upon *RnpS1* knock down but also in the wild type condition. This experiment strongly suggested that flanking sequences were absolutely required for splicing of intron four (Fig 11B', Construct 2).

We next assessed the relative importance of the two flanking introns. To this end, we designed some constructs containing either the upstream or the downstream intron. We found that removing intron 5 had a strong effect on intron 4 excision (50% retention), which was further exacerbated by depletion of RnpS1 (80% retention, Fig. 11B,B', construct 3). Conversely, removing only intron 3 led to a minor effect on intron 4 splicing (20% retention) but this defect was strongly aggravated by RnpS1 depletion (75% retention, Fig. 11B,B', construct 4). Altogether, these results showed that both flanking introns are required for correct splicing of intron 4, with intron 5 having a greater impact. Furthermore, both flanking introns exert their effect in an RnpS1-dependent manner.

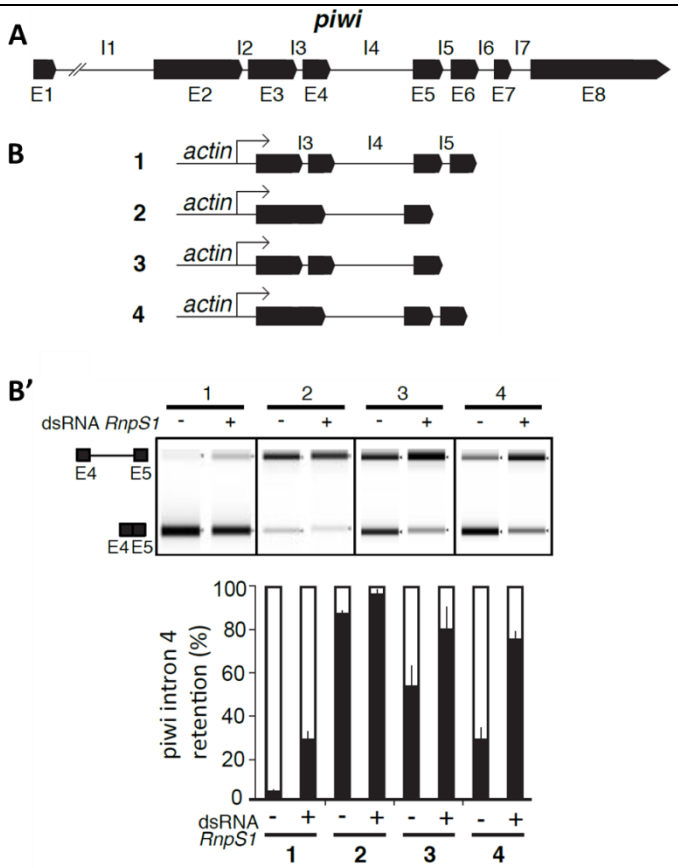

Figure 11: *piwi* intron 3 and 5 influence intron 4 removal via RnpS1. A. Schematic of the *piwi* genomic locus, with introns (I) and exons (E) marked accordingly. B. Various minigene constructs (1-4) containing the *piwi* fourth intron with or without flanking introns. B'. Top: Capillary gels of RT-PCR with primers flanking intron 4 using extracts from control or *RnpS1* depleted *Drosophila* S2R+ cells. Bottom: Graphs showing the ratio of spliced (white) versus unspliced (black) intron fourth.

E. RNPS1-DEPENDENT INTRON REMOVAL REQUIRES THE SPLICING OF FLANKING INTRONS

In order to get more insights into the underlying mechanism, we asked whether the positive effect of the flanking introns was due to their cis-regulatory sequences or to their splicing *per se*. To test whether their splicing was important, we generated point mutation in the 5' splice site of intron 3 and examined the splicing of intron 4 (Fig. 12 A,B). Strikingly, preventing splicing of intron 3 completely inhibited the splicing of the downstream intron (Fig. 12 B,B' construct $3^{I3\text{-}5mut}$). A similar result was observed under the mutation of the 3' splice site (Fig. 12B,B' construct $3^{I3\text{-}3mut}$). Moreover, mutating the 5' splice site of the intron 5 gave similar results (Fig. 12A,B,B' construct $4^{I5\text{-}5mut}$). Together, these data demonstrate that splicing of flanking introns is absolutely required to enable RnpS1 to promote intron 4 splicing.

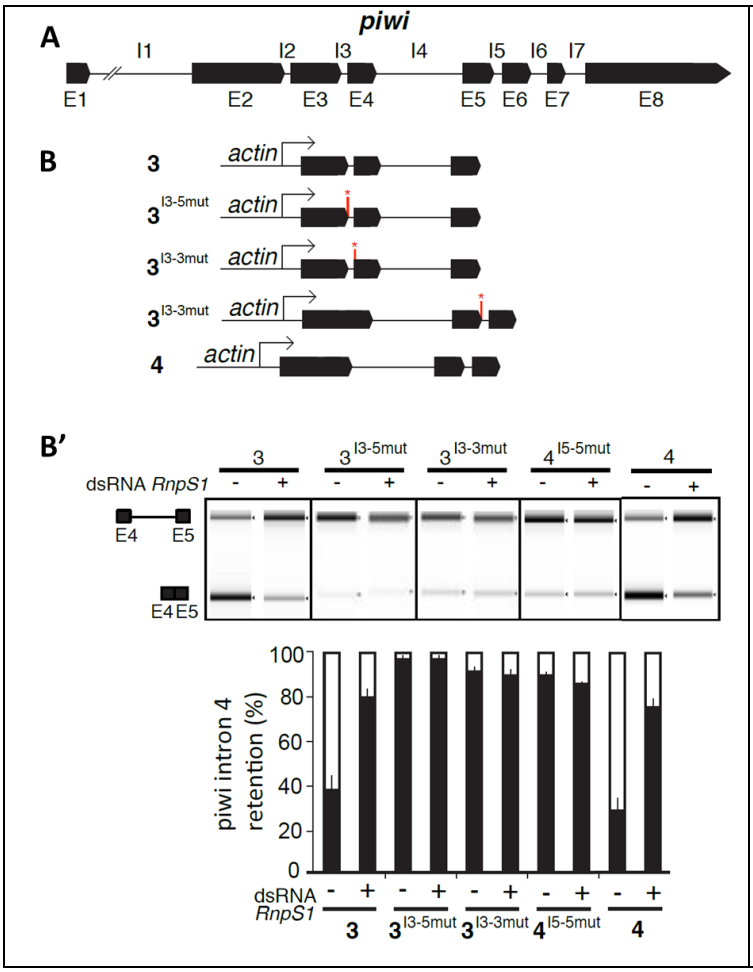

Figure 12: RnpS1-dependent intron removal requires the splicing of flanking introns
A. Schematic of the *piwi* genomic locus, with introns (I) and exons (E) marked accordingly. B. Various minigene constructs (3 or 4) containing the *piwi* fourth intron with or without splice site disruption (red asterisk). B'. Top: Capillary gels of RT-PCR with primers flanking intron 4 using extracts from control or *RnpS1* depleted *Drosophila* S2R+ cells. Bottom: Graphs showing the ratio of spliced (white) versus unspliced (black) intron fourth.

F. THE PRESENCE OF THE TRANSPOSONS DOES NOT CONFER LOCAL HETEROCHROMATIN NUCLEATION IN THE PIWI LOCUS

Our group has previously showed that many genes having large introns and expressed from heterochromatic location required the EJC for their splicing [138,139]. However, *piwi* is located in euchromatin and contains only small introns. We wondered whether a common mechanism in the splicing of these transcripts would exist. Heterochromatic introns are usually enriched for repeated elements and degenerated transposons. These sequences have the ability to nucleate heterochromatin, which can potentially interfere with the splicing process. Interestingly, we observed that the intron 4 of *piwi* also contain transposons (Fig. 13A). Therefore, we asked whether the presence of transposon remnants within the *piwi* locus trigger local heterochromatin. This nucleation would provide a chromatin environment similar to the chromatin structure of the heterochromatic loci and could be an explanation for the EJC dependence. To this

end, we tested the presence of the histone mark H3K9me3, a specific signature of heterochromatin, at the *piwi* locus. We performed chromatin immunoprecipitation experiments followed by qPCR using primers spanning the intron four. While a strong enrichment in H3K9me3 was observed at the heterochromatic *MAPK* locus, no detectable enrichment was found at the *piwi* locus (Fig. 13B). This result indicates that the presence of degenerated transposons in *piwi* intron does not nucleate heterochromatin. Therefore, we concluded that the dependency of *piwi* splicing to the EJC was not conferred by the heterochromatin state of the locus.

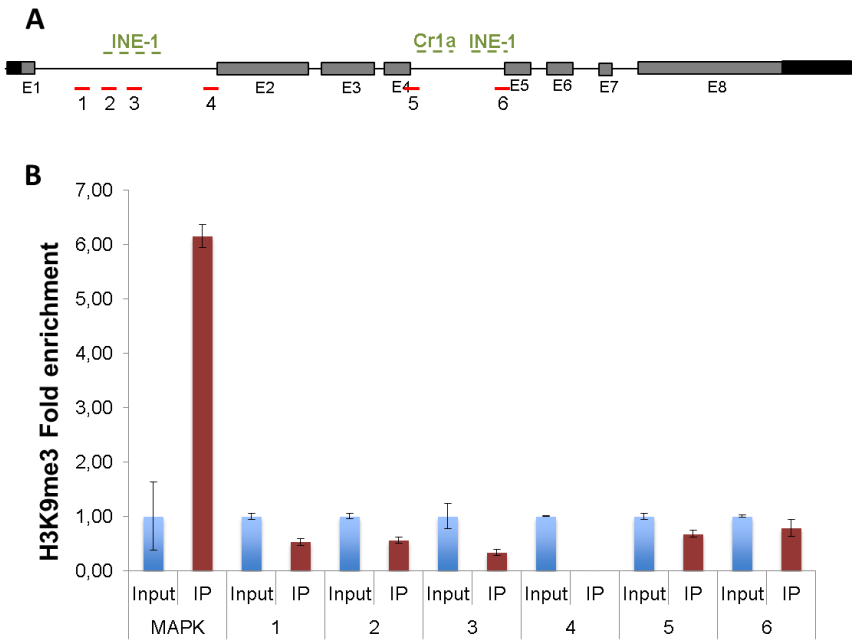

Figure 13:Presence of transposons in *piwi* introns does not confer local heterochromatin nucleation
A. Schematic of the *piwi* genomic locus, with introns and exons. Transposon remnants are indicated in green, probed regions for ChIP assay are indicated in red. B. quantification of the fold enrichment of H3K9me3 at the *piwi* locus. One region in *MAPK* gene serves as a positive control. Despite the presence of transposable elements, the *piwi* locus is devoid of heterochromatin marks.

G. THE STRENGTH OF SPLICE SITES DOES NOT CREATE THE EJC DEPENDENCE

To investigate which cis-sequence elements confer the EJC dependence, I engineered additional versions of the *piwi* minigene (Fig. 11). I first asked whether the strength of splice site was involved. Using the BDGP splice site predictor software we found that the 5' splice site (5'ss) of intron 4 has a low score (0.31) compared to the 5'ss of the other *piwi* introns (Fig. 14A). Thus, we replaced the 5'ss of intron 4 with the 5'ss of the higher scoring intron 2 (Fig. 14 A, B Construct $3^{I45'\text{-}I25'}$). While this substitution improved the splicing efficiency of intron 4 by 30% in control condition, it does not relieve the requirement for RnpS1. In fact, under *RnpS1* knock down, an increase of 20% of intron 4 retention was still observed (Fig. 14 B, Construct $3^{I45'\text{-}I25'}$). This result ruled out the splice site strength as a main determinant of RnpS1 dependence. However, the fact that a better splice site improves splicing may partly account for the inefficient removal observed in control condition (*in vitro* Fig. 11 constructs 1, 3 and *in vivo* Fig.9).

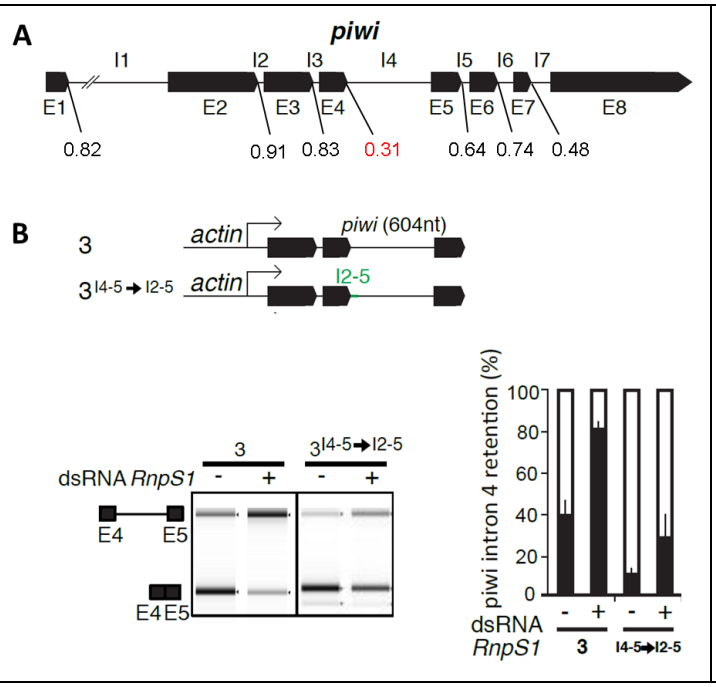

Figure 14: Strength of the splice sites does not create dependence to EJC
A. Schematic of the *piwi* genomic locus, with introns (I) and exons (E) marked accordingly. Strength of the 5' splice sites is indicated. B. Diagrams of the minigene constructs replacing *piwi* fourth intron 5' splice site with that the 5'ss of the *piwi* second intron (green). Capillary gels and graphs showing the ratio of spliced (white) versus unspliced (black) fourth intron.

H. A WEAK POLY-PYRIMIDINE TRACT AT THE 3'END OF THE RETAINED INTRON CONFERS THE EJC DEPENDENCY

Since neither heterochromatin nor splice site strength was responsible for the RnpS1 dependence I made additional minigene constructs to further dissect the mechanim. I replaced the entire intron 4, except the 5'ss, with the sequence of an intron of similar size from the *tubulin* transcript. This substitution completely rescued the splicing defect associated with the lack of RnpS1 (Fig. 15 A construct 3I45'-Itub). This loss of RnpS1 dependence was observed when the endogenous 5'ss of intron 4 was still present, confirming that the 5'ss was not the element conferring the dependence. Furthermore, this experiment demonstrated that the intron 4 was retained due to specific elements that lie within the intron itself rather than to binding of splicing inhibitors at flanking exons.

In order to identify the minimal sequence that confers RnpS1 dependence, I re-introduced portions of the *piwi* intron of different sizes into this construct. (Fig. 15B,C respectively). Interestingly, introducing only the last 100 nucleotides of *piwi* intron 4 was sufficient to alter splicing and reestablish RnpS1 dependence (Fig. 15B construct P3'-100). We noticed that this sequence was highly enriched in adenosine and thymidine in its last 30 nucleotides. This region normally contains the poly-pyrimidine tract that is enriched for thymidine. Therefore, an excess of adenosines might weaken this element. To test this hypothesis, I specifically replaced the *piwi* A-rich sequence with the last 30 nucleotides of the tubulin intron, which displays a canonical poly-pyrimidine tract (Fig. 15C construct P3'+571). This substitution was able to rescue the splicing defect and to alleviate the RnpS1 dependence.

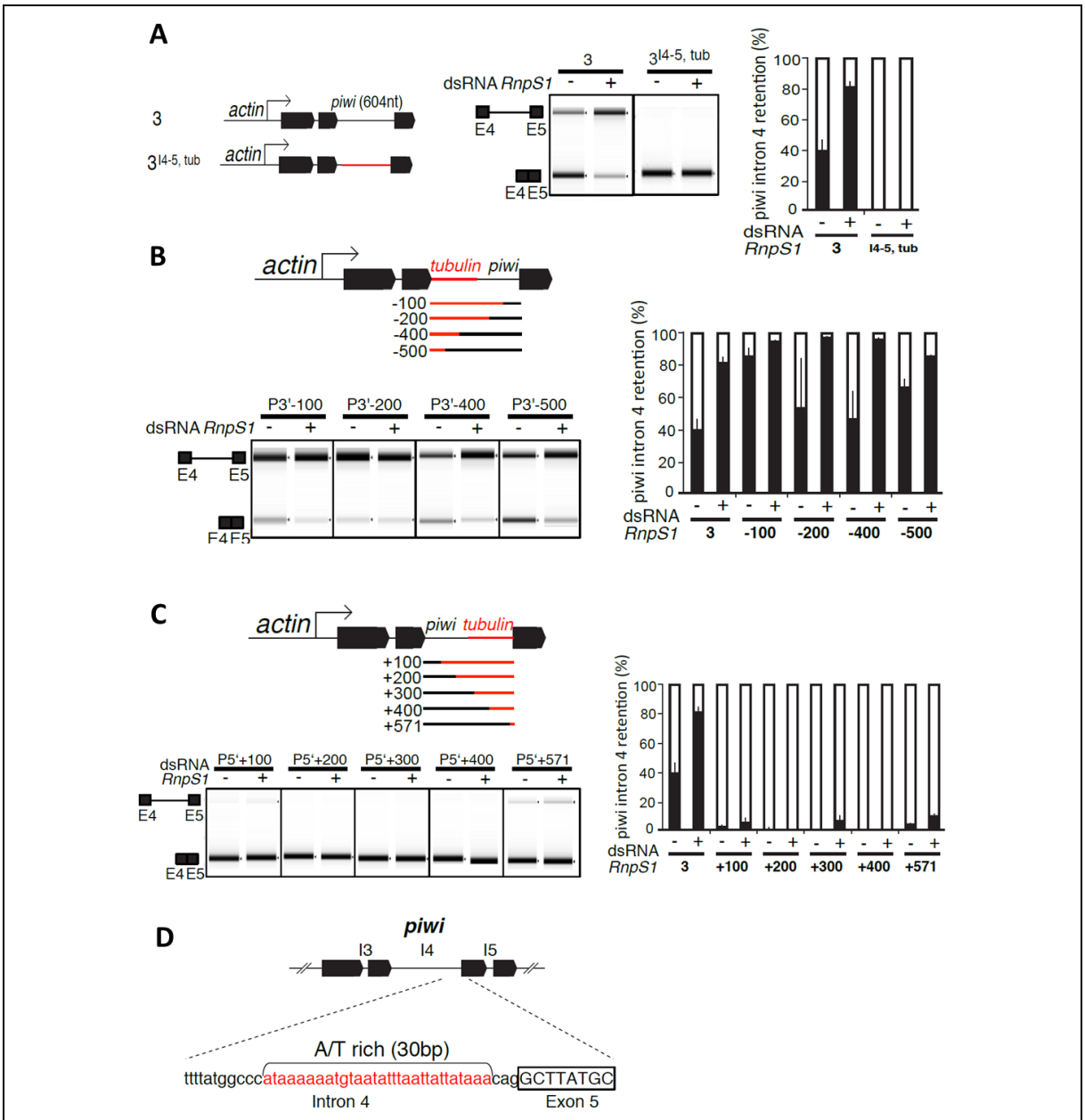

Figure 15: A weak poly-pyrimidine tract at the end of the intron 4 confers the EJC dependence. A-C. Diagrams of the minigene constructs replacing *piwi* fourth intron (A) or portions of the fourth intron (B and C) with the entire (A) or parts (B and C) of *tubulin* intron (Red). Capillary gels and graphs showing the ratio of spliced (white) versus unspliced (black) fourth intron. D. Portrayal of the A/T-rich region (Red) adjacent to the *piwi* fourth intron splice acceptor.

To test whether the 30 A-rich nucleotides of *piwi* intron would be also sufficient to create the RnpS1 dependence in a heterologous intron I engineered a minigene construct in which I inserted this minimal sequence into one intron of the *ftz* transcription factor 1 gene (ftz-f1) (Fig. 16A). I selected the intron 5 because it was similar in size compared to intron 4 of *piwi* and does not require RnpS1 for its splicing (Fig. 16A , Construct ftz-f1). Strikingly, I found that the replacement of its last 30

nucleotides with the A-rich sequence of *piwi* intron led to a retention of 40%, which is increased by 25% upon RnpS1 knock down (Fig. 16A, Construct ftz-f1^{piwiPPT}). This result demonstrated that the poly-pyrimidine tract of the *piwi* intron 4 was sufficient to alter splicing and generate the RnpS1 dependence.

Taken together, our results suggest a model in which the EJC facilitates the splicing of *piwi* intron four after its initial deposition to adjacent exon junctions. The prior splicing of flanking introns would bring the EJC and its splicing subunit in proximity to intron four, that will then facilitates the recognition of this intron by the spliceosome (Fig. 16B). The exact mechanism by which RnpS1 facilitates spliceosome function remains to be determined.

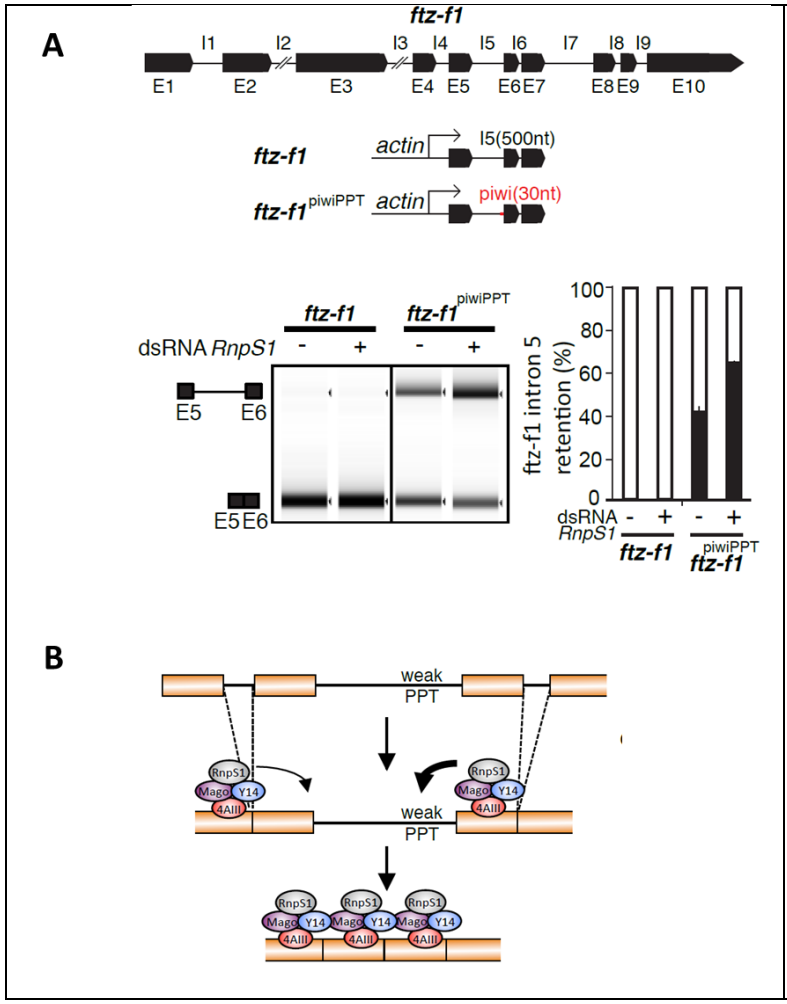

Figure 16: The weak poly-pyrimidine tract at the end of the intron 4 is sufficient to confer the EJC dependence
A. Minigene construct diagrams, capillary gels and quantification of splicing efficiency of the *ftz-f1* fifth intron locus. Both the wild type intron 5 and the intron 5 containing the *piwi* poly-pyrimidine tract (red) are depicted. B. Model showing the pre-EJC and RnpS1 acting at flanking exon junctions to rescue inefficient splicing of the *piwi* fourth intron.

2.2. THE EJC CONTROLS THE REMOVAL OF INTRONS IN OTHER EUCHROMATIC TRANSCRIPTS

We wondered whether the splicing defect observed upon *RnpS1* knockdown was unique to *piwi* or more general. To this end, we examined our RNA sequencing datasets from *RnpS1*-depleted ovaries in search of transcripts that show signs of intron retention. Using a very stringent cutoff (fdr<0.05), we found that 35 transcripts showed a significant increase in intronic-derived reads, with some transcripts having multiple introns retained. We validated these intron retention events by RT-qPCR. Out of 5 tested events, 4 were positive (Fig. 17).

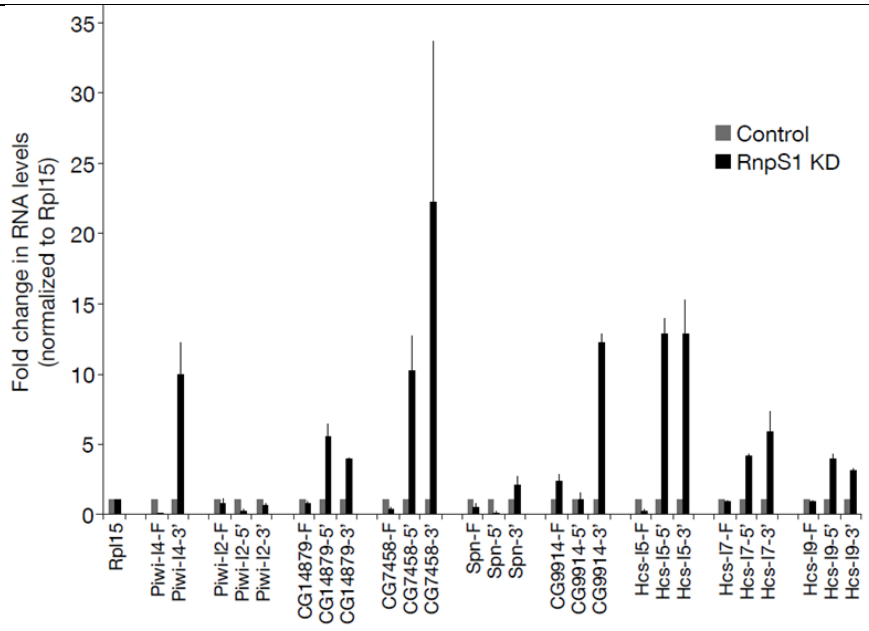

Figure 17: *RnpS1* depletion leads to numerous intron-retention events
Quantitative RT-PCR of control and *RnpS1* knock down ovaries using primers spanning introns (labeled I#). Identified retained introns were found using DEX-Seq (Table 1 in the annexes) or JuncBASE (Brooks et al. 2011) (data not shown)

Next, we searched for common features among the retained introns. We found that out of the 35 retained introns, 9 were enriched in adenosines in their last 30 nucleotides suggesting that weak poly-pyrimidine tracts might be a determinant of RnpS1 dependence in other introns. However, other features that weaken the strength of introns are also likely involved in this dependency.

Taken together, our results suggest that RnpS1 and the EJC play a general role in facilitating the removal of introns that are difficult to splice. These data expand the catalog of regulatory functions of the EJC. In this case, a sub-specialization that is essential for the efficient maturation of diverse mRNAs in the cell.

2.3. PIWI EXON 3 NEGATIVELY INFLUENCE THE SPLICING OF INTRON 4

By examining the splicing efficiency of intron 4 with our minigene constructs we noticed that the presence of the exon 3 was always associated with a increase retention of intron 4. For instance, a construct containing exon 4 to exon 5 was better spliced than a construct encompassing exon 3 to exon 5. This was unanticipated since the EJC is in principle unable to bind the first construct due to the lack of flanking introns (Fig. 18 construct 3 vs construct 5). Similarly, a construct containing exon 3 to 5 but lacking intron 3 strongly impaired intron 4 splicing (Fig. 18 construct 2). I found that this effect was RnpS1-independent since it was not aggravated upon its knock down. To validate this observation, we replaced the exon 3 and intron 3 with the exon 2 and intron 2 of *piwi*, respectively. This replacement clearly led to a better excision of the intron 4. Therefore, our results strongly suggest that exon 3 contains an inhibitory splicing activity.

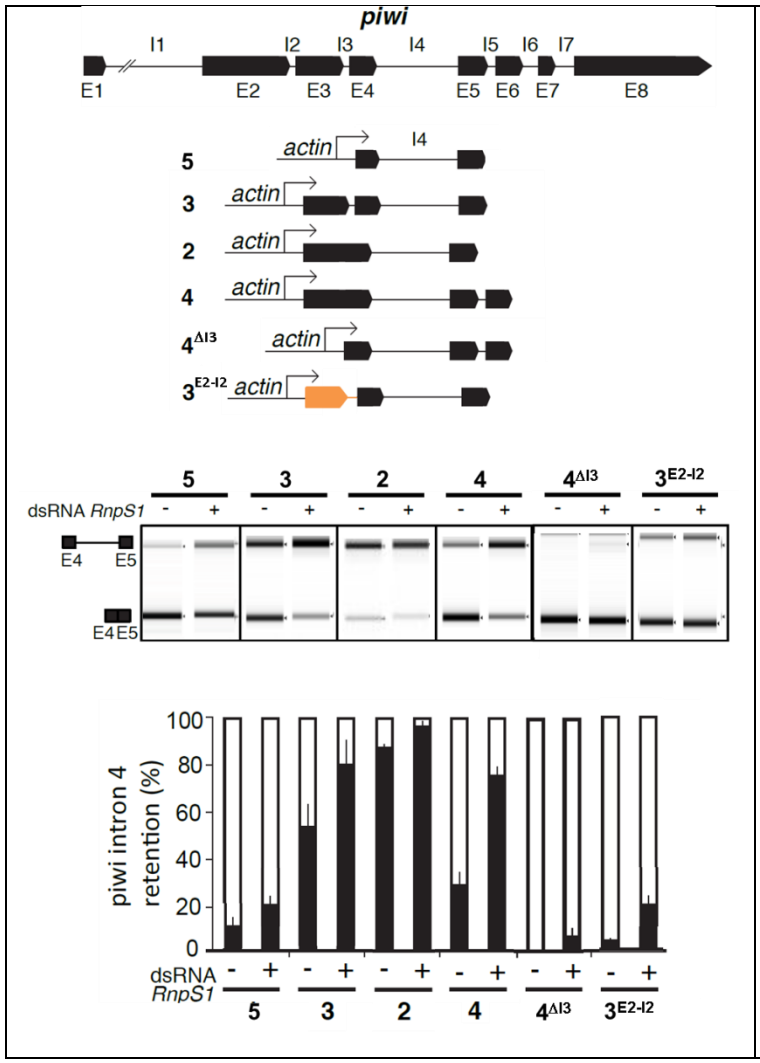

Figure 18: A splicing inhibitory activity of the intron 4 is located in the exon 3 and acts independently of the EJC.
Schematic of the *piwi* genomic locus, with introns (I) and exons (E) marked accordingly. Various minigene constructs containing the *piwi* fourth intron with or without *piwi* exon 3 or with replacement of *piwi* exon 3 and intron 3 by exon 2 and intron 2 (orange). Capillary gels of RT-PCR with primers flanking intron 4 using extracts from control or *RnpS1* depleted *Drosophila* S2R+ cells. Graphs showing the ratio of spliced (white) versus unspliced (black) fourth intron.

To gain more insight into this activity, I made additional constructs to shorten the exon 3 in order to isolate a minimal inhibitory sequence responsible for this effect. Surprisingly, shortening the exon 3 gave rise to a progressive diminution of the intron retention (Fig. 19). This suggested that the inhibitory activity was not due to a specific binding of a splicing inhibitor but was rather provided by the secondary structure of the RNA. Consistent with this, I could not find any known consensus sites for RNA binding proteins mapping to exon 3 using the RBPmap software.

Further, we hypothesized that if a secondary structure in exon 3 was important for this inhibitory effect, placing this exon 3 downstream of intron 4 might alter the structure and alleviate this effect. In contrast, if a splicing inhibitor binds specifically exon 3, the inhibitory effect might still be observed in this condition. We found that placing intron 3 downstream of intron 4 completely rescued the splicing efficiency (Fig.

19). Although several interpretations can be made from this experiment, it strongly suggests that secondary RNA structures negatively influence the splicing of intron 4 independently of EJC activity. However, we do not know whether this regulation also occurs in the context of the full *piwi* transcript.

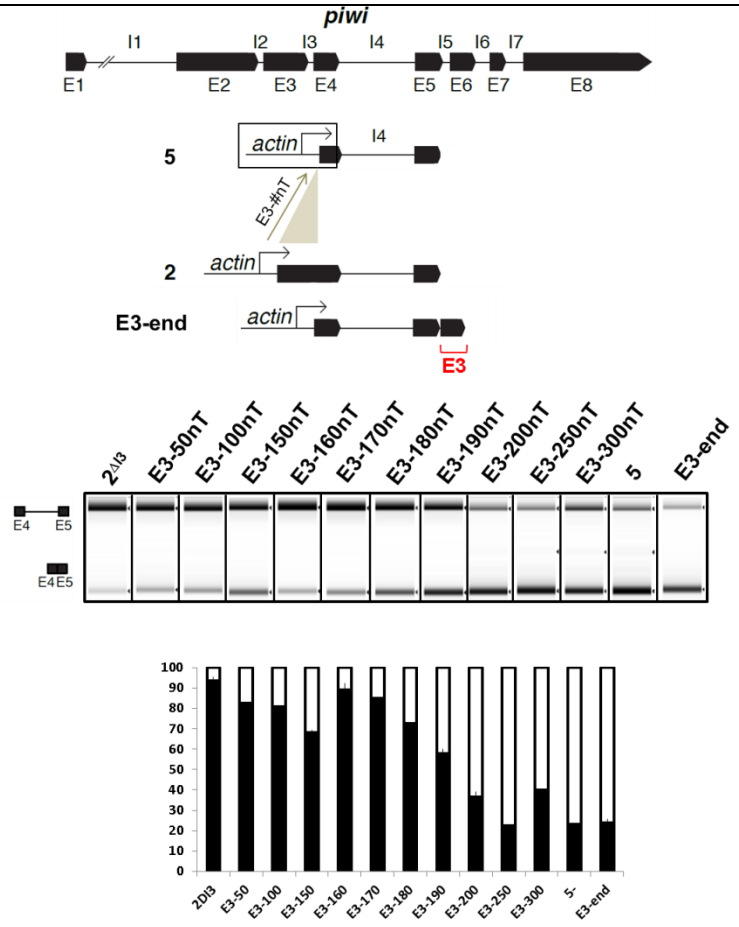

Figure 19: An RNA secondary structure within the exon 3 might be responsible of its inhibitory effect on intron 4 removal.
Schematic of the *piwi* genomic locus, with introns (I) and exons (E) marked accordingly. Various minigene constructs containing the *piwi* fourth intron with *piwi* exon 3 deleted for the indicated number of nucleotides (E3-#nT) or *piwi* exon 3 located at the end of the minigene (red). Capillary gels of RT-PCR with primers flanking intron 4 using extracts from *Drosophila* S2R+ cells. Graphs showing the ratio of spliced (white) versus unspliced (black) fourth intron.

3. KNOCK DOWN OF THE CORE EJC GIVES RISE TO EXON SKIPPING IN THE *PIWI* PRE-mRNA

We found that the retention event of *piwi* intron 4 observed upon *RnpS1* knock down was comparable to the retention observed upon knock down of core components of the EJC, such as *mago*. High level of intron retention was observed in both cases. However, we noticed that in addition to intron retention the depletion of *mago* gave rise to multiple exon skipping, that were not seen upon *RnpS1* knock down. This indicates that the core EJC has an additional splicing activity. This result is reminiscent to the splicing defects observed for large heterochromatic transcripts where many exon skipping were found in *mago* KD but only marginally seen upon RnpS1 depletion. Therefore, it would be interesting to identify the regulatory mechanism underlying the differential requirement for the EJC in both intron and excision definitions.

DISCUSSION

The EJC is recruited onto mRNA concomitant with splicing and remains strongly associated with the transcripts. Following their export to the cytoplasm, it can influence their subsequent cytoplasmic fate. Several studies recently challenged the view that the EJC serves only as a memory of the nuclear history but instead showed some evidence for its involvement in the splicing process itself [84,138,139]. However, since EJC subunits are recruited only at a late step of the splicing reaction it remained unclear how they can influence this process. Here, we provide a mechanism by which the EJC influences the splicing of the *piwi* transcript with the help of its splicing co-factors RnpS1 and Acinus. We found that the excision of the weak intron in the *piwi* transcript (intron 4) is facilitated by the splicing of strong adjacent introns, likely via the deposition of the EJC at spliced junctions. Furthermore, we found additional introns containing non-canonical splice sites that require the EJC and the splicing subunits for their removal strongly suggesting that this mechanism is broadly used. However, we also noticed that the core EJC has an additional activity in exon definition that it is barely observed with the associated splicing subunits suggesting an alternative mechanism in pre-mRNA splicing by the core components. This potential novel mechanism will be discussed below.

1. ROLE OF THE EJC IN THE piRNA PATHWAY

In this study we demonstrate that a main effector of the EJC in the piRNA pathway is the Piwi protein itself. This conclusion is based on several observations.

- First, depletion of *RnpS1* strongly affects *piwi* RNA level but has only minor effect on the level of other known piRNA-pathway components.
- Second, the class of transposons that is up-regulated upon *RnpS1* knockdown is similar to the one affected in the *piwi* knockdown.
- Third, expression of a cDNA encoding for Piwi partially restore a normal level of transposons in the *RnpS1* knockdown context.

However, based on our unpublished data and work from other groups we believe that the EJC may have additional functions in the piRNA pathway.

- It has been shown that the splicing and export factor UAP56 is involved in the nuclear export of dual strand piRNA cluster precursor transcripts in the germline [224]. UAP56 is an EJC peripheral subunit, suggesting that the core components may have a similar function. However, UAP56 is also a member of the TREX complex that can interact with mRNAs. Therefore, it is also possible that UAP56 carry out its piRNA function independently of the EJC core.
- The EJC seems to also regulate AGO3, as we observed a moderate decrease in transcript levels upon *RnpS1* knock down (Results, Fig. 8A) as well as an effect on AGO3 protein localization in *mago* mutant cells (Results, Fig. 4B). Furthermore, *mago* knockdown led to strong intron retention in *Ago3* transcripts (Sup. Fig. S1). However, despite this severe phenotype, we could still detect piRNA ping-pong activity in both *RnpS1* and *mago* knockdowns, indicating that the remaining AGO3 is sufficient to perform its critical functions.
- It has been recently shown that some piRNA cluster transcripts, contained introns (e.g the *flamenco* transcript). We found that upon *mago* knockdown, more exon skipping was observed in the *flamenco* pre-mRNA (Sup. Fig. S2). Since *Flamenco* encodes for a long transcript expressed from heterochromatin location this EJC activity might relate from its previous described role in splicing long heterochromatic transcripts. However, it remains unclear how this defect interferes with the piRNA pathway.
- We have found that Mago, but not RnpS1, controls stellate expression in the testis (Sup. Fig. S3). Interestingly, Piwi is not expressed in the germ cells of the testis indicating that this function must require an additional target. This target could either be Ago3 or another not yet identified protein.

To summarize, despite the strong effect of the EJC on *piwi* splicing, the EJC appears to act a different level in the piRNA pathway to repress transposition in the gonads. Furthermore, the EJC core components seem to play more important roles compared to the splicing subunit RnpS1.

2. NOVEL ROLE FOR THE EJC IN PROMOTING COOPERATION BETWEEN INTRONS

In vivo cooperation between introns for their processing has already been observed at several instances but the mechanism has remained unclear. For instance, introns of the tumor necrosis factor β transcript (TNFβ) are spliced more efficiently in HeLa cells when upstream introns are still present [272]. Similarly, upstream introns influence the efficiency of final intron removal of the human triosephosphate isomerase (TPI) transcript [273]. Importantly, in both cases, mutating the splice sites of the upstream introns abolishes their enhancing effect, indicating that their splicing *per se* is important. It has been proposed that the presence or absence of upstream introns could confer different secondary structures to the transcripts, potentially influencing efficiency of the splicing reaction [274–276]. However, in the light of our results, we favor an alternative hypothesis in which deposition of the EJC at the adjacent spliced junctions could explain the positive effect of the upstream introns on the removal of the downstream ones.

Although some studies showed evidence that intron excision follows an overall 5' to 3' polarity that is established during transcription [277,278] our data clearly show that this is not a strict rule. In the case of the *piwi* transcript, not only the splicing of the intron 3 influences the removal of intron 4 but the splicing of intron 5 appears absolutely required for efficient splicing of the upstream intron (intron 4). These results are consistent with recent genome-wide studies in *Drosophila* demonstrating that first introns are generally less efficiently co-transcriptionally spliced than subsequent introns [63]. As the excision of these introns is kinetically delayed, it would be interesting to test whether the EJC also facilitates their processing.

3. EVOLUTIONARY PERSPECTIVES

Few additional questions can be raised from this work. It is not clear whether the splicing of *piwi* by the EJC is developmentally regulated. In other words, is the splicing of this intron a mechanism to control Piwi level in the ovaries? So far, we have found no evidence for this possibility. The fact that we can reproduce the splicing regulation of *piwi* in S2R+ cells indicates that this mechanism is not gonad specific and is not regulated by other members of the piRNA pathway that would act as a sensor to modulate Piwi level. Second, even though Piwi is conserved across evolution, the structure of its introns is not (Sup. Fig. S4). In particular, the transposon sequences are only present in the *melanogaster* subgroup of Drosophilidae that stretch the intron sequence up to 700 nucleotides, whereas the other species have a shorter intron of approximately 60 nucleotides. Although the piwi intron 4 of nearly all the Drosophilidae is devoid of canonical poly-pyrimidine tract, introns with poorly defined poly-pyrimidine tract can be efficiently spliced *in vitro* if they are smaller than 90 nucleotides [279]. Consistent with this, the group of J. Brennecke examined the splicing of the homologous *piwi* introns in other *Drosophila* species [86] and found that the activity of the EJC on *piwi* splicing is restricted to the *melanogaster* subgroup.

Therefore, it is unlikely that the EJC plays a specific regulatory function in controlling Piwi expression. We favor a model in which the EJC acts as a general facilitator of splicing, helping the removal of introns having non-canonical splice sites or being too large to be efficiently recognized by the splicing machinery. As transposon mobilization within introns can increase their size or disrupt splice sites, the EJC could also be seen as a safe-keeper mechanism against damages caused by transposition.

4. ROLE OF RNPS1 AND ACINUS

Another question raised by this study is to understand how RnpS1 facilitates the splicing of introns once it is recruited to flanking exon junctions. It can either recruit components of the spliceosome directly or via additional splicing regulators such as SR proteins or hnRNP. In agreement with this later possibility, few splicing regulators have been shown to interact with both RnpS1 and components of the spliceosome. One of them is the splicing factor SRp54. One attractive possibility is that RnpS1 exerts its function via its interaction with SRp54, which was shown to bridge components of the spliceosome with non-canonical pyrimidine tracts. In future studies, it would be interesting to inactivate its function and examine its effect on *piwi* splicing. Furthermore, a proteomic screen to identify new partners of RnpS1 might also help to understand its role in the splicing process.

Besides, RnpS1 forms two mutually exclusive sub-complexes with additional EJC subunits, one termed apoptosis and splicing-associated protein (ASAP) and the other termed PSAP [268,269] (see introduction part). They share the Sin3-associated 18 kDa protein (SAP18) and differ by their association with apoptotic chromatin inducer in the nucleus (Acinus) for ASAP and Pinin for PSAP. Several functions have been attributed to these complexes in transcription, programmed cell death and mRNA processing [268,280–284]. Whether all these functions are shared with EJC core components is not known, and nor is the underlying mechanism dictating their specificity. Intriguingly, we observed that *Acinus* depletion, similarly to *RnpS1* knock down, strongly affects *piwi* splicing while neither SAP18 nor Pinin show a detectable requirement. These results not only indicate a functional specificity between the two complexes but also within the ASAP complex.

5. DISTINCT ROLES OF THE EJC IN INTRON AND EXON DEFINITIONS

In a previous study, our group has shown that many genes expressed from heterochromatic loci including *MAPK* are dependent on the EJC for their splicing [138]. These genes are embedded in a highly compact chromatin structure and often contain remarkably long introns (some nearly 150KB) with high density of repetitive elements. An accompanying study presented evidence that many transcripts containing large introns exhibit exon-skipping events upon EJC knockdown [139]. This is in stark contrast to *piwi*, which is expressed from euchromatin and contains relatively small introns. Furthermore, it displays intron retention in the absence of the EJC. Thus, depending of its target, EJC loss results in two types of splicing phenotypes. Another difference is the involvement of the splicing subunits. While EJC core KD gives rise to exon-skipping and intron retention events RnpS1 and Acinus KD only lead to intron retention. So, how does the core EJC prevent exon skipping in large genes?

One clue to these questions came from some observations in our lab that EJC loss is accompanied with a decrease in nucleosome occupancy and a gain in CTD phosphorylation of RNA Polymerase II (Pol II) at the *MAPK* locus, one of the EJC heterochromatic targets. As several studies have previously shown that the rate of transcription elongation can affect exon definition our data suggest that the EJC might prevent exon skipping from large transcripts by decreasing the phosphorylation state of Pol II in order to reduce the rate of transcription elongation. Consistent with this hypothesis, we have found that decreasing the speed of Pol II rescues exon skipping associated with the absence of the EJC *in vivo*. This effect appears specific since mutations in other spliceosome components do not increase Pol II phosphorylation. Therefore, these results support a mechanism in which the EJC influences exon definition through the modulation of transcription elongation. This mechanism might be particularly important to ensure proper splicing of challenging introns such as heterochromatic introns, which are usually large and rich in repeated elements. The mechanism by which the core EJC modulates the rate of transcription elongation remains to be determined.

During the writing my PhD thesis, a genome-wide study aiming to investigate the contribution of the EJC in alternative splicing in human cells has been performed. Interestingly, this study revealed an important role of EJC in alternative splicing [140]. Moreover, this function seems to be independent of the EJC splicing subunits, which is similar to the function we described in exon definition in *Drosophila*. Importantly, the authors found that the addition of drugs that lower the rate of transcription elongation rescued some of the EJC splicing defects pointing out for a connection between the EJC and the transcription machinery. This is in agreement with our observations indicating that this function might be conserved throughout evolution. Therefore, the next exciting challenge in the field of the EJC will be to characterize this novel mechanism.

ANNEXES

1. SUPPLEMENTAL DATA

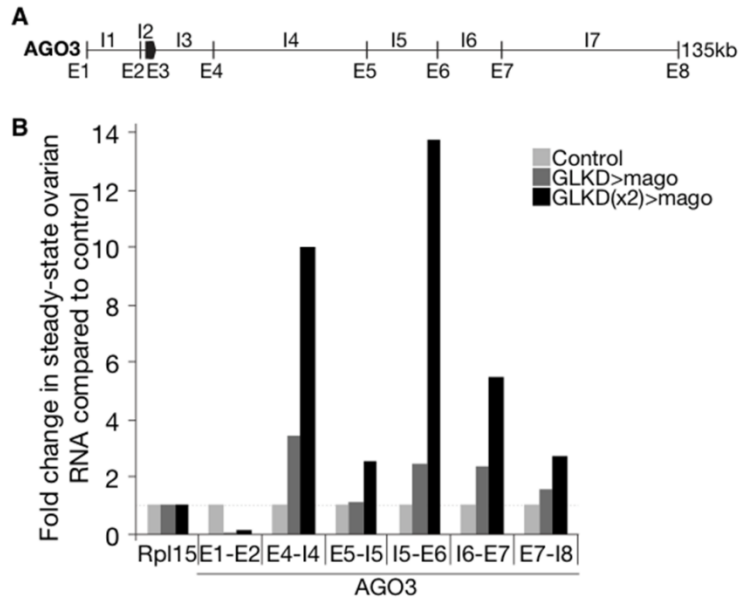

Figure S1: *mago* knock down led to exon skipping in *AGO3* mRNA
A. Schematic of the *AGO3* genomic locus, with introns and exons. B. Quantitative RT-PCR measuring *AGO3* RNA level at each exon (E) junction of the transcript, normalized to *RpL15* (gray) in *mago* knockdown (using one copy of the Gal4 driver (dark gray) or two (black)) ovaries.

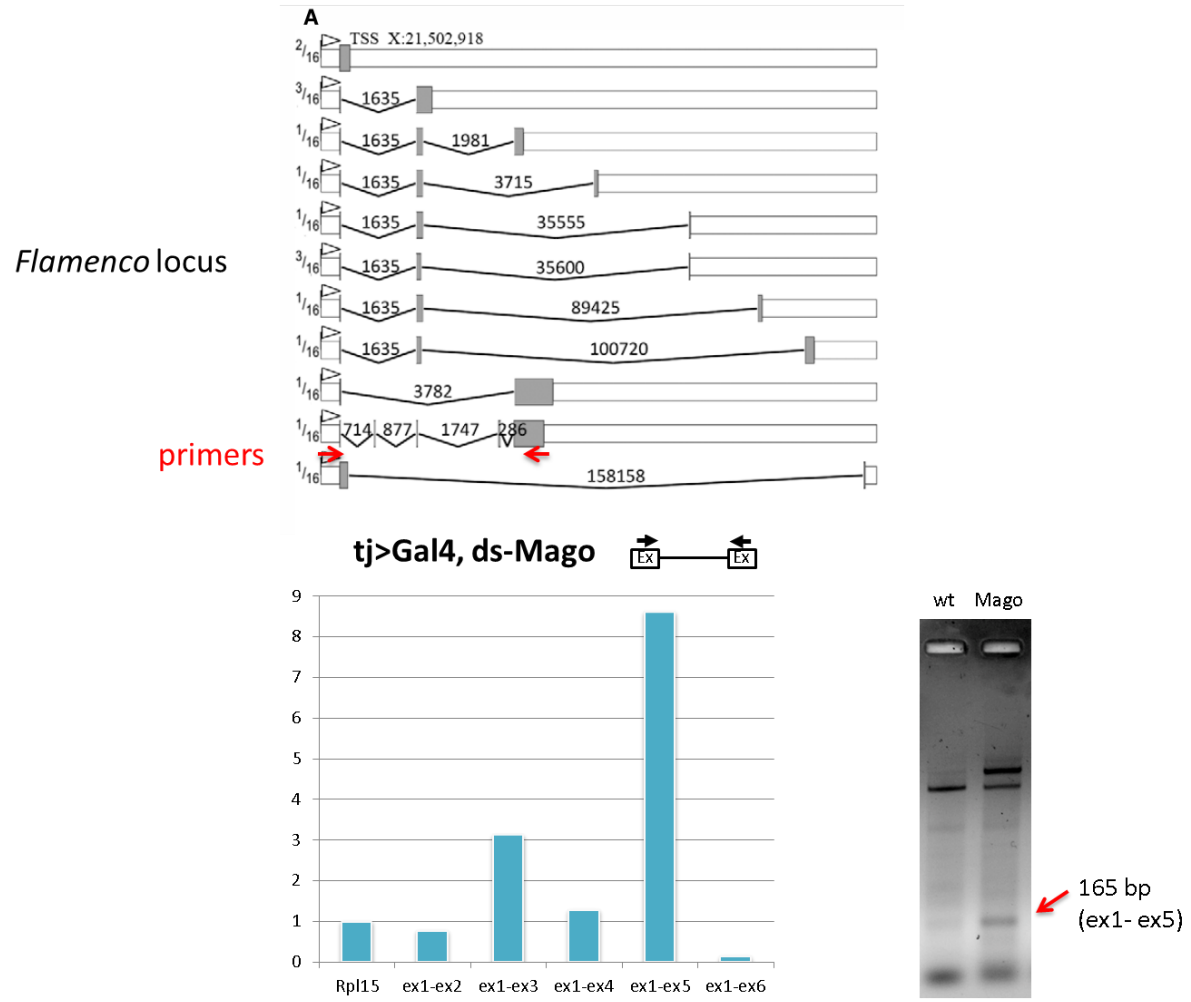

Figure S2: *mago* knock down led to exon skipping in *flamenco* mRNA
Top: depiction of the different isoforms of the *flamenco* transcript. The size of spliced regions is indicated in bp. Bottom: Fold change of exon skipping pattern in ds-*Mago* condition, normalized to Rpl15. Gel of the Semi-qPCR using primers flanking the region from the exon 1 to the exon 5 (represented in red on the upper panel). A 165 bp appears under Mago knock down.

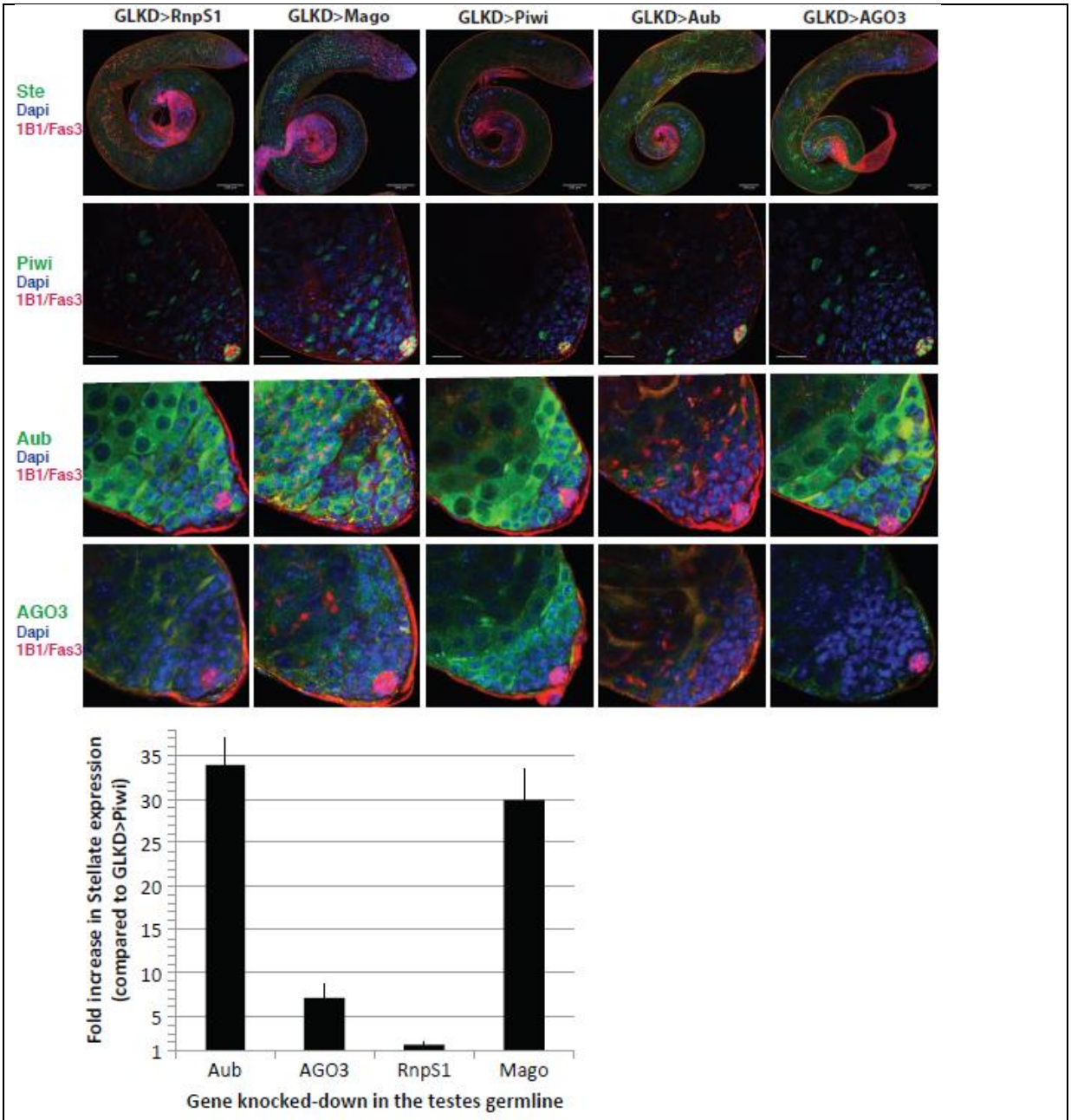

Figure S3: The core EJC but not RnpS1 controls the expression of Stellate in testes
Top. Adult testes from the indicated genotypes are dissected and stained with anti-1B1/Fas3 (red), anti-Stellate/Piwi/Aub/AGO3 (green) and Dapi (blue). Bottom. qRT-PCR of *Stellate* in testis samples depleted for *Aub*, *AGO3*, *RnpS1* or *Mago* compared to *Stellate* expression in *piwi* knockdown (control condition).

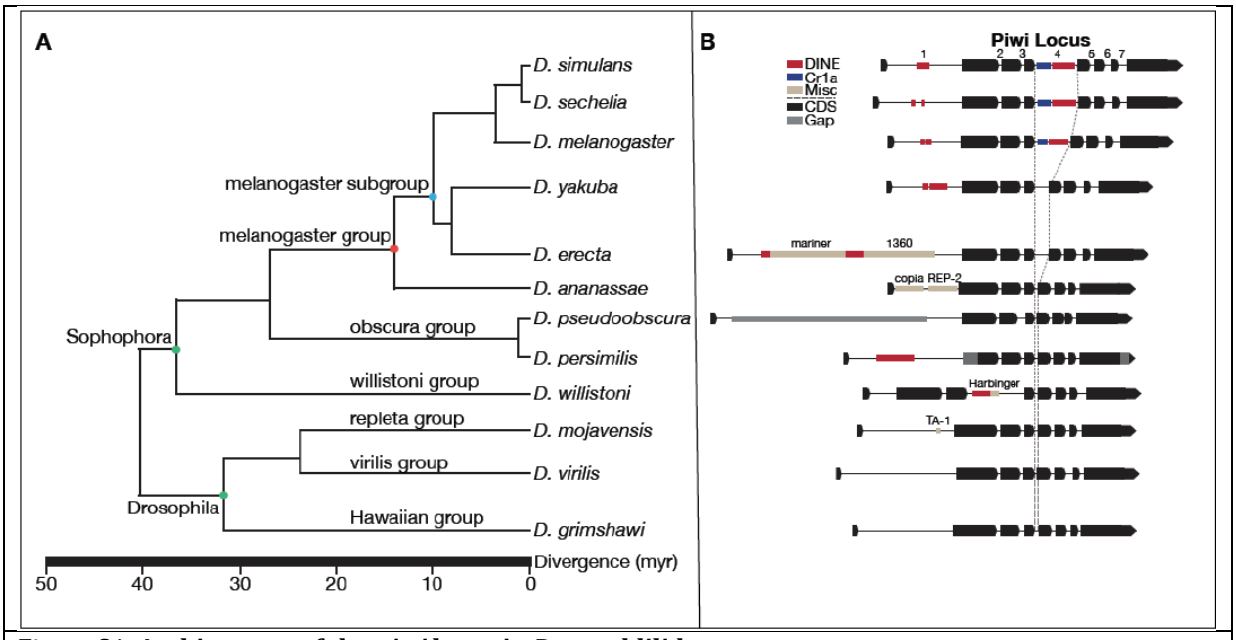

Figure S4: **Architecture of the *piwi* locus in *Drosophlilidae*.**
A. Taxon tree of *Drosophila* species. The divergence is indicated at the bottom. B. Diagram of the *piwi* locus for the species represented on the left. The transposon sequences are annotated.

Table 1: Fly stocks and antibodies used in this study

A. Fly stocks

traffic jam-GAL4 (Kyoto Stock center, DGRC #10455)

nos-GAL4-VP16 (Van Doren et al. 1998)

ey-GAL4 (Hazelett et al. 1998)

$mago^{93D}$/CyO

$tsu^{\Delta 18}$/CyO

btz^{2}/TM6b (Van Eeden et al. 2001)

$piwi^{1}$/CyO

$piwi^{2}$/CyO

hs-hid(Y)/UAS-Dcr-2; *nos*-GAL4-NGT; *nos*-GAL4-VP16, Burdock-LacZ/Tm3,Ser (Handler et al., 2013)

hs-hid(Y); *tj*-GAL4, gypsy-lacZ/CyO (Handler et al, 2013)

w; If/CyO; pUAST-BLRP_TEV_GFP-Piwi/TM3, Ser (kind gift from J. Brennecke)

w; If/CyO; pUAST-BLRP_TEV_GFP-Piwi [DN]/TM3, Ser (kind gift from J. Brennecke)

dsRNA-*piwi* (VDRC KK101658)

dsRNA-*mago* (VDRC GD28132)

dsRNA-*tsu* (VDRC KK107385)

dsRNA-*btz* (VDRC KK108761)

dsRNA-*RnpS1* (VDRC GD51851)

dsRNA-*acinus* (VDRC KK102407)

dsRNA-*pinin* (VDRC KK106336)

dsRNA-*SAP18* (VDRC KK105352)

dsRNA-*Ref1* (VDRC GD12031, KK104471)

shRNA-*RpnS1* (y^{1} sc^{*} v^{1}; P{TRiP.GL00540}attP2)

shRNA-*acinus* (Bloomington stock center)

B. Antibodies

rabbit anti-b Galactosidase (1:5000, Life Technologies™ A11132)

mouse anti-Tubulin (1:5000; Covance® mms-410P)

mouse anti-HA (1:500; WB 1:2000; Covance® mms-101P)

anti-gH2AV (1:500, Rockland™ 600-401-914)

Rat anti-Elav (1:100; Developmental Studies Hybridoma Bank)

Guinea Pig anti-Sens (1:1000) (Nolo et al. 2000)

mouse 1B1 monoclonal supernatant (adducin-like) (1:20; Developmental Studies Hybridoma Bank)

chicken anti-GFP (1:1000; Aves Labs)

mouse anti-b-Galactosidase (1:1000, Promega)

Phalloidin-TRITC (1:500, Promega)

rabbit anti-Vasa (1:5000, Lehmann Laboratory)

rabbit anti-Piwi (1:5000) (kind gift of G. Hannon)

rabbit anti-Aub (1:2000) (kind gift of G. Hannon)

rabbit anti-AGO3 (1:2000) (kind gift of G. Hannon)

Table 2: List of primers used in this study for RT-PCR, qPCR and for generating *piwi* minigene constructs and PCR templates used in RNAi

A. Primers for RT-PCR

Gene	Forward primer	Reverse primer
***piwi* splicing defects**		
5'UTR	TCGCGAGTGCCAAAAAGTAAC	
3'UTR		TGCTGAAGTGCTTAGCAAAAG
EX4-F-Ex5-R	CGAGATGCGCTCAAACTTTC	CAATCGATCAGTGCGTTGTT

B. Primers for real-time qPCR analyses

Gene	Forward primer	Reverse primer
Piwi		
int-4	GTACTAATTCCTGAGCTCTGCC	CTTAAGAAATAGATAGTTTGC
I2 retention	GGTATCGTTTTAAGTGCTGTTTGG	CCAAACAGCACTTAAAACGATACC
E1-2	CAGGCGTCCACTTAACGAAG	CGGGTCACCTGATGAAGATC
E2-3	AATCTCTTTGATCCCCGAGC	TCTCGGTGCGCATAACTTTG
E3-4	GGACAAACTCCGAAATCAAC	CCTTGTCCCTATTTTTGGAG
E4-5	CGAGATGCGCTCAAACTTTC	CAATCGATCAGTGCGTTGTT
E5-6	GGAACATGGAACTGGACAAC	GGTCTCTGAAGTGCCTTTGC
E6-7	TTCCCATGAACTCCGAACTC	CACACAATCATCCATTGCTC
E7-8	GCCTCGTACCCAATGATAAC	TTGGTGGCAATGCTCATAAG
tubulin	CTCAGTGCTCGATGTTGTCC	GCCAAGGGAGTGTGTGAGTT
Rpl15	AGGATGCACTTATGGCAAGC	GCGCAATCCAATACGAGTTC
mapk	AGGAGCTTATGGCATGGTTG	TGCAACTCTTTGGTTTGTTAGC
gypsy	GTTGAGGCAAGGATTGGAAA	TAAGCAGGTCAGCACCCTCT
ZAM	ACTTGACCTGGATACACTCA	GAGTATTACGGCGACTAGGG
idefix	AAACAAATCGTGGCAGGAAG	GCTTCTTTGGTTGGTCTGGA
TART	GCCTGGGCGTATTAGTCAGA	ACTTATGAGACGCCTCTGTT
Het-A	CGCGCGGAACCCATCTTCAG	CGCCGCAGTCGTTTGGTGAG
mdg1	CGTTCCCATGTCCGTTGTGA	AACAGAAACGCCAGCAACAG
roo	CGGCACTCCACTAACTTCTC	CGTCTGCAATGTCCTGGCTC
blood	AGACGTTCATTACAGATCAAGGTACGGA	AGTTCGTATGGGCAATAGTCATGGACT
GATE	GTCAGCCTTTAAGCTTTCGATG	CAGCATGATGTCCCTGTACT
aub(1)	AGCTGGTCTCGCGCTACTAC	TGGCATATCTCGGAACAAAG
aub(2)	AAACCAACTGGGCTGATGTC	TGGAGGGGGATCACTACCTG
ago3	GCAATTGCTTCAGAAAATGG	TGGATTCATCCGTCGTTAGC
vret	CGCTCATCAAGGAGATGGAG	TTGATCAGGTTCGGGTAACG
armi	ACTTTTACAACGGGGGAAGC	GTCGGCTTTTCGTTCTTCAG
fs(1)Yb	GTCAAGTGCTGCTCCTCTGG	TCAAACGCTCCTCGAACTTG

vas	ACCTAGACTGCCGGTCATTG	CGGGAAAGCAACACTCATTT
tej	TCCATTCCGGATACGGTTAC	GGGAGCCTCGATTCTTGGAC
krimp	AGCTGAATGGCTGTTTTAAG	CTGCCGATTTCAAAATGCAC
qin	GGAGCCCCTAAATTCATTCG	ACCGACGAGGTGTCAATAGG
Boyb	AGAATGCCAATCGACCTCTG	TAGGCCCAAGTTTCTCATGG
spn-E	ATGGCCACGAAGTTCAAATC	CCTTGAGGCTGTTCGTTGAG
mael	TCGTGCTAAACGCCAAGATG	ATAGACGTCGGTGGTCAAGG
Hel25E	TGTTCTCTGCCACATTGAGC	CGTCGACGTAGACCTCCATG
zuc	CCTCAGAGGTGATTTGGAAG	AGATCTCGCCCAGTTCATTG
Hen1	TGCGCTTCTTCCAGCTAATG	AACCAGTGGATTAACGCTGG
cuff	ACTTTGGTGCGGGATTTATG	CGCCATAACTTGGGAAATAC
squ	ATGGTCAACTGGCCTACGAC	CCGAATATCACGACCAGTTC
Soyb	GGCAGAACGAGGAGTACCTG	CATAAACTGCGACTCAAATG
papi	GCGGTGGATCCAATATCAAG	TTGTGGCCACTGTCTTTGTC
CG14879	CGTGCACATCCTGGACTTGGC	CGAGTTGCATCTTCTGTGGC
	GCCACAGAAGATGCAACTCG	CGCGCTTCAGTCGCTGAGCATGC
CG7458	CGCGTATCCTGGCTGGTCCG	CCACATTGTTTATAGTGCCCACC
	GGACGCGACTTCAAGTTCAGG	CCATCAATAGAGTCTATCAAC
spn	GCACCCTGCCAATTGAAATCGG	GCACTTGTGATATTCAAAGC
	CGGTCAAGGACGAACTGGATGC	CCAAGATCAACGGCGGCTCC
CG9914	GCAAAGAACGAGAAAGTTGGC	GGTAGCCCACCGAGGCGAACAGC
	GCTGTTATTCGCTAGGATATCC	GGATATCCAGGTGCTGACTGG
Hcs	CTTGTTGAAATCATCGCTTTGTGC	GCACTATTAAGAAGATCGGCAAGG
	CGCGGTGGAGACGGTGTCGACC	CGTCAGGTTCGGCGTTGCTATCACG
	GCACTTCTCCACTGACAGCG	CGGTTCCCAGCACATGGACATCC
	CCACGAAACGCTGCTCAGAGC	GCTAGAATATTGGATTCTGATGC
	CCATTTGTTTTGGCATCAGAATCC	CCATGGTCAGGTGCAGCTGC
	CCTGGCTGTGCTCCCCGTGCAGC	GCGATGATTTCAACAAGAAC

C. Primers for PCR templates used in RNAi

Gene	Forward primer	Reverse primer
RnpS1	TTAATACGACTCACTATAGGGAGACACATCGGACAGCAGCTCTTCG	TTAATACGACTCACTATAGGGAGAGCAATAGCAGTGACAGCTCTCGTTAG
Acinus	TTAATACGACTCACTATAGGGAGAAACGAGAAGACCGAGGTCGAGAC	TTAATACGACTCACTATAGGGAGACGGTCTTTGCTTGCTTCAGATGG
SAP18	TTAATACGACTCACTATAGGGAGACTGCATTCGATTATATAGTTATCG	TTAATACGACTCACTATAGGGAGAGTTCATACTCCATTGAGTGTCAAGT

D. Primers for generating minigene constructs are available on demand

2. MATERIALS AND METHODS

1.1. MATERIALS

STRAINS AND CULTURE MEDIA

Flies

The fly stocks are listed in Table 1 (Supplemental data). When applicable, w^{1118} served as control.

All experiments were performed at 25°C on standard fly medium (33L water; 5400g maize semolina; 300g soya flower; 555g beer yeast powder; 225g agar-agar; 1200g malt extract; 1200g beetroot sugar syrup; 75g Methyl-4-hydroxybenzoate)

Bacteria

All experiments involving bacterial cloning were performed with the DH5a strain (Invitrogen): F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1 (Hanahan et al., 1985)

All experiments were performed at 37°C on LB medium Luria (10g/L tryptone; 5g/L yeast extract; 0.5g/L NaCl; 15g/L agar for a solid medium)

Cells

All cell experiments were carried out in the late embryonic stage cells S2R+ (S2 receptor +) (Schneider et al., 1972)

All experiments were performed at 25°C with Schneider medium (life technologies) with 10% fetal bovine serum (PAA laboratories)

BUFFERS

PBS 10X: NaCl 1,37M; KCl 27mM; KH_2PO_4 11,5mM; Na_2HPO_4 64mM

WB running buffer 10X: 30g Tris; 144g Glycine; 50mL 20% SDS; add H_2O to 1L

WB transfer buffer10X: 22,5g Tris; 103,3g Glycine; 200mL MeOH; 800mL H_2O

TBS 10X: 30,2g Tris; 73g NaCl; 14mL HCl; add H_2O to 1L

Loading Lämmli buffer 6X: 360mM Tris-Cl pH6,8; 12%SDS; 60% glycerol; 30% β-mercaptoethanol; 0,06% bromophenol blue

PLASMIDS

Sequencing of *piwi* cDNA

The different *piwi* cDNA obtained by RT-PCR were cloned into a Topo-TA-vector (Invitrogen) and sent to the sequencing company GATC.

Study of the splicing mechanism

All minigenes were expressed in S2R+ cells and tagged with HA in their C-terminal domain. They derived from the fly expressing vector pPAC modified by PCR to insert an HA tag and additional restriction sites.

PRIMERS

Primers are listed in Table 2.

ANTIBODIES

Antibodies are listed in the Table 1.

1.2. METHODS

OVARY IMMUNOSTAINING

The ovaries were dissected in PBS and fixed 20 minutes at room temperature in PBS added with 5% Formaldehyde. They were rinsed 10 min at room temperature with PBT 1% (PBS containing 1% Triton-X-100) before being permeabilized 1 hour at room temperature in PBT 1%. The ovaries were blocked 1 hour in PBTB 0,2% (PBS containing 0,2% Triton-X-100 and 1% BSA). The incubation with primary antibody has been done over-night at 4°C in PBTB 0,2%. The ovaries were then washed 2 times 30 minutes in PBTB 0,2% at room temperature and blocked 1H in PBTB 0,2% containing 5% normal donkey serum. The incubation with the secondary antibody has been done 2 hours at room temperature in PBTB 0,2% containing 5% serum. The ovaries were washed 5 times 30 minutes with PBT 0,2% and mounted in VECTASHIELD mounting medium (VECTOR). Images were captured using either TCS-SP5 (Leica) or LSM 510 (Zeiss) confocal microscopes.

WESTERN BLOT

Protein extraction

Protein extracts were performed from ovaries or S2R+ cells lysed 20 minutes at 4°C in ice-cold lysis buffer (50mM Tris [pH 8]; 150mM NaCl; 1% Triton X-100; protease inhibitor cocktail [Roche]; 5mM EDTA; 5mM NaF; 1mM Na_3VO_4; 0,1% SDS and 0,5% sodium deoxycholate). The samples were then centrifuged 15 minutes at 4°C. The protein concentration of supernatants was measured with a Bradford assay (BioRad) and these supernatants were then mixed with Lämmli loading buffer before being used for western blot experiments.

Western Blot analysis

Samples were loaded on a SDS-PAGE polyacrylamide gel. gel running was performed in running buffer at 90V during 3 hours. Proteins were transferred on a nitrocellulose membrane 0.45um (BioRad) for 45 minutes at 90V. Ponceau tests were systematically performed to check transfer efficiency. Membranes were washed in TBST (TBS containing 0,2% Tween-20) and blocked over-night at 4°C in TBST containing 5% dry milk. Incubation with primary antibodies was performed 2 hours at room temperature in TBST containing 5% dry milk. Membranes were then washed 3 times 15 minutes with TBST. Incubation with secondary antibodies were performed 2 hours at room temperature in TBST. After washing 4 times 15 minutes with TBST, membranes were revealed using the chemiluminescent HRP substrate West Pico (Thermo). The signal was detected using GelDoc XRS+ from BioRad.

MEASUREMENT OF RNA LEVEL

RNA extraction

S2R+ cells or ovaries were lysed in 0,4 mL TRIzol (Life Technologies) for 5 minutes at room temperature. 107 µL of chloroform was added and the samples were mixed by vortex and incubated at room temperature during 10 minutes. The samples were centrifuged 15 minutes at 4°C at full speed. The upper aqueous phase was transferred in a new tube. The nucleic acids were precipitated by addition of 267µL of isopropanol. The samples were centrifuged 15 minutes at 4°C at full speed. The pellets were washed with 75% Ethanol and air dried 15 minutes. The pellets were resuspended in 20µL of RNase free water (Ambion) and heated 10 minutes at 60°C. Extracted RNAs were treated with Rq1 DNase (Promega) for 45 minutes at 37°C. The samples were then clean using the RNeasy mini Kit from QIAGEN. Elution has been performed with 30µL of RNase free water.

Reverse transcription–qPCR and –semi-qPCR

2µg of clean extracted RNAs were used to perform the reverse transcription reaction. The samples were incubated 1 hour at 42°C with MLV-reverse transcriptase (Promega). A heat inactivation was performed at 80°C for 5 min.

The quantification of RNA level was performed by quantitative-PCR analysis. Total cDNA were mixed with SYBR-Greer reagent (Life Technologies) and the primers of interest. The mixture was loaded on a 384-well-plate ant the measurement has been done by using a ViiA7 Real-Time PCR system (Applied Biosystem). For semi-quantitative PCR analysis, cDNAs were amplified using Phusion High-Fidelity DNA Polymerase (New England BioLabs).

RNA CLONING FOR DEEP-SEQUENCING

Total RNA from control, mago KD and RnpS1 KD ovaries were used to clone small RNA libraries as previously described (Pfeffer et al., 2005). Additionally, 5' and 3' cloning adapters were modified with five randomized nucleotides flanking the small RNA to minimize ligation biases. Strand-specific transcript profiling was performed on 2µg of total RNA from dsRNA-*RnpS1*/*tj*-GAL4; shRNA-*RnpS1*/*nos*-GAL4-VP16 or *tj*-GAL4; *nos*-GAL4-VP16 control ovaries depleted of ribosomal RNAs using two rounds of Epicentre® Ribo-Zero™ magnetic beads, and libraries were generated using an Illumina® TruSeq® kit. Libraries were multiplexed and sequenced on an Illumina® HiSeq® 2000 sequencing platform. All transcript and small RNA sequencing data have been deposited at Gene Expression Omnibus (accession numbers GSE57710 and GSE59327).

SPLICING DEFECT ANALYSIS USING MINIGENE REPORTERS

Construction of plasmids

piwi inserts were amplified from a Piwi BAC (157K19) and cloned into the modified Ppac vector, using the restriction enzymes Xho1 and Kpn1 (BioLabs). To construct chimeric minigenes or minigenes containing a deletion, long primers of 140 nucleotides overlapping the chimeric region were used to amplify the different fragments of the construct, and the products were used for a second round of PCR using flanking primers containing Xho1 and Kpn1 restriction sites.

Mutagenesis

Primers of 60 nucleotides containing point mutations of interest were designed in sense and antisense orientation. They were phosphorylated using T4 Polynucleotide Kinase (BioLabs) and used to perform PCR on a topo vector containing the fragments to mutate using Phusion high-fidelity DNA polymerase (BioLabs). The PCR fragments were

then purified and digested with DpnI (BioLabs). The obtained unligated plasmids containing the point mutation were transformed in DH5α competent cells to be amplified. The insert containing the point mutation was then cut out using restriction enzymes Xho1 and Kpn1 (BioLabs) and cloned into the previously described modified pPAC vector (see section "construction of plasmids").

Cell culture, double stranded RNA treatment and cell transfection

S2R+ cells were maintained in Schneider's medium (life technologies) supplemented with 10% fetal bovine serum (FBS, PAA laboratories).

PCR products obtained after amplification of an EST of interest with primers containing T7 sites were used for *in vitro* transcription using the MEGAscript T7 kit (Ambion). The reaction was then heated at 65°C for 30 minutes and cooled down slowly at room temperature to let hybridization of complementary single stranded RNAs.

S2R+ cells were treated with 15µg dsRNA in 1 mL Schneider's medium without FBS during 6 hours. After incubation, 1 mL of Schneider's medium containing FBS was added to the cells. The ds RNA treatment was repeated 1 and 3 days later.

Cells were transfected 4 days after the first dsRNA treatment with 300ng of minigene constructs, using Effectene kit (QIAGEN). RNA and protein extracts were performed 3 days after transfection in TRIzol (Ambion) and lysis-buffer, respectively.

Capillary analysis

In order to quantify the splicing defects, we used a capillary gel system. It consists in an accurate measurement of the different transcript isoforms present in a sample. 2µL of 5 times-diluted RT-PCR products were mixed with 2µL of reagents (High Sensitivity D1K Reagents), containing fluorophores, which are incorporated in the DNA, and two additional primers of 1500 and 25 nucleotides. The samples were then loaded on a 2200 Tape Station (Agilent Technologies), to carry out electrophoresis along capillary gels (High Sensitivity D1K ScreenTapes, Agilent technologies). The fluorescence in the DNA of each band present on gels were measured and quantified by normalizing to the markers. The markers were also used to determine the exact size of each band. After determination of the band identity, the proportion of spliced versus unspliced isoforms was calculated by percentage referring to the total amount of DNA present in the sample.

3. ARTICLE

The exon junction complex controls transposable element activity by ensuring faithful splicing of the *piwi* transcript

Colin D. Malone,[1,2,5] Claire Mestdagh,[3,5] Junaid Akhtar,[3] Nastasja Kreim,[3] Pia Deinhard,[3] Ravi Sachidanandam,[4] Jessica Treisman,[1] and Jean-Yves Roignant[3]

[1]Kimmel Center for Biology and Medicine at the Skirball Institute of Biomolecular Medicine, Department of Cell Biology, New York University School of Medicine, New York, New York 10016, USA; [2]Howard Hughes Medical Institute, [3]Institute of Molecular Biology (IMB), 55128 Mainz, Germany; [4]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA

The exon junction complex (EJC) is a highly conserved ribonucleoprotein complex that binds RNAs during splicing and remains associated with them following export to the cytoplasm. While the role of this complex in mRNA localization, translation, and degradation has been well characterized, its mechanism of action in splicing a subset of *Drosophila* and human transcripts remains to be elucidated. Here, we describe a novel function for the EJC and its splicing subunit, RnpS1, in preventing transposon accumulation in both *Drosophila* germline and surrounding somatic follicle cells. This function is mediated specifically through the control of *piwi* transcript splicing, where, in the absence of RnpS1, the fourth intron of *piwi* is retained. This intron contains a weak polypyrimidine tract that is sufficient to confer dependence on RnpS1. Finally, we demonstrate that RnpS1-dependent removal of this intron requires splicing of the flanking introns, suggesting a model in which the EJC facilitates the splicing of weak introns following its initial deposition at adjacent exon junctions. These data demonstrate a novel role for the EJC in regulating *piwi* intron excision and provide a mechanism for its function during splicing.

[*Keywords*: EJC; splicing; transposon; piRNA; Piwi; AGO3]

Supplemental material is available for this article.

Received May 19, 2014; revised version accepted July 15, 2014.

mRNA maturation requires pre-mRNA splicing, a process that removes introns from primary transcripts and subsequently ligates adjacent exons. Splicing plays essential roles during development, and its alteration is linked to a variety of diseases, including cancer (Srebrow and Kornblihtt 2006; Orengo and Cooper 2007; Baralle et al. 2009; David and Manley 2010). *Cis*-regulatory sequences within introns and exons regulate splicing by recruiting the spliceosome and associated factors to the primary transcript. Other layers of regulation, such as the influence of chromatin and transcription complexes, can also impact the splicing process (for review, see Long and Caceres 2009; Luco et al. 2011; Braunschweig et al. 2013; De Conti et al. 2013; Iannone and Valcarcel 2013). Despite considerable progress, the understanding of how the spliceosome successfully recognizes and precisely excises diverse introns in a variety of pre-mRNA molecules remains a challenge.

A critical link between pre-mRNA splicing and cytoplasmic functions was elucidated with the discovery of the exon junction complex (EJC) (Le Hir et al. 2000). This ribonucleoprotein complex is deposited onto mRNA 20–24 nucleotides (nt) upstream of exon–exon junctions as a consequence of splicing and remains stably associated during nuclear export to influence subsequent post-transcriptional events. In human cells, the EJC is associated with the majority of exon junctions, and its binding is linked to the adjacent binding of other mRNP components such as serine/arginine-rich (SR) proteins (Sauliere et al. 2012; Singh et al. 2012). The core of the EJC is composed of four proteins. The DEAD-box RNA helicase eIF4AIII directly binds to the RNA and is recruited early in the splicing reaction via its interaction with the spliceosome component CWC22 (Shibuya et al. 2004; Alexandrov et al. 2012; Barbosa et al. 2012; Steckelberg et al. 2012). Mago nashi

[5]**These authors contributed equally to this work.**
Corresponding author: j.roignant@imb-mainz.de
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.245829.114.

(Mago) and Tsunagi (Tsu/Y14) are recruited later during excision of the lariat intron and exon joining to lock the complex onto RNA by inhibiting the ATPase activity of eIF4AIII (Ballut et al. 2005). These three subunits constitute the pre-EJC. In the cytoplasm, the addition of Barentsz (Btz) completes the core EJC and further stabilizes association with the RNA (Bono et al. 2006; Gehring et al. 2009; Bono and Gehring 2011). Numerous other proteins interact transiently with the EJC to mediate additional post-transcriptional functions (Tange et al. 2005).

Despite the association of the EJC with most spliced RNAs, much work has shown its involvement in specific and distinct functions in vivo. For instance, both Mago and Tsu are required for dorsal–ventral axis formation in the *Drosophila* egg chamber (Micklem et al. 1997; Newmark et al. 1997; Hachet and Ephrussi 2001; Mohr et al. 2001) and affect earlier processes in oogenesis, such as germline stem cell (GSC) differentiation and oocyte specification (Parma et al. 2007). In vertebrates, reduction of *mago* affects neural stem cell division and melanoblast development (Silver et al. 2010, 2013). Finally, the three pre-EJC subunits are required for photoreceptor differentiation in flies (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010). Investigation of this function revealed that the pre-EJC is necessary for splicing *MAPK* and other transcripts that contain long introns and are expressed from heterochromatic loci. Why these features create a requirement for the pre-EJC is currently not understood. Other studies recently found a role for the EJC in controlling the alternative splicing of the proapoptotic regulator Bcl-x in human cancer cells and the splicing of the ryanodine receptor in *Xenopus* (Haremaki and Weinstein 2012; Michelle et al. 2012). Since the EJC joins the spliceosome at a late step in the splicing reaction, it remains unclear whether it can directly regulate splicing and, if so, by what mechanism.

In order to gain mechanistic insight into the role of the EJC in the splicing process, we set out to look for additional EJC targets. As mentioned, the EJC plays an important role in controlling axis polarity in *Drosophila*, as *mago* mutants lay ventralized embryos (Micklem et al. 1997; Newmark et al. 1997). Intriguingly, a similar defect is caused by impairment of the small RNA-based Piwi-interacting RNA (piRNA) pathway, suggesting a potential role for the EJC in piRNA production or function (Theurkauf et al. 2006). The piRNA pathway is an evolutionarily conserved RNAi system that acts predominantly in gonads to suppress the activity of transposable elements (Sarot et al. 2004; Malone et al. 2009). In flies, many piRNAs are derived from long precursors expressed from heterochromatic loci termed "piRNA clusters." piRNA clusters are densely packed with transposable elements, often fragmented and no longer active, which produce precursor RNAs that give rise to piRNAs capable of targeting active transposons. piRNAs are loaded into three distinct Argonaute proteins of the PIWI clade: Piwi, Aubergine (Aub), and AGO3 (for review, see Carmell et al. 2002). Piwi is predominantly nuclear and is required in both germline and surrounding somatic follicle cells to guide transcriptional silencing via heterochromatin formation (Cox et al. 1998, 2000; Megosh et al. 2006; Sienski et al. 2012; Le Thomas et al. 2013; Rozhkov et al. 2013). In contrast, Aub and AGO3 are germline-specific and enriched in a perinuclear structure termed "nuage" (Wilson et al. 1996; Harris and Macdonald 2001; Brennecke et al. 2007; Nishida et al. 2007; Nagao et al. 2010). Aub and AGO3 are targeted to cleave transposon or piRNA precursor transcripts, respectively, via their slicer activity. This reciprocal cleavage and subsequent processing amplifies piRNA production and intensifies the germline silencing response in what is called the "ping pong" cycle (Brennecke et al. 2007; Gunawardane et al. 2007; Senti and Brennecke 2010).

In this study, we found that the EJC inhibits transposon activity in both ovarian germline and follicle cells. The cytoplasmic subunit Btz is not involved, implying a nuclear function. Consistent with this, we found that the level of *piwi* transcript is reduced in the absence of the pre-EJC or of the accessory splicing subunit RnpS1 and its cofactor, Acinus (Acn). Although *piwi* is a euchromatic gene without large introns, the splicing of its fourth intron is impaired without the pre-EJC. We found that the main determinant of its retention upon EJC depletion was the presence of a weak polypyrimidine tract (PPT). We further demonstrate that excision of the *piwi* fourth intron requires prior splicing of the flanking introns. This suggests a model in which splicing of strong introns allows deposition of the EJC at their exon junctions, which facilitates the subsequent removal of the adjacent intron that contains a poorly defined PPT. In addition to demonstrating its novel function in the piRNA pathway via *piwi* splicing, these results imply that the EJC may play a more general role in determining the kinetics of intron excision.

Results

The EJC prevents transposon derepression in both germline and follicle cells

To investigate a possible role of the EJC in the piRNA pathway, we took advantage of dsRNA constructs from the Vienna *Drosophila* Resource Center (VDRC) and Transgenic RNAi Project (TRiP) collections to specifically knock down individual EJC components. We first confirmed the functionality of these constructs by driving their expression in larval eye discs using the *eyeless* (*ey*)-GAL4 driver. Expression of *mago* dsRNA or *tsu* dsRNA in the eye recapitulated the phenotypes observed with the loss-of-function mutant alleles, although their effects were weaker (Supplemental Fig. S1). In addition, expression of *btz* dsRNA had no effect on photoreceptor differentiation, consistent with the normal development of eye disc cells homozygous for the btz^2 loss-of-function allele (Supplemental Fig. S1). Expression of *eIF4AIII* dsRNA was cell-lethal, preventing further analysis (data not shown). We therefore expressed dsRNA targeting the other three subunits in either germline or surrounding follicle cells (using *nanos* [*nos*]-GAL4 or *traffic jam* [*tj*]-GAL4 drivers, respectively) (Fig. 1A) and examined the levels of representative transposons and ovarian morphology.

Somatic knockdown (SKD) of *mago* and, to a lesser extent, *tsu* in follicle cells gave rise to rudimentary

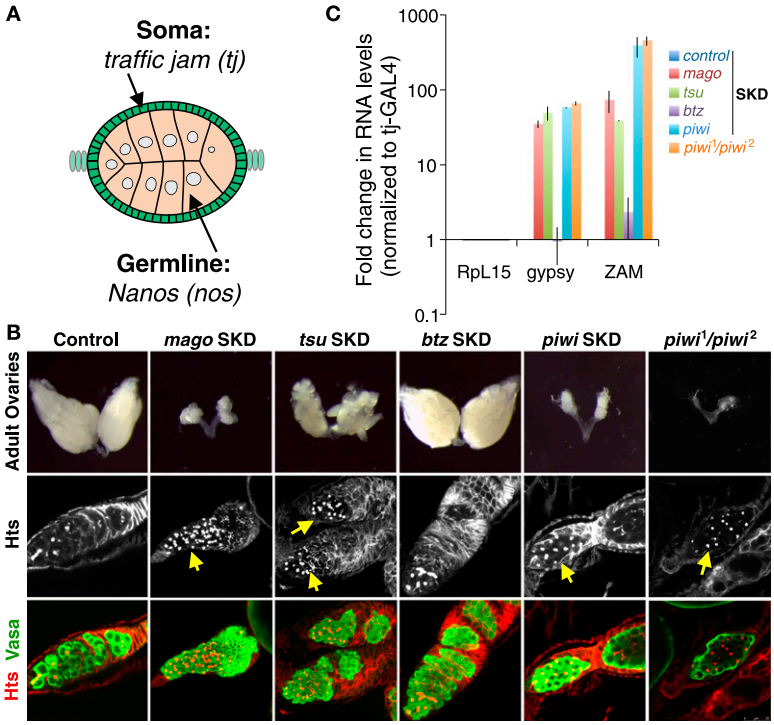

Figure 1. The pre-EJC prevents transposon mobilization in the *Drosophila* ovary. (*A*) Schematic representation of an egg chamber showing somatic (traffic jam) and germline (nanos) cells. (*B*) Adult ovaries from the indicated genotypes were stained with anti-Hts (red) and anti-Vasa (green) and show extra stem cells (yellow arrows). (SKD) Somatic knockdown. (*C*) qRT–PCR of *gypsy* and *ZAM* in ovarian samples depleted for EJC subunits in follicle cells. *piwi* mutant ovaries (orange) are shown as a control.

ovaries, which were devoid of mature egg chambers (Fig. 1B). In these knockdown ovaries, follicle cells formed multicellular layers and failed to encapsulate developing egg chambers, similar to *piwi* knockdown and loss of function (Fig. 1B; Jin et al. 2013). To gain insight into these developmental defects, we stained ovaries with the Hts antibody, which marks the spectrosome and cell cortex. The spectrosome is a spherical, germline-specific organelle found in GSCs, and as cystoblasts differentiate, it becomes increasingly branched to form the fusome. In wild-type ovaries, spectrosomes are present in the two or three stem cells at the tip of the germarium. Ovaries with follicle cells depleted of *mago* and *tsu* showed an increase in the number of spectrosome-like structures, indicating an accumulation of GSCs that had failed to differentiate (Fig. 1B). This is reminiscent of the GSC tumor-like phenotype of *piwi* loss of function (Fig. 1B; Jin et al. 2013). In contrast, depletion of *btz* had no effect on either morphology or GSC differentiation.

Given the similarity of *mago* and *tsu* knockdown phenotypes to that of *piwi*, we tested whether the transposon silencing defects observed in the absence of *piwi* (Sarot et al. 2004) are similarly induced by EJC depletion. Activity of the retrotransposons gypsy and ZAM is restricted to follicle cells, and their proliferation is inhibited by the piRNA pathway. We found that, upon SKD of *mago* and *tsu*, like *piwi*, the levels of *gypsy* and *ZAM* RNA increased dramatically, while *btz* SKD had no effect (Fig. 1C; Supplemental Fig. S2A). Germline knockdowns (GLKDs) of *mago* and *tsu* also displayed significant derepression of the predominant germline transposons *HeT-A*, *GATE*, and *Burdock*, which again remained unchanged upon *btz* depletion (Supplemental Fig. S2B,C). Loss of *mago* and *tsu* led to a marked increase in dsDNA breaks (Klattenhoff et al. 2007), supporting their role in protecting the genome from transposon mobilization (Supplemental Fig. S2D). These results indicate that the core nuclear components of the EJC are required for GSC differentiation and ensure the competency of the piRNA pathway to regulate transposons in both germline and somatic cells of the ovary.

The EJC controls the level of Piwi protein

Defects in the piRNA pathway often lead to a mislocalization and/or reduction in the level of Piwi-clade Argonaute proteins (Malone et al. 2009; Olivieri et al. 2010). In fact, piRNA pathway factors have been categorized based on their effects on the localization pattern of Argonaute proteins and whether they preferentially alter primary (Piwi) or secondary (Aub and AGO3) piRNA biogenesis or both (Olivieri et al. 2012).

To confirm that the pre-EJC regulates transposon expression by promoting the activity of the piRNA pathway, we assessed Piwi-clade protein localization in several EJC mutant contexts. In *mago*- and *tsu*-null mutant clones, the levels of AGO3 and Aub did not appear significantly changed (Fig. 2A,B; Supplemental Fig. S2E). However, these proteins appeared to partially relocalize to cytoplasmic bodies, suggesting that the integrity of nuage or their localization to this structure was disturbed. In contrast, we observed a significant decrease of Piwi levels in both germline and somatic clones of both *mago* and *tsu* (Fig. 2C; Supplemental Fig. S2E). Collectively, these results indicate that the pre-EJC

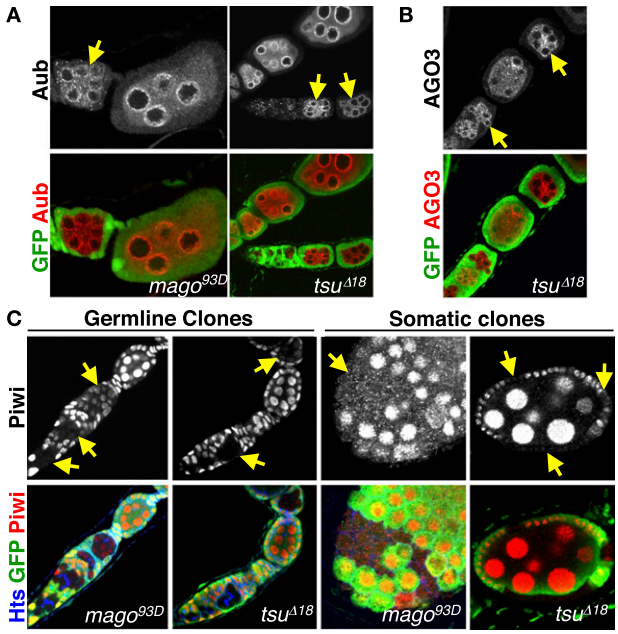

Figure 2. The pre-EJC specifically controls Piwi levels. (*A–C*) Ovaries containing clones homozygous for the *Y14*-null allele $tsu^{\Delta 18}$ or for the *mago* loss-of-function allele ($mago^{93D}$) are marked by the absence of GFP (green; yellow arrows) and were stained with antibodies against Aub (*A*), AGO3 (*B*), or Piwi (red) and Hts (blue) (*C*).

is required for normal Piwi levels but has only a minor effect on AGO3 and Aub localization.

Loss of the splicing subunit RnpS1 recapitulates EJC core subunit phenotypes

Since knockdown of cytoplasmic Btz has no impact on transposon levels, the pre-EJC is likely to regulate the piRNA pathway through a nuclear mechanism. Two nuclear functions have previously been proposed for the pre-EJC: one in splicing (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010; Michelle et al. 2012) and one in the control of mRNA export (Le Hir et al. 2001; Gatfield and Izaurralde 2002; Shiimori et al. 2013). In order to investigate which post-transcriptional function of the EJC was involved in the piRNA pathway, we knocked down a representative set of EJC accessory factors and analyzed both Piwi and transposon levels. Interestingly, we found that depletion of the splicing subunit *RnpS1* reduced Piwi levels in both follicle and germline cells (Fig. 3A,B). This effect was further confirmed by Western blot analysis (Supplemental Fig. S3A). In contrast, knocking down the export factor Aly/Ref1 had no impact on Piwi, while an additional RnpS1 cofactor, Acinus (Acn), was also required (Fig. 3A; Supplemental Fig. S4A). Interestingly, other direct RnpS1 partners, including Sin3-associated 18-kDa protein (SAP18) and Pinin, were not required for Piwi protein accumulation and localization (Fig. 3A).

To further characterize the role of RnpS1 in the piRNA pathway, we investigated the effect of its knockdown on transposon levels using several complementary approaches. First, we drove expression of dsRNA targeting *RnpS1* in either germline or follicle cells in the presence of Burdock-GFP or Gypsy-LacZ transposon reporters, respectively (Handler et al. 2013). While no signal was detected in control ovaries, clear expression was observed in both cell types upon *RnpS1* knockdown (Supplemental Fig. S3B). Second, we performed quantitative RT–PCR (qRT–PCR) and observed that, upon depletion of *RnpS1*, both germline and somatic transposons were strongly derepressed in their respective tissues (Supplemental Fig. S3C,D). Finally, we took a genome-wide approach to obtain a comprehensive view of the impact of *RnpS1* knockdown on the level of transposons. Total RNAs derived from either control or *RnpS1* knockdown ovaries (in both germline and follicle cells) were depleted of ribosomal RNAs, cloned, and sequenced. We confirmed that many transposons were significantly up-regulated upon *RnpS1* depletion. All classes of elements, which include those predominantly expressed in the soma (Idefix) or germ cells (HeT-A and GATE), showed increased expression (Fig. 3C–E). Altogether, these results indicate that RnpS1 is required to silence transposons and acts similarly to core piRNA pathway components (Czech et al. 2013).

EJC depletion and loss of Piwi have similar effects on small RNA profiles

Since loss of Piwi-clade proteins can lead to dramatic and specific alterations in piRNA populations (Malone et al. 2009), we sought to investigate whether EJC disruption also affects small RNA levels. We profiled small RNAs from control, *mago*, and *RnpS1* tissue-specific knockdown ovaries and analyzed small RNA production from piRNA clusters. For both *mago* and *RnpS1*, GLKD preferentially reduced piRNAs from the germline-specific 42AB cluster, while SKD reduced piRNA expression from the soma-specific flamenco cluster (Fig. 3F). An expanded analysis revealed that *mago* and *RnpS1* GLKDs reduced piRNAs from all germline clusters, similar to *piwi*, while another somatic cluster from the *tj* 3′ untranslated region (UTR) was unaffected or even slightly increased (Fig. 3G). Conversely, SKD reduced *flamenco* and *tj* piRNAs, leaving germline clusters mostly unaffected (Fig. 3G). These results indicate that nuclear components of the EJC are required for either piRNA production or stability.

Since piRNAs are essential for germline transposon regulation, we looked for effects on the levels and patterns of piRNAs that target transposons. As with piRNA clusters, transposable elements display preferential expression in either germline or somatic tissues (Brennecke et al. 2008). After sorting transposons by class and piRNA abundance, we assessed whether piRNAs that mapped to these transposons were capable of engaging in germline ping-pong, a signature of Aub and AGO3 function. Our data indicate that Aub and AGO3 remain engaged in ping-pong to regulate transposon expression during *RnpS1* GLKD despite an overall reduction in piRNA levels (Supplemental Fig. S3F). *mago* knockdown reduced transposon targeting piRNAs to an extent similar to or even more severely than

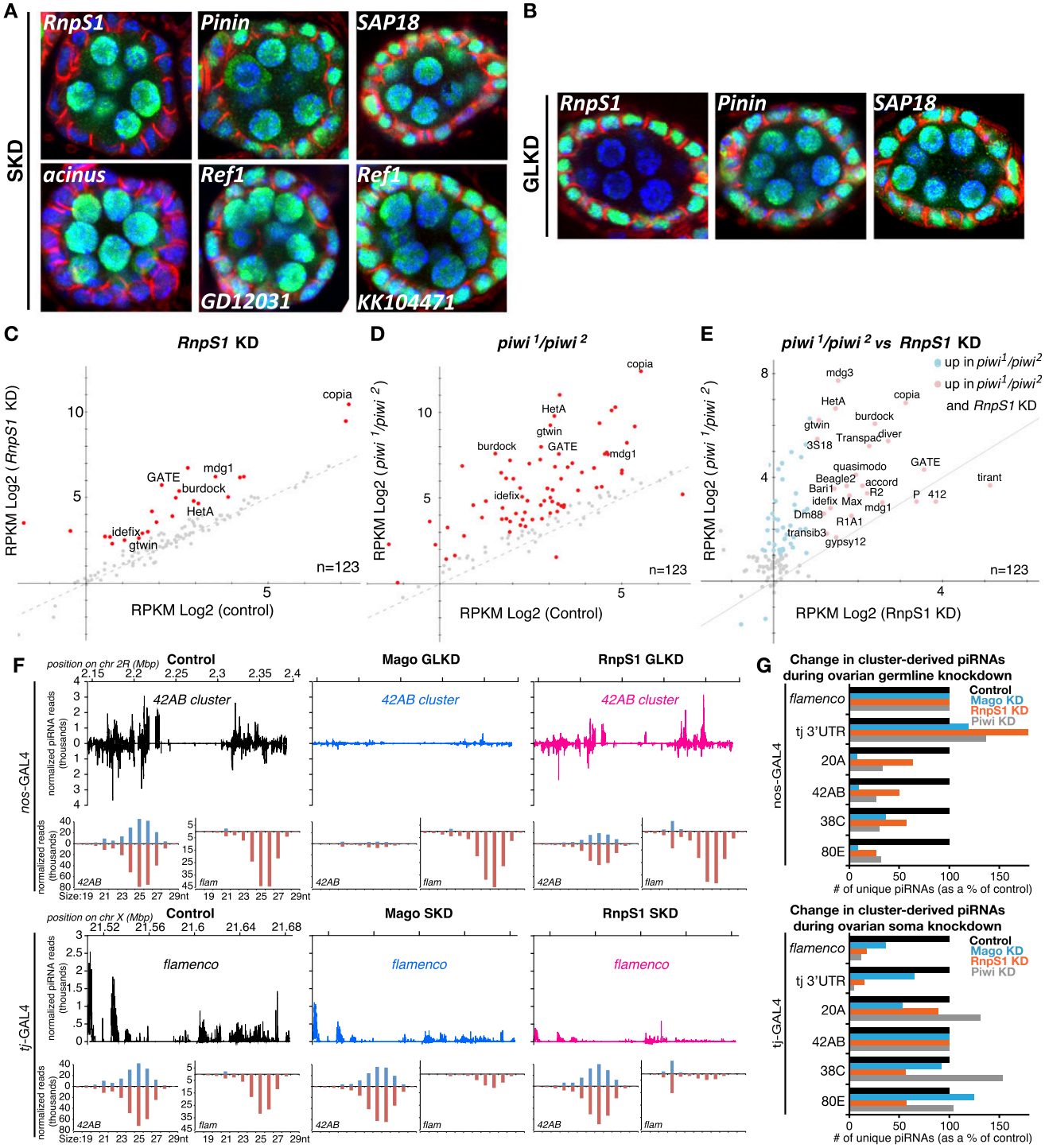

Figure 3. The EJC accessory subunit RnpS1 is required for Piwi accumulation and transposon regulation. (*A*,*B*) Egg chambers stained for Piwi (green), Hts (red), and DNA (blue) during SKD (*A*) or GLKD (*B*) of *RnpS1* and accessory factors. (*C*–*E*) Twofold or greater, statistically significant transposon derepression (red and blue) in *RnpS1* knockdown and *piwi* mutant ovaries compared with controls and each other. No transposons showed a more than twofold change only in *RnpS1* knockdown. (*F*) Unique small RNA mapping and size distribution over canonical germline (42AB) and somatic (flamenco) piRNA clusters in ovaries with EJC component knockdowns. (*G*) Normalized representation of cluster piRNA levels in *mago* (blue), *RnpS1* (orange), and *piwi* (gray) knockdown ovaries.

RnpS1, which could explain its stronger morphological phenotype. In conclusion, the fact that ping-pong remains intact indicates that secondary piRNA biogenesis is not wholly dependent on the EJC. Our results support an essential role for the EJC in the Piwi-dependent primary piRNA pathway.

RnpS1 is required to facilitate piwi transcript splicing

A failure to load small RNAs into Piwi triggers a defect in Piwi nuclear accumulation and its subsequent destabilization (Olivieri et al. 2010, 2012). Therefore, the observed decrease in Piwi levels during *RnpS1* depletion could be an indirect consequence of diminished piRNA levels. Alternatively, since the EJC plays critical roles in mRNA metabolism, a more direct effect on *piwi* RNA processing is also plausible. To distinguish between these possibilities, we examined *piwi* mRNA levels by qRT–PCR using several amplicons spanning the transcript. Interestingly, we found that *RnpS1*-depleted ovaries showed a 40%–80% decrease in *piwi* transcript levels (Fig. 4A), with the strongest effect observed on the amplicon spanning exons 4 and 5. We confirmed that *piwi* transcript levels were also reduced by *mago* GLKD using RNA fluorescent in situ hybridization (FISH) (Fig. 4B). This effect appears specific, since, of 20 validated components of the piRNA pathway, only *piwi* mRNA (and, to a lesser extent, *AGO3*) displayed a significant decrease upon *RnpS1* knockdown (SKD and GLKD) (Supplemental Fig. S5A). Further qRT–PCR displayed a significant increase in *AGO3* intron retention, indicating partial dependence on the EJC for efficient transcript processing (Supplemental Fig. S6). This may relate to the previously described function of the EJC in regulating large heterochromatic genes.

To search for additional piRNA pathway genes targeted by the EJC, we used our RNA sequencing (RNA-seq) data to compare the expression in control versus *RnpS1* knockdown ovaries of all genes identified in three screens for regulators of the piRNA pathway (Czech et al. 2013; Handler et al. 2013; Muerdter et al. 2013). In total, 467 genes were examined, and we found that five in addition to *piwi* showed a decrease in expression of at least 1.5-fold (Supplemental Fig. S5B). Of these candidates, *piwi* was most strongly down-regulated, suggesting that *piwi* is the main target of RnpS1 in the piRNA pathway. To further confirm this hypothesis, we compared the genome-wide changes of transposon levels in *RnpS1* knockdown ovaries with those of *piwi* mutant ovaries. Piwi regulates only a subset of transposons, so derepression of these transposons provides a clear signature of its effect. Strikingly, while the effect on transposon levels was stronger in *piwi* mutant ovaries (Fig. 3C,D), every transposon up-regulated in *RnpS1* knockdown ovaries was also up-regulated in *piwi* mutants (Fig. 3E). These results indicate that *piwi* reduction is partially sufficient to explain the effects of RnpS1 on the piRNA pathway.

To test whether the EJC regulates *piwi* transcriptionally or post-transcriptionally, we took advantage of an enhancer trap P element inserted in the first intron of *piwi* that expresses *lacZ* under *piwi* regulatory sequences (Lin and Spradling 1997). In *tsu* mutant clones, *lacZ* levels were not

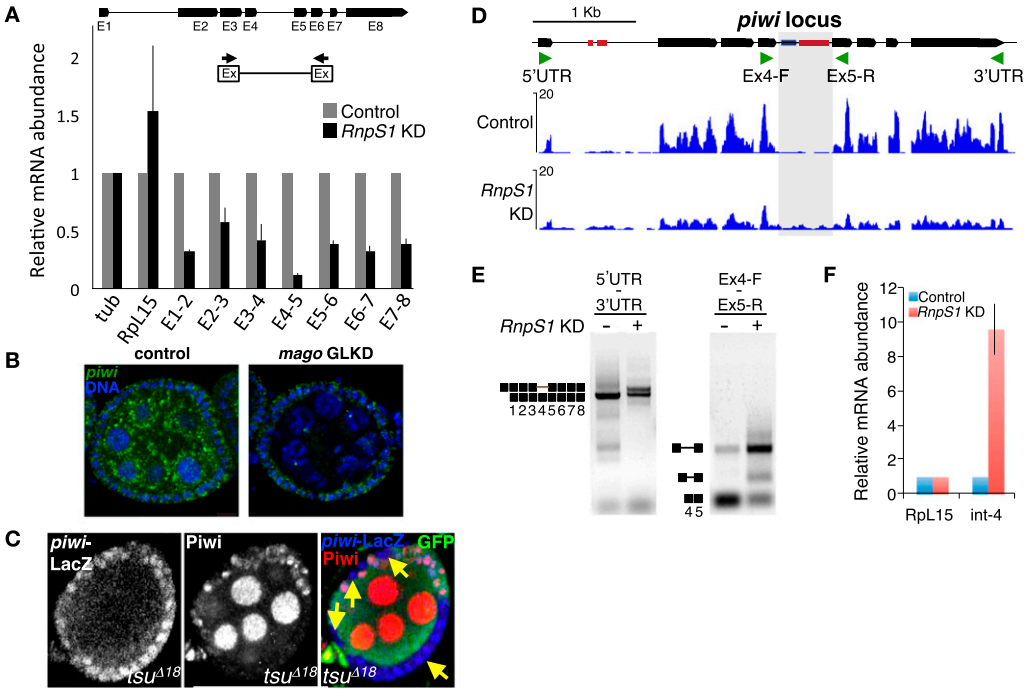

Figure 4. Efficient splicing of the *piwi* fourth intron is dependent on the EJC. (*A*) qRT–PCR spanning each exon (E) junction of the *piwi* transcript, normalized to *tubulin*, and with *RpL15* as a control. (*B*) *piwi* transcripts are visualized using RNA FISH in control and *mago* GLKD egg chambers (green). (*C*) Assessment of *piwi* transcription using a LacZ reporter (blue) in *tsu* mutant somatic clones marked by a lack of GFP (green; yellow arrows). Piwi protein is stained in red. (*D*) RNA transcript mapping over the *piwi* locus from control and *RnpS1* knockdown ovaries. Intronic transposons Cr1A (blue) and DINE (red) and intron read accumulation (gray box) are shown. (*E*) RT–PCR of full-length *piwi* and intron 4 (primers shown by green arrows in *D*) with and without *RnpS1* knockdown. (*F*) Quantification by RT–PCR of *piwi* fourth intron retention.

decreased, suggesting that EJC loss does not reduce *piwi* transcription (Fig. 4C).

We then sought to investigate whether *piwi* splicing remained intact. Using the *RnpS1* knockdown RNA-seq data, we observed a general decrease in the number of reads mapping to *piwi* exons, which is consistent with the reduction detected by qPCR and immunofluorescence. In contrast, the number of reads mapping to intron 4 was significantly increased, indicating an accumulation of this intron in the absence of RnpS1 (Fig. 4D). To confirm these observations, primers spanning the entire transcript as well as each exon–exon junction were used to amplify *piwi* from wild-type and *RnpS1*-depleted ovarian RNA. We detected a major defect in the splicing of the fourth intron (Fig. 4E), leading to a dramatic accumulation of unspliced RNA (Fig. 4F). Interestingly, we also observed some retention of this intron in the wild-type condition, indicating that the intron is normally inefficiently spliced. Furthermore, we found a less striking defect in the splicing of the first intron, where activation of a cryptic splice site occurs in wild-type ovaries at a low level and more frequently in *RnpS1* knockdown conditions (Fig. 4D; data not shown). Thus, these data show that *piwi* intron 4 and, to a lesser extent, intron 1 are suboptimally spliced under normal conditions and that RnpS1 and the EJC must support efficient transcript splicing in order to maintain Piwi protein levels sufficient for its function.

Piwi cDNA partially rescues transposon derepression caused by EJC knockdown

After determining that the predominant defect observed during EJC loss is in the splicing and subsequent accumulation of Piwi, we asked whether we could restore transposon regulation by expressing Piwi from a cDNA template. We expressed integrated, inducible Piwi-GFP cDNA and a control cDNA that lacks a nuclear localization signal and produces a cytoplasmic protein in vivo (Fig. 5A; Sienski et al. 2012). We expressed these constructs in somatic cells together with dsRNA targeting *armitage* (*armi*), a piRNA component that is post-transcriptionally required for Piwi nuclear localization (Haase et al. 2010; Olivieri et al. 2010; Saito et al. 2010). As expected, even wild-type Piwi is incapable of entering the nucleus during *armi* depletion (Fig. 5B). In contrast, Piwi produced from this cDNA localized properly to the nucleus in the absence of RnpS1, indicating that the EJC does not play a role in Piwi localization (Fig. 5B).

To examine whether reconstitution of Piwi is sufficient to restore transposon regulation in the absence of RnpS1, we examined the steady-state levels of several transposable elements. Using qRT–PCR, we detected a somatic restoration of *piwi* transcript levels from cDNA despite persistent depletion of *RnpS1* (Fig. 5C). We then assessed transposon levels and found a partial suppression of *ZAM* transcription and, to a lesser extent, *Blood* (Fig. 5C). We used RNA FISH recognizing *ZAM* transcripts to test whether *piwi* cDNA could also rescue the effects of loss of a core EJC subunit. We observed a marked restoration of *ZAM* silencing in ovaries with SKD of *tsu* (Supplemental Fig. S7A). We noticed that the cDNA constructs were expressed in a variegated pattern in follicle cells (Fig. 5B; Supplemental Fig. S7B), potentially preventing a more robust rescue of transposon silencing. Taken together, we conclude that Piwi is a critical effector of the EJC in preventing transposon mobilization.

RnpS1-dependent intron removal requires splicing of flanking introns

To gather more insight into the molecular mechanism by which the EJC regulates *piwi* splicing, we designed several minigene constructs containing portions of the *piwi* locus encompassing the EJC target intron (intron 4) and transfected them into *Drosophila* embryonic Schneider ($S2R^+$) cells, which have no detectable endogenous *piwi*. We found that expression of a short genomic construct from exon 3 to exon 6 (construct-1) recapitulated the RnpS1 dependence of intron 4 removal observed in the gonads (Fig. 6A,B,B′). While intron 4 was almost completely spliced in construct-1, its retention was increased by 20% upon *RnpS1* depletion (Fig. 6B,B′). Interestingly, we observed that removing the flanking introns from this construct had a severe impact on intron 4 excision under control conditions, and this was not aggravated by *RnpS1* depletion (Fig. 6B,B′, construct-2). We next assessed the relative importance of the two flanking introns. We found that removing only intron 5 had a strong effect on intron 4 removal (50% retention), which was further exacerbated by depletion of *RnpS1* (85% retention) (Fig. 6B,B′, construct-3). Conversely, removing only intron 3 led to a minor effect on intron 4 splicing (20% retention), but this defect was strongly aggravated by *RnpS1* depletion (70% retention) (Fig. 6B,B′, construct-4). Altogether, these results show that both flanking introns are required for correct splicing of intron 4, with intron 5 having a greater impact. Furthermore, both flanking introns exert their effect through RnpS1.

The positive effect of the flanking introns could be due to either their *cis*-regulatory sequences or their splicing per se. To test whether their splicing was important, we generated point mutations in the 5′ or 3′ splice sites (5′ss or 3′ss) of intron 3 and examined the splicing of intron 4. Strikingly, prevention of intron 3 splicing completely inhibited the splicing of the downstream intron (Fig. 6C,C′). Furthermore, knocking down *RnpS1* did not aggravate this effect. A similar effect was observed when the 5′ss of intron 5 was mutated in construct-4. In this case, preventing intron 5 splicing leads to a complete inhibition of the splicing of the upstream intron (Fig. 6C,C′). Together, these results demonstrate that splicing of flanking introns is necessary to enable RnpS1 to promote splicing, likely via the deposition of the EJC at exon junctions.

A weak PPT at the 3′ end of the retained intron confers RnpS1 dependence

We wondered whether the splicing defect observed in *RnpS1* knockdown was unique to *piwi* or more globally prevalent. To this end, we reanalyzed our RNA-seq data sets from *RnpS1*-depleted ovaries in search of transcripts

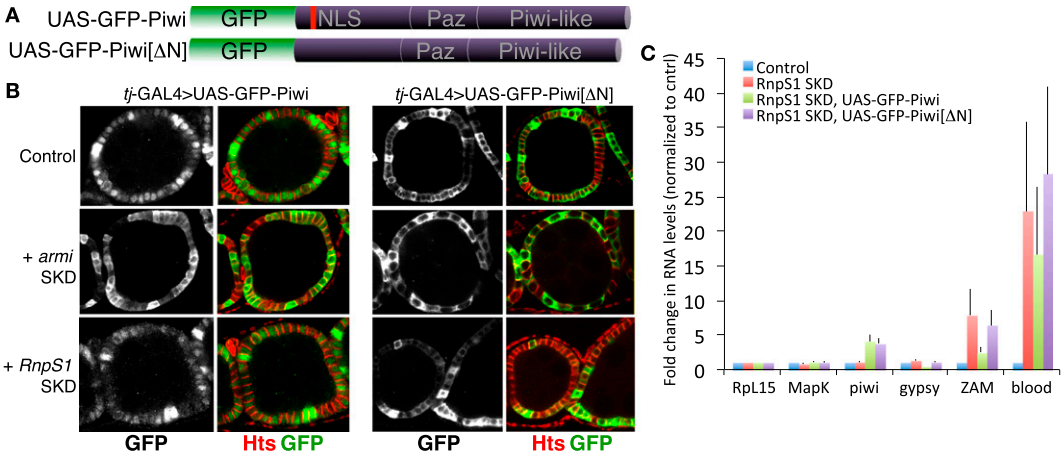

Figure 5. Expression of Piwi cDNA partially restores transposon up-regulation caused by EJC knockdown. (*A*) Schematic of UAS promoter-driven GFP-Piwi and GFP-Piwi with the nuclear localization signal removed to serve as a negative control. (*B*) Somatic expression of Piwi constructs in control and *armi* or *RnpS1* knockdown egg chambers stained for anti-Hts (red) and anti-GFP (green). (*C*) qRT–PCR of somatic transposons in *RnpS1* SKD ovaries without and with the coexpression of Piwi constructs.

that showed signs of intron retention. Interestingly, we found that 33 transcripts showed a significant increase in intronic-derived reads, with some transcripts having multiple introns retained (Supplemental Table S1). Out of seven tested directly, six intron retention events were validated by qPCR (Supplemental Fig. S8).

To investigate the type of sequence elements able to generate RnpS1 dependence, we engineered several versions of the *piwi* minigene construct-3. We first asked whether the strength of the splice sites could explain the EJC dependency. Using the Berkeley *Drosophila* Genome Project (BDGP) splice site predictor software, we noticed that the 5′ss of intron 4 has a low score (0.31) compared with other 5′ss of the *piwi* primary transcript. Therefore, we replaced the 5′ss of intron 4 with the 5′ss of the higher-scoring intron 2. While this substitution improves the splicing efficiency of intron 4, it does not relieve the requirement for RnpS1 (Fig. 7A), ruling out splice site strength as a main determinant of RnpS1 dependence. Next, we replaced the entire intron 4 with an intron of similar size from a *tubulin* transcript. Interestingly, this substitution completely rescued the splicing defect associated with the lack of RnpS1 (Fig. 7A). Loss of RnpS1 dependence was also observed when the endogenous 5′ss of intron 4 was still present. This indicates that retention of the *piwi* intron is due to specific sequence elements that lie within the intron itself rather than binding of splicing inhibitors at flanking exons. In order to identify the minimal sequence that confers RnpS1 dependence, we reintroduced portions of the *piwi* intron into this construct (Fig. 7B,C). Interestingly, introducing only the last 100 nt of *piwi* intron 4 was sufficient to alter splicing and reestablish RnpS1 dependence. This sequence is highly enriched in adenine in its last 30 nt, suggesting that the PPT of this intron is altered (Fig. 7D). To confirm this hypothesis, we specifically replaced this A-rich sequence with the last 30 nt of the *tubulin* intron, which displayed a complete rescue of the splicing defect (Fig. 7B).

In order to investigate whether this sequence is sufficient to generate RnpS1 dependence, we engineered a minigene construct in which we inserted this minimal sequence into a heterologous intron derived from the *ftz transcription factor 1* (*ftz-f1*) gene. We selected intron 5 of *ftz-f1* because it is similar in size to intron 4 of *piwi* and does not require RnpS1 for its splicing (Fig. 7E; data not shown). Strikingly, we found that the replacement of its last 30 nt with the A-rich sequence of the *piwi* intron led to a retention of 40%, which is increased by 25% upon *RnpS1* knockdown. This result demonstrates that the PPT of *piwi* intron 4 is sufficient to alter splicing and generate RnpS1 dependence. In addition, we found that, out of the 35 retained introns in *RnpS1* knockdown, nine are enriched in adenine in their last 30 nt (Supplemental Table S1), suggesting that weak PPTs may be a determinant of RnpS1 dependence in other introns, although additional features are likely to be involved.

Taken together, our results suggest that RnpS1 and the EJC play a general role in facilitating the removal of introns that are relatively difficult to splice. These data expand the catalog of regulatory functions of the EJC and, in this case, a subspecialization essential for the efficient maturation of diverse mRNAs in the cell.

Discussion

Convergence of the EJC and piRNA pathways

EJC subunit depletion causes a decrease in piRNA levels and Piwi function, leading to broad-scale transposon mobilization and demonstrating its essential role in the piRNA pathway. Consistent with our findings, EJC subunits have recently been identified as regulators of the piRNA pathway in genome-wide screens in *Drosophila* (Czech et al. 2013; Handler et al. 2013; Muerdter et al. 2013). We now demonstrate that the EJC controls the piRNA pathway by facilitating the splicing of a poorly

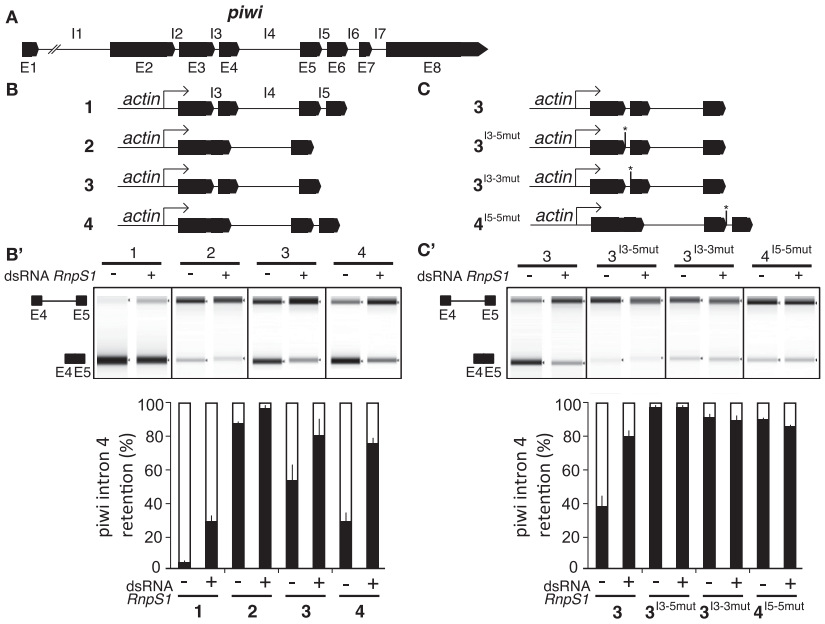

Figure 6. The EJC requires neighboring introns to support efficient splicing of the *piwi* fourth intron. (*A*) Schematic of the *piwi* genomic locus, with introns (I) and exons (E) marked accordingly. (*B*,*C*) Various minigene constructs (1–4) containing the *piwi* fourth intron with or without flanking introns or splice site disruptions (asterisk). (*B*′,*C*′, *top*) Capillary gels of RT–PCR with primers flanking intron 4 using extracts from control or *RnpS1*-depleted *Drosophila* S2R$^+$ cells. (*Bottom*) Graphs showing the ratio of spliced (white) versus unspliced (black) fourth intron.

structured intron within the *piwi* pre-mRNA. In the absence of the EJC, the unspliced *piwi* transcript is unable to support sufficient Piwi protein production to maintain transposon regulation in the gonads. This indicates an additional role for splicing factors in the piRNA pathway, as it was previously shown that the splicing and export factor UAP56 is implicated in the processing of dual-strand piRNA cluster transcripts (Zhang et al. 2012). The EJC may also regulate AGO3, as we observed a moderate decrease in transcript levels on *RnpS1* knockdown (Supplemental Fig. S5), an effect on AGO3 localization in *mago* mutants (Supplemental Fig. S2E), and partial intron retention (Supplemental Fig. S6). However, in both cases, we still detected piRNA ping-pong activity, indicating that AGO3 maintains enzymatic function upon depletion of the EJC.

A novel function for the pre-EJC and splicing factor RnpS1

The EJC is recruited to mRNAs concomitant with splicing and remains strongly associated following export to the cytoplasm, where it can influence their subsequent fates. Several studies have recently challenged the view that the EJC serves only as a memory of the nuclear history of mRNAs and provided evidence for its involvement in the splicing process itself (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010; Michelle et al. 2012). However, since EJC subunits are recruited only at a late step of the splicing reaction, just prior to exon ligation, it remained unclear how they could influence this process. Here we provide a mechanism by which the EJC regulates the splicing of the *piwi* transcript. We found that excision of the weak intron in the *piwi* transcript (intron 4) is facilitated by the splicing of strong adjacent introns, likely via the deposition of the EJC at neighboring splice junctions. In addition to the core pre-EJC, the EJC splicing subunit RnpS1 appears to be necessary to perform this function.

In vivo cooperation between introns for their processing has been previously observed, but the mechanism has remained unclear (Neel et al. 1993; Nesic and Maquat 1994; Romano et al. 2001). For instance, introns of the *tumor necrosis factor β* (*TNFβ*) transcript are spliced more efficiently in HeLa cells when upstream introns are still present (Neel et al. 1993). Similarly, upstream introns influence the efficiency of removal of the final intron of the human *triosephosphate isomerase* (*TPI*) transcript (Nesic and Maquat 1994). Importantly, in both cases, mutating the splice sites of the upstream introns abolishes their enhancing effect, indicating that their splicing per se is important. It has been proposed that the presence or absence of upstream introns can confer different secondary transcript structures, potentially influencing the efficiency of the splicing reaction (Tomizawa and Itoh 1981; Wong and Polisky 1985; Watakabe et al. 1989; Clouet d'Orval et al. 1991). In light of our results, we favor the hypothesis that deposition of the EJC at adjacent splice junctions could explain the positive influence of upstream introns.

Although some studies provided evidence that intron excision follows a 5′-to-3′ polarity established during transcription (Wetterberg et al. 1996; Pandya-Jones and Black 2009), our data clearly show that this is not a strict rule. In the case of the *piwi* transcript, not only does third intron splicing influence the removal of the fourth intron, but the splicing of intron 5 appears absolutely required for efficient splicing of the upstream intron (4). These results are consistent with recent genome-wide studies in *Drosophila* demonstrating that first introns are generally less efficiently cotranscriptionally spliced than subsequent introns (Khodor et al. 2011). As the excision of these introns is kinetically delayed, it would be interesting to test whether the EJC also facilitates their processing.

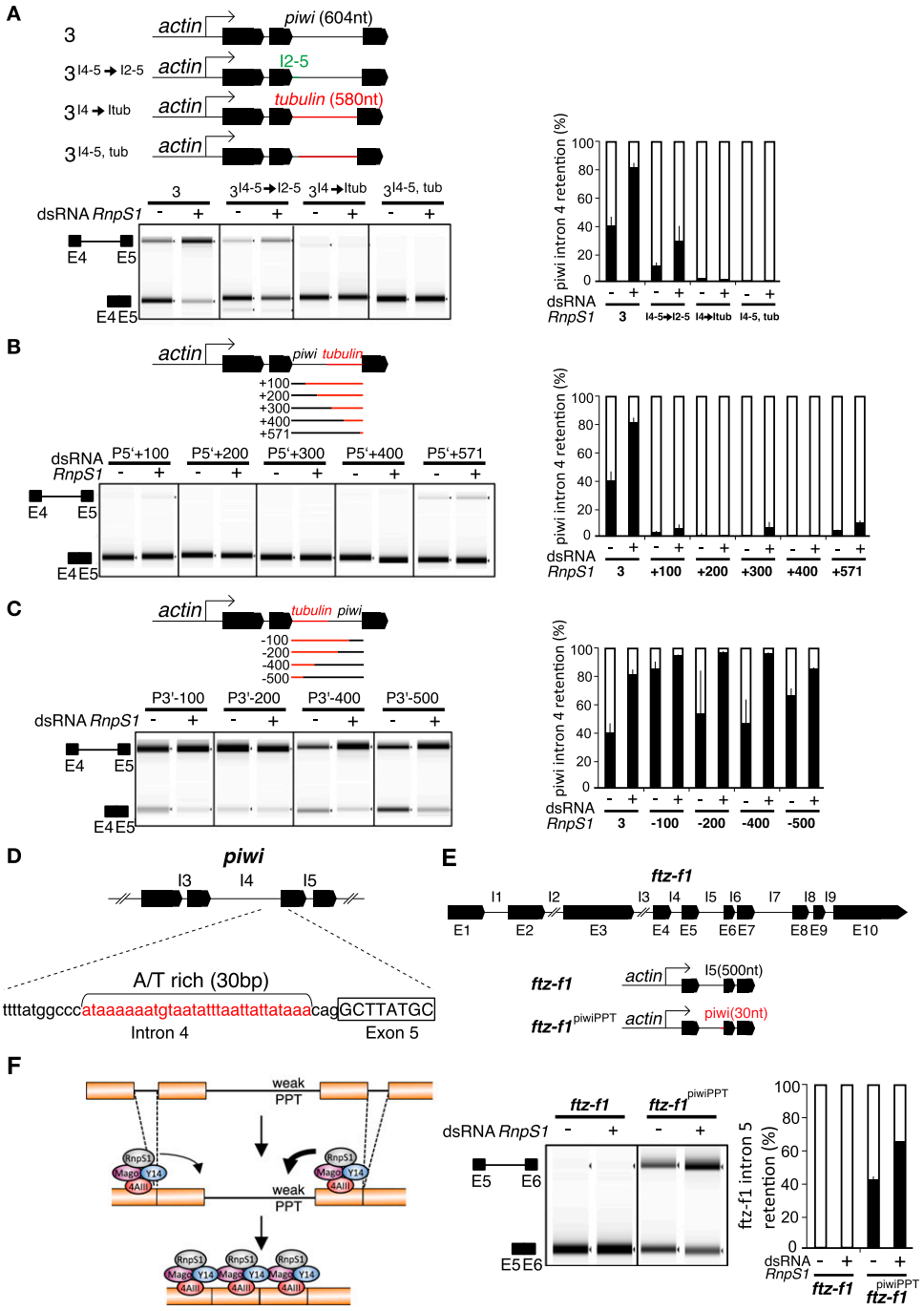

Figure 7. An abnormal PPT in the fourth intron of *piwi* confers sensitivity to the EJC. (*A–C*) Diagrams of minigene constructs replacing the *piwi* fourth intron 5′ss with that of the *piwi* second intron (green) or portions of the fourth intron with the *tubulin* intron (red). Capillary gels and graphical depictions of qRT–PCR of minigene constructs with and without *RnpS1* knockdown in *Drosophila* S2R$^+$ cells showing the ratio of spliced (white) versus unspliced (black) fourth intron. (*D*) Portrayal of the A/T-rich region (red) adjacent to the *piwi* fourth intron splice acceptor. (*E*) Minigene construct diagrams, capillary gels, and graphical depictions of splicing of the *ftz-f1* fifth intron locus. Both the wild-type intron 5 and one swapped with the *piwi* PPT (red) were tested. (*F*) Model showing the pre-EJC and RnpS1 acting at flanking exon junctions to rescue inefficient splicing of the *piwi* fourth intron.

In a previous study, we had shown that many genes expressed from heterochromatic loci, including *MAPK,* are dependent on the pre-EJC for their splicing (Roignant and Treisman 2010). These genes are embedded in a highly compact chromatin structure and often contain remarkably long introns occupied by a high density of repetitive elements. An accompanying study presented evidence that many transcripts containing large introns exhibit exon-skipping events upon pre-EJC knockdown (Ashton-Beaucage et al. 2010). This is in stark contrast to *piwi*, which

is expressed from a euchromatic domain, contains relatively small introns, and shows intron retention in the absence of the pre-EJC. Despite these differences, we found that its first and fourth introns, which are both dependent on the EJC for their faithful splicing, contain transposable element fragments. While these fragments might influence the efficiency of splicing by stretching the size of the intron, their presence is not necessary to generate EJC dependence. We identified 41 euchromatic introns of a size similar to the piwi fourth and containing transposable elements. However, those introns do not require RnpS1 for their splicing (Supplemental Fig. S9). In contrast, this dependence is due to a small region within the retained intron that contains a highly degenerate PPT (Fig. 7D). Replacing this region with a canonical PPT from the *tubulin* intron completely rescues splicing and abolishes EJC dependence. Conversely, replacing the canonical PPT from the *ftz-f1* intron with the degenerate PPT from the *piwi* intron is sufficient to confer EJC dependence (Fig. 7E). Therefore, our results support a model in which deposition of the EJC at adjacent splice junctions brings RnpS1 in proximity to the weak intron such that it facilitates spliceosome function by circumventing the absence of a canonical PPT (Fig. 7F).

RnpS1 forms two mutually exclusive subcomplexes with additional EJC subunits—one termed apoptosis and splicing-associated protein (ASAP), and the other termed PSAP (Schwerk et al. 2003; Murachelli et al. 2012). They share the SAP18 and differ by their association with an apoptotic chromatin inducer in the nucleus (Acinus) for ASAP and Pinin for PSAP. Several functions have been attributed to these complexes in transcription, programmed cell death, and mRNA processing (Zhang et al. 1997; Mayeda et al. 1999; Sahara et al. 1999; Li et al. 2003; Schwerk et al. 2003; Sakashita et al. 2004; Joselin et al. 2006; Vucetic et al. 2008; Singh et al. 2010). Whether all of these functions are shared with EJC core components is not known, nor is the underlying mechanism dictating their specificity. Intriguingly, we observed that *Acinus* depletion, similarly to *RnpS1* knockdown, strongly affects *piwi* splicing, while neither SAP18 nor Pinin show a detectable requirement (Fig. 3A,B). These results indicate a functional specificity not only between the two complexes but also within the ASAP complex.

To summarize, we describe a role for the EJC in maintaining genome integrity by preventing transposon mobilization in the *Drosophila* germline. This is accomplished through the regulation of *piwi* transcript splicing, uncovering a novel mechanism of EJC-based rescue of inefficient splicing. Given the broad conservation of the EJC and the extent to which it can influence splicing events, we envision an extensive role that will continue to be uncovered.

Materials and methods

Drosophila *stocks and genetics*

All experiments were performed at 25°C on standard medium. When applicable, w^{1118} flies served as controls. The fly stocks used are listed in Supplemental Table S2. $mago^{93D}$ and $tsu^{\Delta 18}$ mutant clones in the ovaries were generated by crossing FRT42D, $mago^{93D}$ or FRT42D, $tsu^{\Delta 18}$ males to FRT42D, ubi-GFP; hsFLP122 females. Larvae were heat-shocked for 1 h at 38.5°C in both the first and second instar.

Immunohistochemistry and Western blot analysis

Ovaries were dissected from 4- to 5-d-old flies into ice-cold PBS. Tissues were fixed in 5% FA in PBS for 20 min, permeabilized in 1% Triton X-100 in PBS, blocked in 0.2% Triton X-100 containing 1% BSA in PBS, stained with primary antibody, washed, stained with secondary antibody, washed, and mounted in VectaShield with DAPI (Vector Laboratories). Images were captured using either a TCS-SP5 (Leica) or LSM 510 (Zeiss) confocal microscope.

Western blots were performed as previously described (Miura et al. 2006). Protein extracts were generated from ovaries lysed in ice-cold lysis buffer (50 mM Tris at pH 8, 150 mM NaCl, 1% Triton X-100, complete protease inhibitor cocktail [Roche], 5 mM EDTA, 5 mM NaF, 1 mM Na_3VO_4, 0.1% SDS, 0.5% sodium deoxycholate).

Primary antibodies used are listed in Supplemental Table S2. Secondary antibodies were coupled to Alexa 488 (1:1000), Cy3 (1:200), Cy5 (1:500), and Alexa 647 (1:500) (Jackson ImmunoResearch).

Measurement of RNA levels

Total RNA was extracted from cells or ovaries using TRIzol (Life Technologies) and treated as described (Roignant and Treisman 2010). For qRT–PCR analysis, cDNA was generated and then assayed using a ViiA7 real-time PCR system (Applied Biosystems). Primer sequences are listed in Supplemental Table S3. For semi-quantitative PCR analysis, cDNA was amplified using Phusion high-fidelity DNA polymerase (New England BioLabs).

Mutagenesis

Point mutation-containing primers were designed in sense and antisense orientations to generate mutant PCR fragments and were subsequently phosphorylated using T4 polynucleotide kinase (New England BioLabs). Mutant fragments were cloned into TOPO vectors (Life Technologies) and then amplified using Phusion high-fidelity DNA polymerase. Fragments were then digested with DpnI (New England BioLabs) and transformed into DH5α-competent cells (Life Technologies).

RNA FISH

RNA FISH was performed using Stellaris probes as previously described (Raj et al. 2008) with slight modification. Forty-eight 20-nt Quasar 670-labeled probes were designed against the consensus ZAM retrotransposon, excluding LTR sequences.

Cell culture, RNAi, and transfection

$S2R^+$ cells were maintained in Schneider's medium supplemented with 10% fetal calf serum. dsRNAs were generated using the MEGAscript T7 (Life Technologies). $S2R^+$ cells were treated with 15 μg of dsRNA, transfected 3 d afterward using Effectene (Qiagen), and retreated with 15 μg of dsRNA. RNA and protein extractions were performed after 7 d in TRIzol or lysis buffer, respectively.

RNA cloning and sequencing

Total RNA from control and *mago* and *RnpS1* GLKD and SKD ovaries were used to clone small RNA libraries as previously described (Brennecke et al. 2007). Additionally, 5′ and 3′ cloning adapters were modified with five randomized nucleotides flank-

ing the small RNA to minimize ligation biases (Jayaprakash et al. 2011).

Strand-specific transcript profiling was performed on 2 μg of ribosomal RNA-depleted total RNA from *RpnS1* knockdown (GLKD and SKD), $piwi^1/piwi^2$, or control ovaries. All transcript and small RNA-seq data has been deposited at Gene Expression Omnibus (accession nos. GSE57710 and GSE59327).

Computational analysis

Raw small RNA sequences were demultiplexed, and randomized nucleotides were removed by trimming the first 5 nt as well as the last five before the sequenced cloning adapter. Reads were then mapped to the *Drosophila* genome and annotated features, including miRNAs and transposons, as previously described (Brennecke et al. 2007). Ping-Pong analysis was performed as previously described (Brennecke et al. 2008). SKD and GLKD libraries were normalized to reads derived from the 42AB or flamenco piRNA clusters, respectively (Preall et al. 2012).

Raw RNA-seq data were demultiplexed, and the first 7 nt were trimmed because of low sequencing quality. Reads were then mapped against the *Drosophila* genome BDGP5 (ensemble release 73). TopHat2 (Kim et al. 2013) and bowtie2 (Langmead and Salzberg 2012) were used to map, allowing two mismatches. Tags were assigned to features using HTSeq-count (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html), and differential gene expression was determined using DESeq (Anders and Huber 2010). Differentially expressed genes were filtered using false discovery rate (FDR) 0.1.

Differential intron expression was determined using DEXSeq (Anders et al. 2012), version 1.8.0. Read counting was performed using HTSeq-count 0.5.3p9 (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html) with custom Python scripts provided by the DEXSeq.

For transposon analysis, reads were mapped against the transposon consensus reference (FlyBase release FB2013.05) using bowtie2, and multimapping reads were excluded. Differential expression analysis was performed using DESeq and determined using FDR 0.1, and only transposons with average coverage greater than one read per kilobase per million (RPKM) and a fold change greater than two were considered.

Acknowledgments

We thank members of the Roignant laboratory for helpful discussions, Felipe Karam Teixeira and Tatjana Treck for assay design assistance, and Anna Lena Leifke for technical support. We thank Institute of Molecular Biology facilities for their support, laboratory members and Rene Ketting for critical reading of the manuscript, the Bloomington *Drosophila* Stock Center and the Vienna *Drosophila* Resource Center for fly strains, and the Developmental Studies Hybridoma Bank for antibodies. Part of this work was performed in the laboratory of Dr. Ruth Lehmann, with support from the Howard Hughes Medical Institute. C.D.M. is supported by a post-doctoral fellowship from the Helen Hay Whitney Foundation. This project was funded in part by grants CIG 334288 (Marie Curie) to J.-Y.R, 1R21HG007394-01 (National Institutes of Health) to R.S., and MCB 1051022 (National Science Foundation) and EY013777 (National Institutes of Health) to J.T.

References

Alexandrov A, Colognori D, Shu MD, Steitz JA. 2012. Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. *Proc Natl Acad Sci* **109:** 21313–21318.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11:** R106.

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

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

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

Baralle D, Lucassen A, Buratti E. 2009. Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep* **10:** 810–816.

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

Bono F, Gehring NH. 2011. Assembly, disassembly and recycling: the dynamics of exon junction complexes. *RNA Biol* **8:** 24–29.

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

Braunschweig U, Gueroussov S, Plocik AM, Graveley BR, Blencowe BJ. 2013. Dynamic integration of splicing within gene regulatory pathways. *Cell* **152:** 1252–1269.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128:** 1089–1103.

Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322:** 1387–1392.

Carmell MA, Xuan Z, Zhang MQ, Hannon GJ. 2002. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev* **16:** 2733–2742.

Clouet d'Orval B, d'Aubenton Carafa Y, Sirand-Pugnet P, Gallego M, Brody E, Marie J. 1991. RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts. *Science* **252:** 1823–1828.

Cox DN, Chao A, Baker J, Chang L, Qiao D, Lin H. 1998. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev* **12:** 3715–3727.

Cox DN, Chao A, Lin H. 2000. piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127:** 503–514.

Czech B, Preall JB, McGinn J, Hannon GJ. 2013. A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol Cell* **50:** 749–761.

David CJ, Manley JL. 2010. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev* **24:** 2343–2364.

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

Gatfield D, Izaurralde E. 2002. REF1/Aly and the additional exon junction complex proteins are dispensable for nuclear mRNA export. *J Cell Biol* **159:** 579–588.

Gehring NH, Lamprinaki S, Hentze MW, Kulozik AE. 2009. The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. *PLoS Biol* **7:** e1000120.

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* **315:** 1587–1590.

Haase AD, Fenoglio S, Muerdter F, Guzzardo PM, Czech B, Pappin DJ, Chen C, Gordon A, Hannon GJ. 2010. Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*. *Genes Dev* **24:** 2499–2504.

Hachet O, Ephrussi A. 2001. *Drosophila* Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. *Curr Biol* **11:** 1666–1674.

Handler D, Meixner K, Pizka M, Lauss K, Schmied C, Gruber FS, Brennecke J. 2013. The genetic makeup of the *Drosophila* piRNA pathway. *Mol Cell* **50:** 762–777.

Haremaki T, Weinstein DC. 2012. Eif4a3 is required for accurate splicing of the *Xenopus laevis* ryanodine receptor pre-mRNA. *Dev Biol* **372:** 103–110.

Harris AN, Macdonald PM. 2001. Aubergine encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. *Development* **128:** 2823–2832.

Iannone C, Valcarcel J. 2013. Chromatin's thread to alternative splicing regulation. *Chromosoma* **122:** 465–474.

Jayaprakash AD, Jabado O, Brown BD, Sachidanandam R. 2011. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res* **39:** e141.

Jin Z, Flynt AS, Lai EC. 2013. *Drosophila* piwi mutants exhibit germline stem cell tumors that are sustained by elevated Dpp signaling. *Curr Biol* **23:** 1442–1448.

Joselin AP, Schulze-Osthoff K, Schwerk C. 2006. Loss of Acinus inhibits oligonucleosomal DNA fragmentation but not chromatin condensation during apoptosis. *J Biol Chem* **281:** 12475–12484.

Khodor YL, Rodriguez J, Abruzzi KC, Tang CH, Marr MT 2nd, Rosbash M. 2011. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev* **25:** 2502–2512.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14:** R36.

Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA, Theurkauf WE. 2007. *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev Cell* **12:** 45–55.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359.

Le Hir H, Izaurralde E, Maquat LE, Moore MJ. 2000. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon–exon junctions. *EMBO J* **19:** 6860–6869.

Le Hir H, Gatfield D, Izaurralde E, Moore MJ. 2001. The exon–exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J* **20:** 4987–4997.

Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Toth KF. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev* **27:** 390–399.

Li C, Lin RI, Lai MC, Ouyang P, Tarn WY. 2003. Nuclear Pnn/DRS protein binds to spliced mRNPs and participates in mRNA processing and export via interaction with RNPS1. *Mol Cell Biol* **23:** 7363–7376.

Lin H, Spradling AC. 1997. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* **124:** 2463–2476.

Long JC, Caceres JF. 2009. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* **417:** 15–27.

Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. 2011. Epigenetics in alternative pre-mRNA splicing. *Cell* **144:** 16–26.

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* **137:** 522–535.

Mayeda A, Badolato J, Kobayashi R, Zhang MQ, Gardiner EM, Krainer AR. 1999. Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. *EMBO J* **18:** 4560–4570.

Megosh HB, Cox DN, Campbell C, Lin H. 2006. The role of PIWI and the miRNA machinery in *Drosophila* germline determination. *Curr Biol* **16:** 1884-1894.

Michelle L, Cloutier A, Toutant J, Shkreta L, Thibault P, Durand M, Garneau D, Gendron D, Lapointe E, Couture S, et al. 2012. Proteins associated with the exon junction complex also control the alternative splicing of apoptotic regulators. *Mol Cell Biol* **32:** 954–967.

Micklem DR, Dasgupta R, Elliott H, Gergely F, Davidson C, Brand A, Gonzalez-Reyes A, St Johnston D. 1997. The mago nashi gene is required for the polarisation of the oocyte and the formation of perpendicular axes in *Drosophila*. *Curr Biol* **7:** 468–478.

Miura GI, Buglino J, Alvarado D, Lemmon MA, Resh MD, Treisman JE. 2006. Palmitoylation of the EGFR ligand Spitz by Rasp increases Spitz activity by restricting its diffusion. *Dev Cell* **10:** 167–176.

Mohr SE, Dillon ST, Boswell RE. 2001. The RNA-binding protein Tsunagi interacts with Mago Nashi to establish polarity and localize oskar mRNA during *Drosophila* oogenesis. *Genes Dev* **15:** 2886–2899.

Muerdter F, Guzzardo PM, Gillis J, Luo Y, Yu Y, Chen C, Fekete R, Hannon GJ. 2013. A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in *Drosophila*. *Mol Cell* **50:** 736–748.

Murachelli AG, Ebert J, Basquin C, Le Hir H, Conti E. 2012. The structure of the ASAP core complex reveals the existence of a Pinin-containing PSAP complex. *Nat Struct Mol Biol* **19:** 378–386.

Nagao A, Mituyama T, Huang H, Chen D, Siomi MC, Siomi H. 2010. Biogenesis pathways of piRNAs loaded onto AGO3 in the *Drosophila* testis. *RNA* **16:** 2503–2515.

Neel H, Weil D, Giansante C, Dautry F. 1993. In vivo cooperation between introns during pre-mRNA processing. *Genes Dev* **7:** 2194–2205.

Nesic D, Maquat LE. 1994. Upstream introns influence the efficiency of final intron removal and RNA 3′-end formation. *Genes Dev* **8:** 363–375.

Newmark PA, Mohr SE, Gong L, Boswell RE. 1997. mago nashi mediates the posterior follicle cell-to-oocyte signal to organize axis formation in *Drosophila*. *Development* **124:** 3197–3207.

Nishida KM, Saito K, Mori T, Kawamura Y, Nagami-Okada T, Inagaki S, Siomi H, Siomi MC. 2007. Gene silencing mechanisms mediated by Aubergine piRNA complexes in *Drosophila* male gonad. *RNA* **13:** 1911–1922.

Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J. 2010. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J* **29:** 3301–3317.

Olivieri D, Senti KA, Subramanian S, Sachidanandam R, Brennecke J. 2012. The cochaperone shutdown defines a group

of biogenesis factors essential for all piRNA populations in *Drosophila*. *Mol Cell* **47:** 954–969.
Orengo JP, Cooper TA. 2007. Alternative splicing in disease. *Adv Exp Med Biol* **623:** 212–223.
Pandya-Jones A, Black DL. 2009. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15:** 1896–1908.
Parma DH, Bennett PE Jr, Boswell RE. 2007. Mago Nashi and Tsunagi/Y14, respectively, regulate *Drosophila* germline stem cell differentiation and oocyte specification. *Dev Biol* **308:** 507–519.
Preall JB, Czech B, Guzzardo PM, Muerdter F, Hannon GJ. 2012. shutdown is a component of the *Drosophila* piRNA biogenesis machinery. *RNA* **18:** 1446–1457.
Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. 2008. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* **5:** 877–879.
Roignant JY, Treisman JE. 2010. Exon junction complex subunits are required to splice *Drosophila* MAP kinase, a large heterochromatic gene. *Cell* **143:** 238–250.
Romano M, Marcucci R, Baralle FE. 2001. Splicing of constitutive upstream introns is essential for the recognition of intra-exonic suboptimal splice sites in the thrombopoietin gene. *Nucleic Acids Res* **29:** 886–894.
Rozhkov NV, Hammell M, Hannon GJ. 2013. Multiple roles for Piwi in silencing *Drosophila* transposons. *Genes Dev* **27:** 400–412.
Sahara S, Aoto M, Eguchi Y, Imamoto N, Yoneda Y, Tsujimoto Y. 1999. Acinus is a caspase-3-activated protein required for apoptotic chromatin condensation. *Nature* **401:** 168–173.
Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev* **24:** 2493–2498.
Sakashita E, Tatsumi S, Werner D, Endo H, Mayeda A. 2004. Human RNPS1 and its associated factors: a versatile alternative pre-mRNA splicing regulator in vivo. *Mol Cell Biol* **24:** 1174–1187.
Sarot E, Payen-Groschene G, Bucheton A, Pelisson A. 2004. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster* flamenco gene. *Genetics* **166:** 1313–1321.
Sauliere J, Murigneux V, Wang Z, Marquenet E, Barbosa I, Le Tonqueze O, Audic Y, Paillard L, Roest Crollius H, Le Hir H. 2012. CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat Struct Mol Biol* **19:** 1124–1131.
Schwerk C, Prasad J, Degenhardt K, Erdjument-Bromage H, White E, Tempst P, Kidd VJ, Manley JL, Lahti JM, Reinberg D. 2003. ASAP, a novel protein complex involved in RNA processing and apoptosis. *Mol Cell Biol* **23:** 2981–2990.
Senti KA, Brennecke J. 2010. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet* **26:** 499–509.
Shibuya T, Tange TO, Sonenberg N, Moore MJ. 2004. eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat Struct Mol Biol* **11:** 346–351.
Shiimori M, Inoue K, Sakamoto H. 2013. A specific set of exon junction complex subunits is required for the nuclear retention of unspliced RNAs in *Caenorhabditis elegans*. *Mol Cell Biol* **33:** 444–456.
Sienski G, Donertas D, Brennecke J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151:** 964–980.
Silver DL, Watkins-Chow DE, Schreck KC, Pierfelice TJ, Larson DM, Burnetti AJ, Liaw HJ, Myung K, Walsh CA, Gaiano N, et al. 2010. The exon junction complex component Magoh controls brain size by regulating neural stem cell division. *Nat Neurosci* **13:** 551–558.
Silver DL, Leeds KE, Hwang HW, Miller EE, Pavan WJ. 2013. The EJC component Magoh regulates proliferation and expansion of neural crest-derived melanocytes. *Dev Biol* **375:** 172–181.
Singh KK, Erkelenz S, Rattay S, Dehof AK, Hildebrandt A, Schulze-Osthoff K, Schaal H, Schwerk C. 2010. Human SAP18 mediates assembly of a splicing regulatory multiprotein complex via its ubiquitin-like fold. *RNA* **16:** 2442–2454.
Singh G, Kucukural A, Cenik C, Leszyk JD, Shaffer SA, Weng Z, Moore MJ. 2012. The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* **151:** 750–764.
Srebrow A, Kornblihtt AR. 2006. The connection between splicing and cancer. *J Cell Sci* **119:** 2635–2641.
Steckelberg AL, Boehm V, Gromadzka AM, Gehring NH. 2012. CWC22 connects pre-mRNA splicing and exon junction complex assembly. *Cell Reports* **2:** 454–461.
Tange TO, Shibuya T, Jurica MS, Moore MJ. 2005. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* **11:** 1869–1883.
Theurkauf WE, Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA. 2006. rasiRNAs, DNA damage, and embryonic axis specification. *Cold Spring Harb Symp Quant Biol* **71:** 171–180.
Tomizawa J, Itoh T. 1981. Plasmid ColE1 incompatibility determined by interaction of RNA I with primer transcript. *Proc Natl Acad Sci* **78:** 6096–6100.
Vucetic Z, Zhang Z, Zhao J, Wang F, Soprano KJ, Soprano DR. 2008. Acinus-S′ represses retinoic acid receptor (RAR)-regulated gene expression through interaction with the B domains of RARs. *Mol Cell Biol* **28:** 2549–2558.
Watakabe A, Inoue K, Sakamoto H, Shimura Y. 1989. A secondary structure at the 3′ splice site affects the in vitro splicing reaction of mouse immunoglobulin μ chain pre-mRNAs. *Nucleic Acids Res* **17:** 8159–8169.
Wetterberg I, Bauren G, Wieslander L. 1996. The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. *RNA* **2:** 641–651.
Wilson JE, Connell JE, Macdonald PM. 1996. aubergine enhances oskar translation in the *Drosophila* ovary. *Development* **122:** 1631–1639.
Wong EM, Polisky B. 1985. Alternative conformations of the ColE1 replication primer modulate its interaction with RNA I. *Cell* **42:** 959–966.
Zhang Y, Iratni R, Erdjument-Bromage H, Tempst P, Reinberg D. 1997. Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell* **89:** 357–364.
Zhang F, Wang J, Xu J, Zhang Z, Koppetsch BS, Schultz N, Vreven T, Meignin C, Davis I, Zamore PD, et al. 2012. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* **151:** 871–884.

4. BIBLIOGRAPHY

1. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 3171–3175 (1977).
2. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12,** 1–8 (1977).
3. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. U. S. A.* **75,** 4853–4857 (1978).
4. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136,** 701–718 (2009).
5. Steitz, J. A. Snurps. *Sci. Am.* **258,** 56–60, 63 (1988).
6. Patel, S. B. & Bellini, M. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res.* **36,** 6482–6493 (2008).
7. Du, H. & Rosbash, M. The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature* **419,** 86–90 (2002).
8. Mount, S. M. & Steitz, J. A. Sequence of U1 RNA from Drosophila melanogaster: implications for U1 secondary structure and possible involvement in splicing. *Nucleic Acids Res.* **9,** 6351–6368 (1981).
9. Branlant, C. *et al.* The conformation of chicken, rat and human U1A RNAs in solution. *Nucleic Acids Res.* **9,** 841–858 (1981).
10. Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A. & Steitz, J. A. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. *Cell* **33,** 509–518 (1983).
11. Brosi, R., Gröning, K., Behrens, S. E., Lührmann, R. & Krämer, A. Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science* **262,** 102–105 (1993).
12. Behrens, S. E. & Lührmann, R. Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes Dev.* **5,** 1439–1452 (1991).
13. Zaric, B. *et al.* Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. *J. Biol. Chem.* **280,** 16066–16075 (2005).
14. Bell, M., Schreiner, S., Damianov, A., Reddy, R. & Bindereif, A. p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *EMBO J.* **21,** 2724–2735 (2002).
15. Rader, S. D. & Guthrie, C. A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. *RNA N. Y. N* **8,** 1378–1392 (2002).
16. Bach, M., Winkelmann, G. & Lührmann, R. 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 6038–6042 (1989).
17. Patterson, B. & Guthrie, C. An essential yeast snRNA with a U5-like domain is required for splicing in vivo. *Cell* **49,** 613–624 (1987).
18. Winkelmann, G., Bach, M. & Lührmann, R. Evidence from complementation assays in vitro that U5 snRNP is required for both steps of mRNA splicing. *EMBO J.* **8,** 3105–3112 (1989).
19. Fetzer, S., Lauber, J., Will, C. L. & Lührmann, R. The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that can be phosphorylated by the snRNP-associated protein kinase. *RNA N. Y. N* **3,** 344–355 (1997).
20. Jurica, M. S. & Moore, M. J. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12,** 5–14 (2003).
21. Ruskin, B., Zamore, P. D. & Green, M. R. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52,** 207–219 (1988).
22. Zamore, P. D. & Green, M. R. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 9243–9247 (1989).
23. Wu, S., Romfo, C. M., Nilsen, T. W. & Green, M. R. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402,** 832–835 (1999).

24. Makarova, O. V. *et al.* A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J.* **23,** 2381–2391 (2004).

25. Song, E. J. *et al.* The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. *Genes Dev.* **24,** 1434–1447 (2010).

26. Chanarat, S. & Sträßer, K. Splicing and beyond: the many faces of the Prp19 complex. *Biochim. Biophys. Acta* **1833,** 2126–2134 (2013).

27. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3,** (2011).

28. Chen, W. & Moore, M. J. The spliceosome: disorder and dynamics defined. *Curr. Opin. Struct. Biol.* **24,** 141–149 (2014).

29. Reed, R. Protein composition of mammalian spliceosomes assembled in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 8031–8035 (1990).

30. Bennett, M., Piñol-Roma, S., Staknis, D., Dreyfuss, G. & Reed, R. Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Mol. Cell. Biol.* **12,** 3165–3175 (1992).

31. Gozani, O., Feld, R. & Reed, R. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev.* **10,** 233–243 (1996).

32. Valcárcel, J., Gaur, R. K., Singh, R. & Green, M. R. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science* **273,** 1706–1709 (1996).

33. Makarov, E. M. *et al.* Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298,** 2205–2208 (2002).

34. Konarska, M. M., Vilardell, J. & Query, C. C. Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol. Cell* **21,** 543–553 (2006).

35. Marais, G., Nouvellet, P., Keightley, P. D. & Charlesworth, B. Intron size and exon evolution in Drosophila. *Genetics* **170,** 481–485 (2005).

36. Bartolomé, C., Maside, X. & Charlesworth, B. On the abundance and distribution of transposable elements in the genome of Drosophila melanogaster. *Mol. Biol. Evol.* **19,** 926–937 (2002).

37. Mattick, J. S. & Gagen, M. J. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.* **18,** 1611–1630 (2001).

38. Berget, S. M. Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270,** 2411–2414 (1995).

39. Fox-Walsh, K. L. *et al.* The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 16176–16181 (2005).

40. Jackson, I. J. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* **19,** 3795–3798 (1991).

41. Hall, S. L. & Padgett, R. A. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.* **239,** 357–365 (1994).

42. Patel, A. A. & Steitz, J. A. Splicing double: insights from the second spliceosome. *Nat. Rev. Mol. Cell Biol.* **4,** 960–970 (2003).

43. Kelemen, O. *et al.* Function of alternative splicing. *Gene* **514,** 1–30 (2013).

44. Schaal, T. D. & Maniatis, T. Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol. Cell. Biol.* **19,** 261–273 (1999).

45. Schaal, T. D. & Maniatis, T. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.* **19,** 1705–1719 (1999).

46. Matlin, A. J., Clark, F. & Smith, C. W. J. Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6,** 386–398 (2005).

47. Pozzoli, U. & Sironi, M. Silencers regulate both constitutive and alternative splicing events in mammals. *Cell. Mol. Life Sci. CMLS* **62,** 1579–1604 (2005).

48. Boukis, L. A. & Bruzik, J. P. Functional selection of splicing enhancers that stimulate trans-splicing in vitro. *RNA N. Y. N* **7,** 793–805 (2001).

49. Tacke, R. & Manley, J. L. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* **14,** 3540–3551 (1995).

50. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456,** 470–476 (2008).

51. Nussinov, R. Conserved signals around the 5' splice sites in eukaryotic nuclear precursor mRNAs: G-runs are frequent in the introns and C in the exons near both 5' and 3' splice sites. *J. Biomol. Struct. Dyn.* **6,** 985–1000 (1989).

52. McCullough, A. J. & Berget, S. M. G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. *Mol. Cell. Biol.* **17,** 4562–4571 (1997).

53. Chen, Y. *et al.* RNA secondary structure and compensatory evolution. *Genes Genet. Syst.* **74,** 271–286 (1999).

54. Graveley, B. R. Sorting out the complexity of SR protein functions. *RNA N. Y. N* **6,** 1197–1211 (2000).

55. Wu, J. Y. & Maniatis, T. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75,** 1061–1070 (1993).

56. Makarova, O. V., Makarov, E. M. & Lührmann, R. The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *EMBO J.* **20,** 2553–2563 (2001).

57. Cartegni, L. & Krainer, A. R. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat. Genet.* **30,** 377–384 (2002).

58. Buratti, E. & Baralle, F. E. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell. Biol.* **24,** 10505–10514 (2004).

59. Shepard, P. J. & Hertel, K. J. Conserved RNA secondary structures promote alternative splicing. *RNA N. Y. N* **14,** 1463–1469 (2008).

60. Beyer, A. L. & Osheim, Y. N. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* **2,** 754–765 (1988).

61. Martin, R. M., Rino, J., Carvalho, C., Kirchhausen, T. & Carmo-Fonseca, M. Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell Rep.* **4,** 1144–1155 (2013).

62. Ameur, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol.* **18,** 1435–1440 (2011).

63. Khodor, Y. L. *et al.* Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev.* **25,** 2502–2512 (2011).

64. Khodor, Y. L., Menet, J. S., Tolan, M. & Rosbash, M. Cotranscriptional splicing efficiency differs dramatically between Drosophila and mouse. *RNA N. Y. N* **18,** 2174–2186 (2012).

65. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15,** 163–175 (2014).

66. Cramer, P., Pesce, C. G., Baralle, F. E. & Kornblihtt, A. R. Functional association between promoter structure and transcript alternative splicing. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 11456–11460 (1997).

67. Cramer, P. *et al.* Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol. Cell* **4,** 251–258 (1999).

68. Pagani, F., Stuani, C., Zuccato, E., Kornblihtt, A. R. & Baralle, F. E. Promoter architecture modulates CFTR exon 9 skipping. *J. Biol. Chem.* **278,** 1511–1517 (2003).

69. Robson-Dixon, N. D. & Garcia-Blanco, M. A. MAZ elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon IIIb. *J. Biol. Chem.* **279,** 29075–29084 (2004).

70. Allemand, E., Batsché, E. & Muchardt, C. Splicing, transcription, and chromatin: a ménage à trois. *Curr. Opin. Genet. Dev.* **18,** 145–151 (2008).

71. De la Mata, M., Lafaille, C. & Kornblihtt, A. R. First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. *RNA N. Y. N* **16,** 904–912 (2010).

72. Ip, J. Y. *et al.* Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Res.* **21,** 390–401 (2011).

73. De la Mata, M. *et al.* A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell* **12,** 525–532 (2003).

74. Nogues, G., Kadener, S., Cramer, P., Bentley, D. & Kornblihtt, A. R. Transcriptional activators differ in their abilities to control alternative splicing. *J. Biol. Chem.* **277,** 43110–43114 (2002).

75. Dujardin, G. *et al.* How slow RNA polymerase II elongation favors alternative exon skipping. *Mol. Cell* **54,** 683–690 (2014).

76. Fong, N. *et al.* Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* **28,** 2663–2676 (2014).

77. Hsin, J.-P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* **26,** 2119–2137 (2012).

78. De la Mata, M. & Kornblihtt, A. R. RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat. Struct. Mol. Biol.* **13,** 973–980 (2006).

79. Gu, B., Eick, D. & Bensaude, O. CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Res.* **41,** 1591–1603 (2013).

80. Luco, R. F. *et al.* Regulation of alternative splicing by histone modifications. *Science* **327,** 996–1000 (2010).

81. Le Hir, H., Moore, M. J. & Maquat, L. E. Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev.* **14,** 1098–1108 (2000).

82. Le Hir, H., Izaurralde, E., Maquat, L. E. & Moore, M. J. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.* **19,** 6860–6869 (2000).

83. Saulière, J. *et al.* The exon junction complex differentially marks spliced junctions. *Nat. Struct. Mol. Biol.* **17,** 1269–1271 (2010).

84. Michelle, L. *et al.* Proteins associated with the exon junction complex also control the alternative splicing of apoptotic regulators. *Mol. Cell. Biol.* **32,** 954–967 (2012).

85. Malone, C. D. *et al.* The exon junction complex controls transposable element activity by ensuring faithful splicing of the piwi transcript. *Genes Dev.* **28,** 1786–1799 (2014).

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

87. Ashton-Beaucage, D. & Therrien, M. The exon junction complex: a splicing factor for long intron containing transcripts? *Fly (Austin)* **5,** 224–233 (2011).

88. Bono, F. & Gehring, N. H. Assembly, disassembly and recycling: the dynamics of exon junction complexes. *RNA Biol.* **8,** 24–29 (2011).

89. Shibuya, T., Tange, T. Ø., Sonenberg, N. & Moore, M. J. eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat. Struct. Mol. Biol.* **11,** 346–351 (2004).

90. Ballut, L. *et al.* The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat. Struct. Mol. Biol.* **12,** 861–869 (2005).

91. Tange, T. Ø., Shibuya, T., Jurica, M. S. & Moore, M. J. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA N. Y. N* **11,** 1869–1883 (2005).

92. Andersen, C. B. F. *et al.* Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* **313,** 1968–1972 (2006).

93. Chan, C. C. *et al.* eIF4A3 is a novel component of the exon junction complex. *RNA N. Y. N* **10,** 200–209 (2004).

94. Gehring, N. H. *et al.* Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol. Cell* **20,** 65–75 (2005).

95. Lorsch, J. R. & Herschlag, D. The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry (Mosc.)* **37,** 2180–2193 (1998).

96. Cordin, O., Banroques, J., Tanner, N. K. & Linder, P. The DEAD-box protein family of RNA helicases. *Gene* **367,** 17–37 (2006).

97. Svitkin, Y. V. *et al.* The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5′ secondary structure. *RNA N. Y. N* **7,** 382–394 (2001).

98. Pause, A. & Sonenberg, N. Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J.* **11,** 2643–2654 (1992).

99. Pause, A., Méthot, N. & Sonenberg, N. The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Mol. Cell. Biol.* **13,** 6789–6798 (1993).

100. Nielsen, K. H. *et al.* Mechanism of ATP turnover inhibition in the EJC. *RNA N. Y. N* **15,** 67–75 (2009).

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

102. Parma, D. H., Bennett, P. E. & Boswell, R. E. Mago Nashi and Tsunagi/Y14, respectively, regulate Drosophila germline stem cell differentiation and oocyte specification. *Dev. Biol.* **308,** 507–519 (2007).

103. Lewandowski, J. P., Sheehan, K. B., Bennett, P. E. & Boswell, R. E. Mago Nashi, Tsunagi/Y14, and Ranshi form a complex that influences oocyte differentiation in Drosophila melanogaster. *Dev. Biol.* **339,** 307–319 (2010).

104. Fribourg, S., Gatfield, D., Izaurralde, E. & Conti, E. A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat. Struct. Biol.* **10,** 433–439 (2003).

105. Gong, P., Zhao, M. & He, C. Slow co-evolution of the MAGO and Y14 protein families is required for the maintenance of their obligate heterodimerization mode. *PloS One* **9,** e84842 (2014).

106. Kataoka, N. *et al.* Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol. Cell* **6,** 673–682 (2000).

107. Le Hir, H., Gatfield, D., Braun, I. C., Forler, D. & Izaurralde, E. The protein Mago provides a link between splicing and mRNA localization. *EMBO Rep.* **2,** 1119–1124 (2001).

108. Le Hir, H., Gatfield, D., Izaurralde, E. & Moore, M. J. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20,** 4987–4997 (2001).

109. Dostie, J. & Dreyfuss, G. Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr. Biol. CB* **12,** 1060–1067 (2002).

110. Mingot, J. M., Kostka, S., Kraft, R., Hartmann, E. & Görlich, D. Importin 13: a novel mediator of nuclear import and export. *EMBO J.* **20,** 3685–3694 (2001).

111. Bono, F., Cook, A. G., Grünwald, M., Ebert, J. & Conti, E. Nuclear import mechanism of the EJC component Mago-Y14 revealed by structural studies of importin 13. *Mol. Cell* **37,** 211–222 (2010).

112. Kim, V. N. *et al.* The Y14 protein communicates to the cytoplasm the position of exon-exon junctions. *EMBO J.* **20,** 2062–2068 (2001).

113. Shi, H. & Xu, R.-M. Crystal structure of the Drosophila Mago nashi-Y14 complex. *Genes Dev.* **17,** 971–976 (2003).

114. Lau, C.-K., Diem, M. D., Dreyfuss, G. & Van Duyne, G. D. Structure of the Y14-Magoh core of the exon junction complex. *Curr. Biol. CB* **13,** 933–941 (2003).

115. Degot, S. *et al.* Metastatic Lymph Node 51, a novel nucleo-cytoplasmic protein overexpressed in breast cancer. *Oncogene* **21,** 4422–4434 (2002).

116. Luo, M. L. *et al.* Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* **413,** 644–647 (2001).

117. Stutz, F. & Izaurralde, E. The interplay of nuclear mRNP assembly, mRNA surveillance and export. *Trends Cell Biol.* **13,** 319–327 (2003).

118. Ishigaki, Y., Li, X., Serin, G. & Maquat, L. E. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* **106,** 607–617 (2001).

119. Lejeune, F., Ishigaki, Y., Li, X. & Maquat, L. E. The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J.* **21,** 3536–3545 (2002).

120. Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H. & Lührmann, R. Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452,** 846–850 (2008).

121. Mishler, D. M., Christ, A. B. & Steitz, J. A. Flexibility in the site of exon junction complex deposition revealed by functional group and RNA secondary structure alterations in the splicing substrate. *RNA N. Y. N* **14,** 2657–2670 (2008).

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

123. Daguenet, E. *et al.* Perispeckles are major assembly sites for the exon junction core complex. *Mol. Biol. Cell* **23,** 1765–1782 (2012).

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

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

126. Barbosa, I. *et al.* Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nat. Struct. Mol. Biol.* **19,** 983–990 (2012).

127. Ideue, T., Sasaki, Y. T. F., Hagiwara, M. & Hirose, T. Introns play an essential role in splicing-dependent formation of the exon junction complex. *Genes Dev.* **21,** 1993–1998 (2007).

128. Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N. & Moore, M. J. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA N. Y. N* **8,** 426–439 (2002).

129. Jurica, M. S. & Moore, M. J. Capturing splicing complexes to study structure and mechanism. *Methods San Diego Calif* **28,** 336–345 (2002).

130. Lykke-Andersen, J., Shu, M. D. & Steitz, J. A. Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* **293,** 1836–1839 (2001).

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

132. Degot, S. *et al.* Association of the breast cancer protein MLN51 with the exon junction complex via its speckle localizer and RNA binding module. *J. Biol. Chem.* **279,** 33702–33715 (2004).

133. Macchi, P. *et al.* Barentsz, a new component of the Staufen-containing ribonucleoprotein particles in mammalian cells, interacts with Staufen in an RNA-dependent manner. *J. Neurosci. Off. J. Soc. Neurosci.* **23,** 5778–5788 (2003).

134. Palacios, I. M., Gatfield, D., St Johnston, D. & Izaurralde, E. An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. *Nature* **427,** 753–757 (2004).

135. Bono, F. *et al.* Molecular insights into the interaction of PYM with the Mago-Y14 core of the exon junction complex. *EMBO Rep.* **5,** 304–310 (2004).

136. Diem, M. D., Chan, C. C., Younis, I. & Dreyfuss, G. PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat. Struct. Mol. Biol.* **14,** 1173–1179 (2007).

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

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

139. Ashton-Beaucage, D. *et al.* The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in Drosophila. *Cell* **143,** 251–262 (2010).

140. Wang, Z., Murigneux, V. & Le Hir, H. Transcriptome-wide modulation of splicing by the Exon Junction Complex. *Genome Biol.* **15,** 551 (2014).

141. Dimitri, P., Junakovic, N. & Arcà, B. Colonization of heterochromatic genes by transposable elements in Drosophila. *Mol. Biol. Evol.* **20,** 503–512 (2003).

142. Luo, M. J. & Reed, R. Splicing is required for rapid and efficient mRNA export in metazoans. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 14937–14942 (1999).

143. Carmody, S. R. & Wente, S. R. mRNA nuclear export at a glance. *J. Cell Sci.* **122,** 1933–1937 (2009).

144. Hsu, I.-W. *et al.* Phosphorylation of Y14 modulates its interaction with proteins involved in mRNA metabolism and influences its methylation. *J. Biol. Chem.* **280,** 34507–34512 (2005).

145. Zhou, Z. *et al.* The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407,** 401–405 (2000).

146. Stutz, F. *et al.* REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA N. Y. N* **6,** 638–650 (2000).

147. Rodrigues, J. P. *et al.* REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 1030–1035 (2001).

148. Schmidt, U., Richter, K., Berger, A. B. & Lichter, P. In vivo BiFC analysis of Y14 and NXF1 mRNA export complexes: preferential localization within and around SC35 domains. *J. Cell Biol.* **172,** 373–381 (2006).

149. Abruzzi, K. C., Lacadie, S. & Rosbash, M. Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *EMBO J.* **23,** 2620–2631 (2004).

150. Gatfield, D. & Izaurralde, E. REF1/Aly and the additional exon junction complex proteins are dispensable for nuclear mRNA export. *J. Cell Biol.* **159,** 579–588 (2002).

151. Longman, D., Johnstone, I. L. & Cáceres, J. F. The Ref/Aly proteins are dispensable for mRNA export and development in Caenorhabditis elegans. *RNA N. Y. N* **9,** 881–891 (2003).

152. Braddock, M. *et al.* Intron-less RNA injected into the nucleus of Xenopus oocytes accesses a regulated translation control pathway. *Nucleic Acids Res.* **22,** 5255–5264 (1994).

153. Matsumoto, K., Wassarman, K. M. & Wolffe, A. P. Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. *EMBO J.* **17,** 2107–2121 (1998).

154. Lu, S. & Cullen, B. R. Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. *RNA N. Y. N* **9,** 618–630 (2003).

155. Nott, A., Meislin, S. H. & Moore, M. J. A quantitative analysis of intron effects on mammalian gene expression. *RNA N. Y. N* **9,** 607–617 (2003).

156. Noé, V., MacKenzie, S. & Ciudad, C. J. An intron is required for dihydrofolate reductase protein stability. *J. Biol. Chem.* **278,** 38292–38300 (2003).

157. Wiegand, H. L., Lu, S. & Cullen, B. R. Exon junction complexes mediate the enhancing effect of splicing on mRNA expression. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 11327–11332 (2003).

158. Nott, A., Le Hir, H. & Moore, M. J. Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes Dev.* **18,** 210–222 (2004).

159. Richardson, C. J. *et al.* SKAR is a specific target of S6 kinase 1 in cell growth control. *Curr. Biol. CB* **14,** 1540–1549 (2004).

160. Mamane, Y., Petroulakis, E., LeBacquer, O. & Sonenberg, N. mTOR, translation initiation and cancer. *Oncogene* **25,** 6416–6422 (2006).

161. Ma, X. M., Yoon, S.-O., Richardson, C. J., Jülich, K. & Blenis, J. SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. *Cell* **133,** 303–313 (2008).

162. Chazal, P.-E. *et al.* EJC core component MLN51 interacts with eIF3 and activates translation. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 5903–5908 (2013).

163. Maquat, L. E. When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. *RNA N. Y. N* **1,** 453–465 (1995).

164. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* **23,** 198–199 (1998).

165. Thermann, R. *et al.* Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J.* **17,** 3484–3494 (1998).

166. Wagner, E. & Lykke-Andersen, J. mRNA surveillance: the perfect persist. *J. Cell Sci.* **115,** 3033–3038 (2002).

167. Schell, T., Kulozik, A. E. & Hentze, M. W. Integration of splicing, transport and translation to achieve mRNA quality control by the nonsense-mediated decay pathway. *Genome Biol.* **3,** REVIEWS1006 (2002).

168. Conti, E. & Izaurralde, E. Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol.* **17,** 316–325 (2005).

169. Lejeune, F. & Maquat, L. E. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.* **17,** 309–315 (2005).

170. Singh, G., Rebbapragada, I. & Lykke-Andersen, J. A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol.* **6,** e111 (2008).

171. Lykke-Andersen, J. mRNA quality control: Marking the message for life or death. *Curr. Biol. CB* **11,** R88–91 (2001).

172. Czaplinski, K. *et al.* The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes Dev.* **12,** 1665–1677 (1998).

173. Kashima, I. *et al.* 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,** 355–367 (2006).

174. Lykke-Andersen, J. Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay. *Mol. Cell. Biol.* **22,** 8114–8121 (2002).

175. Lejeune, F., Li, X. & Maquat, L. E. Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. *Mol. Cell* **12,** 675–687 (2003).

176. Grimson, A., O'Connor, S., Newman, C. L. & Anderson, P. SMG-1 is a phosphatidylinositol kinase-related protein kinase required for nonsense-mediated mRNA Decay in Caenorhabditis elegans. *Mol. Cell. Biol.* **24,** 7483–7490 (2004).

177. Gatfield, D., Unterholzner, L., Ciccarelli, F. D., Bork, P. & Izaurralde, E. Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. *EMBO J.* **22,** 3960–3970 (2003).

178. Mohr, S. E., Dillon, S. T. & Boswell, R. E. The RNA-binding protein Tsunagi interacts with Mago Nashi to establish polarity and localize oskar mRNA during Drosophila oogenesis. *Genes Dev.* **15,** 2886–2899 (2001).

179. Hachet, O. & Ephrussi, A. Drosophila Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. *Curr. Biol. CB* **11,** 1666–1674 (2001).

180. Palacios, I. M. RNA processing: splicing and the cytoplasmic localisation of mRNA. *Curr. Biol. CB* **12,** R50–52 (2002).

181. Hachet, O. & Ephrussi, A. Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. *Nature* **428,** 959–963 (2004).

182. Ghosh, S., Marchand, V., Gáspár, I. & Ephrussi, A. Control of RNP motility and localization by a splicing-dependent structure in oskar mRNA. *Nat. Struct. Mol. Biol.* **19,** 441–449 (2012).

183. Silver, D. L., Leeds, K. E., Hwang, H.-W., Miller, E. E. & Pavan, W. J. The EJC component Magoh regulates proliferation and expansion of neural crest-derived melanocytes. *Dev. Biol.* **375,** 172–181 (2013).

184. Saulière, J. *et al.* CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat. Struct. Mol. Biol.* **19,** 1124–1131 (2012).

185. Theurkauf, W. E. *et al.* rasiRNAs, DNA damage, and embryonic axis specification. *Cold Spring Harb. Symp. Quant. Biol.* **71,** 171–180 (2006).

186. Saito, K. *et al.* Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev.* **20,** 2214–2222 (2006).

187. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313,** 320–324 (2006).

188. Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Curr. Biol. CB* **11,** 1017–1027 (2001).

189. Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev.* **12,** 3715–3727 (1998).

190. Deng, W. & Lin, H. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* **2,** 819–830 (2002).

191. Houwing, S. *et al.* A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129,** 69–82 (2007).

192. Aravin, A. *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442,** 203–207 (2006).

193. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442,** 199–202 (2006).

194. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128,** 1089–1103 (2007).

195. Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. *Science* **315,** 1587–1590 (2007).

196. Malone, C. D. *et al.* Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* **137,** 522–535 (2009).

197. Lau, N. C. Small RNAs in the animal gonad: guarding genomes and guiding development. *Int. J. Biochem. Cell Biol.* **42,** 1334–1347 (2010).

198. Saito, K. & Siomi, M. C. Small RNA-mediated quiescence of transposable elements in animals. *Dev. Cell* **19,** 687–697 (2010).

199. Senti, K.-A. & Brennecke, J. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet. TIG* **26,** 499–509 (2010).

200. Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12,** 246–258 (2011).

201. Levine, M. T. & Malik, H. S. Learning to protect your genome on the fly. *Cell* **147,** 1440–1441 (2011).

202. Prud'homme, N., Gans, M., Masson, M., Terzian, C. & Bucheton, A. Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. *Genetics* **139,** 697–711 (1995).

203. Mével-Ninio, M., Pelisson, A., Kinder, J., Campos, A. R. & Bucheton, A. The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. *Genetics* **175,** 1615–1624 (2007).

204. Desset, S., Buchon, N., Meignin, C., Coiffet, M. & Vaury, C. In Drosophila melanogaster the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways. *PloS One* **3,** e1526 (2008).

205. Pélisson, A. *et al.* Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. *EMBO J.* **13,** 4401–4411 (1994).

206. Chalvet, F. *et al.* Proviral amplification of the Gypsy endogenous retrovirus of Drosophila melanogaster involves env-independent invasion of the female germline. *EMBO J.* **18,** 2659–2669 (1999).

207. Leblanc, P. *et al.* Life cycle of an endogenous retrovirus, ZAM, in Drosophila melanogaster. *J. Virol.* **74,** 10658–10669 (2000).

208. Brasset, E. *et al.* Viral particles of the endogenous retrovirus ZAM from Drosophila melanogaster use a pre-existing endosome/exosome pathway for transfer to the oocyte. *Retrovirology* **3,** 25 (2006).

209. Kalmykova, A. I., Klenov, M. S. & Gvozdev, V. A. Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. *Nucleic Acids Res.* **33,** 2052–2059 (2005).

210. Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318,** 761–764 (2007).

211. Carmell, M. A. *et al.* MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* **12,** 503–514 (2007).

212. Batista, P. J. *et al.* PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Mol. Cell* **31,** 67–78 (2008).

213. Das, P. P. *et al.* Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. *Mol. Cell* **31,** 79–90 (2008).

214. Houwing, S., Berezikov, E. & Ketting, R. F. Zili is required for germ cell differentiation and meiosis in zebrafish. *EMBO J.* **27,** 2702–2711 (2008).

215. Saito, K. *et al.* A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. *Nature* **461,** 1296–1299 (2009).

216. Robert, V., Prud'homme, N., Kim, A., Bucheton, A. & Pélisson, A. Characterization of the flamenco region of the Drosophila melanogaster genome. *Genetics* **158,** 701–713 (2001).

217. Klattenhoff, C. *et al.* The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138,** 1137–1149 (2009).

218. Li, X. Z., Roy, C. K., Moore, M. J. & Zamore, P. D. Defining piRNA primary transcripts. *Cell Cycle Georget. Tex* **12,** 1657–1658 (2013).

219. Gu, W. *et al.* CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as C. elegans piRNA precursors. *Cell* **151,** 1488–1500 (2012).

220. Rangan, P. *et al.* piRNA production requires heterochromatin formation in Drosophila. *Curr. Biol. CB* **21,** 1373–1379 (2011).

221. Pane, A., Jiang, P., Zhao, D. Y., Singh, M. & Schüpbach, T. The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. *EMBO J.* **30,** 4601–4615 (2011).

222. Mohn, F., Sienski, G., Handler, D. & Brennecke, J. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. *Cell* **157,** 1364–1379 (2014).

223. Zhang, Z. *et al.* The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* **157,** 1353–1363 (2014).

224. Zhang, F. *et al.* UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* **151,** 871–884 (2012).

225. Huang, H.-Y. *et al.* Tdrd1 acts as a molecular scaffold for Piwi proteins and piRNA targets in zebrafish. *EMBO J.* **30,** 3298–3308 (2011).

226. Watanabe, T. *et al.* MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. *Dev. Cell* **20,** 364–375 (2011).

227. Handler, D. *et al.* A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J.* **30,** 3977–3993 (2011).

228. Pane, A., Wehr, K. & Schüpbach, T. zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. *Dev. Cell* **12,** 851–862 (2007).

229. Olivieri, D., Sykora, M. M., Sachidanandam, R., Mechtler, K. & Brennecke, J. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *EMBO J.* **29,** 3301–3317 (2010).

230. Haase, A. D. *et al.* Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. *Genes Dev.* **24,** 2499–2504 (2010).

231. Nishimasu, H. *et al.* Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature* **491,** 284–287 (2012).

232. Ipsaro, J. J., Haase, A. D., Knott, S. R., Joshua-Tor, L. & Hannon, G. J. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* **491,** 279–283 (2012).

233. Darricarrère, N., Liu, N., Watanabe, T. & Lin, H. Function of Piwi, a nuclear Piwi/Argonaute protein, is independent of its slicer activity. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 1297–1302 (2013).

234. Kirino, Y. & Mourelatos, Z. 2'-O-methyl modification in mouse piRNAs and its methylase. *Nucleic Acids Symp. Ser. 2004* 417–418 (2007). doi:10.1093/nass/nrm209

235. Kirino, Y. & Mourelatos, Z. The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA N. Y. N* **13,** 1397–1401 (2007).

236. Kirino, Y. & Mourelatos, Z. Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nat. Struct. Mol. Biol.* **14,** 347–348 (2007).

237. Saito, K. *et al.* Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev.* **21,** 1603–1608 (2007).

238. Kawaoka, S., Izumi, N., Katsuma, S. & Tomari, Y. 3' end formation of PIWI-interacting RNAs in vitro. *Mol. Cell* **43,** 1015–1022 (2011).

239. Kamminga, L. M. *et al.* Hen1 is required for oocyte development and piRNA stability in zebrafish. *EMBO J.* **29,** 3688–3700 (2010).

240. Szakmary, A., Reedy, M., Qi, H. & Lin, H. The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in Drosophila melanogaster. *J. Cell Biol.* **185,** 613–627 (2009).

241. Zamparini, A. L. *et al.* Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. *Dev. Camb. Engl.* **138,** 4039–4050 (2011).

242. Sienski, G., Dönertas, D. & Brennecke, J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151,** 964–980 (2012).

243. Lim, A. K. & Kai, T. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 6714–6719 (2007).

244. Olivieri, D., Senti, K.-A., Subramanian, S., Sachidanandam, R. & Brennecke, J. The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in Drosophila. *Mol. Cell* **47,** 954–969 (2012).

245. Preall, J. B., Czech, B., Guzzardo, P. M., Muerdter, F. & Hannon, G. J. shutdown is a component of the Drosophila piRNA biogenesis machinery. *RNA N. Y. N* **18,** 1446–1457 (2012).

246. Wang, S. H. & Elgin, S. C. R. Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 21164–21169 (2011).

247. Klenov, M. S. *et al.* Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 18760–18765 (2011).

248. Klenov, M. S. *et al.* Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. *Nucleic Acids Res.* **35,** 5430–5438 (2007).

249. Huang, X. A. *et al.* A major epigenetic programming mechanism guided by piRNAs. *Dev. Cell* **24,** 502–516 (2013).

250. Brower-Toland, B. *et al.* Drosophila PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev.* **21,** 2300–2311 (2007).

251. Guzzardo, P. M., Muerdter, F. & Hannon, G. J. The piRNA pathway in flies: highlights and future directions. *Curr. Opin. Genet. Dev.* **23,** 44–52 (2013).

252. Lau, N. C. *et al.* Characterization of the piRNA complex from rat testes. *Science* **313,** 363–367 (2006).

253. Lau, N. C., Ohsumi, T., Borowsky, M., Kingston, R. E. & Blower, M. D. Systematic and single cell analysis of Xenopus Piwi-interacting RNAs and Xiwi. *EMBO J.* **28,** 2945–2958 (2009).

254. Lau, N. C. *et al.* Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. *Genome Res.* **19,** 1776–1785 (2009).

255. Post, C., Clark, J. P., Sytnikova, Y. A., Chirn, G.-W. & Lau, N. C. The capacity of target silencing by Drosophila PIWI and piRNAs. *RNA N. Y. N* **20,** 1977–1986 (2014).

256. Rajasethupathy, P. *et al.* A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell* **149,** 693–707 (2012).

257. Ross, R. J., Weiner, M. M. & Lin, H. PIWI proteins and PIWI-interacting RNAs in the soma. *Nature* **505,** 353–359 (2014).

258. Rouget, C. *et al.* Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. *Nature* **467,** 1128–1132 (2010).

259. Livak, K. J. Organization and mapping of a sequence on the Drosophila melanogaster X and Y chromosomes that is transcribed during spermatogenesis. *Genetics* **107,** 611–634 (1984).

260. Kotelnikov, R. N. *et al.* Peculiarities of piRNA-mediated post-transcriptional silencing of Stellate repeats in testes of Drosophila melanogaster. *Nucleic Acids Res.* **37,** 3254–3263 (2009).

261. Nagao, A. *et al.* Biogenesis pathways of piRNAs loaded onto AGO3 in the Drosophila testis. *RNA N. Y. N* **16,** 2503–2515 (2010).

262. Hirano, T. *et al.* Small RNA profiling and characterization of piRNA clusters in the adult testes of the common marmoset, a model primate. *RNA N. Y. N* **20,** 1223–1237 (2014).

263. Shpiz, S., Ryazansky, S., Olovnikov, I., Abramov, Y. & Kalmykova, A. Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the drosophila germline. *PLoS Genet.* **10,** e1004138 (2014).

264. Micklem, D. R. *et al.* The mago nashi gene is required for the polarisation of the oocyte and the formation of perpendicular axes in Drosophila. *Curr. Biol. CB* **7,** 468–478 (1997).

265. Jin, Z., Flynt, A. S. & Lai, E. C. Drosophila piwi mutants exhibit germline stem cell tumors that are sustained by elevated Dpp signaling. *Curr. Biol. CB* **23,** 1442–1448 (2013).

266. Lin, H. & Spradling, A. C. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. *Dev. Camb. Engl.* **124,** 2463–2476 (1997).

267. Shiimori, M., Inoue, K. & Sakamoto, H. A specific set of exon junction complex subunits is required for the nuclear retention of unspliced RNAs in Caenorhabditis elegans. *Mol. Cell. Biol.* **33,** 444–456 (2013).

268. Schwerk, C. *et al.* ASAP, a novel protein complex involved in RNA processing and apoptosis. *Mol. Cell. Biol.* **23,** 2981–2990 (2003).

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

270. Czech, B., Preall, J. B., McGinn, J. & Hannon, G. J. A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. *Mol. Cell* **50,** 749–761 (2013).

271. Saito, K. *et al.* Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. *Genes Dev.* **24,** 2493–2498 (2010).

272. Neel, H., Weil, D., Giansante, C. & Dautry, F. In vivo cooperation between introns during pre-mRNA processing. *Genes Dev.* **7,** 2194–2205 (1993).

273. Nesic, D. & Maquat, L. E. Upstream introns influence the efficiency of final intron removal and RNA 3'-end formation. *Genes Dev.* **8,** 363–375 (1994).

274. Tomizawa, J. & Itoh, T. Plasmid ColE1 incompatibility determined by interaction of RNA I with primer transcript. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 6096–6100 (1981).

275. Wong, E. M. & Polisky, B. Alternative conformations of the ColE1 replication primer modulate its interaction with RNA I. *Cell* **42,** 959–966 (1985).

276. Watakabe, A., Inoue, K., Sakamoto, H. & Shimura, Y. A secondary structure at the 3' splice site affects the in vitro splicing reaction of mouse immunoglobulin mu chain pre-mRNAs. *Nucleic Acids Res.* **17,** 8159–8169 (1989).

277. Wetterberg, I., Baurén, G. & Wieslander, L. The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. *RNA N. Y. N* **2,** 641–651 (1996).

278. Pandya-Jones, A. & Black, D. L. Co-transcriptional splicing of constitutive and alternative exons. *RNA N. Y. N* **15,** 1896–1908 (2009).

279. Guo, M., Lo, P. C. & Mount, S. M. Species-specific signals for the splicing of a short Drosophila intron in vitro. *Mol. Cell. Biol.* **13,** 1104–1118 (1993).

280. Mayeda, A. *et al.* Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. *EMBO J.* **18,** 4560–4570 (1999).

281. Sahara, S. *et al.* Acinus is a caspase-3-activated protein required for apoptotic chromatin condensation. *Nature* **401,** 168–173 (1999).

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

283. Joselin, A. P., Schulze-Osthoff, K. & Schwerk, C. Loss of Acinus inhibits oligonucleosomal DNA fragmentation but not chromatin condensation during apoptosis. *J. Biol. Chem.* **281,** 12475–12484 (2006).

284. Singh, K. K. *et al.* Human SAP18 mediates assembly of a splicing regulatory multiprotein complex via its ubiquitin-like fold. *RNA N. Y. N* **16,** 2442–2454 (2010).